Università degli Studi di Palermo

AREA RICERCA E TRASFERIMENTO TECNOLOGICO
SETTORE DOTTORATI E CONTRATTI PER LA RICERCA
U. O. DOTTORATI DI RICERCA

Dottorato in Studi Umanistici – D081
Dipartimento di Scienze Umanistiche
L-LIN/01

THE CORE PROPERTIES OF LANGUAGE IN NONHUMAN ANIMALS: INSIGHTS FROM BOTTLENOSE DOLPHINS

IL DOTTORE
GABRIELE GANAU

IL COORDINATORE
MARINA CALOGERA CASTIGLIONE

IL TUTOR
ANNAMARIA BARTOLOTTA

IL CO TUTOR
JAN CASALICCHIO

CICLO XXXVII (2022-2024)
ANNO CONSEGUIMENTO TITOLO 2024/2025

Abstract

Il linguaggio umano è uno dei fenomeni più complessi del regno animale, e la sua origine e la sua evoluzione rimangono poco chiare. Alcuni studiosi (ad es. Berwick e Chomsky 2016) ritengono che le sue proprietà fondamentali siano unicamente umane, e scaturiscano da una singola mutazione molto favorevole, che rappresenta un balzo evolutivo in discontinuità con il percorso delle altre specie. Altri studiosi (ad es. Tomasello 2008) propongono che il linguaggio si sia evoluto gradualmente e in continuità con altri animali, suggerendo che i suoi tratti principali possano essere presenti in specie diverse dall'uomo, in forme più rudimentali. Secondo Chomsky (2001), la proprietà più importante alla base del linguaggio è l'operazione *merge*, che genera strutture sintattiche gerarchiche e complesse combinando ricorsivamente due elementi alla volta. Berwick e Chomsky (2019) sostengono che il *merge* non si possa separare dalla ricorsività, costituendo dunque un tratto "tutto o niente", in cui le combinazioni sono sempre potenzialmente infinite. Tuttavia, Rizzi (2016) suggerisce che il *merge* abbia diversi livelli di complessità. La forma più semplice, definita *merge-1*, consiste nella combinazione non ricorsiva di due unità dotate di significato (ad es. *mela rossa*). Negli ultimi vent'anni, vari studi sulla comunicazione e sulla cognizione animale (ad es. Arnold e Zuberbühler 2006, Ouattara *et al.* 2009, Suzuki e Matsumoto 2022, Leroux *et al.* 2022) hanno identificato in certe specie di uccelli e primati delle combinazioni di richiami che si possono interpretare come forme di *merge-1*, mettendo dunque in discussione l'unicità umana di questo tratto.

Questa tesi mira a portare chiarezza nel dibattito sull'evoluzione del linguaggio, comparando le teorie che derivano dalle prospettive opposte di unicità e non unicità, e individuando quali proprietà fondamentali del linguaggio siano realmente uniche (se ve ne sono), e quali siano condivise con altre specie. Sebbene il *merge* rappresenti il focus dell'analisi, si considerano anche altre proprietà rilevanti, che vengono classificate in caratteristiche linguistiche e non linguistiche. Le caratteristiche linguistiche includono proprietà relative alla struttura, al significato, e all'uso contestuale dei segnali; le caratteristiche non linguistiche si riferiscono ai tratti di dominio generale (come organizzazione cerebrale, cognizione, struttura sociale e comportamenti) che, nonostante non siano propriamente linguistici, sono fondamentali per lo sviluppo e il funzionamento del linguaggio. La letteratura relativa agli studi su animali in natura e in cattività viene quindi analizzata per verificare la presenza o l'assenza di queste caratteristiche, con un'attenzione particolare ai cetacei, fra cui spiccano i delfini dal naso a bottiglia. Questi animali presentano una combinazione di complessità cognitiva e produzione vocale avanzata che è raro trovare

in animali diversi dall'uomo (ad es. i primati mostrano abilità cognitive sviluppate ma hanno sistemi di produzione vocale limitati). L'analisi indica che diverse caratteristiche cruciali per il linguaggio sono presenti in altri animali, fra cui – apparentemente – anche la capacità di combinare unità dotate di significato, in un'operazione simile al *merge-1*. Tuttavia, queste combinazioni sono limitate a due elementi, indicando l'assenza di ricorsività. Altre proprietà assenti in specie diverse dall'uomo – come produttività, struttura gerarchica, lessico esteso – sembrano dipendere da operazioni ricorsive. Di conseguenza, la distinzione fondamentale fra gli umani e gli altri animali non si troverebbe nell'abilità di produrre semplici operazioni combinatorie binarie, ma piuttosto nella capacità di applicare ricorsivamente queste operazioni per un numero illimitato di volte. Le implicazioni evolutive di questa conclusione vengono discusse, fra cui la possibile origine della ricorsività, in continuità o in discontinuità con le altre specie. Inoltre, viene dedicata particolare attenzione ai tratti e alle dinamiche condivise dalle specie dotate di sistemi di comunicazione complessi, con un'enfasi sul ruolo centrale della socialità – una caratteristica presente in tutti questi animali, inclusi gli umani. Queste osservazioni vengono quindi applicate ai modelli teorici sull'evoluzione del linguaggio, individuando gli elementi più convincenti di ciascuna prospettiva e tentando di conciliare le differenze. Infine, viene proposta una ricostruzione speculativa sull'origine del linguaggio umano, che tiene in considerazione l'interazione dei fattori: 1) una dotazione biologica geneticamente determinata; 2) il ruolo dell'esperienza e dell'interazione sociale; 3) uno sviluppo graduale nel tempo. Si spera che questo lavoro possa servire sia come risorsa per i ricercatori che vogliono esplorare le capacità linguistiche degli animali, sia come guida al complesso e variegato dibattito relativo alla natura e all'origine del linguaggio.

Human language is one of the most complex phenomena in the animal kingdom, and its origin and evolution remain unclear. Some scholars (e.g. Berwick & Chomsky 2016) argue that its fundamental properties are uniquely human and arose from a single favourable mutation, representing a saltational leap, in discontinuity with the evolutionary trajectory of communicative skills of other animals. Other scholars (e.g. Tomasello 2008) propose that language evolved gradually and in continuity with other species, suggesting that its core features may be present in more rudimentary forms in nonhuman animals. According to Chomsky (2001), the most essential property of language is the operation known as *merge*, which generates complex hierarchical syntactic structures by recursively combining two meaningful elements at a time. Berwick & Chomsky (2019) contend that merge cannot be separated from recursion, constituting an "all or nothing" trait, whereby combinations are inherently open-ended and potentially infinite. However, Rizzi (2016) suggests that merge

can occur at varying levels of complexity. The simplest form – referred to as 1-merge – entails the non-recursive combination of two meaningful units (e.g. *red apple*). Over the past two decades, studies in animal communication and cognition (e.g. Arnold & Zuberbühler 2006, Ouattara *et al.* 2009, Suzuki & Matsumoto 2022, Leroux *et al.* 2022) have identified call combinations in certain bird and primate species that could be interpreted as instances of 1-merge, thus challenging the notion of human uniqueness of this trait.

This dissertation aims to bring clarity to the debate of language evolution by comparing the theories rooted in the opposite perspectives of uniqueness versus non uniqueness, and by investigating which core properties of language are indeed unique (if any), and which are shared with other species. This approach helps to reevaluate the preexisting theories of language evolution in light of the recent findings from animal studies. While merge is a central focus, other relevant properties are also considered and classified into linguistic and non-linguistic features. Linguistic features include properties related to structure, meaning, and contextual usage of signals; non-linguistic features refer to domain-general traits – such as brain organisation, cognition, social structure, and behaviours – that, although not linguistic per se, are essential for the emergence and functioning of language. The literature on both captive and wild animal studies is reviewed to assess the presence or absence of these features, with a particular focus on cetaceans, and especially bottlenose dolphins. These animals exhibit a combination of cognitive complexity and advanced vocal production and perception rarely found together in other species (e.g. primates show high cognitive abilities but limited vocal production systems). The analysis suggests that many language-related features are present in nonhuman animals, apparently including the capacity to combine meaningful units in pairs, akin to 1-merge. However, these combinations are consistently limited to two elements, indicating an absence of recursion. Other properties absent in nonhuman species – such as productivity, hierarchical structure, and expansive lexicons – also appear to depend on recursive operations. Hence, the key distinction between humans and other animals may lie not in the ability to perform basic compositional operations, but in the capacity for recursive application of these operations. The evolutionary implications of this conclusion are explored, including whether recursion arose in continuity with, or as a discontinuity from, other species. Furthermore, special attention is given to traits and dynamics shared by species with complex communication systems, with an emphasis on the central role of sociality – a common feature in all such species, including humans. These findings are then related back to theoretical models of language evolution, identifying the most compelling elements of each and attempting to reconcile their differences. Consequently, a speculative account of language evolution is

proposed, one that acknowledges the interplay of: (1) a genetically determined biological endowment; (2) experience and social interaction; and (3) gradual development over time. It is hoped that this work will serve both as a resource for researchers investigating linguistic capacities in nonhuman animals and as a guide through the complex and multifaceted debate surrounding the nature and evolution of language.

Table of contents

Introduction **8**

Part I

1. Language "uniqueness" and fundamental properties **13**

1.1 From spoken languages to the faculty of language 13

1.2 The main features of language 15

1.3 The key role of syntax 21

2. Main theories about the nature of human language **24**

2.1 The nature of language according to innatist frameworks 25

2.2 The nature of language according to emergentist frameworks 28

2.2.1 Cognitive linguistics 29

2.2.2 Functional linguistics 33

2.2.3 General overview 37

2.3 Innatism vs emergentism: a comparison 38

2.4 Points of contact and evo-devo approaches 41

2.5 Bridging generative merge with gradual development 44

3. Theories about the origin and evolution of human language **50**

3.1 The faculty of language in the broad (FLB) and narrow (FLN) sense 51

3.2 Innatism in language evolution 54

3.2.1 Strong innatism: language as an evolutionary leap 56

3.3 Emergentism in language evolution 59

3.3.1 Strong emergentism: language as a form of social interaction 59

3.4 Intermediate approaches to language evolution 63

3.4.1 "Moderate" emergentism: language as a multimodal system 64

3.4.2 "Moderate" emergentism: language as a complex adaptive system 67

3.4.3 "Moderate" innatism: language in the language-ready brain 72

3.5 Theories of language evolution: a comparison 76

3.6 Bridging generative merge and gradual evolution 79

4. The properties of language **88**

4.1 Signal 89

4.2 Signal learning 89

4.3 Structure 90

4.3.1 Duality of patterning 91

4.3.2 Phonology 91

4.3.3 Morphosyntax 92

4.4 Meaning 94

4.5 Contextual usage 97

4.5.1 Intentionality 97

4.5.2 Signal exchanges 98
4.5.3 Pragmatic functions 98
5. The general traits related to language 102
5.1 Biological traits 102
5.2 Brain and cognition 103
5.3 Sociality 105
6. Back to the core traits: reflections and research questions 109

Part II

Introduction 112
1. Two approaches: non-linguistic domains and hominins 114
1.1 The research of the core language properties in human non-linguistic domains ... 114
1.2. The research of the core language properties in extinct hominins 118
1.2.1 The limitations in the reconstruction of the language of extinct hominins 118
1.2.2 The protolanguage 128
2. The research of language properties in nonhuman animals 131
3. The properties of language in nonhuman animals 138
3.1 Signal 138
3.2 Signal learning 139
3.3 Structure 141
3.3.1 Duality of patterning 141
3.3.2 Phonology 142
3.3.3 Morphosyntax 144
3.4 Meaning 153
3.5 Contextual usage 159
3.5.1 Intentionality 159
3.5.2 Signal exchanges 161
3.5.3 Pragmatic functions 163
4. The general traits related to language in nonhuman animals 168
4.1 Biological traits 168
4.2 Brain and cognition 170
4.3 Sociality 175
5. Overview: the most promising "linguistic" animals 179

Part III

Introduction 184
1. Biology of cetaceans 187
1.1 Anatomy of cetaceans 190

1.2 Physiology of cetaceans 193
1.3 Brain structure of cetaceans 195
1.4 Vocal production and perception in cetaceans 202
1.4.1 The problems with the definition of "dialect" 211
2. Humpback whales, killer whales, and sperm whales 213
2.1 Humpback whales 215
2.1.1 Vocalisations of mysticetes 217
2.1.2 The songs of humpback whales 221
2.2 Killer whales 233
2.2.1 The "dialects" of the orcas 237
2.3 Sperm whales 242
2.3.1 The clicks of sperm whales 243
3. The bottlenose dolphin 247
3.1 Society and behaviour of bottlenose dolphins 247
3.2 Cognition of bottlenose dolphins 250
3.3. Language-like skills emerged in captive dolphins 255
4. Vocalisations of bottlenose dolphins 263
4.1 Signature whistles 265
4.2 Bray calls 272
4.3 Possible analyses 276
5. The properties of language in cetaceans 285
5.1 Signal 285
5.2 Signal learning 285
5.3 Structure 286
5.3.1 Duality of patterning 286
5.3.2 Phonology 287
5.3.3 Morphosyntax 288
5.4 Meaning 290
5.5 Contextual usage 292
5.5.1 Intentionality 292
5.5.2 Signal exchanges 293
5.5.3 Pragmatic functions 294
6. The general traits related to language in cetaceans 298
6.1 Biological traits 298
6.2 Brain and cognition 298
6.3 Sociality 300
7. Cetaceans, other animals, and humans: a comparison 302
7.1 Answers to the research questions 307
Conclusion 311
References 323

Introduction

Human language is considered the most complex communicative phenomenon in the animal kingdom. As far as we know, no other species is capable to produce meaningful structures as diverse and hierarchically organised as humans do, and no other communicative system conveys a similar amount of information. It has been argued that language has been the engine of human evolution (see for example Berwick & Chomsky 2016, Bickerton 2014, Corballis 2011, Deacon 1997), and that most of our complexity resides in our ability to cooperate, think, and transmit knowledge through it.

Although we might not have the most complex vocal apparatus (other species have more complex traits in terms of vocal production and perception, which allow them to produce and perceive sounds beyond the human capability), it seems that we do the most productive use of it: as observed by Wilhelm von Humboldt (1836), by combining a limited number of discrete elements, humans generate a potentially limitless array of words, sentences, and meanings, while other animals do not show this level of complexity. It can be rightfully objected that we do not know enough about communication and cognition of the other species to say that humans have *the most* complex language. However, everyone should agree that no other animal has reached a level of civilisation, nor an impact on the environment, comparable to that of humans, which stably occupy the top of the food chain in an unprecedented way, by modifying the planet through cooperation, coordination, and shared knowledge. Putting aside any qualitative judgement, it is in this sense that I say that human language is the most complex communicative phenomenon in the animal kingdom.

Whether it is considered as a cognitive power, a communicative system, or a complex set of different skills, language certainly is the tool that led us to our society, to poetry and technology, and to questions like those of this dissertation, which is written through the very system it tries to analyse. All of this is unique – no matter the origin or the real properties of this phenomenon – in the sense that it is not found in other species, which display, at best, some rudimental traces of it. Hence, there must be something which allows humans, and only humans, to master this form of language. For the same reason, during the course of evolution, something must have prompted humans, and only humans, to develop this form of language.

This raises some profound and long-lasting questions:

- What is language?
- Where has it come from?
- Is it uniquely human?

In this dissertation I address these fundamental issues, seeking to answer these questions by resorting to the most recent discoveries in different disciplines, ranging from theoretical linguistics to neuroscience, palaeoanthropology, biology, informatics, ethology, and animal cognition. The only way to encompass the phenomenon of language is to consider multiple, sometimes conflicting approaches, and to try to bring order to the vast array of theories and the debates they have originated. A special attention is paid to the topic of human uniqueness, which requires the search for language properties in animal communication and cognition, to investigate whether the core traits of language, its most rudimental and possibly ancient principles, exist in some form in other species. Apart from traditional studies on primates and birds, I focus on the order of cetaceans, animals which have not been thoroughly examined from the perspective of language – also because of some technical limitations to the studies – but which show many promising traits that might indicate a "linguistic" complexity even superior to that of our closest living relatives. Among cetaceans, the main subject of my analysis are the bottlenose dolphin (*Tursiops truncatus*), which stands out for its complex cognition, sociality, and communication, which include some language-related properties that apparently have never been found in any other nonhuman species. Hence, the question "is language uniquely human?" will first become "which properties of language are present in nonhuman animals?", and then will be "which properties of language are present in bottlenose dolphins?". More specifically, I aim to identify the core properties of language, typically described as a combinatorial operation (referred to as "merge" in some frameworks, though generally recognised under different names across the field). I then investigate whether these core properties are present in nonhuman animals, with a particular focus on bottlenose dolphins. This kind of comparative study will also be useful to analyse the accounts of the origin of language proposed by theoretical linguists and to evaluate them in light of the data coming from animal sciences – which are sometimes overlooked by the scholars –, with the final aim to determine which hypotheses are compatible with the most recent empirical discoveries, and which are not. This work can thus lead to a new account of language evolution that has the potential to mitigate the differences between the two theoretical poles, and to advocate for a convergence of some aspects of the origin of language by deflating some of the most extreme postulations presented in the literature.

The dissertation is divided into three parts: Part I focuses on human language, identifying its core properties according to the traditional literature, and discussing the issue of uniqueness. Chapter 1 describes the main features of language and highlights the key role of syntax, while chapter 2 is dedicated to a presentation of the main theories regarding the nature of language, which can be divided into innatist, emergentist and some approaches that we might

define as “intermediate”. After outlining the differences and the dichotomies between these approaches, I describe their points of contact, in search for a view that can take the best insights from both the opposite poles. Then, chapter 3 focuses on the theories of language evolution, which – following the theories on the nature of language – is divided into innatist and emergentist. Each of them is, in turn, be further classified into strong and moderate positions. To better describe language, in chapter 4 I propose an original classification of its properties, divided in the layers of signal, signal learning, structure, meaning and contextual usage. Then, in chapter 5, I present another classification, this time focused on the general characteristics related to language, which are not linguistic but are necessary requirements for language to exist: I thus describe the most relevant aspects of biological traits, brain and cognition, and sociality. These two classifications will allow me to obtain a clearer picture of the phenomenon of language, and to identify the most important features for my analysis: merge, recursion, and lexicon. All these considerations will lead me to present my research questions, related to the nature and the uniqueness of these elements, in chapter 6. Part II focuses on nonhuman animals, to determine whether the core features of language, identified in my classifications, are uniquely human as some scholars propose. Chapter 1 briefly outlines the research of these core features in non-linguistic domains, and in extinct hominins, which are not the main topic of my dissertation. Then, chapter 2 describes the advantages of the comparative method, and its effectiveness in exploring the similarities and the differences between humans and other species. Hence, in chapter 3 I propose the same classification of the properties of language presented in Part I, but now focused on nonhuman animals: I thus review the existing literature to identify which features are shared – even in more rudimental forms – and which ones appear to be unique. The same approach is reserved to the general traits related to language, which are discussed in chapter 4, following the same classification of Part I. After this, in chapter 5 I will be able to isolate the most promising animals, which show some interesting cases of call combinations that can be considered a simpler form of human merge. Part III is entirely dedicated to cetaceans, a group of animals that stand out for having a series of skills that put them closer to humans than other species. This part is the most original contribution of this dissertation. Chapter 1 describes their fundamental anatomical, physiological, cerebral, cognitive, and vocal characteristics; while chapter 2 focuses more in detail on the species of humpback whales (*Megaptera novaeangliae*), killer whales (*Orcinus orca*), and sperm whales (*Physeter macrocephalus*), describing their complex social systems, behaviours, and vocalisations, underlining the properties related to language that emerge from their analysis. Subsequently, chapters 3 and 4 are devoted to bottlenose dolphins, the most promising cetacean species for several reason

related to peculiar traits and data accessibility. Chapter 3 outlines their behaviour, society, and complex cognitive and linguistic abilities emerged in captivity, while chapter 4 describes and analyses their vocalisations in the wild, discussing the possibility of call combinations comparable to those observed in Part II for other animals. After this detailed inspection, I summarise the findings in chapter 5 and 6, by presenting the same classifications of language properties and general traits related to language of Part I and Part II. Then, in chapter 7, I compare all the species discussed in this dissertation, putting together humans, hominins, cetaceans, and other complex species like nonhuman primates and birds. This work will allow me to find differences and similarities, to identify the uniquely human traits, and to propose an answer to my research questions. Furthermore, I will be able to isolate some shared dynamics and selective pressures which might have led to human language and complex animal communication. I thus conclude the thesis by evaluating the theoretical accounts of language evolution described in Part I, considering the animal data from Part II and Part III, and indicating which ones appear more convincing. Finally, I propose my speculative reconstruction of the origin of language, which tries to reduce the distance between the opposite theoretical poles, highlighting their best intuitions, without the need of completely excluding each other.

Part I

1. Language "uniqueness" and fundamental properties

To explore human language, the first fundamental thing to do is to define it and to understand its main features, the elements which make it the complex phenomenon that we know and instinctively utilise, without having to think about its rules or its nature. In this chapter I focus on introducing this subject and its main features. First, I give some definitions of language, showing how the phenomenon is too complex and variegated to be captured by a few sentences, and I show the conceptual step from the research of the origins of spoken languages to the research of the origin of the faculty of language (the ability to produce it and understand it). Then, I identify some of the most debated issues, especially the concept of human uniqueness, which becomes more controversial when we focus not on the full, complex phenomenon, but on its most rudimental (and possibly non-unique) traits. To further examine its complexity, I then describe and evaluate the key features of language proposed by Charles C. Hockett, which represent one of the most influential starting points for the analysis of language, and a subsequent re-elaboration by Gaetano Berruto and Massimo Cerruti which is more effective in highlighting the generative power of human language. Then I focus on other properties that have been considered core traits of language, as the compositional ability to combine meaningful units to create complex syntactic structures (called merge in the generative framework proposed by Noam Chomsky, see e.g. Chomsky 1995, 2001). Because syntax is traditionally considered the stronghold of human uniqueness, it receives a special attention throughout this dissertation. This first chapter serves as an introduction to the topics of the nature, evolution, and uniqueness of human language.

1.1 From spoken languages to the faculty of language

Language is an elusive concept, which encompasses many (if not most) aspects of human cognition, communication, culture, and civilisation. For this reason, it is not easy to define it clearly. The answer to the question "what is language?" depends on the perspective that is adopted: is it an instrument of thought or communication? Is it instinctive or learned? As we will see during the course of this dissertation, the topic of language nature and evolution revolves around several dichotomies. I start with a classical definition of language, coming from the Oxford Dictionary:

> The system of spoken or written communication used by a particular country, people, community, etc., typically consisting of words used within a regular grammatical and

> syntactic structure; (also) a formal system of communication by gesture […]. (sense 1.a) (*Oxford English Dictionary*, s.v. language)

Another definition, perhaps more complete, comes from the Encyclopedia Britannica:

> A system of conventional spoken, manual (signed), or written symbols by means of which human beings, as members of a social group and participants in its culture, express themselves. The functions of language include communication, the expression of identity, play, imaginative expression, and emotional release. (Robins & Crystal 2024, s.v. language).

Both definitions highlight the key functions of expression and communication of language and the socio-cultural dynamics in which it is used; however, only the Oxford Dictionary highlights the structured nature of language, while the Encyclopedia Britannica mentions the importance of symbols, pairings of form and meaning that represent the essential units of language. None of the two definitions is sufficient to describe the complexity of this phenomenon, because no short description is suited to encompass its many facets. In this dissertation, I will describe the main properties and functions of language, and it will become clear how there is no simple answer to the many interrogatives which revolve around it. What is certain is that language is a peculiar and composite phenomenon not observed in other species (at least not at the human level), so it is more than legitimate to wonder where it came from.

The question of the origin of language is very ancient, and it interested theologists and philosophers since the beginning of history (Di Sparti 2006, Graffi 2019). During the nineteenth century, inspired by the discovery and further study of the Indo-European language, some scholars focused on the research of a primordial language from which all other human languages derived (see e.g. Villar 1997). The theoretical possibilities were (and still are) two: language first emerged in a single population, and then it gradually spread to all mankind, differentiating itself during the millennia until reaching the diversity of languages we see today (monogenesis); or language emerged independently in different communities, which developed different systems (polygenesis), so that a primordial language did never exist (see Graffi 2019). After a first moment of enthusiasm, it became clear that the linguistic methods could not go beyond a few thousands of years in the past, and so the focus shifted on another aspect: the biological capability of humans to produce language, defined *faculty of language*. The interest on this topic arose in the second half of the twentieth century, and further increased in the new millennium, enhanced by

interdisciplinary approaches which involve the study of cognition, brain structure, social dynamics, and evolution, both in humans and nonhuman animals, also including hominins thanks to advances in genetics and palaeoanthropology. Furthermore, new innovations in computational sciences and statistical analysis brought new insights on the matter, and their contribution is destined to increase due to the current progresses in machine learning and artificial intelligence.

Therefore, in this dissertation, when I refer to language, I always refer to the faculty of language, or ability to produce language shared by all humans and involved in the acquisition and use of all specific languages spoken by the populations, whether it is English, Chinese, Italian, and so on. Hence all the properties which are described in the next sections apply to any language.

1.2 The main features of language

Several theories have been proposed about the "uniqueness", origin, and evolution of language (see below), which differ in many aspects that I discuss in the next pages. One of the main questions is whether language evolved in discontinuity or in continuity with other species. In the first case, language would be totally different from any other phenomenon observed in the animal kingdom, and would likely be the consequence of a unique trait or event which represented a clear break with the path of human ancestors (see Berwick & Chomsky 2016). In the second case, the evolution of human language would have started from traits shared with other species, probably in a gradual path of increasing complexity, moved by selective pressures and social dynamics (see Tomasello 2008, Pinker & Bloom 1990). In this second, opposite stance, there isn't necessarily a unique trait, and thus it is believed that some rudimental properties connected with language can be found in other species (see Townsend *et al*. 2018).

Most positions agree that language is obtained through the combination of several traits and abilities, both physical and cognitive, which might be present in other animals (and in some cases they are certainly present, e.g. complex vocal apparatus). However, the views differ on the already cited idea, proposed by some scholars (e.g. Berwick & Chomsky 2016), that there is a linguistic core property that is uniquely human, without which language would never be produced (see Wacewicz *et al.* 2020, Fitch 2010 for a review of the debate).

To verify the claim that there is a unique core property in human language, a useful approach is to identify and define this property (or set of properties) from a theoretical perspective, and then search for the same feature, or simpler forms of it, in nonhuman

species. The first step towards this goal is to identify the main features of language. A starting point is the analysis of the fundamental characteristics of human language made by Charles F. Hockett, who describes a total of sixteen key properties: duality, productivity, arbitrariness, interchangeability, specialisation, displacement, cultural transmission (Hockett 1958, 1959), vocal auditory channel, broadcast transmission and directional reception, rapid fading, total feedback, semanticity, discreteness (Hockett 1960a, [1960b] 1977), prevarication, reflexiveness, learnability (Hockett 1966, 1968). Table 1 gives a definition for each feature[1]:

Language feature	**Description**
Duality	Presence of two levels: one of meaningless signalling units (e.g. phonemes), and one of meaningful units (e.g. words)
Productivity	Ability to produce novel utterances by combining meaningless units
Arbitrariness	There is no natural correspondence between the physical form of the words and the meanings associated with them
Interchangeability	Humans can give and receive the same linguistic signals: all messages that can be heard can be also produced
Specialisation	There is no particular relationship between the trigger and the direct physical consequences of the messages
Displacement	Ability to refer to elements outside of the context
Cultural transmission	Involvement of learning and teaching in the development of language
Vocal auditory channel	Importance of systems of vocal production and perception
Broadcast transmission and directional reception	A linguistic signal is "openly" transmitted, so that anyone in the surroundings can hear it, and locate its source
Rapid fading	Linguistic signals are transitory: they disappear after they are uttered
Total feedback	The speaker can hear the totality of his utterances
Semanticity	The elements that compose a message (signifier) are associated with elements of the real world (signified)
Discreteness	The language units are functionally distinct: the system is not scalar (each element is discrete with no "mixed" units)
Prevarication	Messages can be false or meaningless (relies on semanticity)
Reflexiveness	Language can be used to talk about language itself
Learnability	An individual can learn to speak more than one language

Table 1. Description of the sixteen key properties of language by Hockett (1958, 1959, 1960a, [1960b] 1977, 1966, 1968).

[1] These features will be described more in detail in the following chapters.

As of today, these features are not considered sufficient to capture the nature of language, and some critiques have been moved: according to Slawomir Wacewicz & Przemysław Żywiczyński (2015), despite its value as a descriptive tool, Hockett's classification is not effective in the comparison with other species, and thus is unfit for language evolution research. The scholars discuss two main reasons. The first is the "focus on the means at the expense of the content" (Wacewicz & Żywiczyński (2015: 37): many of Hockett's key properties focus only on the vocal aspect of communication, excluding other modalities, especially gestures. As Wacewicz & Żywiczyński (2015) observe, oral and signed languages have the same properties, and the same message can be transmitted in both systems, because language is independent from the mean used to express it. Furthermore, communication is multimodal (see also Kendon 2004), and speech is usually accompanied by other signals (e.g. gestures, facial expression). Hockett's features do not consider these important aspects, and thus they cannot be used to search for linguistic properties in animal gestures, nor to explore the idea that human language evolved from gestural communication. The second reason discussed by Wacewicz & Żywiczyński (2015: 38) is the "focus on the code itself, rather than cognitive abilities of its users". Hockett describes only the "external" aspects of language, the properties which emerge superficially in the code. The mere presence of these features does not explain their origin: behind these superficial properties there must be some deeper cognitive mechanisms, which had evolved in some way, and the research should focus on them. There cannot be a rise in structural complexity without an increase in cognitive complexity to support it. This focus on cognitive mechanisms facilitates the comparison with other species, which can be pursued by searching for rudimental "linguistic" cognitive abilities in nonhuman animals.

It should be considered that Wacewicz & Żywiczyński (2015) maintain an "individualistic-internalistic" perspective of language, meaning that they give a major importance to the internal level of cognition rather than to its external communicative aspects. For this reason, they suggest that language is based on traits such as "cooperation with non-relatives, shared intentionality, metarepresentation and Theory of Mind, mimesis and intentional imitation, enhanced memory and executive function, symbolic representation, open endedness, and recursion – all of which are better understood as cognitive skills rather than features of language in the E-sense." (Wacewicz & Żywiczyński 2015: 42). The considerations of Wacewicz & Żywiczyński point to the importance of a series of cognitive traits and dynamics, some of them strictly connected with sociality and interaction, which will be discussed in the following chapters of this dissertation. About

Hockett's features, however, I believe that they can still be useful in language evolution studies, for three reasons. First, the finding of "superficial" properties of language (e.g. productivity) in nonhuman animals can be a starting point for a deeper analysis of the species that display it, and thus can help determine which animals are worthy of further examination. Second, the presence in nonhuman communication systems of a feature described by Hockett (e.g. displacement in bee dancing) tells us that this feature is not uniquely human, and that can be obtained by other species, likely through convergent evolution. Independently from the cognitive skills of two animals – which might be very different, like in humans and bees – the presence of a shared trait suggests that some properties of language can be mastered by a species without having "the whole package". Third, Hockett's classification has been a starting point for several scholars who further analysed, modified and updated his features, adding new properties that better describe language, and also repairing some of the flaws identified by Wacewicz & Żywiczyński (2015).

In this direction, a more effective classification is that proposed by Berruto & Cerruti (1997), in which the characteristics, despite being sixteen like in Hockett, present several differences. I have chosen this account as an example among many others because it manages to capture many aspects of the generative capability of human language, but it does not aim to include nor summarise all the proposals present in the literature. Table 2 describes the features of Berruto & Cerruti (1997).[2]

Language feature	**Description**
Biplanarity	Presence of a signifier (physical level of the linguistic sign) and of a signified (the associated semantic information) (similar to Hockett's semanticity)
Arbitrariness	There is no natural correspondence between the physical form of the words and the meanings associated with them (same as Hockett 1958)
Double articulation	Presence of two levels of articulation: a signifier can be divided in meaningful units (morphemes) and in meaningless units (phonemes) (same as Hockett's duality)
Modality-independence	The same message can be transmitted in different forms (oral, written)
Linearity	The elements composing a message are ordered in time in a linear way, one signifier following the other, due to the physical nature of sound
Discreteness	The difference between the language signs is absolute with clear separations between them: two words can be identical or different, there is no continuum (like Hockett 1966)

[2] The translation of the names of their features, originally in Italian, is mine.

Semantic omnipotence	Theoretical potential to express through language any meaning, even from other codes and domains
Multifunctionality	Language can be used for different functions (e.g. represent thought, share information and feelings, interact with others)
Reflexiveness	Language can be used to talk about language itself, in a metalinguistic way (same as Hockett 1966)
Productivity	Ability to produce novel messages by combining or giving new values to existing units (also defined *regular creativity*, allows to obtain infinite results starting with a finite number of rules, recursively applied) (similar to Hockett 1958)
Recursion	Ability to perform the same finite operation for a potentially infinite number of times, applying the operation (e.g. combining two elements), and then applying the same operation to its result, and so forth: the only limit of recursive application is represented by memory capacity
Displacement	Ability to refer to elements outside of the context (same as Hockett 1958)
Detachment from stimuli	Ability to recall concepts independently from the external stimulus which generates them (connected with displacement)
Cultural transmission	Involvement of learning and teaching in the development of language: each language is transmitted in a cultural context (same as Hockett 1958)
Syntactic complexity and structure dependency	Language has a complex hierarchical syntactic structure, organised in a net of long-distance dependencies between elements: the meaning of a message does not depend only on word meanings, but also on structural relations
Equivocity	There is no univocal correspondence between signifier and signified (a signifier can have more than one meaning and a signified can have more than one word to describe it)

Table 2. Description of the properties of human language according to Berruto & Cerruti (1997).

As Table 2 shows, some of the most relevant features proposed by Hockett are maintained, with minor differences in the definitions (arbitrariness, discreteness, reflexiveness, productivity, displacement, cultural transmission). Others are given a different name, but resemble some of the original properties: biplanarity is similar to Hockett's semanticity, and double articulation (Martinet 1960) is almost identical to duality.[3] The most evident difference is the absence of references to the vocal auditory channel and its related properties (which are reduced to linearity and discreteness), thus this classification is more

[3] Both properties are probably influenced by the work of Louis Hjelmslev (1943), who identified two "planes" of language: *plereme*, which contains meaningful units, and *ceneme*, which contains meaningless units.

open to other forms of communication. This seems to be confirmed by the new feature modality-independence, referred to the fact that language uses different means: apart from oral and written forms, this definition can include also gestures and sign languages. Other new properties are detachment from stimuli (connected with displacement, because both rely on the ability to evade the context), equivocity (which expands an aspect of arbitrariness), semantic omnipotence and multifunctionality (both focused on the theoretical potential of language to express every meaning and many functions), and, finally, recursion and syntactic complexity. The latter two are the most innovative and relevant, because they reflect the crucial aspects of generative capability and structural complexity present in human language, which seem to be the main difference with other species. Recursion allows to expand the language structures to a level which is potentially infinite, by adding new elements to words and sentences (see also § 2.5). This creates hierarchically organised structures that reach a level of syntactic complexity unparalleled in the animal kingdom (Berwick & Chomsky 2016). Another aspect of syntactic complexity is structure dependency: the sense of the words depends not only on their meanings, but also on their position in the structure of the sentence. Recursion and syntactic complexity are thus two core features of human language, and the research on these properties in nonhuman animals should be encouraged.

Although the classification of Berruto & Cerruti (1997) solves some of the "problems" present in Hockett's features (i.e., the "focus on the means at the expense of the content", by excluding the vocal auditory channel and including structural and generative complexity), it does not address other issues raised by Wacewicz & Żywiczyński (2015), namely the "focus on the code itself, rather than cognitive abilities of its users": the perspective is still on what they define as "external" features of language, i.e., the properties that appear on the surface of this phenomenon, and it does not focus on the deeper cognitive operations from which all the features derive. In the generative theory, properties such as recursion and structure dependency are part of the I-language, and are thus considered "internal", while the E-language is referred to the externalised sentences, which rise on the surface (Chomsky 1957, 1986, see also § 2.1). However, Wacewicz & Żywiczyński (2015) are more interested in identifying the most rudimental cognitive principles that allow the production and comprehension of language (and are thus *behind* language),[4] and to determine their presence in nonhuman species, to assess whether they are uniquely human or not.[5]

[4] However, the property of recursion could be defined as "internal" even from the perspective of Wacewicz & Żywiczyński (2015), if we consider it a cognitive skill independent from language (see chapters 2 and 3).
[5] The search for these principles in the animal kingdom (and especially in bottlenose dolphins) is the topic or Part II and Part III of this dissertation.

Furthermore, Berruto & Cerruti (1997) do not consider some cognitive traits connected with sociality and interaction (e.g., cooperation, shared intentionality, Theory of Mind, intentional imitation) deemed important by Wacewicz & Żywiczyński (2015) – among many other scholars. These characteristics, which are not linguistic per se, can be hardly separated by the language phenomenon, especially while searching for its cognitive and evolutionary origins. This leads us to one of the main dichotomies in the conception of human language: is it a "mosaic", i.e. a conjunction of several equally important distinct features (more or less related), possibly with different origins (Hurford 2003), or has it a central core property, uniquely human, without which no language could exist (Fitch *et al.* 2005, Hauser *et al.* 2002)? Without dwelling on the topic, I believe that language cannot be considered as a whole block, because it is a multimodal phenomenon that clearly involves several different systems and abilities, some of which are present in other species, as scholars from different frameworks have acknowledged (see § 1.3). However, it is possible that there are some core cognitive properties which produce language by using and connecting different pre-existing systems, and the human uniqueness could be represented by these properties (or by their higher level of complexity). For this reason, in the next section I start by focusing on the core elements which, according to most scholars, make language unique.

1.3 The key role of syntax

In the generativist approach, syntax is considered at the centre of language because it mediates between the phonetic-articulatory interface (Phonetic Form) and the semantic-conceptual interface (Logical Form), representing the "heart" of the whole system (Chomsky 1965). According to Chomsky (see e.g. Berwick & Chomsky 2016), it is uniquely human, and this strong postulation probably influenced the view of scholars from different frameworks, to the extent that the generative power of human syntax tends to be considered the most elusive aspect of animal communication (see e.g. Townsend *et al*. 2018, Züberbuhler 2019). Syntax is described as the study of the principles that govern the combination of words to form grammatical sentences in a language (Chomsky & Halle 1968). It aims to understand the underlying principles that determine the order, hierarchy, and relationships between linguistic elements within a sentence. Syntax thus explores how humans combine and organise these linguistic elements in complex meaningful structures (phrases, clauses, sentences), by following a set of rules. It is an unconscious phenomenon, because the rules are not deliberately established by conventions between people: humans spontaneously acquire the syntactic structures of the language to which they are exposed,

and instinctively apply them (Pinker 1994, Chomsky 1957), starting at the age of 2-3 years (Hoff 2009). Furthermore, humans can generalise those rules by creating structures and meanings that they have never heard before[6] (see e.g. Goldberg 2006, O'Grady 2005, Chomsky 1965).

While some scholars believe that the basic rules of syntax are genetically transmitted (see e.g. Chomsky 1957), others propose that they are just learned through experience (see e.g. Tomasello 2003). In the first case, according to the Principles & Parameters model (Chomsky 1981, see also below), humans have both innate principles common to all languages and more flexible parameters, which vary across languages and are "set" by the children after linguistic exposure. Later, in the Minimalist Program, Chomsky (1995, see also below) abandons the idea of parameters and proposes that there are only some generic principles. On the other hand, scholars like Tomasello (2003, 2008) believe that there are no pre-existing genetically transmitted principles nor parameters: only general cognitive skills and experience are involved in language acquisition (see also Kirby *et al*. 2015). This leads us to the Poverty of stimulus argument (Chomsky 1965), according to which children are exposed to a minimal part of a language, but nevertheless they are able to abstract and generalise all its rules and master them totally. This idea is a milestone in generative frameworks, and it is used as an argument for the innateness of the language faculty, because experience alone would not account for all the structures which are generated, and thus there should be a genetic component which allows this productivity (Chomsky 1959). However, other scholars believe that the poverty of stimulus is not a real issue, because it is not proven that children are not exposed to a sufficient "amount" of language, and the ability to generalise could rely on more general cognitive properties, without the need of a genetically transmitted "linguistic package" (Vakili 2023).

Regardless of this debate, most scholars agree that syntax (innate or learned) is based on the ability to combine discrete elements. This ability has two layers, which use the same principles: combinatoriality and compositionality. **Combinatoriality** (sometimes called phonological syntax) is the ability to combine meaningless discrete units into more complex structures, which can be meaningful or not (Engesser & Townsend 2019, Hurford 2011, 2007). The units combined are phonemes and create morphemes and words. **Compositionality**, a concept generally attributed to the philosopher Gottlob Frege[7], is the ability to combine meaningful units into more complex structures, whose meaning is the sum

[6] The idea of the ability to generalise rules is shared also in frameworks which do not believe in the innateness of syntax (see e.g. O'Grady 2005, Tomasello 2003).
[7] However, some scholars have questioned Frege's authorship (see Pleyer *et al*. 2022).

of their parts (Szabó 2020, Hurford 2011, 2007). A common definition, among many similar others, is that of Klaus Zuberbühler (2019: 2): "the meaning of a complex expression is determined by the meanings of its constituents and the rules used to combine them". In this case, the units combined are words and morphemes (obtained through combinatoriality), and the products are compound words (e.g *hunter-gatherer*), phrases (e.g. *red apple*) or – through the recursive application of the combination – sentences (e.g. *I eat the red apple*).

These properties share some similarities with the double articulation of André Martinet (1949), present in both Berruto & Cerruti's (1997) and Hockett's (1958, 1959) classifications (in Hockett it is called duality): the fact that language can be decomposed into two levels, one of meaningless and one of meaningful units, reflects the ability to combine meaningless units (combinatoriality) and to combine meaningful units (compositionality). For this reason, combinatoriality and compositionality can be considered "behind" double articulation, and they are closer to what Wacewicz & Żywiczyński (2015) identify as deeper cognitive skills, which should be investigated in nonhuman animals. In some theories, the ability to recursively combine meaningful units is considered the core property of human language, and it is called merge (Chomsky 1995, see below). In other theories, the compositional ability is less restrictive than Chomsky's merge (e.g. it can be non-recursive, see Rizzi 2016), and it can depend on general cognitive skills, instead of purely linguistic ones (Kirby *et al.* 2015, Hurford 2011, Fitch 2010, Goldberg 2006, Tomasello 2003, Brighton 2002, Kirby 2000). In any case, most frameworks accept the idea that language is created through the combination and composition of elements, and that there must be a cognitive ability that allows humans to perform these combinations and compositions. Furthermore, the concept of combinatoriality and compositionality has been extended to non-linguistic domains, including the cognitive processes behind tool use, other complex behaviours and aspects of communication that require the combination or composition of elements (see Gontier *et al.* 2024). About the origin of this ability, and its possible presence in other animals, there are several different proposals, that depend on the theory of language from which they are derived. As Ray Jackendoff cleverly observed in the title of his 2010 paper, "Your theory of language evolution depends on your theory of language". In the next chapter I describe and compare some of the main theories about the nature of human language, focusing on generative and emergentist approaches.

2. Main theories about the nature of human language

Theories of language can be divided into two main groups, which represent the opposite poles of a continuum that includes several accounts with "intermediate" positions, more or less pointed towards one of the extremes. On the one side there are innatist theories, according to which our ability to produce language is innate in humans and written in our DNA (Berwick & Chomsky 2016, Chomsky 1995); on the other side there are emergentist theories, which (in their strongest version) propose that language emerges spontaneously through experience and interaction, relying on general cognitive abilities only; as a consequence, it is more similar to a cultural object than to a biological phenomenon (Goldberg 2006, Tomasello 2003). Among the innatist approaches, the most relevant position is surely that of Chomsky's generative framework (Chomsky 1957, 1965, 1981, 1995). Among emergentist approaches, I include different theories which share some points of view; in particular, they distance themselves from generativism and consider the role of experience and general cognitive mechanisms behind language. They can be divided into four groups: cognitive, usage-based, constructionist and functional approaches (see e.g. Goldberg 2006, Tomasello 2003, Langacker 1987, Halliday 1984). It should be made clear that "innatism" and "emergentism" are not the only terms used for this classification: for example, Tecumseh Fitch (2010: 106) describes an opposition between formalism and functionalism, and Michael Pleyer & Stefan Hartmann (2019) distinguish biolinguistics from usage-based approaches. I have chosen these terms to highlight the distinction between innateness and emergence, but there are other dichotomies that focus more on other aspects, and are equally valid.

In the next sections, I focus first on innatist frameworks, describing the main aspects of the generative theory, which is not the unique innatist approach, but certainly the most influential. Then I describe the main aspects of the emergentist approaches, which are more variegated, offering an overview of cognitive and functional linguistics, highlighting the complementarities and the diverse aims of their analyses. After that, I compare the innatist and emergentist frameworks, identifying the main dichotomies and showing their differences and points of contact, also mentioning the theories that do not clearly fall under one of the two categories. Finally, I describe some new promising perspectives that could overcome some of the oppositions: evo-devo approaches and Rizzi's (2016) proposal of a non-recursive merge, which could bridge generative frameworks and the idea of a gradual development of language (from birth to maturation).

2.1 The nature of language according to innatist frameworks

In the generative theory proposed by Chomsky (1957, 1986), language is viewed as an entirely innate biological phenomenon, primarily a tool of inner thought (I-language) that is then externalised in communication (E-language). A major point in Chomsky's approach is the existence of a Universal Grammar: a set of innate principles and rules shared by all humans, which govern the generation of the structures of a language and define the limits to variation (Chomsky 1957). The theory originated as an attempt to explain some aspects related to language acquisition by children, including the aforementioned poverty of stimulus, whereby children would be exposed to a limited quantity of language, yet they master it entirely and are able to generalise its rules (Chomsky 1965).

The central focus in generativism is syntax, which within the framework of Universal Grammar becomes the study of the innate linguistic principles that govern the common underlying structure of all languages, and the investigation of their manifestation in the structure of sentences across different languages (Chomsky 2005). For this reason, generative studies believe that the theory does not only describe and classify the structure of languages, but it also reflects what happens in the human brain: the syntactic structures, visually represented by the well-known tree diagrams, would correspond to "real" combinatorial cognitive processes that occur in the mind when we produce or interpret language (Jackendoff 2002, Pinker 1994, Chomsky 1980). The aim of many generative studies, in fact, is to describe different languages through the same syntactic structures, identifying similarities that can confirm the presence of "universal" rules and support the idea that these are biologically predetermined (see, for example, Baker 2008, Roberts 2005, Chomsky 1986).

Among the arguments in favour of the innateness of syntax is the fact that humans have an instinctive knowledge of the grammatical and ungrammatical structures of their native language, and they are able to understand the structure of sentences even with invented words, if these are conjugated correctly (Chomsky 1957). Furthermore, all languages seem to have basic structural similarities, suggesting the presence of fundamental constraints that limit their variation, and cannot depend on experience (Pinker 1994). Chomsky reformulates the classic concepts of *langue* and *parole* (Saussure 1916) in *competence* and *performance*, whereby competence refers to the underlying knowledge or mental representation of language that all speakers possess, encompassing its grammatical rules, principles, and structures; and performance refers to the actual use of language in real

life situations, involving the application of linguistic competence to produce speech or comprehend language in specific contexts (Chomsky 1965).

In the Principles & Parameters model (1981), Chomsky proposes that there is a sort of linguistic "package" encoded in the genome, which includes fixed and immutable elements present in all languages (the principles of Universal Grammar), and variable elements, set as binary choices (parameters, e.g., "pro-drop language: yes/no"). According to Chomsky, at birth all possible parameters of the various languages are present in children, but then, following exposure to the mother tongue during the first period of life, only the parameters of the language in question would instinctively be "set". For this reason, on one hand, any language can be acquired quickly and spontaneously in youth, obeying to specific biologically determined acquisition phases, and on the other hand, in adulthood, it becomes difficult if not impossible to master a new language, because its parameters are no longer accessible in the brain (Lightfoot 1999).[8]

According to Chomsky (1957, 1965, 1981), language is acquired effortlessly because of a Language Acquisition Device (LAD) in the human brain, an innate cognitive mechanism that is posited to contain the universal principles of language and facilitates the process of language acquisition by guiding children's language learning based on the linguistic input they receive.

More recently, Chomsky developed the Minimalist Program (1995), in which he seeks to reduce the syntactic operations to a single principle, a biologically simple and optimised operation which could be credible from an evolutionary perspective: recursive merge. Chomsky (1995) describes merge as a fundamental operation in syntax, which takes two syntactic objects α and β and forms a new object {α, β}. Merge is considered a simple, intuitive, and efficient way to build structure, with no inherent limitations on the size of the objects that can be merged or on the number of steps that can apply. Merge is therefore a finite binary operation, which combines two meaningful units at a time, creating a more complex structure whose meaning is equal to the sum of its parts (e.g., *red apple*). It bears resemblance with the concept of compositionality (Gil 2022, Szabó 2020, Hurford 2007). The difference lies, however, in recursion: the merge operation can be applied for an indefinite number of times to its own result, generating hierarchically ordered structures which become progressively more complex. "Red apple" can be merged with new elements using new instances of merge: "The red apple", "I eat the red apple", "I eat the red apple that

[8] There is an ongoing debate on the accessibility of UG in adults: according to the Full Transfer/ Full Access hypothesis (Schwartz & Sprouse 1996), when adults acquire an L2 they transfer to it the structures of their L1 (Full Transfer), and then they restructure them by resorting to UG (Full Access) (see also White 2003).

I bought" etc. The whole human linguistic complexity would therefore derive from this combinatorial capacity.

Chomsky (2001) describes merge as a series of operations, which share the same property of producing an infinite set of objects by using a finite set of elements. Its simplest form is the unordered combination of two elements, and it is called set merge.[9] Set merge has a further distinction: external merge combines two separate objects, or an object already merged in the structure with a new one; internal merge creates a copy of an object already present in the structure, and merges it at another point in the sentence (such as the old "move", which is now achieved through the formation of copies, that leave a trace in the syntactic positions previously occupied). A more complex operation is pair merge, which generates ordered pairs of elements.[10] Both operations show traits in common with previous formulations: set merge resembles substitution (Chomsky 1986) and pair merge resembles adjunction (Chomsky 1995). According to Chomsky, merge can only be recursive (Berwick & Chomsky 2019), and it cannot exist in intermediate forms: the combinatorial operation cannot be separated from recursion, nor can it be described as part of a continuum of increasing complexity. By excluding the possibility of a gradual evolution of merge, its presence in rudimental forms in other species is excluded as well, and it is even considered useless to search for it. This claim is at the centre of a debate which is far to be solved (see Berwick & Chomsky 2019, Martins & Boeckx 2019).

This overview of the main points of generative linguistics focused on Chomsky's work; however, there is a whole field of studies pursued by other scholars, which share the fundamental assumption of the central importance of syntax, merge and recursion in human language, and further develop the theory by adding different perspectives. An example is the approach regarding emergent parameters and parameter hierarchies, pursued by the scholars of the Rethinking Comparative Syntax (ReCoS) project, such as Ian Roberts, Theresa Biberauer, and Michelle Sheehan (see Roberts 2012, Biberauer *et al.* 2014, Biberauer 2018). In brief, these approaches explore syntactic variation within generative grammar, updating the traditional view of syntactic parameters. Within this model, parameters are not viewed as innately specified features of Universal Grammar. Rather, they are considered to emerge in a hierarchically structured manner – from broad, cross-linguistically salient generalisations (macroparameters) to increasingly fine-grained, language-specific distinctions (micro and nanoparameters). Critically, this hierarchy guides the child's

[9] Set merge is even more similar to the concept of compositionality, and its unordered nature brings it closer to Gil's (2022) simple bare compositionality.
[10] However, it has been proposed to eliminate pair merge, considering that the simplest form of set merge is sufficient to describe the language operations (see Collins 2002, Oseki 2015).

unconscious parsing of the linguistic input, allowing her to construct the grammar of their target language through a principled and constrained process, rather than via arbitrary or purely inductive mechanisms. Therefore, from this perspective UG has no parameters at all at childbirth, thus resulting more simple and biologically plausible (see Roberts 2012, Biberauer *et al.* 2014, Biberauer 2018).

More difficult to classify are other positions that, while distancing themselves from Chomsky's stronger views, share the idea of an internal perspective on language (and thus as a tool of thought, an internal cognitive phenomenon rather than an external and communicative one). However, these perspectives do not necessarily adhere to the idea of merge or human uniqueness, and they suggest that a Universal Grammar could have evolved gradually by following selective pressures, so they cannot be fully included in this group (see e.g. Jackendoff 2002, Deacon 1997, Pinker 1994). According to Pier Marco Bertinetto (2003: 1), there are other approaches in open contrast with generativism that can also be considered formalist, like lexical functional grammar or head driven phrase structure grammar: both postulate the existence of constraints that make some aspects of the linguistic structures biologically predetermined, so that language constructions do not rely only on experience.

2.2 The nature of language according to emergentist frameworks

Under the label of emergentism[11] I include different approaches that start from a common ground, because they share the basic idea that language does not depend on pre-determined principles in the human brain, but it rather emerges gradually through interaction with the environment and other cognitive processes. In the emergentist view, languages don't have a fixed structure, which is in constant change pushed by different factors and pressures like interaction between individuals, social dynamics, needs, and non-linguistic cognitive and physical skills, in opposition to the formalist idea of a fixed Universal Grammar (MacWinney & O'Grady 2015). In this view there is no language "organ": the complexity relies in the combination of several elements which produce the final result, which is involuntary and somehow inevitable, but it is not a purely biological process. There is not a set of pre-established grammatical rules which humans instinctively apply, and not even a grammar in the traditional sense; but instead there is a series of conventions determined by culture and experience that "assemble" together sequences of words

[11] Some scholars prefer the term emergentism, while others define these frameworks as functionalism or usage-based, so that it is not clear which term should include the other two (see Pleyer & Hartmann 2019). I chose emergentism to highlight the emergent nature of language shared by all these approaches.

(Tomasello 2003). Grammar is simply our way to describe this process. Therefore, language structures can be defined "always provisional, always negotiable, and in fact epiphenomenal" (Hopper 1987:142).

Language acquisition within the emergentist framework is seen as a dynamic and interactive process. Children learn language by actively engaging with their environment, observing linguistic input from caregivers and other speakers, and gradually internalising the patterns and regularities present in the language they are exposed to. Emergentist theories of language acquisition emphasise the role of input frequency, distributional information, and social interaction in shaping children's linguistic development (Wulff & Ellis 2018). In these frameworks, the problem of the poverty of stimulus is not considered a real issue: because all language emerges from experience and interaction, children must acquire language only relying on these factors combined with their general cognitive skills. Bertinetto (2003) observed that these approaches are better suited to describe language change, on a diachronic perspective, while generative frameworks better capture the complexity of language acquisition at the synchronic level.

It must be made clear that not all the theories that I describe in this section totally rely on experience: this is the most "extreme" view, and obviously no theory can really exclude the presence of species-specific biological traits, that – at least – facilitate language learning and acquisition. The main difference with innatism and formalism is that the biological traits which allow language are considered general cognitive abilities, used in different domains, and did not specifically evolve for language. The best example in this direction is cognitive linguistics, which I describe in the next paragraph, with some of its sub-branches: usage-based approaches, construction grammar, and connectionism. Then I briefly focus on functional linguistics, sketching the theories of functional grammar, functional discourse grammar and systemic functional grammar.

2.2.1 Cognitive linguistics

Cognitive linguistics focuses on the internal aspect of language, but it polemically differentiates from Chomsky's generative approach because it believes that language is rooted in general cognitive skills, which are used also for many other non-linguistic processes. It was started at the beginning of the 1980s with an agreement between George Lakoff and Ronald Langacker, which unified their models of conceptual metaphor theory (see Lakoff & Johnson 1980) and cognitive grammar (see Langacker 1987). Cognitive linguistics challenges the Chomskyan idea of the modularity of language, according to which

there is a clear separation between the conceptual-intentional interface and the sensory-motor system. Cognitive frameworks propose the concept of *embodiment*: our understanding of abstract concepts and linguistic expressions is shaped by our bodily experiences and interactions with the physical world; thus, the low sensory-motor processes cannot be separated from the so-called higher processes of language and thought (Ferretti *et al*. 2018). Instead of having a central core system which connects the two interfaces (conceptual-intentional and sensory-motor), it is the opposite: our complexity is grounded on our physical nature, which filters and determines the innermost aspects of cognition, so that even thinking is seen as an automatic and mainly unconscious process (see Dancygier 2017, Croft & Cruse 2004, Lakoff & Johnson 1999). There is thus a "deep link between social cognition, action, and language" (Ferretti *et al.* 2018: 227).

Cognitive linguistics is connected to the so-called **usage-based approaches**:[12] similarly to the concept of emergentism, the usage-based approach to language proposes that linguistic knowledge emerges from patterns of language usage, with a special attention to frequency and statistical processing. It emphasises the importance of actual language use in shaping linguistic structure and acquisition, viewing language as a dynamic, complex adaptive system that evolves through interaction. It focuses on how language is learned, used, and processed in real-life communicative contexts: linguistic structures arise from repeated patterns of usage, and speakers' knowledge of language is based on their experience with these usage patterns. Some of the key principles of the usage-based approach include: 1) *frequency and entrenchment*, according to which linguistic forms that occur more frequently in language use become more entrenched in speakers' minds, leading to stronger associations between form and meaning; 2) *constructions* (see below), which are the central units of analysis and represent form-meaning pairings at various levels of abstraction, from simple word combinations to complex syntactic patterns; 3) *generalisation and productivity*, because speakers generalise from specific instances of language use to create new constructions and apply existing ones in novel contexts (productivity arises from this process of generalisation and adaptation); 4) *social interaction*, in the measure that language learning and use are inherently social activities, which provide opportunities for learners to observe and participate in communicative exchanges, which contribute to their acquisition of language knowledge; 5) *language change*, considered the result of cumulative processes of

[12] However, it is not clear which term should contain the other one: Pleyer & Hartmann (2019), for example, suggest that the term usage-based is almost interchangeable with emergentism, and thus can be applied to both cognitive and functional linguistics. This is surely true, because both frameworks are rooted in use and experience, although they have different nuances: for this reason, I believe that usage-based approaches should fall under cognitive linguistics, although they are fully compatible with functionalism.

language use and transmission across generations, that occurs through the replication and modification of linguistic forms in response to communicative needs and usage pattern (see Tomasello 2003, Ellis & Wullf 2015, Wullf & Ellis 2018). Today, these processes are explored by iterated learning studies (see e.g. Kirby *et al.* 2015), whose experiments involve people that have to find a way to communicate in an artificial language by starting with a few shared elements, from which they spontaneously build more complex constructions through repeated interactions: this approach is widely used also in the field of language evolution, although it is criticised by Chomsky (see e.g. Berwick & Chomsky 2016), because it involves humans already endowed with the modern language faculty.

A very important aspect of cognitive linguistics and usage-based approaches, which can be applied also to functional linguistics, is **construction grammar**, also abbreviated as CxG (see Goldberg, 1995 2006). Construction grammar emphasises the role of constructions in language comprehension, production, and acquisition: a construction is a form-meaning pairing that encapsulates the syntactic, semantic, and pragmatic properties of a linguistic expression. Unlike traditional grammar models that focus on abstract rules or structures, Construction Grammar treats constructions as the basic units of language: sentences are formed by combining various constructions in a hierarchical and compositional manner (Goldberg 2005, Ellis & Wullf 2015). As Nick Ellis and Stephanie Wullf (2015: 64) explain, "constructions blur the traditional distinction between lexicon and grammar", because a sentence, instead of being seen as the consequence of the "application of grammatical rules to put a number of words obtained from the lexicon in the right order" (Ellis & Wullf 2015: 65), is considered "a combination of constructions, some of which are simple and concrete while others are quite complex and abstract" (Ellis & Wullf 2015: 65). These constructions can indeed range from simple patterns like "subject-verb-object" to more complex structures involving multiple clauses and discourse-level phenomena. Basically, all linguistic elements are constructions: a word, e.g. *cat*, is a construction, and the same goes for each morpheme, even those that codify a grammatical function, like the *-s* of plural: *cats* is thus a construction formed by the combination of two different pairings of meaning and form, *cat* and *-s*. The same applies for morphemes like *-able*, which convey the meaning "it can be done" (*doable*, *manageable*, *acceptable*). Following this reasoning, also idiomatic expressions like *pain in the neck*, with the meaning of "something very disturbing", are constructions, and the same goes for any kind of syntactic construction, like for example *Paul gives the kid a candy*, or *John gives me a dollar*. These two sentences share the form Subject-Verb-Object-Object, which remains the same in all the constructions of these kind (no matter the specific words), conveying the common meaning of something being transferred from somebody to

somebody else (Ellis & Wullf 2015). Even more abstract constructions, like the difference between an active or a passive form of the same message, are not described with syntactic movements starting from an original, unmarked structure, but as combinations of constructions with the function to shift the attentional focus, on the agent or on the patient (Goldberg 2006, Ellis & Wullf 2015). An effective example of how sentences are built through combinations of constructions is given by Ellis & Wullf (2015: 65), with *What did Nick give the squirrel?* which, according to them, contains the following constructions: 1) *Nick*, *squirrel*, *give*, *what*, *do* constructions, 2) VP, NP constructions, 3) Subject-Verb-Object-Object construction, 4) Subject-Auxiliary inversion construction.

The language knowledge is thus seen as a large deposit of constructions empirically acquired, stored in the mental lexicon and activated and assembled dynamically during language processing and production. Not all the constructions can be combined, but this depends on the compatibility between their meanings or functions (Ellis & Wullf 2015). Furthermore, usage contributes to fix the constructions, so that the most frequent ones become more rooted in cognition. It has been proposed that this approach can effectively describe the process of grammaticalisation (Benitez-Burraco 2017, see also Pleyer & Hartmann 2019).

Units in construction grammar are combined through several mechanisms: 1) *constructional schemas* (abstract templates that represent recurring patterns in language, which capture the structural and functional properties of constructions and serve as the basis for constructing sentences); 2) *constructional networks* (organised based on their similarities and relationships, allowing speakers to generalise across constructions and apply learned patterns to new contexts); 3) *constructional operations* (speakers manipulate constructions through various operations such as blending, coercion, and extension to create novel expressions or adapt existing ones to fit different contexts); 4) *constructional meaning* (constructions encode not only syntactic structures but also semantic and pragmatic information, so that meaning arises from the interaction between form and function within constructions) (see Golberg 1995, 2006, Wullf & Ellis 2018). Ellis & Wullf (2015) also describe other processes connected with learning and establishment of constructions: 1) *associative learning* (learning constructions through the association between form and meaning, which is easier with "fixed" associations like *cat* and harder with variable associations like *-ing*, which can have different functions); 2) *rational cognitive processing* (related to the already discussed frequency of use, which entrenches the more utilised constructions); 3) *exemplar-based learning* (connected with the fact that language is learned implicitly through distributional analyses automatically performed by the brain); 4)

emergent relations and patterns (which highlights the graduality of the process of learning, with language that emerges as a complex adaptive system from the interaction of cognitive processes, interaction and experience, as discussed in the emergentism paragraph).

Another interesting approach is that of **connectionism** (see Ellis 1998, Westermann *et al.* 2006). Inspired by principles of neural networks, connectionist models propose that language arises from the interaction of simple processing units that simulate the functioning of the human brain. According to Ellis (1998: 638), "although language behaviour can be described as being rule-like, this does not imply that language behaviour is rule-governed". Connectionist approaches thus use computer models to explore "how simple learning mechanisms in artificial neural network are able to acquire the associations between, for example, forms and meanings, […] and then use these associations to produce novel responses by "on-line" generalization." (Ellis 1998: 638). To sum up, "connectionist models demonstrate how subsymbolic associative systems, where there are neither given nor identifiable rules, nevertheless simulate rule-like grammatical behavior" (Ellis 1998: 638). In other words, in this view the connections and combinations observed in language reflect the natural connections and combinations of neurons in the human brain, so that even the most complex operations that might seem to be governed by specific innate linguistic rules in fact emerge spontaneously following more physiological and general dispositions (Westermann *et al.* 2006).

2.2.2 Functional linguistics

While cognitive linguistics is more focused on the internal aspect of language, and its cognitive processes, **functional linguistics** – a term proposed with the following meaning by William Croft in the 1980s in opposition to formalism and structuralism[13] – focuses on its external perspective, identifying communication as the main feature of language. For this reason, this framework emphasises the communicative functions of language and the role of language in social interactions. It views language as a system of meaning-making that serves various functions in human communication, such as expressing thoughts, conveying emotions, negotiating social relationships, and coordinating actions. Functional linguistics

[13] The term "functionalism" has been used with different and sometimes opposite meanings, also to embrace theories which can be considered closer to formalism. For example, Croft (2015) believes that functionalism should include only usage-based approaches and cognitive linguistics, while Newmeyer (1999) calls "functionalist" the approaches which propose the existence of a Universal Grammar gradually obtained through natural selection (see e.g. Pinker 1994, Jackendoff 2002). In this brief overview, I only outline the "external" communicative perspective shared by most of these frameworks, including those which can be considered more formalist because they postulate some kind of internal fixed structure.

seeks to uncover the underlying principles that govern how language is used to achieve these communicative goals (see e.g. Dik 1981, Hurford 1990, Halliday & Matthiessen 2014).
Like cognitive linguistics, functional linguistics takes a usage-based approach to language, but it focuses on how linguistic forms are shaped by communicative needs and situational contexts. It prioritises the analysis of language in actual use, examining the ways in which speakers and listeners deploy linguistic resources to achieve specific communicative purposes. Central to functional linguistics is the notion that linguistic forms are motivated by their communicative functions and that meaning arises from the interaction between linguistic forms and their context of use (Dik 1981, Hurford 1990, Halliday 1984, Halliday & Matthiessen 2014).

Among the main topics studied by functional linguistics we can mention 1) the analysis of the already cited *communicative acts* (which include referential, expressive, directive and metalinguistic functions), examining how different linguistic structures are used to perform these functions in various communicative contexts; 2) the analysis of language in terms of *systems of choices for expressing meaning*, focusing on the functional components of language, such as mood, modality, and transitivity, and how they interact to convey meaning in discourse (see Halliday & Matthiessen 2014); 3) analysis of *register and genre*, investigating the ways in which language varies according to different social contexts and communicative purposes, which lead to conventionalised patterns of communicative activity with characteristic linguistic features; 4) analysis of *text and discourse*, focusing on how language is organised and structured to create coherent and cohesive communication, by examining features such as theme-rheme structure, information structure, and discourse markers to uncover patterns of meaning (Dik 1981, Hurford 1990, Halliday & Matthiessen 2014). Like in formalism, some branches of functional linguistics adopt a typological perspective, in which several functional principles grounded on pragmatics have been identified and analysed, the most relevant being: 1) *order theme-rheme*, with the subject (theme) usually preceding the object; 2) *animate first*, with subjects, usually animate, that precede the object; 3) *given before new*, with the known information coming before the unknown data; 4) *urgency principle*, according to which the most important or urgent information is communicated first (see Song 2012).

The umbrella term of functional linguistics contains several approaches. The most influential are Simon Dik's functional grammar (Dik 1981), its further development by Kees Hengeveld and Lachlan Mackenzie functional discourse grammar (see Hengeveld *et al.* 2008) and Michael Halliday's systemic functional grammar (see Halliday 1984, Halliday & Matthiessen 2014). In short, Dik's functional grammar assigns to the constituents of a

sentence three types of functions: 1) *semantic* (Agent, Patient, etc.), which defines the role of participants in the actions or state of affairs expressed; 2) *syntactic* (Subject and Object), identifying their different perspectives; 3) *pragmatic* (Theme and Tail, Topic and Focus), which describes the informational level of the constituents, related to the context of the interaction (Dik 1997). Functional grammar primarily focuses on the analysis of sentence-level structures and the grammatical relations within them. It deals with how linguistic elements such as words, phrases, and clauses function within the boundaries of a single sentence, aiming to understand how grammar serves various communicative functions and organises meaning at the level of individual utterance.

Functional discourse grammar extends the principles of Dik's functional grammar to analyse discourse beyond the sentence level, focusing on how language works in larger communicative contexts. It offers a more comprehensive account of discourse structure, incorporating insights from pragmatics, discourse analysis, and sociolinguistics (Hengeveld *et al.* 2008). In functional discourse grammar, discourse is seen as a hierarchical structure consisting of various levels, including the clause, the proposition, and the illocutionary act. These levels interact to create coherent and meaningful communication. More specifically, functional discourse grammar considers the pragmatic, semantic, morphosyntactic and phonological aspects of the utterances, which reflect four levels of the grammatical component: *interpersonal* (pragmatics), *representational* (semantics), *morphosyntactic*, and *phonological*. Furthermore, four components are involved in the building of an utterance, namely 1) *conceptual* (in which the communicative intention arises); 2) *grammatical* (formulation of the utterance); 3) *contextual* (related to pragmatics); 4) *output* (practical realisation of the utterance) (Hengeveld *et al.* 2008). Hence the main difference between functional grammar and functional discourse grammar regards the level of analysis, which is focused on the internal structure of the sentences in the former, and on the larger units of discourse in the latter.

Finally, systemic-functional grammar (see Halliday 1984, Halliday & Matthiessen 2014) views grammar as a resource for making meaning in social situations, and it considers language as a system of choices for expressing meaning in different contexts of use. It analyses how linguistic structures are organised and used to achieve various communicative functions, such as representing experiences, expressing attitudes, and negotiating social relationships. Halliday identifies five main dimensions of language: 1) *structure* (syntagmatic order, including clause, phrase, word, morpheme); 2) *system* (paradigmatic order, referred to the choice of expression); 3) *stratification* (semantics and lexicogrammar, term which refers to the continuity between lexicon and grammar); 4) *instantiation*

(relationship between "potential" and "actual" expression); 5) *metafunction* (ideational, interpersonal and textual). The most original contribution from Halliday is that of metafunctions, which are considered groups of semantic systems common in all languages that organise how language is used to convey meaning in communication (Halliday & Matthiessen 2014). The *ideational* metafunction deals with the expression and representation of experiences, entities, and events in the world. It is concerned with conveying information about the external world and the relationships between participants (agents and entities affected by actions), processes (actions, events, states), and circumstances (conditions, locations, times). It enables speakers and writers to communicate about the world and to convey their experiences, perceptions, and interpretations of reality through language. The *interpersonal* metafunction focuses on the interactional and interpersonal aspects of language use, including how language is used to establish and negotiate social relationships, roles, and identities. Whitin the interpersonal metafunction, language is used to enact social actions, such as making requests, giving commands, asking questions, and expressing attitudes and emotions. It involves the negotiation of power, solidarity, politeness, and relationships between interlocutors. This function thus enables speakers and writers to engage in social interaction, to express their intentions and attitudes, and to manage interpersonal relationships through language. The *textual* metafunction is concerned with the organisation and coherence of language in discourse and text. It deals with how language is structured and arranged to create cohesive and coherent communicative texts. It also involves the structuring of information, the management of thematic progression, and the signalling of discourse relations, enabling speakers and writers to produce well-structured and coherent texts that are accessible and understandable to recipients, facilitating effective communication and comprehension (Halliday & Matthiessen 2014).

As it can be observed, functional grammar (together with functional discourse grammar) and systemic functional grammar present some differences: while both are functional frameworks that analyse language in discourse, they differ in their scope, unit of analysis, and theoretical assumptions. Systemic functional grammar provides a comprehensive model for analysing the systemic organisation of language in discourse, while functional grammar and functional discourse grammar offer a more detailed analysis of discourse structure across multiple levels of linguistic organisation.

2.2.3 General overview

After this brief description of the main instances of cognitive and functional linguistics, it can be stated that the two frameworks are complementary, because they focus on two different, non-conflicting aspects of language (internal and external perspective), and they are both rooted in experience and general cognitive mechanisms. Both theoretical frameworks lead to usage-based approaches, which share some principles and methodologies. Construction grammar, although it can be considered more connected with cognitive linguistics, is compatible with functional linguistics as well: however, while cognitive linguistics focuses on the inner mechanisms that allow humans to categorise concepts and create linguistic structures, most functional approaches are more interested in analysing the wider level of discourse, strictly connected with the communicative context, instead of its anatomical parts.

Despite their many differences, it is fundamental to highlight that the emergentist approaches also acknowledge the importance of some of the properties considered hallmarks of innatist and generative frameworks. An example is recursion, which is considered a key trait of generative grammar (see Berwick & Chomsky 2016). Pleyer & Hartmann (2019) observe that the main difference in this case is the role which is attributed to this property: instead of considering it of linguistic domain (with all of its other applications, e.g. calculus, "exapted" from the original linguistic function), emergentist theories suggest that recursion can arise "from combined activities of memory, lexicon, discourse, and role activation" (Pleyer & Hartmann 2019: 6), citing the usage-based proposal of Christiansen & Chater (2015) according to which recursion emerges following skills of general domain. They also underscore that even some biolinguistics account (like Boeckx 2013, 2021) criticise the idea of the presence of a uniquely human core trait (identified by Hauser *et al.* 2002 in recursion), because it "distracts" from the complex nature of language as a mosaic formed by several interacting elements. Furthermore, although being considered important, recursion is not always considered present in all aspects of language: Simon Townsend *et al.* (2018), for example, highlight that there are linguistic constructions that are simple, non-productive and non-hierarchical, and thus do not require the level of subordination and embedding that Chomsky potentially assigns to all linguistic operations (see Berwick & Chomsky 2016).

Even more importantly, emergentist approaches do not deny that a key property of language is the ability to combine meaningful structures, although they do not consider this skill necessarily recursive. I will discuss this idea more in detail later, but for now I point out that most scholars agree that an ability similar to Chomskyan merge has to be present not

only in language but also in cognition (see Pleyer & Hartmann 2019). It seems likely that the ability of combining concepts, before words, is present in the human mind, and is used also in non-linguistic operations like "basic action processing" (MacWhinney 2015, see also Pleyer & Hartmann 2019). Even Chomsky (2007) admits that this operation can be important for other cognitive systems apart from language. A similar reasoning can be applied to hierarchical structuring, which seems to be present in human cognition beyond the sole linguistic activity (ranging from many fields from categorisation to social dynamics). Despite these points of contact, innatist and emergentist frameworks present several differences, which I compare and discuss in the next section.

2.3 Innatism vs emergentism: a comparison

After the brief overview of the main theories of language, ordered under the categories of innatism and emergentism, we can proceed with a comparison of their main assumptions. I make clear again that the terms "innatism" and "emergentism" are a simplification, and that there are other distinctions sometimes preferred by scholars (e.g. formalism vs functionalism, see Croft 2015, Fitch 2010, or biolinguistics vs usage-based, see Pleyer & Hartmann 2019). The difference in terminology contributes to create confusion, because not all the terms are perfectly interchangeable and – as we are about to see – some theories can belong to one side or the other depending on the criterium used to differentiate the frameworks. I thus focus on the main dichotomies, starting with innatism and emergentism, but the discussion will not be limited to these two. As we discussed in § 2.1, innatism refers to the idea that language is an instinctive, innate phenomenon, which is genetically encoded in our DNA: this implies that humans have a biological endowment – a series of species-specific traits – which allows language to be produced. This idea is mainly proposed in generative frameworks (see e.g. Berwick & Chomsky 2016), but it is not limited to them: there are also theories inspired by evolutionary biology that believe that humans obtained a language faculty through processes of natural selection (e.g. Pinker 1994, Jackendoff 2002), which can be considered innatist (see chapter 3 below for a detailed description of several accounts of language evolution). In both views, the biological endowment is not limited to general principles, but contains what Chomsky defined Universal Grammar (see § 2.1 above), a specific set of linguistic rules that allow the formation of grammatical sentences.[14] In this case, the difference between Chomsky's

[14] However, as I stated in § 2.1, in the Minimalist Program Chomsky (1995) tries to reduce the complex rules of UG to a single, simple merge operation.

generativism and Pinker's or Jackendoff's evolutionary approaches is the evolution of this set of rules (saltational or gradual), which I discuss below.

On the other hand, according to emergentism language spontaneously emerges without the need of an innate biological endowment, through the interaction of the several factors discussed in the previous pages (see e.g. Tomasello 2003, MacWhinney & O'Grady 2015). This view thus recognises a greater role to experience in creating the linguistic structures and can include both cognitive[15] and functional linguistics. Therefore, according to emergentist and usage-based frameworks, there is no fixed structure of language, nor a Universal Grammar: the rules of language are simply cognitive constructions and cultural conventions, that become entrenched through repeated usage and statistical repetitions (see construction grammar in § 2.2.1 above). For this reason, language uses general cognitive mechanisms and properties which are utilised also for other tasks (see e.g. Tomasello 2003), a perspective shared by both cognitive and functional linguistics (see e.g. Croft & Cruse 2004, Halliday & Matthiessen 2014). On the other hand, generative and other innatist frameworks propose that language has a specific cognitive domain dedicated to it, which evolved with the primary function of processing and creating language structures (Berwick & Chomsky 2016).

Another important difference between innatism and emergentism is the perspective adopted: in innatism language is seen as an internal phenomenon, thus as an instrument of thought, with communication being a secondary consequence of this cognitive power. This view is shared not only by generative frameworks (Chomsky 2005) and by "innatist-evolutionary" theories (e.g. Pinker 1994, Jackendoff 2002), but also by cognitive linguistics (see e.g. Lakoff & Johnson 1980, Langacker 1987), which in opposition to generativism believes that there is no language faculty in the strict sense, but nonetheless focuses on the internal cognitive processes of the mind. The external perspective of language, which sees it as a tool for communication, is the focus of other usage-based approaches, namely the functional linguistic frameworks, which clearly state that their interest relies in the communicative function of language, which determines its structure (see e.g. Dik 1981, Hengeveld *et al.* 2008, Halliday & Matthiessen 2014).

Another dichotomy related to the focus of the analysis is that between form and function, from which the already cited distinction between formalism and functionalism derives (see Croft 2015, Fitch 2010). In this case, the generative framework perfectly falls

[15] However, as Pleyer & Hartmann (2019) point out, some scholars like Adger (2013) claim that even cognitive linguistics and construction grammar could be considered innatist, because they assume the default presence of a set of cognitive skills, although they do not consider them purely linguistics, leading to a sort of "general domain" innatism.

under formalism, with its focus on the structure of language and especially on syntax. However, the "innatist-evolutionary" theories are more focused towards the functions and the pragmatic aspects of language, like functional and cognitive linguistics. To further complicate the issue, some functional frameworks like head-driven phrase structure grammar (which I did not discuss in this dissertation), despite being interested in the external, communicative perspective of language, propose a rigid syntactic structure, and thus can be considered formalist (see Bertinetto 2003).

Finally, a crucial difference between innatist and emergentist frameworks, but especially between generative and non-generative approaches, is the human uniqueness of the core traits of language. This issue, fundamental for the aims of this dissertation, will be better discussed in the next sections, in light of the theories of language evolution. For now, it is important to keep in mind that, although most – if not all – scholars recognise that the human language is unique for its complexity, this uniqueness greatly varies from the generativist idea of a qualitatively novel core trait (Berwick & Chomsky 2016) to the complex "mosaic" of several shared traits and selective pressures which, together, produce a unique result, despite being probably present in other species in more simple forms (Fitch 2010, Tomasello 2003). In the middle, there are the views like that of Steven Pinker (1994) and Jackendoff (2002), in which the final result of human uniqueness is reached gradually by starting from shared, non-unique traits. The table below summarises these observations.

Dichotomies	
Innatism *Language is innate in humans which instinctively apply it*	**Emergentism** *Language emerges through interaction of several factors*
Biological endowment *A series of species-specific traits allow language*	**Role of experience** *Experience has a major role in the formation of linguistic structures*
Predetermined rules (UG) *There exist a genetically transmitted Universal Grammar*	**Gradual development (no UG)** *There is no fixed structure: language is learned through interaction*
Domain specificity (linguistic) *There are some cognitive modules specifically designed for language*	**Domain general** *Language is produced by using general cognitive skills*
Internal perspective *Language is a tool of thought*	**External perspective** *Language is a tool of communication*
Form *Focus on structure/syntax*	**Function** *Focus on function/pragmatics*
Uniqueness *The core traits of language are exceptional and present only in humans*	**Non uniqueness** *There is no unique core trait: language is triggered by many (maybe shared) factors*

Borderline theories
Innatism with a gradual evolution → UG achieved through natural selection
Functionalism (focus on communication) with a formal fixed structure (e.g. head driven phrase structure grammar)
Usage-based approaches with an internalistic perspective (cognitive linguistics)

Table 3. Comparison of the main dichotomies in the theories about the nature of human language.

2.4 Points of contact and evo-devo approaches

As simplified by the table, some of the dichotomies appear difficult to overcome (e.g. domain specificity vs domain general, or uniqueness vs non uniqueness), while others could be mainly considered different perspectives on the analysis of language (focus on form vs function). We saw that not all the approaches described in the previous sections clearly fall under one category: at the most extreme opposite poles we can put Chomsky's generative grammar (see Berwick & Chomsky 2016), which shares all the assumptions of the left column (innatism), and Michael Tomasello's usage-based account of language, on the right column of emergentism (see Tomasello 2003). The several other accounts, as shown in the previous pages, take mixed elements from the two columns.

Without considering the single theories, but focusing on the dichotomies, I believe that the distance between some aspects can be reduced. First, despite the obvious difference between innatism and emergentism, it has to be clarified what kind of innate linguistic capacity humans have: if the innate biological endowment is not considered as a complete set of rules, but as a genetic predisposition to acquire linguistic structures (called by Fitch 2010 *instinct to learn*, and by Fitch 2014 *dendrophilia*, referring to the human attitude to create tree-like hierarchical structures to organise concepts), then this innate ability can "guide" and facilitate the learning of language and its constructions through practice and interaction. In this case the rules would not be completely emergent, but not even completely pre-determined. Furthermore, it is clear that language (like most complex abilities) relies on both biology *and* experience: what changes is the role and the importance that the two opposite poles give to these factors. And even if we wanted to postulate a full-blown Universal Grammar, then we should consider that several scholars like Pinker (1994), Jackendoff (2002), Pinker & Bloom (1990) have already shown that this does not automatically imply a non-gradual development: interaction and social pressures might have shaped this property during the course of evolution. The issue of domain specificity versus domain general is more complicated, because the two aspects mutually exclude each other.

However, there are scholars (see Ambridge & Lieven 2011 in Pleyer & Hartmann 2019) who are convinced that "language can neither be completely domain general nor an entirely modular system" (Pleyer & Hartmann 2019: 5), thus suggesting an interaction between some domain specific traits and general domain features. It is reasonable to believe that, even if there was a specific module dedicated to language in the brain, a "language organ", this could not operate in complete isolation to other fundamental domain general traits. In any case, the presence of a domain specific trait (even if it is only a small part of the whole system), would mean that language is not completely produced with general domain traits, and vice versa, so it seems impossible to find a middle ground. The view of language as an internal instrument of thought or as an external tool of communication (like the focus on form and syntactic structure versus function and pragmatic aspects) is more a matter of distinct perspectives than irreconcilable differences, and both can offer interesting insights of the nature of language. Finally, I believe that also the debate between uniqueness and non-uniqueness might be partly solved or at least mitigated, by acknowledging that a unique trait or phenomenon does not have to be necessarily considered an exception to the rules of natural selection (as already proposed by several scholars, e.g. Pinker & Bloom 1990, and further elaborated in § 3.4), and that several examples of this kind of uniqueness are present in the adaptations of many species, making peculiar and astonishing phenomena like language, although rare, not so inconceivable on evolutionary scales.

In any case, for an effective reconciliation between the many dichotomies described in this section it is necessary that the theories of both sides abandon some of their most extreme postulations, and actively seek a fruitful middle ground as a starting point for new elaborations. An interesting view on this issue comes from Pleyer & Hartmann (2019), which analyse the different frameworks (under the names of biolinguistics and usage-based theories) showing that new interdisciplinary approaches which tap into biological sciences and genetics have put the opposite poles of innatism and emergentism much closer than in the past.

According to Pleyer & Hartmann (2019), on the "moderate" emergentist side language is being considered more and more as a complex adaptive system, which is thus a consequence of a series of diverse pressures and mechanisms, regarding both biology and experience (see e.g. Dabrowska 2015). On the other side, some biolinguistics approaches (e.g. Martins *et al.* 2016) are focusing on the evolutionary developmental biology, which analyses the interactions between genotype and phenotype (genetic structure and development), thus unifying nature and nurture. These two approaches are very similar and fully compatible, and it seems that the opposite innatist and emergentist frameworks have

been independently converging towards this common ground (although the "strong" positions that I previously described are still very rooted and probably more common). More specifically, the evolutionary developmental biology, or evo-devo, starts from the hypothesis that genes are not invariable blueprints of traits, skills, processes or behaviours which are then invariably expressed in the organism, in a sort of genetic determinism, but have the more "flexible" role of creating "proteins that interact in complex, environmentally modulated networks, to build and maintain brains" (Bowling 2017, cited in Pleyer & Hartmann 2019: 8). In this view, the information codified in the genes interacts with a series of complex internal and external factors related to development, context, experience, and thus the biology and physiology of an organism change and develop even after birth following several pressures (Gilbert 2003, Pigliucci & Müller 2010; Laland *et al.* 2015). In other words, by starting from a certain genetic endowment, an organism can have subsequent modifications depending on other factors, leading it to take slightly different possible ways on the ontogenetic level. As a consequence, there is no opposition anymore between nature and nurture, innateness and experience, because they are interacting factors which lead to the final result. To use the words of Martins *et al.* (2016), reported by Pleyer & Hartmann (2019: 8):

> From an evo-devo perspective, it is not possible to distinguish relevantly between the influence of the genes and the influence of the environment in development, since the end product is the result of the interaction of the information from both levels. In light of Evo-Devo, few dichotomies (e.g. I-Language/E-Language, Nature/Nurture, FLN/FLB, gradualism/saltationism and even adaptation/ exaptation) make perfect sense (Martins *et al.*, 2016, p. 161).

The evo-devo approach can mitigate also other concepts highly debated in linguistics, like the poverty of stimulus (see § 1.3): if experience contributes to shape the organism and the brain, it can also lead to the formation of new neuronal connections and cognitive processes which could almost be compared with the fixation of linguistic parameters; furthermore, some aspects of the critical periods could be explained with a progressive decaying of cerebral plasticity (see Pleyer & Hartmann 2019). The same reasoning can be applied to the modularity issue: some elements of linguistic domain could indeed exist, but instead of being innate, they would develop through experience and interaction, thanks to the cerebral plasticity (see Pleyer & Hartmann 2019). Finally, this view seems to point towards the idea that language is made by several non-unique building blocks, evolved for

different and more general tasks, which together create the unique result of this complex human phenomenon, as suggested by Fitch (2010, 2017).

2.5 Bridging generative merge with gradual development

Despite the many points of contact that I reviewed in the previous section, and the promising perspective of the evo-devo approach, innatist and emergentist frameworks have still many differences which will not be easily overcome. In my opinion, the most irreconcilable issue is Chomsky's idea of recursive merge, as proposed in his Minimalist Program (Chomsky 1995) and discussed in many subsequent publications (see also § 2.1 above). Robert Berwick and Noam Chomsky (2016) acknowledge some aspects of evolutionary developmental biology, and focus on the fact that minor modifications in the genome can yield very big changes; however their interpretation of evo-devo (also presented in Chomsky 2010) has been criticised and defined by Antonio Benítez-Burraco and Victor Longa (2010) too deterministic and simplistic, not accounting for the variegated complexity of the processes involved (see also Pleyer & Hartmann 2019). We have already seen that emergentist approaches recognise both the importance of merge (or a similar operation) and recursion (see § 2.2.3). However, many scholars believe that recursion can emerge from general domain traits (Christiansen & Chater 2015), or that it can be present outside language and used in other tasks like visual and musical processing (see e.g. Martins 2012, 2014, Jackendoff & Pinker 2005). Berwick & Chomsky (2016), on the other hand, believe that recursion cannot be separated from merge, because it is an intrinsic aspect of it. For them merge can only be recursive, and they explicitly claim it (see Berwick & Chomsky 2019) answering to some critiques moved by Pedro Tiago Martins and Cedric Boeckx (2019), which highlight the "no half-merge fallacy". In the view of Berwick & Chomsky (2016, 2019) recursive merge is the core operation of human language, and it cannot exist in intermediate forms: its theoretically infinite generative power is always present in all language constructions, even the simplest, which should always be conceived as hierarchically organised. Chomsky's merge is a language-specific, qualitatively unique, unprecedented cognitive operation which works differently from any other process observed in nature, and thus it cannot be decomposed, it cannot have a gradual development, nor different levels of complexity (see Berwick & Chomsky 2019). This strong view makes it clearly incompatible with any other framework. However, as I hinted in § 2.2.3, most scholars agree that some kind of combinatorial operation has to be present in human cognition, to combine concepts and the linguistic elements which represent them. These

views generally come from other frameworks and are mostly explored in the field of animal cognition (see e.g. Schlenker *et al.* 2016, 2023), which is the topic of the second chapter of this dissertation. Nonetheless, they are worth considering even from the perspective of humans. As previously cited, Townsend *et al*. (2018: 1) observe that "beneath the staggering complexity of human syntax, there is an extensive layer of nonproductive, nonhierarchical syntax", and they give examples of idiomatic expressions (*pain in the neck*), simple coordination (*duck and cover*), elements that are still interpreted in a hierarchical way by generative frameworks. Furthermore, Jackendoff & Wittenberg (2014) present a classification of different levels of language complexity, which starts from one-word utterances and becomes increasingly complex (two-words, concatenations, simple phrases, recursive phrases): their point is that many aspects of language are not complex, recursive, nor hierarchical, and that many messages can be communicated without recurring to the full generative power claimed to be always present by Chomsky. In my opinion, it seems reasonable that humans do not use all their cognitive computational capability for their most elementary expressions, because it would require more energy than necessary, and this contradicts the concept of optimisation emphasised in Chomsky's Minimalist Program (1995). Jackendoff & Wittenberg (2014) add that when there is no syntactic complexity, the message is facilitated by more contextual and pragmatic signals, which become more and more unnecessary when the hierarchical recursive structures come to help in conveying complex messages. A parallel of this can be seen when speakers of different languages try to communicate, and rely more on context, producing a higher level of iconic and indexical signals (my observation). In addition, this can reflect the development of language in children, which start from simple, one-word utterances, which become more complex with their growth and their increasing experience and interactions (see e.g. the Growing Trees hypothesis, Friedman *et al.* 2021).

These proposals show that the idea of "full complexity or nothing" is limited to some generative theories, although they are very influential. In contrast, there are several accounts which propose simpler, non-recursive formulations of the merge operation, which is usually defined compositionality (see § 1.3 above). An example is the concept of bare and constructional compositionality proposed by David Gil (2022). In brief, Gil gives a typological classification of different forms of compositionality, adopting a semantic perspective: the focus is not on the structures generated by the combinations, but on the elements that are combined. Although admitting that there could be more complex constructions (e.g. tripartite), Gil focuses on the combinations of two meaningful units; what changes is the nature of these units. The main division is between bare and constructional

compositionality. In bare compositionality, "the meaning of a complex expression is determined solely by the meanings of its constituents" (Gil 2022: 1). It is thus referred to the combinations of two lexical items, with no relevance of the order of the elements: the message is understood no matter which element comes first. This formulation appears similar to what Chomsky (2001) defines set merge (see § 2.1 above). Constructional compositionality refers to the combinations of a lexical element with a grammatical or configurational one. While grammatical elements (e.g. the plural *-s*) are considered material, linguistic items, configurational elements include a range of "immaterial" items – like word order, repetition, duration, intensity – that convey additional information despite not being purely linguistic (for example, the repetition of the same lexical item can add a meaning of urgency). The extension of the notion of compositionality to non-linguistic items is a very interesting perspective, because it suggests the presence of an underlying cognitive combinatorial operation which goes beyond the elements traditionally considered part of language. While the simple levels of bare and constructional compositionality are considered symmetric, in the most complex levels, which he defines relational compositionality, the relationship between the elements is asymmetric, and determines further constraints in the interpretation: this is the most common type of compositionality of human language, in which, for example, the verb *ate* is not a simple combination of the concepts of *eat* and *past*, but it also codifies that the past time is connected to the specific action of eating, performed by an agent (Gil 2022). Although there are many similarities with the concept of generative merge, Gil explicitly distances himself from it for two reasons: first, the Association Operator – as he calls this compositional process – is semantic and not syntactic, and second, although being binary, it is not (necessarily) recursive. Again, it is evident how Chomsky's idea of intrinsic and inalienable recursion separates this theory from most of the others.

As we have seen, most of the attempts of "mitigating" the recursive merge come from other linguistic frameworks or even from other disciplines; but to make a clear argument against this view it is necessary to start from the generative perspective. The most interesting example in this direction comes from Luigi Rizzi, a generative scholar who investigated an alternative possibility to that of Berwick & Chomsky (2019). Rizzi (2016, 2021) proposes a typological classification of merge systems of increasing complexity, considering the generative potential and the resources needed to produce it. He starts from the consideration that a language needs at least a lexicon (a finite repertoire of meaningful units) and a

combinatorial device (which takes two units and combines them in a complex expression[16]). Following this reasoning, Rizzi identifies four types of systems:

- **0-merge** systems: absence of combination. There is no combinatorial device, and thus a language is limited to the lexicon, a repertoire of signals with associated meanings (e.g. *apple*).
- **1-merge** systems *(item – item)*: simple non-recursive combination of two meaningful units (e.g. *red* + *apple*). It expands the repertoire, but it does not go beyond this operation. This system requires a lexicon and a simple combinatorial device.
- **2-merge** systems *(item – phrase)*: combination of an item with a phrase, which is obtained by a previous combination of meaningful units. One element is a simple word, the other one is complex: it requires a lexicon, a combinatorial device and a temporary memory storage, in which the information of the previous combination is stored (e.g. *he + will meet the girl*). The system is recursive, because "it can generate a potentially unlimited set of linguistic expressions" (Rizzi 2016: 143), although the specifier always remains simple.
- **3-merge** systems *(phrase – phrase)*: combination of two phrases, two elements both obtained by previous combinations of meaningful units. It requires a lexicon, a combinatorial device and two temporary memory storages, which store the information of both previous combinations (e.g. *the boy + will meet the girl*, or *the men you told me about + know how to address the problem*). This system allows the formation of complex specifiers and can generate all the most complex recursive structures of human language. Rizzi further observes that in 3-merge "each complex entity is formed by several applications of different types of merge but, if syntactic structures are binary (Kayne 1983), no more than two temporary storage spaces are needed for each application of phrase-phrase merge" (Rizzi 2016: 143).

The most complex systems include the simplest ones, so that 3-merge also allows item-item and item-phrase combinations, 2-merge allows item-item combinations, and all three can use lexical elements without combinations at all. Rizzi (2016) observes that human language has the full power of 3-merge, because all existing languages can create these complex constructions, but it does not exclude that other animals have more rudimental forms of computation (see chapter 2). A very important aspect in this theory is the memory

[16] Although Chomsky (1995) argues that merge combines unordered pairs, Rizzi (2021) follows Kayne (2011) which argues that the pairs created by merge are ordered.

available, because, if the operation is the same for all the levels of merge, then the difference resides in the different memory capacity. According to Rizzi, the temporary memory storages could be specific, dedicated areas in the brain, so that there might be a qualitative difference from 1-merge to the more complex types: the spaces would allow the merge operation to create recursive combinations. However, in my opinion they could also be reinterpreted – partly departing from Rizzi's position – as more memory and computational power in general, with no specific dedicated areas: in this case, the difference between 1-merge and 3-merge could be considered quantitative. This perspective will be explored in detail in the chapter dedicated to the evolution of language (see § 3.6). In any case, this view of increasing complexity is in open contrast to the Chomskyan idea (Berwick & Chomsky 2016, 2019) of "no half-merge". The concept of 1-merge, in fact, is similar if not identical to the definition of compositionality (see Szabó 2020), and represents an interesting parallel with Gil's bare (and also constructional) compositionality. The two formulations seem almost complementary, with Gil's compositionality focused on the semantic perspective, and Rizzi's 1-merge focused on the syntactic perspective. Furthermore, this hierarchy of complexity might reflect language development in children, who go through different phases (one-word stage, two-word stage, etc.): in this view, the increase of complexity could depend on a gradual development of brain and memory that little by little allows more complex structures. This could also be compatible with the already cited proposal of Jackendoff & Wittenberg (2014), and it seems aligned to the idea that not all the complexity of language appears at the same time, and that not all the structures of language require the same complexity. A first attempt to verify the idea of Rizzi (2016) was made by Giuseppe Samo (2021), which analysed several corpora from a quantitative perspective, to check whether all languages have 3-merge constructions, and if there are some preferences between 2-merge and 3-merge systems. What emerged from the study is that all languages analysed have 3-merge constructions, and that there is no particular preference between 2-merge and 3-merge systems in adult corpora. However, the preliminary results of child corpora showed that children indeed preferred 2-merge constructions, leading to the possibility that – after a 0-merge (or one-word) phase and a 1-merge (or two-word) phase – the use of 2-merge might precede that of 3-merge.

Interestingly, Rizzi (2016) also discusses the possibility that more 1-merge systems could work in "cascades", producing non-recursive combinations of more than two items. An example would be the structure of the syllable, in which a nucleus merges with a coda to create a rhyme, and then the rhyme merges with an onset to obtain the whole syllable. Following this reasoning, it would be possible to propose hierarchical combinations of a

limited number of elements, without the requirement of recursion. However, it is not clear what the difference between cascade 1-merge systems and 2-merge systems would be, considering that recursion is only potentially infinite: if recursion leads to a high but limited number of combinations, then the difference with a non-recursive cascade system should be quantitative, and not qualitative (unless we postulate that the items merged in cascades are of a different type, e.g. the syllable is not composed by autonomous meaningful units).

Overall, despite the several proposals of a kind of compositionality capable to bridge the Chomskyan recursive merge (and the innatist perspective) with the idea of graduality (and possibly emergence), I believe that Rizzi's hierarchy is the most relevant one for two reasons: 1) it is grounded in the generative framework, and thus considers all the inner aspects of the theory it elaborates on; and 2) it manages to effectively separate merge and recursion, two properties which are generally taken into account by scholars from different frameworks, but they are hardly accepted when put together in an indivisible form. For this reason, the proposal of Rizzi (2016) will have a crucial role in the next sections and chapters of my dissertation, where I focus on the theories of the origin and the evolution of human language, and on the presence of language properties in nonhuman animals.

3. Theories about the origin and evolution of human language

As the reader might have noticed, in the previous sections I tried to describe the theories about the nature of language without focusing on the aspects of language evolution and its origin, issues which deserve to be treated separately in detail, being the main topic of this dissertation. However, the nature and the evolution of language are strongly intertwined, and to keep them separated has not been a simple task. Theories of language can influence the perspective about the theories of its evolution (see Jackendoff 2010), as Pleyer & Hartman (2019) observe, but the opposite process is true as well: studying the evolution of language can help us describe its nature. In this section, I outline the main theories which have been proposed to explain the origin of language: all of them are related to the theories of language that I described (and some are proposed by the same authors), and the previous sections serve as a background to understand the reasoning and the motivations of the scholars who reconstruct, with an inevitable degree of speculation, how language evolved.

When focusing on language evolution, some new dichotomies emerge, mainly related to the aspects of graduality versus non-graduality in its development. First, there is the difference of views between saltationism, which refers to the generativist idea that language emerged through a sudden and unprecedented evolutionary leap, as proposed by Berwick & Chomsky (2016), discussed in section 3.1, and gradualism, which suggest that language evolve gradually through several phases (e.g. Pinker & Bloom 1990, Tomasello 2008), an idea that includes different theories from both innatist and emergentist perspectives (§§ 3.2-3.5). A related issue concerns protolanguage, which is the hypothesis that, before human complex language existed, a more rudimental form of communication evolved in hominins, which was simpler than language (maybe linear, non-compositional, or in any case limited) but more complex than other communication systems, enough to give a concrete selective advantage to the species that utilised it. Some scholars strongly reject it (see Berwick & Chomsky 2016), while many others agree on this possibility, although with different hypotheses about its nature (lexical, musical, gestural, see e.g. Bickerton 2014, Fitch 2010, Kendon 2011). Finally, another issue concerns the contrast between discontinuity versus continuity. On one hand there are the scholars who believe that language evolved in discontinuity with the path that led other animals to develop their communication systems, possibly starting from unique traits[17] (again Berwick & Chomsky 2016). On the

[17] It has to be considered that discontinuity do not necessarily imply a saltation, although scholars that are in favour of discontinuity generally also support the idea of saltation (e.g. Berwick & Chomsky 2016). There could be discontinuity even if a simple, non-linguistic unique trait slowly develops until it becomes language, or if a non-unique trait (present, for example, in birds), gradually evolves in humans (and maybe also in birds),

other hand, there are the scholars who believe that language gradually evolved starting from traits present in a common ancestor between humans and other animals, and thus shared with other species (e.g. Boeckx 2021, Townsend *et al.* 2018). All these ideas are strongly connected with the topic of uniqueness or non-uniqueness of the core traits of language.

In this chapter I try to follow the same division between innatism and emergentism that I adopted in the previous one. However, being language evolution a field even more intricated – if possible – than that of the nature of language, in some cases I will have to adopt different distinctions. First, I present the influential and well-known description of the faculty of language proposed by Hauser *et al.* (2002), in which the authors propose a famous theoretical framework with a distinction between shared and unique traits. Then I focus on the "strong" innatist side of the theories, which is mainly referred to Chomsky's generative position, and then I describe some of the "strong" emergentist accounts which see language as a phenomenon of social interaction (as proposed by Tomasello 2008, 2014). Subsequently, I focus on the emergentist and innatist positions that I define "moderate", because they show some points of contact, and might be considered a middle ground between the two extremes; they also show many compatibilities with the evo-devo approaches and the concept of co-evolution of genes and culture, of language and the brain. Among the emergentist "moderate" accounts I include the theories that view language as a multimodal system, or as a complex adaptation, among the innatist "moderate" accounts I describe the theories which refer to the concept of "language-ready brain". After that, I summarise the descriptions with a table similar to that of section 2.3, which is enriched with the new insights and issues that emerge when language is seen on the diachronic perspective of its evolution. Finally, I explore the idea of a gradual evolution of the Chomskyan merge operation, following the insights of the already discussed classification of Rizzi (2016), and the contributions of other scholars. This work – although inevitably incomplete – should help readers to navigate the intricate scenario of the theories about the evolution of language.

3.1 The faculty of language in the broad (FLB) and narrow (FLN) sense

In the famous publication by Marc D. Hauser, Noam Chomsky, and Tecumseh Fitch (2002)[18], the authors outline "two more restricted conceptions of the faculty of language, one broader and more inclusive, the other more restricted and narrow" (Hauser *et al.* 2002:

but this evolution did not start in a common ancestor (convergent evolution). This second case could be considered a "deep homology".

[18] The content of the publication was further elaborated upon in the subsequent clarification of 2005, in response to some criticisms raised by Jackendoff & Pinker (2005).

1570). The faculty of language in the broad sense, or FLB, "includes an internal computational system (FLN, below) combined with at least two other organism-internal systems, which we call 'sensory-motor' and 'conceptual-intentional'" (Hauser *et al.* 2002: 1571). The sensory-motor system refers to the apparatus of vocal and gestural production, and to acoustic and visual perception; the conceptual-intentional system concerns the cognitive sphere, intentions, and the conceptual world: according to the authors, most or all of these elements are shared with other animals, at least with vertebrates. In FLB, in addition to these linguistic elements shared with other animals, is also included the biological capacity which allows humans – and only humans – to master language, the faculty of language in the narrow sense (FLN). However, all the elements of non-linguistic domain that favor it (such as memory) are excluded from the classification. Citing Hauser *et al.* (2002),

> FLN is the abstract linguistic computational system alone, independent of the other systems with which it interacts and interfaces. FLN is a component of FLB, and the mechanisms underlying it are some subset of those underlying FLB (Hauser *et al.* 2002: 1571).

In this view, the linguistic computation system (which constitutes the FLN) is the engine that connects and coordinates the two systems of communication and thought.

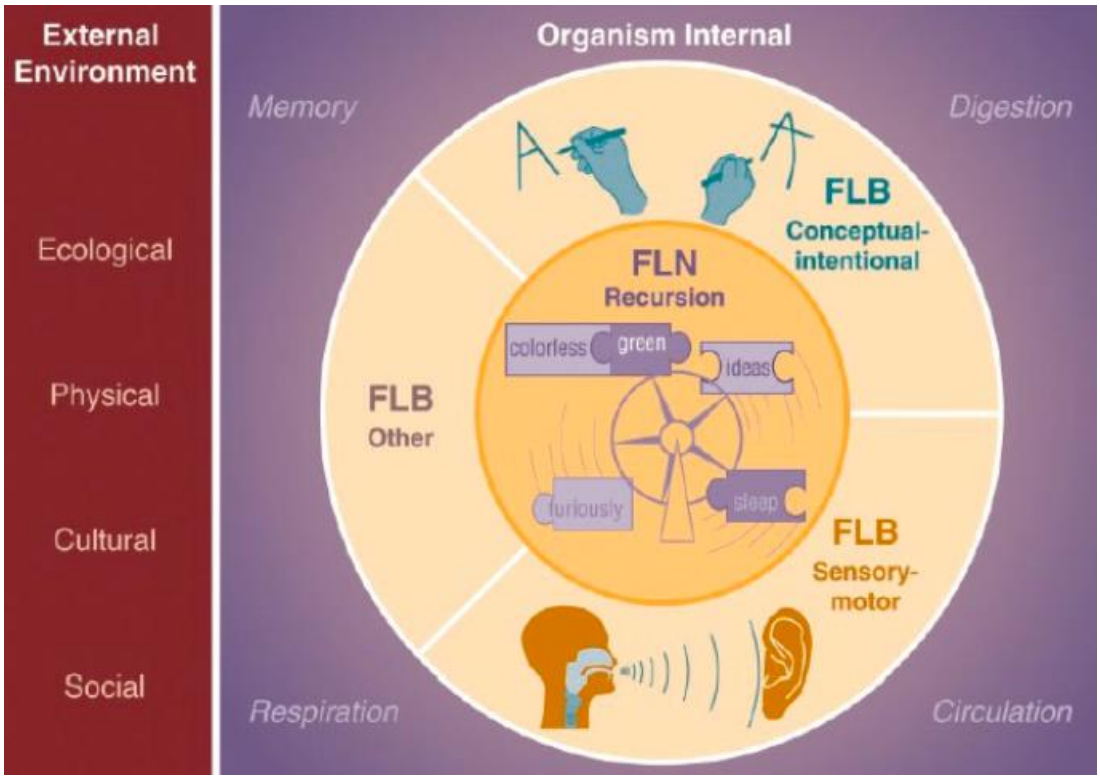

Fig. 1. Scheme that resumes the division in FLB and FLN (Hauser *et al*. 2002: 1570).

The diagram shown in Fig. 1 illustrates how FLN[19], with the property of recursion, serves as the computational core interfacing with the other systems composing FLB. The third component of FLB, "Other", has been left open to accommodate any new systems that may be identified. As can be observed, other physiological activities within the organism (memory, digestion, circulation, respiration) are external to FLB; furthermore, elements related to the ecological, physical, cultural, and social context are represented as external – albeit influential. Regarding the nature of the faculty of language, three possible hypotheses are presented:

- Hypothesis 1: "FLB is strictly homologous to animal communication" (Hauser *et al.* 2002: 1572). Every element of it is therefore shared with other animals, including FLN, whose principles may be present in other less developed or modified forms.
- Hypothesis 2: "FLB is a highly complex adaptation for language." (Hauser *et al.* 2002: 1572). Many of its elements, including FLN, have thus been shaped by natural selection for that specific function over the course of human evolution, in continuity with previous communication systems; but "Although homologous mechanisms may exist in other animals, the human versions have been modified by natural selection to the extent that they can be reasonably seen as constituting novel traits, perhaps exapted from other contexts" (Hauser *et al.* 2002: 1572).
- Hypothesis 3: "Only FLN is uniquely human." (Hauser *et al.* 2002: 1572). In this hypothesis, most of the complexity manifested by language derives from the complexity of FLB interfaces (not exclusive to humans and evolved before language), and the only recently evolved and unique element in our species, constituting FLN, is recursion, and the connections to the interfaces it governs.

This proposal is useful because it presents different hypotheses that can fit in innatist, emergentist and "intermediate" frameworks, and it represents a good theoretical starting point. However, the favourite hypothesis of the authors is number 3. In Hauser *et al.* (2002) and their subsequent publication Fitch *et al.* (2005), therefore, it is stated that there is only one novel and species-specific capacity, recursion, and all the complex syntactic, generative, hierarchical and non-linear constructions that determine our language derive from it.

[19] Wacewicz *et al*. (2020: 64-66) point out a difference between this definition and that of 2005, where indeed FLN is described as "what is uniquely human" within FLB, taking for granted adherence to Hypothesis 3 (see below) regarding its nature. As Wacewicz *et al.* (2020: 66) observe, "The 2002 paper defines FLN as a computational core and hypothesises its human uniqueness; conversely, the 2005 paper defines FLN as uniquely human and hypothesises that it includes a computational core.

Recursion indeed allows for discrete infinity[20], that is, the ability to generate infinite results by combining and recombining finite elements, which leads to an almost limitless and unparalleled expressive power in the animal kingdom. The operation that allows for combining elements is the already discussed merge (see § 2.1), but despite being a cornerstone of other works (see Berwick & Chomsky 2016), it is not mentioned in either the 2002 or 2005 papers. Infinite productivity, rather than being solely dependent on recursion, should also depend on the operation of recursive merge, that is, the repeated combination – always binary – of two elements, generating a new element from both a syntactic and conceptual perspective, which is then combined again applying the same operation to the result, and so on, with the only limit being memory capacity (Berwick & Chomsky, 2016: 17-18). In the next section I discuss how this formulation is important for the innatist perspective, especially in the most extreme position proposed by Chomsky's generative framework.

3.2 Innatism in language evolution

The reasoning of Hauser *et al*. (2002) has implications for the "strong" innatist position of the evolution of language. Reducing FLN to recursion alone responds to the needs of Chomsky's Minimalist Program (see § 2.1), which seeks to identify the most elementary principles underlying language, because in evolutionary terms the emergence of small simple modifications is much more justifiable than the appearance of complex systems (Berwick & Chomsky 2016: 73-74). Before the Minimalist Program, Chomsky had devised the Principles and Parameters model to explain the poverty of the stimulus and to justify the idea that languages have superficial differences but share their basic structures, (Chomsky 1981, see also § 2.1). In this framework, Universal Grammar – the set of innate rules in humans that allow the production of any language – contained Principles, fixed structural elements present in every language that constrained syntax construction, and Parameters, a series of unset elements that consequently varied depending on the language and that were responsible for linguistic diversity. Universal Grammar contained both Principles and Parameters, which were genetically transmitted: while the Principles were shared and manifest, the Parameters consisted of a sort of package of possible alternatives already included in the genome. During acquisition, the children would select the parameters corresponding to the language they were learning, setting them on the basis of the context to

[20] The term can be considered similar to the concept of productivity, or regular creativity, described by Berruto & Cerruti (1997), see § 1.2.

which they were exposed (Berwick & Chomsky 2016: 71-72). This idea implied that the innate and genetically transmitted component was extremely vast, complex, and articulated, because it had to include all possible alternatives of all existing or past languages. Derek Bickerton (2014: 44) notes that the Principles and Parameters model is hardly reconcilable with evolution and is incompatible with the recursive merge of the Minimalist Program. Language could be the consequence of a single cognitive operation (recursive merge) or of a complex set of linguistic rules, not both: this idea would be redundant from an evolutionary standpoint and therefore unnecessary. In Bickerton's reasoning, it is useful to conceive Universal Grammar as the result of a series of operations or algorithms that allow language to be realised like many other completely unconscious processes, such as digestion, vision, or circulation (Bickerton, 2014: 33). In the same book, Bickerton describes his idea of Universal Grammar: he accepts the existence of some fundamental operations genetically transmitted,[21] which we can trace back to the Principles, but he criticises and excludes the Parameters, arguing that although there are many unspecified elements, the possible solutions to set them are still limited. Hence linguistic variation depends on choices made from a narrow range of options, which do not need a genetic component to be performed (Bickerton, 2014: 146-184). Unfortunately, we cannot delve into Bickerton's view in detail here, but we refer to his 2014 text (especially regarding Universal Grammar, to chapters 2, 5, and 6).

Today, as Longa (2013: 8) notes, the Minimalist Program seeks to simplify, and it has added a third factor to the previously considered genetic endowment and experience: "non-specific principles of the faculty of language". According to the "Strong Minimalist Thesis", "the generative process is optimal: language principles are determined by efficient computation, and language adheres to the simplest recursive operation, intended to meet interface conditions in accordance with independent principles of efficient computation." (Berwick & Chomsky, 2016: 73). Therefore, the principles of language correspond to the most effective solution, in terms of practicality and energy expenditure, that the brain adopts for the connection between interfaces and the decoding of information. This idea seems to point to general principles, instead of linguistic ones:

> If the thesis [the Minimalist Program] were proven to be right, FL would have a minimal structure, the simplest one, and many of its mechanisms would derive from conditions

[21] Bickerton's Universal Grammar operations are Attach (related to asymmetric merge), Close (associated with the inevitable need to close phrases), and algorithms that generate noun phrases and verb phrases. Apart from Attach, the others still seem to depend on fundamental conditions of necessity, somewhat like the solutions the author proposes for unspecified elements.

> of efficient computation which 'come for free' (i.e. are not pre- specified by the genes) [...] This leads to the elimination of the purely linguistic principles of Universal Grammar (Longa, 2013: 8).

In the following section I describe the account of the evolution of language which stems from the considerations of the Minimalist Program.

3.2.1 Strong innatism: language as an evolutionary leap

In subsequent works following Hauser *et al.* (2002) and Fitch *et al*. (2005), Chomsky revised his idea about the origin of the faculty of language, making it more coherent with the Minimalist Program, and he described it in detail in an essay dedicated to the subject (Berwick & Chomsky 2016). Three dichotomies regarding the evolution of language were already highlighted in Hauser *et al.* (2002: 1572: 1): unique phenomenon in the human genus or shared with other animals; 2) evolved gradually or in a leap; 3) developed in continuity or discontinuity with previous communicative systems. In the 2016 essay, Berwick & Chomsky take a clear stance, outlining a framework that I briefly summarise. According to Berwick & Chomsky (2016: 68-70), the faculty of language would have emerged alongside the *Homo sapiens* species following a single mutation with macroscopic effects that led to new connections in the brain, closing a sort of circuit whose parts had evolved over time for other reasons, not related to linguistic necessity. This brain reconnection would have resulted in the emergence of recursive merge, an operation absolutely unprecedented and unparalleled in the animal kingdom, and therefore species-specific and linguistically domain-specific (Berwick & Chomsky 2016: 88). Before then, there was not any form of language nor protolanguage (and neither could exist), since there was no operation capable of generating it (Berwick & Chomsky 2016: 73). The mutation would have been a completely exceptional event, an evolutionary big bang comparable to the appearance of eukaryotic cells, or to the birth of life itself: one of the great leaps forward that occasionally occur because evolution, in addition to the gradual levelling of natural selection, is also determined by stochastic processes that can lead to saltation in the evolutionary path, due to luck, chance, and a confluence of factors that can accelerate development, or lead it in completely new and unforeseen directions (Berwick & Chomsky 2016: 33). Berwick & Chomsky (2016: 57-88) spend numerous pages describing the evo-devo approach[22],

[22] However, many scholars (see e.g. Boeckx 2021) are convinced that this interpretation of the evo-devo approach is too simplistic, because it postulates a mutation that, albeit theoretically possible, is very unlikely,

explaining that most of the genome is shared by all living beings, and in many cases, the difference in development (such as the choice between a wing or a limb) does not depend on the gene in question, which is the same, but on the regulators that "interpret" the information it contains and determine the final product. In fact, there is no specific language gene[23], unique to humans (or has not yet been identified). If there is an innate, exclusive, and genetically transmitted component, theoretically it should be found somewhere in the genetic code. In their view, unproven but still possible, the mutation could be attributed to enhancers, DNA sequences involved in genetic transcription of RNA: essentially, a "defective" transcription could have led to a small modification in the genome (without altering its structure), and this small modification would have turned out to be incredibly advantageous (Berwick & Chomsky 2016: 46-50).

Being a random mutation, the innovation initially would have involved a single individual – the first anatomically modern *Homo sapiens* – or at most a small group of relatives (Berwick & Chomsky 2016: 81). The operation and the new connections would have given a significant advantage to those who possessed them, but at first it would have been an internal one: as there was no complex form of communication to start from, language would have developed as a tool of thought (Berwick & Chomsky 2016: 76). This entailed greater intelligence, planning ability, imagination, and therefore would have endowed individuals with a general, non-linguistic benefit, enough to make their group thrive, generate offspring, and spread their fortunate genes (Berwick & Chomsky 2016: 81-82). After the random emergence of the language faculty, in fact, the modification would still have been subjected to the great filter of natural selection, and it would have managed to survive and prevail. Only much later – in a period between 200,000 and 60,000 years ago, probably gradually – language would have externalised into communication, exploiting the pre-existing components that make up the sensory-motor system.[24] Considering that humans have not biologically changed compared to their earliest ancestors, since that moment language would have developed, articulated, and differentiated through culture, but not evolved: the principles of language, genetically and structurally determined, remained the

and thus it is more rational to believe in other more gradual and statistically likely possibilities (following Occam's razor: if there is a simpler explanation, it should be considered first).

[23] Apparently, it is not the well-known FOXP2 gene, which, although it has become famous as the language gene, seems to be rather connected to externalisation (besides being present in other animals as well). Berwick & Chomsky (2016: 45) liken it to the printer in communication with the computer, and not to its central processor.

[24] Miyagawa *et al.* (2025) argue that linguistic capacity must have been present in *Homo sapiens* at least 135,000 years ago, based on genomic evidence from population divergences. Their analysis suggests that all descendant populations retained full linguistic ability, implying that language predated the earliest known human genetic split. They further propose that the emergence of language may have played a crucial role in the later adoption of symbolic and modern human behaviours around 100.000 years ago.

same as those of that hypothetical Prometheus (Berwick & Chomsky 2016: 57-58). As noted by the authors themselves, what remains unexplained in this theory is where the lexicon comes from: recursive merge, to create its recursive combinations, needs elements to combine, and without a protolanguage capable of providing the raw materials, it is difficult to hypothesise the origin of atomic word-like elements[25]. Berwick & Chomsky openly admit this:

> The atomic elements pose deep mysteries. The minimal meaning-bearing elements of human languages – wordlike, but not words – are radically different from anything known in animal communication systems. Their origin is entirely obscure, posing a very serious problem for the evolution of human cognitive capacities, language in particular. (Berwick & Chomsky, 2016: 90).

Despite the obscure points, the scenario outlined by Berwick & Chomsky takes a clear stance on the dichotomies mentioned earlier, raised by Hauser *et al*. (2002): 1) language is a unique phenomenon; 2) it evolved through a leap; 3) it is in total discontinuity with previous communicative systems, to the point that any form of protolanguage is excluded. Additionally, we can add that 4) in the debate between internal language and external language, the authors lean towards the former, believing that language emerged as a tool of thought, externalised in a second moment.

Bickerton, although he shares the idea of a Universal Grammar (albeit different, as I mentioned), strongly disagrees with Berwick & Chomsky regarding the condition of hominids before *Homo sapiens*, especially about the absence of a protolanguage. Along with him, many other scholars from various disciplines disagree with Berwick & Chomsky's view, which in my opinion appears very clear-cut and difficult to be accepted in its entirety. Now that I have described the faculty of language in the division into FLB and FLN, and I have indicated the strongest hypothesis regarding its nature and origin, I can draw an ideal divide between Berwick & Chomsky's saltation hypothesis and the proposals of the scholars who acknowledge the possible existence of a protolanguage and a gradual development (regardless of their relative position on the Universal Grammar).

[25] Bickerton (2014), for example, resolves the issue by theorising a lexical protolanguage, which would have provided the elements to combine well before the emergence of the Universal Grammar, perhaps developed from verses and onomatopoeic sounds.

3.3 Emergentism in language evolution

In this section I illustrate a view that diverges from that of Chomsky, i.e. the emergentism, which includes different approaches. In the next subsection (3.3.1) I describe the most "extreme" ideas, that I defined "strong" emergentism, which focus on the social aspect of language and totally deny the existence of any form of Universal Grammar. However, even these "strong" theories recognise that humans evolved by following some selective pressures and that this evolution led them to acquire new species-specific skills, which allowed humans to understand and produce language, although they are considered non-linguistic. Furthermore, the strongest emergentist theories show many points of contacts with other "intermediate" frameworks, which consider both innate and emergent traits. Although it is impossible to describe all the theories proposed by the many scholars, I will give a brief overview aided by the classification of Wacewicz *et al.* (2020), which identify four approaches that are being pursued in contemporary research: 1) language as a multimodal system; 2) language as a form of social interaction; 3) language as a complex adaptation; 4) language in the already predisposed brain (language-ready brain). I will not thoroughly follow their division; however, I will make reference to them frequently, and describe some of the approaches presented in their paper, integrating their description with some contributions from studies that they do not consider. I share many of their views, starting from the belief that by moving away from the overly strict categories imposed by the terminology of FLN and FLB and abandoning the most extreme positions, the various approaches do not necessarily exclude each other completely, and in certain cases even tend to overlap: all of them result in valid inquiries that focus on different aspects of the same multifaceted phenomenon; a phenomenon that perhaps can be better understood precisely thanks to the flexibility and integration of different analyses (Wacewicz *et al*. 2020: 67). In the next subsection I first focus on what they call "language as a form of social interaction", and then I describe the other perspectives in a separate section (3.4.3), dedicated to the complexity of the intermediate approaches.

3.3.1 Strong emergentism: language as a form of social interaction

Among the emergentist and usage-based approaches, some of the most influential scholars which proposed an account of the origin and evolution of language are Michael Tomasello (2008) and Stephen Levinson (2006) (see also Wacewicz *et al*. 2020 for a review). Both see language as a form of social interaction, and therefore primarily as a communication system. Tomasello's position, in particular, can be considered the furthest

from Chomsky's view, as it is irreconcilable with the idea of a Universal Grammar. As we have seen in section 2.2, according to Tomasello, language is a form of social action, a set of capacities and strategies aimed at coordinating and directing attention, but also at representing objects and events from a particular perspective (Tomasello 2008, see also Wacewicz *et al.* 2020: 78). Language is thus seen as a form of encoded communication, specific to humans, which uses conventional signs, in contrast to the uncoded communication that uses spontaneous signs like imitation or pantomime (Wacewicz *et al*. 2020: 78). In fact, there is no clear boundary between these two communication systems, so they cannot be discussed separately, and in this regard, Tomasello's view approaches the multimodal one (Wacewicz *et al*. 2020: 78, see § 3.4.1 below). In short, for Tomasello, language derives from a series of socio-cognitive capacities of general domain, specific to humans, linked to the concept of *shared intentionality*: this consists in the ability to collaborate, plan, and coordinate with other members of the species to achieve common goals, and (for this purpose) to understand the intentions of other beings, and thus share attention, experiences, thoughts, and emotions, in an awareness of community that transcends individuality. A fundamental feature of shared intentionality is motivation: for this capacity to occur, individuals must be driven by the will to cooperate, which is not a trivial element, considering that other animals apparently do not know how to move beyond an egocentric perspective[26] (see Tomasello 2008). Based on these capacities and needs, language would have taken shape through gestural communication, particularly indicating actions and pantomime. Shared intentionality would thus be found not only at the basis of language but also of all typical human social behaviours: language would be a cultural construct elaborated from general competences. Consequently, for Tomasello, the concept of Universal Grammar should be abandoned[27], since there are no rules or genetically transmitted species-specific linguistic principles, but what is unique and species-specific in humans is the set of capacities that allow complex social and communicative relationships, of which language is the crucial expression (see Tomasello, 2008). The real question for him, therefore, is: what is the origin of the socio-cognitive prerequisites from which language (but also thought) derived?

In his book of 2014, Tomasello explains his theory about the origins of both phenomena, arising from the same communicative necessity: language and thought would

[26] In my opinion, it is necessary to consider that certain animals actually seem capable of transcending the egocentric perspective, regardless of their level of cognition: there are species that live in well-organised communities, where each member contributes or even sacrifices itself for the collective (think of bees and ants, but also colonies of rats, where one individual has the role of 'taster' and is sent ahead to try the food found in a new location, to ensure it is not poisoned or toxic).

[27] Tomasello (2009) goes as far as to say that "the UG is dead".

have gradually formed through repeated interaction with conspecifics in order to coordinate themselves: this would have involved a transition from individual intentionality (a form of relatively complex but egocentric thought, present for example in non-human primates and in the last ancestor shared with them) to shared intentionality, articulated in two steps that led first to joint intentionality, linked to the need to coordinate with individual members of the group for a purpose of common survival or gathering of resources, and then to collective intentionality, connected to a larger sharing with the entire group, leading to a system of broader relationships with a sense of identity (Tomasello 2014: 5-6). These interactions would have emerged around 400,000 years ago, with *Homo heidelbergensis*, because they could not compete with other primate species and were therefore forced to organise themselves: this would have led to gestures and pantomime, at the root of language, pushed by the need to empathise with others and to understand their intentions for pursuing a common goal (Tomasello 2014: 59). In a second phase, with the expansion of the species and the division into various competing communities, a sort of *us-them* conceptual division would have developed to distinguish members of one's own group, together with a form a collective intentionality. A vocal language, more efficient in terms of energy expenditure and functionality, would have emerged from those more articulated relationships (Tomasello 2014: 82). This, in turn, would have determined an expansion of expressive capacity (and also of thought), allowing arbitrary connections between signs and referents and allowing abstract conceptualisations: these conceptualisations, once elaborated, were transmitted to subsequent generations, which did not have to invent them from scratch, so the development of complex languages (and conceptual representations of the world) from that point onwards would have depended only on cultural development (Tomasello 2014: 115). Language would thus be a purely human achievement, which would not have determined the speciation of *sapiens*: what gradually emerged during the evolution of hominids are the skills that lead to shared intentionality. Unfortunately, we cannot delve further, but we refer again to Tomasello (2008 and 2014) for more interesting perspectives.

An important criticism presented to Tomasello by generative scholars is that he does not account for the poverty of stimulus during language acquisition, and the fundamental structural similarities among languages, which have led to the idea of a Universal Grammar[28] (see Jacobs 2011). Tomasello responds to this critique by claiming that syntax is acquired

[28] In the same vein as Tomasello's work is the study by Evans and Levinson (2009): based on a typological comparison, the two authors argued that the existence of a language lacking recursion - such as, according to Everett (2005), Pirahã, spoken by some Amazonian tribes - would demonstrate that not all languages share the same structure, and thus there would be no UG, as Tomasello claims. However, the study has proven controversial both for the methods used and for the conclusions drawn, so the issue remains open.

by the child through a process of categorisation and analogy linked to the ability to identify the patterns and sequences to which they are exposed, which is not linguistic and is no different from how lexicon[29] is learned, reflecting its usage-based perspective (Tomasello 2003: 30). However, this explanation appears quite weak, because it implies that all the children of the world are exposed to inputs that have similar patterns and sequences, or that are at least based on the same linguistic universals. This does not explain where these universals come from, unless we want to postulate that all languages culturally descend from a single (qualitatively modern) human protolanguage. Furthermore, it is usually believed that lexicon is acquired through declarative/explicit memory, while grammatical structures depend on procedural/implicit memory (Ullman 2001, see also § 5.2 below). In this case, one should wonder whether – according to Tomasello – even grammatical structures are acquired through declarative memory.

Another supporter of language as a social phenomenon is Levinson, the proposer of the human interaction engine hypothesis (Levinson 2006), "a model of language evolution in which elements characteristic of modern language incrementally grew upon each other" (Wacewicz *et al.* 2020: 80). For Levinson, language is part of the typical human multimodal communication, understood as a set of socio-cognitive abilities and communicative behaviours, very similar to Tomasello's account. These various abilities would therefore have developed in different moments and ways over time. At their base, there would be so-called "'ethological elements', such as mutual gaze or turn-taking, including vocal turn-taking, which can be found in many primate clades" (Wacewicz *et al*. 2020: 81). Communication, therefore, would have developed from these elements, to adapt to new selective pressures, in clear continuity with other primates.

Finally, an important model which is gaining particular success is the one that sees language through the lens of Iterative Learning, which is fundamentally the process whereby an individual learns a behaviour by observing and imitating someone, who in turn learned it by observing and imitating someone else[30] (Kirby *et al.* 2015). This kind of chain entails that certain elements are retained while others are lost in transmission: in Iterative Learning, there is therefore a sort of mutual adaptation, where on the one hand human speakers adapt and adjust to the characteristics of language, and on the other hand, languages themselves adapt so that they can continue to exist, to be transmitted, and to be spoken. As Wacewicz *et al.*

[29] Not all of these competences would be species-specific: many would instead be quite ancient and shared at least among other primates (Tomasello 2003: 4).

[30] One could observe that this approach seems somewhat circular because, while representing the learning mechanism, it says little about the origin of both language and the ability to learn and imitate. Furthermore, it does not explain how linguistic innovations occur, which cannot be attributed solely to the mere inaccuracy of imitations.

note (2020: 75), this approach can be extended to many cultural products, and in my opinion it would be applicable to numerous other systems as well. Moreover, Wacewicz *et al*. also see similarities between Iterative Learning and the Construction Grammar perspective (see § 2.2.1), which describes language as a collection of form-meaning pairs, constituting the fundamental units of this type of analysis (Wacewicz *et al*. 2020: 75).

In summary, both Tomasello and Levinson (but also Kirby *et al*.) see language as a multifaceted set of various characteristics, whose main function is communication and social action; consequently, even the evolution of language is a process that has involved various factors, many of which developed before the emergence of language itself. As noted by Wacewicz *et al*. (2020: 81), "On this view, language (evolution) is neither strictly biological nor cultural, but instead characterized by an interplay of both evolutionary and cultural-historical processes". For the reasons listed, this view of language, which emphasises the social component, is compatible with both the theories of multimodal language (see § 3.4.1) and of language as a complex adaptation (see § 3.4.2), and – in my opinion – also with the idea of the existence of a biological component,[31] albeit not exclusively dedicated to language.

3.4 Intermediate approaches to language evolution

In this section I describe the approaches that do not fit completely into the categories of innatism and emergentism, because they present mixed elements which could be considered part of one framework or the other. These theories are more numerous and variegated than those that belong to the two extremes, and show many elements of convergence that I will discuss. However, some of them can be considered closer to innatism and others to emergentism, so I have decided to define them instances of "moderate" innatism and "moderate" emergentism. Some of them take into account the modern perspective of the evolutionary developmental biology and the idea of coevolution of genes and culture, which seems a promising approach to understand the process which brought to the absence of language to its full human complexity (see § 2.4). First, I describe some theories that I define "moderate" emergentism, namely those which consider language a multimodal system, and those that consider it a complex adaptive system (see also Wacewicz *et al*. 2020). Multimodal theories primarily emphasise the role of social interaction in the evolution of language. Their view of language as a combination of both gesture and speech

[31] In this case, the biological component should be intended as a set of species-specific features which allow humans – and only humans – to produce language and other complex phenomena.

offers a complementary perspective to other emergentist and innatist approaches, that enriches – and challenges – the more limited view of language as a purely vocal phenomenon. This perspective should be taken into account across all theoretical frameworks. The theories of language as a complex adaptive system propose that humans achieved some sort of unique biological endowment related to language, which could be defined a form of Universal Grammar (although not as complex as Chomsky's). Nonetheless, these theories claim that this is the consequence of a gradual development rooted in natural selection, so that language is first of all a product of the evolution determined by social dynamics and interactions. Second, I describe the theories of "moderate" innatism, which focus on the concept of language-ready brain (see also Wacewicz *et al*. 2020): in this case there is not a Universal Grammar, but a brain which is biologically predisposed to understand and produce language. Like in "moderate" emergentist accounts, this peculiar brain has been developed gradually through evolution, but the main focus here is on the inner cognitive processes which govern it. In any case, because of their minor differences, the two labels of moderate innatism and emergentism could be even removed, and these approaches could be considered under the same category (although, for clarity of exposition, I prefer to maintain them divided). To conclude, I add a table in which I compare the main dichotomies that emerge in language evolution (discontinuity vs continuity, saltation vs gradual evolution, etc.), and the approaches that present "mixed" traits and cannot fall under the categories of emergentism and innatism.

3.4.1 "Moderate" emergentism: language as a multimodal system

A perspective that should always be taken into account is the idea of language as a multimodal phenomenon, which simultaneously utilises different channels and cannot be reduced solely to oral communication.

> Since gesture and occasionally other forms of non-verbal communication are inseparable from (spoken) language, it follows that even though the nonverbal component and the verbal component may be analytically distinct, their evolutionary origins constitute an indivisible explanatory target. In other words, at least for the purposes of explaining its evolutionary origins, gesture must be considered as an integral part of language (Wacewicz *et al*. 2020: 69).

Among the various supporters of this position, Adam Kendon, David McNeill, and Jordan Zlatev are mentioned, whose views differ on some aspects. In Kendon's theory

(2011), language, rather than being an abstract set of rules, is seen as a dynamic and usage-oriented phenomenon: according to him, language involves "the mobilization of several different semiotic systems in different modalities and deployed in an orchestrated relationship with one another" (Kendon, 2014, cited in Wacewicz *et al.* 2020: 69). In my opinion, his argument in favour of multimodal language is based on facts that are difficult to dispute:

> Whenever a person speaks he employs, in a completely integrated fashion, patterns of voicing and intonation, pausings and rhythmicities, which are manifested not only audibly, but kinesically as well. Always there are movements of the eyes, the eyelids, the eyebrows, the brows, as well as the mouth, and patterns of action by the head. In addition there are, from time to time, variously conspicuous hand and forearm actions or 'gestures' (as they are often called), as well as postural and orientational changes. All of this is produced in full orchestration with speech - complex and variable, to be sure, but always orchestrated - and must be seen as inseparable components of the utterance as the utterer produces it (Kendon, 2011: 102).

Among Kendon's various arguments, his observation stands out: "Should we not be impressed by the fact that verbal language, when produced by speakers, is never produced as only words?" (Kendon 2011: 102). Consequently, Kendon (2011: 103) suggests that "the 'natural' state of spoken language is a speech-kinesis ensemble", and that "presumably, this has always been the case". The fundamental units of this ensemble, which he also defines as languaging, are termed *utterances*, which entail the combination of speech and visible (non-oral) action, aimed at translating ideas into "observable behavior, which may be read by others as reportive of those ideas" (Kendon 1980: 208, cited in Wacewicz *et al*. 2020: 70). The reason why this perspective has long been neglected, according to Kendon, is the development and dominance of writing, which ended up imposing a dominant idea of purely verbal (or written) language, excluding gestures and other aspects of expressions that were no longer considered integral parts of language (Kendon 2011: 103). Regarding its origin, he also attributes it to the emergence of multimodal signs, without a predominance of gestures or vocalisations: these signs would have developed in hominins from concrete actions, specifically chewing and handling food, which according to MacNeilage (2008, cited in Wacewicz *et al*. 2020: 70) were controlled by a single executive system, later

dedicated (or we could say exapted[32]) to communicative action (Kendon 2014, cited in Wacewicz *et al*. 2020). According to Kendon, there is continuity with certain principles present in primates, but not in their communication systems. Before the emergence of language, articulated vocalisation would have developed in hominins to manage and maintain complex social relationships,[33] while gestures would have emerged from the practical function of manipulating the surrounding reality through hands and arms (Wacewicz *et al*. 2020: 70).

On the other hand, McNeill (2012) believes that "speech and gesture synergistically express the same overall meanings while remaining semiotically distinct and responsible for the transmission of different aspects of the message: speech for propositional content and gestures for imagistic content" (Wacewicz *et al*. 2020: 71). In his view, therefore, the "peak" of a gesture (its most pronounced part) accompanies the most relevant element of the expression, allowing for the externalisation of basic units of thought, which he defines as Growth Points. The difference with Kendon lies precisely in the division of expressive tasks between gestures and vocalisation; otherwise, they both share the idea that language is an indivisible union of the two elements, and that it has been so since its origins. Another difference is in the view of gestures, which McNeill limits to arm movements accompanying speech, to which they are integrated, whereas Kendon, as we have seen, extends the gestural component to every communicative movement of the body (Wacewicz *et al*. 2020: 71).

For Zlatev (2014), whom I mention briefly, language and gestures are two distinct semiotic systems, composed of different characteristics.[34] Language can use various modalities (vocal in speech, material in writing, bodily in sign languages), and therefore face-to-face communication is always both polysemiotic, as it uses different semiotic systems, and multimodal, as it uses different communication channels (Wacewicz *et al*. 2020: 72). In his view, language is cognitively distinct from the gestural system, but in actual communication, it is inseparable. The key concept in Zlatev's analysis is mimesis, a form of imitation that translates into voluntary communicative or cognitive acts, for which I refer to Wacewicz *et al*. (2020: 72-73) or Zlatev himself (2014).[35]

As noted by Wacewicz *et al*. (2020: 74), the three authors described consider language to be species-specific, but hypothesise continuity with the cognitive and communicative

[32] The term exaptation refers to the phenomenon whereby a trait evolved for a specific function is used for a new or additional purpose, different from the one for which it was originally developed, leading to a new adaptation.
[33] Note the resemblance to the idea that language is the result of social interaction.
[34] For instance, linguistic signs are mainly conventional, gestural signs mainly iconic; the syntagmatic relations between linguistic signs are compositional, between gestural signs, they are linear; language uses double articulation, gesture does not." (Wacewicz et al., 2020: 72).
[35] For his model of language evolution, the mimesis hierarchy, see also Zlatev (2008).

systems of primates, albeit with different nuances. They also share the idea of a watershed of semiotic nature in the evolution of language, connected to the emergence of iconic gestures. Regarding the division between biology and culture, it is emphasised that

> In multimodal approaches, the bulk of biological preadaptations for language, mainly related to the organisation of the neural infrastructure, happened prior to the semiotic breakthrough and facilitated it. The later course of language evolution was almost exclusively the domain of cultural evolution, which led to the emergence of arbitrary symbols and grammar (Wacewicz *et al.* 2020:74).

Regarding the origin and development of these biological preadaptations (but also of subsequent stages, including possible protolanguages and modern 'complete' language), "the multimodal approaches emphasise the importance of cognitive and social factors, for example, the development of complex forms of Theory of Mind, intentionality or cooperation, and treat modern language as integrated in the human socio-cognitive niche." (Wacewicz *et al*. 2020: 74). Wacewicz *et al*. thus conclude that "from this perspective, they are highly compatible with theories that treat language both as a complex adaptive system and as a form of social interaction." (Wacewicz *et al*. 2020: 74).

3.4.2 "Moderate" emergentism: language as a complex adaptive system

A mixed approach that can be considered in part emergentist but also admits the presence of some innate elements (reached through a gradual process), is the one that sees language as a complex adaptation. This idea, which can integrate the perspective of social interaction (§ 3.3.1) and a multimodal system (§ 3.4.1), sees language as a complex adaptation, shaped by natural selection and developed in function of communication, or at least in response to specific selective pressures, "that emerge from social interaction across the timescales of biological evolution, cultural evolution, and ontogenetic development" (Wacewicz *et al.* 2020: 74). Here, language is considered a dynamic set of many interacting factors, of cognitive, physiological, and cultural nature, and its evolution therefore depends on a multifaceted series of circumstances related to various contexts, which developed at different periods (Wacewicz *et al*. 2020: 74). Although the elements that compose language may be shared with animals, in whole or in part, their combination and/or their level of development lead to a result that can be considered qualitatively different from any other communication system, in line with the Hypothesis 2 of Hauser *et al.* (2002) (see § 3.1).

A famous and influential view of this kind, although now outdated, is that of Steven Pinker and Paul Bloom (1990), who seek to demonstrate how language can be entirely ascribed to the laws of natural selection, showing that its development is perfectly coherent with communicative needs. According to them, language would have gradually specialised for this purpose, in continuity with previous systems, without the need to hypothesise a discontinuity that goes beyond the 'rules' of nature: the need to coordinate in the social context and the strong competition among hominids of various species and at various levels of development[36] would have led to an evolutionary 'arms race' that selected the individuals and the groups which were more skilled in communication, more intelligent, capable of collaborating but also of deceiving, and avoiding being deceived (Pinker & Bloom 1990: 753-754). Although the scholars admit the presence of a biological component, perhaps specifically linguistic, they are very clear in asserting that language developed in continuity with the original communicative need that generated it, and therefore is not 'more than nature needs', as Bickerton (2014) would say. Pinker and Bloom's paper presents many observations and arguments that are still valid and interesting today, making its reading stimulating even thirty years after its publication. What is missing is a specific triggering need, beyond the generic need to coordinate, that can justify the entire process.

Another very influential account is that of Bickerton (2014). According to him, between the initial phase of total absence of human language, and the current phase, there would have been a transitional phase in which hominids used a more rudimentary and qualitatively different communication system, or protolanguage, from which the language faculty descends only indirectly. This protolanguage would have arisen following selective pressures and adaptations during the construction of a new ecological niche, represented by the need of hominins to coordinate to gather resources.[37] Bickerton identifies this specific need in the necessity of recovering and plundering megafauna carcasses, dating it back to about two million years ago.[38] In his reconstruction, hominins wandering the savanna in search of food occasionally came across large and valuable carcasses, which a single individual would never have been able to transport or defend from other predators. Since they represented an immense source of nutrition, in these individuals – always struggling against hunger – arose the precise need to involve other conspecifics in their recovery, for a benefit that was primarily individual and selfish, than collective (Bickerton 2014: 86-88).

[36] The authors trace the origin of these dynamics back to at least 2 million years ago, and they believe that many of the hominid species involved in what they define as an "arms race" have never been discovered because fossils have not been found (Pinker and Bloom, 1990: 719).
[37] In this is not different from Tomasello, see § 3.3.1.
[38] The identified period is the same of Pinker & Bloom (1990), although Bickerton distances himself from them in other aspects.

However, first of all, the hominins needed to signal the presence of the carcasses to other group members, who had no direct experience of it: sharing the same selfishness[39] of their conspecifics, they needed to be convinced before being involved. This would have led to the development of displaced reference, i.e., the ability to refer to elements external to the context (e.g., the prey that could not be seen); according to Bickerton (2014: 79), this would have activated what he defines as offline thinking, freeing their cognition from the here and now. This capacity would have gradually been refined through adaptation, passing from spontaneous gestures, such as possibly indicating or dragging, and pantomimes (the imitation of sounds or characteristics of the animal to be recovered), to the vocal production of a sort of proto-words,[40] giving rise to a protolanguage devoid of syntax and with linear order. In a sort of chain reaction, offline thinking would have gradually modified the brain structure of hominins, but their skills and their protolanguage would have remained fundamentally identical until anatomically modern humans. In this idea, Bickerton gets close to Chomsky, because he believes that the language faculty, which emerged exclusively in *sapiens*, stemmed from a mutation which caused a brain rewiring that produced something completely new, in discontinuity: the difference is that this mutation occurred in a species that already used a form of complex communication, and so it was selected because it gave an immediate advantage. The brain reorganisation would have been boosted by the complexity due to the presence of protolanguage and offline thinking, which had been reached gradually, in continuity and through natural selection, starting from latent abilities already existing in the species. Interestingly, by discussing the need of hominins to cooperate to exploit the megafauna carcasses, Bickerton provides the evolutive trigger that is lacking in Pinker & Bloom's account. However, Bickerton's idea cannot be entirely traced back to Pinker and Bloom's line of thought: although he justifies the initial phase of language emergence through purely adaptive mechanisms, he believes that qualitatively modern language emerged following a mutation, under conditions of discontinuity, so this second phase of his theory can be consider closer to the saltational approach and to the innatist idea according to which language would already be biologically predisposed in the brain.

A perspective similar to that of Bickerton, more strongly focused on the concept of coevolution of genes and culture, is that of Terrence Deacon (1997). According to this scholar, language emerged as a result of the interplay between the constraints of human brain

[39] We could say that they lacked that shared intentionality described by Tomasello (2008).
[40] With proto-words, Bickerton identifies a genesis for the fundamental elements that are combined by syntax, a crucial issue that remains unresolved within the framework of Berwick & Chomsky (2016).

and the cultural evolution of symbolic communication.[41] His theory, outlined in his book "The Symbolic Species: The Co-evolution of Language and the Brain," suggests that language arose not solely from genetic mutations or environmental pressures, but rather from the dynamic interaction between biology and culture. Deacon posits that the human brain underwent significant evolution, particularly in the development of the prefrontal cortex, which allowed for symbolic representation and abstract thought. However, he emphasises that the emergence of language cannot be attributed solely to biological changes. Instead, he highlights the importance of cultural evolution, where early humans developed symbolic systems to communicate and cooperate, leading to the gradual emergence of language.

Furthermore, Deacon proposes the concept of "absent referents" (similarly to Bickerton's displaced reference), suggesting that language involves the ability to refer to things absent from immediate perception, which is a unique aspect of human communication. He argues that this ability arose from the interplay between brain structure and cultural practices, rather than being solely genetically predetermined. Overall, Deacon's perspective offers a nuanced understanding of the origin of human language, emphasising the co-evolutionary relationship between biology and culture, combining insights from neuroscience, anthropology, and linguistics.

Another fascinating hypothesis is that of self-domestication, supported for example by Boeckx (2021) and Benítez-Burraco (2020), according to which humans "went through a process similar to that which morphed wolves into dogs" (Boeckx 2021: 32), with a reduction of aggressiveness and an increased propensity for collaboration. In this scenario, humans would have gradually become more docile, tolerant towards others, and inclined to cooperate. These elements, particularly the reduced aggressiveness compared to other primate species, would have changed the context in which humans communicated, learned, and shared knowledge, favouring the teaching and learning of language, ultimately leading to an increase in the complexity of linguistic forms. The slow process of evolution and adaptation would have nonetheless brought also changes to the brain structure, "either directly, as a consequence of domestication processes, or indirectly, via a feedback effect triggered by the new social environment and the new language forms and uses" (Wacewicz *et al*. 2020: 78). For Boeckx (2021), self-domestication is an effective way to understand the interaction between biological and cultural evolution, because it could have contributed to our "ultra social phenotype", thus boosting the evolution of language.

[41] In this central role of symbolism there is a similarity with Berwick & Chomsky, although the processes which led to this ability are completely different.

A different approach, difficult to classify, is that of Michael Corballis (2011), which relies on some traits of innatism intertwined with gradual evolution: for him, the key of human and linguistic complexity is recursion, as in Berwick & Chomsky (2016). However, he believes that recursion is a general domain skill, which evolved gradually and in continuity as an instrument of processing thought, following the dynamics of natural selection – although possibly aided by some advantageous mutations. Crucially, Corballis argues that recursive thinking predates the emergence of language and played a key role in its evolution, because language "parasitically"[42] exploited the cognitive skill of recursion. He suggests that the ability to think recursively allowed early humans to plan and coordinate their actions more effectively, unlocking fundamental features as Theory of Mind, mental time travel and mindreading, and leading to increased social cohesion and survival advantages. The evolution of recursion, and consequently language, was thus driven by the need for social coordination and cooperation. Furthermore, Corballis (2011) suggests that language initially developed through gestural communication (showing similarities with Kendon 2011 and other approaches presented in § 3.4.1), but as hominins began to walk upright and their hands became freed for other tasks, vocalisations gradually gained prominence.[43] Overall, while sharing some of Chomsky's key assumptions about uniqueness of recursion and internal perspective of language, his theory is grounded in graduality and interaction, and strongly rejects the idea of saltation. For this reason, I have cited his theory in this section, although it also shows several points of contact with the "moderate" innatist positions that follow in the next section.

In sum, the idea of language as an adaptive phenomenon aligns well with various approaches that view human evolution in general and language evolution in particular "as the result of a complex feedback loop between our biological endowment and our cultural practices, instead of as the outcome of a linear evolutionary process, with modern cognition appearing first and modern culture/behavior happening later" (Wacewicz *et al*. 2020: 77). Consequently, like in evo-devo approaches, the boundaries between the evolution of the language faculty and language change become blurred and indistinct, because "the biological processes that lead to the emergence (and further development) of language cannot be neatly separated from the cultural ones" (Wacewicz *et al.* 2020: 77).

In this regard, Wacewicz *et al.* (2020: 77) highlight Smith's observation (2018) that this process of coevolution between genes and culture could impose biological limits on

[42] This idea of language as a "parasite" is shared also by Deacon (1997).

[43] This appears counter-intuitive: in Corballis' scenario, hominins used gestures when they walked with four legs, and then, once their hands were free, they reduced gestural signalling in favour of vocalisations.

subsequent cultural developments, and this would therefore determine which systems and structures can be learned and which cannot: this argument, as they note, would have important implications for the evolution of syntax. For all the reasons described, understanding language as a complex adaptation is an approach that is compatible and integrable with numerous other theories, whether focused on multimodality, social interaction, or even a precisely linguistic biological predisposition, an idea that I examine in greater detail in the next paragraph.

3.4.3 "Moderate" innatism: language in the language-ready brain

In this section I outline the theories which advocate for the presence of a biological endowment related to language. From a certain point of view, we could consider the account of Berwick & Chomsky (2016) as part of this group of theories, because they clearly believe in the existence of a brain that is genetically predisposed for language. However, their strong innatist position presents more differences than similarities with the views reported here, which can be considered a mitigation and an evolution of the classical generative theories, and even of the view proposed in the Minimalist Program: these hypotheses do not necessarily exclude a protolanguage and are more reconcilable with other perspectives that I have already described. In fact, even Chomsky's view, with the reduction of FLN to recursive merge, has partly approached the framework of complex adaptation, at least to the extent that the principles are now much more elementary and evolutionarily plausible. Bickerton's position, on the other hand, could be considered a middle ground because, as I have explained (see § 3.4.1), despite his protolanguage is believed to have developed in response to specific evolutionary pressures, the language faculty would have emerged in discontinuity with it,[44] and would have been due to a brain reorganisation that produced novel and seemingly unique processes in the animal kingdom, in a perspective similar in this respect to Berwick & Chomsky (2016). Regarding this, Bickerton writes that "if there is any kind of innate, Universal Grammar in the human brain, it can consist only of structures and processes enshrined in the architecture of that brain" (Bickerton, 2014: 108).

While describing the language-ready brain, Wacewicz *et al.* (2020) focus in particular on Michael Arbib's approach, which, while focusing on how the language is realised at the brain level (Wacewicz *et al*. 2020: 82), embraces various perspectives such as multimodality

[44] Evolution, rather than being driven by protolanguage, would have been stimulated by the new connections that protolanguage had formed in the brain, and would therefore be due to reasons of internal optimisation, without any linguistic purpose. The outcome of language would thus be the byproduct of another process (see Bickerton, 2014).

and social interaction, and recognises the importance of both biological evolution, seen as the basis of the cognitive infrastructure necessary for protolanguage, and cultural evolution, responsible for the subsequent transition to modern language. Arbib, author of the Mirror System Hypothesis (2005) and Cognitive Neuroprimatology (2018), presents an "Evo-Devo-Socio" (Evolutionary, Development-related, Social) approach:

> What evolved (Evo) was a language-ready brain – not a brain with an innate mechanism encoding a universal grammar [...] but rather one enabling a child to acquire language (Devo), but only if raised in a milieu in which language is already present, something which, it is claimed, required tens of millennia of cultural evolution after the emergence of *Homo sapiens* (Socio) (Arbib, 2018: 7, cited in Wacewicz *et al*. 2020: 83).

Therefore, according to Arbib, the faculty of language does not correspond to a Universal Grammar, but to a predisposition to acquire language itself, which is expressed only when exposed to the linguistic context. This predisposition, which he defines as a language-ready brain, evolved due to a series of biological, social, and cultural factors, starting from gestural communication, which emerged thanks to the cognitive infrastructure that – despite not being linguistic in nature – supports language. Arbib identifies this infrastructure with the fundamental abilities of parity and imitation. Parity, in Arbib terms, refers to the ability of interlocutors to agree on the content of a message, where "what counts for the speaker (or producer) must count for the listener (or receiver)" (Arbib 2005: 108), allowing "the same signal [to] count for more or less the same meaning to both the producer and the receiver of this signal" (Wacewicz *et al*. 2020: 83). Regarding imitation, Wacewicz *et al.* acknowledge Arbib's merit in describing, like no other, the stages of its development, which involves the transition from a system of mirror neurons (see chapter 4) – which were used for simple imitation tasks such as grasping and manipulating – to the subsequent ability to recognise complex actions and imitate them, leading to pantomime, which introduced the property of productivity into communication (referred to as open-endedness, which can be traced back to the infinite productivity or regular creativity of Berruto & Cerruti 1997, see § 1.1), namely "the ability to create an open-ended set of complex messages exploiting the primates' open-ended manual dexterity" (Arbib 2012, cited in Wacewicz *et al*. 2020: 83). This "potential to flexibly introduce novel signals for novel messages" (Wacewicz *et al*. 2020: 82) laid the foundation for two other fundamental properties of language, multifunctionality and displacement (see § 1.1), as pantomime allowed one to "communicate about many semantic domains" (Wacewicz *et al*. 2020: 82) and to express meanings disconnected from the immediate context. At the origin of pantomime and imitation,

according to Arbib, are processes “first involved in non-communicative actions such as tool production, much in line with Kendon” (Wacewicz *et al*. 2020: 83): consequently, “the complex imitation of hand movements predates language learning and use and evolved as a form of social sharing of practical skills” (Wacewicz *et al.* 2020: 83).

Another perspective derived from Arbib is that of Denis Bouchard (2015), mentioned by Wacewicz *et al*. (2020: 84), which I briefly mention: according to Bouchard, the brain's predisposition to language is a consequence of the emergence of “offline systems," which can be activated not only by external factors but also by events internal to the brain, without the need to be connected to the context.[45] These systems allow for a more abstract level of representation, enabling unprecedented connections between the conceptual and perceptual-production spheres (which we can identify with Chomsky's conceptual-intentional and sensory-motor interfaces). Mirror neurons are activated in the same way both when something is seen and when something is imagined: for the brain, there is no real difference in these processes; what changes is that in the second case, the signal does not come from (or does not reach) the sensory-motor system (see Arbib 2005). Bouchard explains the emergence of these offline systems with “an increase in synaptic interactions triggered by a number of interacting developments” (Wacewicz *et al.* 2020: 84), such as a larger brain volume, capable of favouring or stimulating these processes, the globular shape of the brain, allowing for more interactions between different modules, and alleles that facilitate synaptic repairs, further increasing the module interactions. Moreover, Bouchard recognises the importance of coevolution between biology and culture: since humans during infancy depend on parental care for a longer period than other animals, during the growth phase, they are constantly exposed to numerous cultural elements, which in turn influence their ontogenetic development (Wacewicz *et al*. 2020: 84).

The last two scholars that I discuss in this section are Benítez-Burraco and Boeckx, who revisit Arbib's model of the language-ready brain, considering Bouchard's insight into the importance of the globular brain in promoting connections between modules. Their studies (Benítez-Burraco & Boeckx, 2014 and 2015, see also Boeckx 2021) focus on how the brains of hominids would have changed (becoming more globular) over the course of our evolutionary history, to support processes involved in language (Wacewicz *et al*. 2020: 84). To explore this issue, they rely on recent studies on biological differences between *sapiens* and Neanderthals (*Homo neanderthalensis*), which show that the greatest diversities are found precisely in the globular aspect of the human brain: according to Benítez-Burraco

[45] In this, it is very similar to Bickerton's (2014) offline thought, where displaced reference represents the basis for the evolution of human thought and language.

and Boeckx, this would have led to a rewiring[46] that enhanced connections between cortical and subcortical areas, allowing the 'hermetic' limits of the various zones dedicated to specific tasks to be overcome, thus allowing the combination and fusion of conceptual units belonging to different domains. This unprecedented capacity, according to Benítez-Burraco and Boeckx, can be equated with the fundamental combinatorial operation underlying the biological basis of language: in other words, Chomsky's merge. Furthermore, in their idea, the entire reorganisation would depend on elementary neurobiological mechanisms, presumably not species-specific nor within the linguistic domain. This interesting hypothesis, as noted by Wacewicz *et al*. (2020: 85), brings neuroscience closer to syntax studies, demonstrating once again that different views can be integrated, and with the progress of science and research, more and more points of contact are identified.

This idea is supported also by Longa (2013), who argues that the energy cost required to manage connections in the nervous system cannot be specified in the genome, citing studies by Christopher Cherniak on non-genomic nativism (2009): "although the connection resources are limited, the brain has finely minimized the connection costs between neurons, in such a way that optimization emerges in the development of neural connections in several levels of the nervous systems" (Longa, 2013: 13). Therefore,

> optimization derives from physical principles, and leads to self-organization of biological matter through structures that, in spite of being innate, do not lie in the genome and do not require the workings of the genes. [...] Cherniak's research suggests that binary structure robustly arises in the connections established by both dendrites and axons [...]. Therefore, binary branching in biological tree structures seems an optimal solution. This also applies to the binary branching, via *Merge*, of linguistic tree structures, which in the same way suggest a nongenomic nativism (Longa, 2013: 13).

In any case, all the approaches described in this section[47] are founded on the belief that the most important characteristic of language is "the presence of compositional, open-ended, and domain-general semantics", which is "thought to rely not on an enhanced, language-specific computational ability, but on an unbounded basic combinatorial ability capable of transcending the limits of core conceptual systems" (Wacewicz *et al.* 2020: 86). As we will see in the next sections, in my opinion, the difference with Chomsky's recursive merge (quite

[46] "Rewiring" is the same term used by Bickerton (2014). There are also many points of contact with Berwick & Chomsky (2016).
[47] With the exception of Chomsky and Bickerton, who despite the similarities continue to see an explicitly linguistic biological capacity, which we can still define as UG.

different from the Principles and Parameters model) is indeed subtle: it seems to be the same operation, both qualitatively and quantitatively, diverging only in nomenclature and domain of membership (general or language-specific). Other similarities with Chomsky include 1) the idea that language is primarily an internal tool of thought, developed in the process of organising and connecting concepts; and 2) the belief that there is an evident predominance of biology over culture: both considerations move us away from the three approaches previously described (see §§ 3.3.1, 3.4.1, 3.4.2), but only partially. The protolanguage is indeed considered more than a mere probability by all authors (except, of course, Chomsky), and the question of the nature of language and its principles – species-specific or not? - opens up a very relevant discussion, which I address in the next sections.

3.5 Theories of language evolution: a comparison

In this section I briefly compare the theories that I have outlined in this chapter. It is impossible to clearly divide them into two groups, because several aspects of innatism and emergentism are shared by different frameworks, which also tend to share the distance with the Chomskyan idea of an evolutionary leap which endowed humans of the operation merge (Berwick & Chomsky 2016). In the following chart I summarise the main dichotomies that are found in language evolution. Some of them (e.g. saltation vs gradual evolution) present no points of contact, although some scholars proposed an intermediate position even between the most irreconcilable ones (see Bickerton 2014). On the left column of innatism there are the key points of what I defined strong innatism, and all of them form the position of Berwick & Chomsky (2016) of language as an evolutionary leap. On the right column we find the traits that can be attributed to the position of Tomasello (2008) of language as a phenomenon of social interaction. All the other theories and frameworks described above (multimodal system, complex adaptation and language ready brain) have some traits of innatism and some traits of emergentism, and I discuss them below.

Dichotomies	
Innatism *Brain modifications are the key of language evolution (merge operation)*	**Emergentism** *Social interaction is the key of language evolution (shared intentionality)*
Saltation (evolutionary leap) *Merge operation is the consequence of a recent mutation that appeared "complete"*	**Gradual evolution** *The linguistic abilities emerged gradually following social selective pressures*
Discontinuity *Language evolution represents a clear cut with respect to other species*	**Continuity** *Language evolved starting from general traits shared with other species*
Qualitative difference *Language is totally different than nonhuman communicative systems*	**Quantitative difference** *Language shares some basic principles with nonhuman communicative systems*
Protolanguage: no *Language had no intermediate phase and emerged at his full potential*	**Protolanguage: yes** *Language evolved in several gradual steps of increasing complexity*
Uniqueness *The evolution of language is a unique and exceptional process, hardly repeatable*	**Non uniqueness** *The processes which led to language are complex but in line with evolution*
Borderline theories	
Social interactions and competition → innate faculty of language (gradual evolution, continuity, with different steps of protolanguage)	
Social interactions → protolanguage in continuity subsequent mutation in discontinuity → merge operation	
Gradual evolution → language-ready brain (operation merge of general domain)	
Coevolution of genes and culture (evo-devo approaches) There are no dichotomies: nature and nurture influenced each other leading to language	

Table 4. Comparison of the main dichotomies in language evolution theories.

As the table shows, these dichotomies are not easy to be overcome. For example, it is not possible to find a middle ground between the absence or presence of a protolanguage. It has to be clarified that the strong innatist position (Berwick & Chomsky 2016) is the one which presents the most adamant claims, and the most critiqued by other scholars. However, the general perspectives that see the key trait of language evolution in brain modification (innatism) or in social interactions (emergentism) can be considered two different views of the same phenomenon, and they might have proceeded together as complementary processes, as Wacewicz *et al*. (2020) suggest. The theories related to the language-ready brain (e.g. Boeckx 2021, Arbib 2005) might be considered more focused on the innatist position, and the theories related to multimodality and adaptation (e.g. Kendon 2011, Pinker & Bloom 1990) might be viewed as more grounded in social interaction, but all of them

agree that external dynamics lead to internal modifications. Apart from Berwick & Chomsky's (2016) saltational view, all the other theories converge on the idea of a gradual development. A middle ground here is Bickerton (2014), for which, as we have seen, a first gradual evolution of protolanguage in hominins pushed by social selective pressures was followed by a "rewiring" that can be considered a mutation, although much less exceptional than that proposed by Chomsky. These two are the only theories considered here that propose a form of discontinuity, while the others recognise a continuist approach. This does not imply that all the basic traits of language were present in the last common ancestor with the other primates, because there is no clear proof of that (see Part II of this dissertation): language might have started to evolve in hominins, as Bickerton (2014) suggests, and so it would be in continuity with them. On the other hand, discontinuity implies a total divergence from the path of our ancestors. This concept is connected with that of qualitative or quantitative difference: language might be a qualitative distinct phenomenon, which uses novel principles or dedicated brain areas, totally different from the communicative system of other animals, or could be the consequence of a gradual development of pre-existing traits, e.g. more brain capacity, more memory, more social interactions, etc. In this case there would be a quantitative difference: this idea is further explored in the next section. Another peculiarity of Berwick & Chomsky (2016) is that they deny the presence of any form of complex communication in hominins, thus excluding a protolanguage: for them language is a very recent conquer, and this is denied by most of the other frameworks, which propose gestural (e.g. Kendon 2011), lexical (e.g. Bickerton 2014) or musical (e.g. Fitch 2010) protolanguages as necessary steps toward human complexity. Finally, the question of uniqueness not only of language, but also of the process that led to it, is controversial: while Berwick & Chomsky (2016) are convinced of that, calling it an evolutionary big bang, Tomasello (2008) is equally convinced that each animal that will face the same selective pressures faced by our ancestors will develop language. Furthermore, a qualitatively unique trait could be also reached gradually, starting from shared traits. This possibility makes perfect sense: following the reasoning of Fitch (2010, 2017), if all the elements of language are present in other species in different forms, but they are combined in humans and only in humans in a unique way that produces a unique result, then there would be a unique phenomenon composed by several non-unique traits, without the need to have an evolutionary leap.

I already partly cited the borderline theories: first, the idea that social interactions and competition gradually led to the evolution of an innate faculty of language shaped by natural selection, in continuity and with different levels of protolanguage, can be well

represented by Pinker & Bloom (1990). Second, the theory of social interactions which led to protolanguage in continuity, and of a subsequent mutation which led to merge and language in discontinuity, is that of Bickerton (2014). Third, the framework of the language-ready brain, which involves a gradual evolution of general cognitive principles and neuronal connections which led to an operation comparable to merge, is shared, among others, by Boeckx (2021) and Arbib (2005)

Finally, most of these theories converge on the idea that language and culture more likely coevolved in a series of interacting processes, in which social interactions gradually led to a more complex communicative system, which led to a more complex brain, which in turn led to a more complex form of language, which led to a more complex culture, and so on. In this perspective, which I already discussed, cultural processes and learning can influence the ontogenesis of the individuals, and thus the dichotomy of nature versus nurture is not a dichotomy anymore, because the two elements are equally important in language evolution and basically indivisible (see Wacewicz *et al*. 2020).

In the next section I discuss the possibility that the generative operation of merge evolved gradually following similar principles, without the need to claim an evolutionary leap.

3.6 Bridging generative merge and gradual evolution

I have presented and compared numerous approaches on the nature and origin of language, starting from Chomsky and embracing the main alternative theories, showing how the points of contact between the various ideas are numerous and progressively increasing over time along with the advancements of science in its various disciplines, particularly genetics and neuroscience, which tend to bring together views once considered irreconcilable. What emerges, in my opinion, is that now all theories converge at least in seeing language as the result of a set of different characteristics (which we could identify as FLB), interacting to create the complex language product that we know. Moreover – even though the approaches have differences, and scholars focus more on some aspects or others – the idea that there are physical operations of the brain underlying language is not only shared, but above all logical.[48] If we now return to the definition by Hauser *et al*. (2002),[49] the FLN would precisely be that central computational system that 'contains' the aforementioned operations, which would coincide with recursive merge (Berwick & Chomsky 2016), which coincides, according to Benítez-Burraco & Boeckx (2015), with the

[48] Even Tomasello (2008) recognises that the brain does some kind of computations to produce language.
[49] But not to that of 2005, in which, as Wacewicz et al. (2020) notice, they take for granted the uniqueness of this system.

combinatorial operations generated by the new connections due to the globularity of the brain. Beyond the differences, the relevant question for all approaches then becomes: where do these operations come from?

The logical possibilities appear to be only four: 1) they are species-specific and of linguistic domain; 2) they are species-specific but not of linguistic domain; 3) they are not species-specific but are of linguistic domain; 4) they are not species-specific and not of linguistic domain. In the first case, which coincides with Hypothesis 3 by Hauser *et al*. (2002) (see § 3.1), also supported in Berwick & Chomsky (2016), the computations would be novel and recently emerged, thus the origin would be traced back to anatomically modern humans. In the second case, the set of operations or abilities that allow language would be exclusive to humans, but this should be attributed to general domain capacities gradually acquired throughout evolution: here we find similarities with Tomasello's view (2014), and with other emergentist approaches. In the third case, the fundamental principles would not be species-specific, but would still be of linguistic domain; consequently, they would exist – or could exist – also in other animals, which would have developed them through convergent evolution, but only in humans they would be used for language: this idea shows similarities with Fitch's hypothesis (2010). Cases 2 and 3 could fit into Hypothesis 2 by Hauser *et al*. (2002) (see § 3.1), according to which the original principles would be shared with other animals, but in the evolutionary path that led to language, they would have changed to the point of becoming something completely new, either because limited to humans, and therefore species-specific (as in case 2), or because limited to a specific brain area dedicated to language (as in case 3); a third possibility is that the principles have become exclusive in both senses, evolving in a set of unique abilities of general *and* linguistic domain (partially approaching case 1 as well). Finally, in the fourth case the fundamental operations would neither be species-specific nor of linguistic domain, and therefore would be present in other animals in a nearly unchanged manner, both in general and (potentially) linguistic domains. This would coincide with Hypothesis 1 by Hauser et al. (2002) (see § 3.1).

As I have extensively shown, Chomsky is convinced of the uniqueness of the principles of the FLN, and because of his influence, this idea is always taken into account, and often used as a starting point for further analyses. However, revisiting old and new studies, the idea that the operations constituting language (or an ancestral form of it) are also present in other domains, and in other animals, seems more than legitimate. Bickerton (2014) admits this possibility, and commenting Chomsky's position according to which “If there was nothing like language until the “syntactic innovation” appeared, are we to assume that

language emerged abruptly, in essentially its modern form" (Chomsky, cited in Bickerton, 2014: 54), he goes on to state that

> The notion that component parts of language assembled gradually over time until their mere collective presence triggered language suggests a progressive, teleological view of evolution more consistent with creationism or intelligent design than with naturalistic science" (Bickerton, 2014: 55).

Even Fitch *et al*. (2005), in response to Jackendoff & Pinker (2005), admit that "the contents of FLN are to be empirically determined, and could possibly be empty, if empirical findings showed that none of the mechanisms involved are uniquely human or unique to language, and that only the way they are integrated is specific to human language". (Fitch *et al.* 2005: 181). In fact, they encourage comparative studies in this direction, and welcome attempts to refute their Hypothesis 3, showing to be ready to reject it if necessary (Fitch *et al*. 2005: 201). Regarding the domain to which the fundamental operations belong, on the other hand, they affirm that

> If, however, one entertains the hypothesis that recursion evolved to solve other computational problems such as navigation, number quantification, or social relationships, then it is possible that other animals have such abilities, but our research efforts have been targeted at an overly narrow search space [...]. If we find evidence for recursion in animals, but in a noncommunicative domain, then we are more likely to pinpoint the mechanisms underlying this ability and the selective pressures that led to it (Hauser *et al.* 2002: 1578).

Regarding this, they present the possibility that recursion may be present in other animals, but hermetically confined to a specific domain developed for a particular function, impenetrable to other systems: what would have happened in human evolution would therefore be a passage of this specific capacity to a more general domain, somehow overcoming the boundary and allowing new connections (Hauser *et al*. 2002: 1578).

Jackendoff & Pinker (2005: 217-218), on the other hand, argue that recursion is not the only necessary element for the development of language, and describe how it is not even exclusively linguistic in domain, since in humans it is present at least in vision and music (see Part II, § 1.1); a point also mentioned by Bickerton (2014: 124), without delving into it. We have already mentioned Tomasello (2008), who believes that any animal possessing the right requirements could develop language, since according to him it is based on principles

of general domain. Also, Benítez-Burraco and Boeckx, from another perspective (see § 3.4.3), argue in detail how they believe that these brain principles are neither species-specific nor specific to language domain:

> On Benítez-Burraco and Boeckx's view, the emergence of a language-ready brain does not entail any drastic changes to the generalised primate brain, nor the evolution of entirely new neural devices. Rather, interareal cerebral communication via the synchronizing of spatially distributed oscillations is a generic strategy of the brain, specific neither to humans nor to language. Likewise, most if not all brain areas supporting language are most likely present in other species. Accordingly, the evolution of the language-ready brain essentially involved a change in the dynamic connectivity of the brain resulting from a new anatomical context (Wacewicz *et al*. 2020: 86).

On the same page is Fitch, who from a comparative perspective affirms:

> Among mammals, the brain is remarkably conservative: detailed aspects of functional neuroanatomy and neural development remain constant. The main differences among species seem to depend more on the relative sizes and connections among brain regions than the presence or absence of qualitatively novel circuits or regions. Short- and long-term memory, basic associative abilities, conditioning of various sorts, and basic navigation skills seem to be broadly similar among species, and rest upon similar neural mechanisms. Many aspects of cognition [...] are directly relevant to language acquisition. All of these capacities form a widespread "cognitive toolkit" shared by most vertebrates and at least some invertebrates (Fitch 2010: 150).

Finally, Longa (2013) provides what, in my opinion, is a very relevant insight: starting from a perspective close to that of Chomsky's, he deviates from it by hypothesising that the FLN is empty, and that what determined the emergence of our language faculty, rather than the property of recursive merge, was a development of cortical areas and an increase in memory capacity, which would have allowed for better management of the complex long-distance relationships that syntax entails, which are very energy-intensive. According to Longa

> the biological machinery of language is neither specifically linguistic nor specifically human, although language itself can still be considered a species-specific innate trait. From such a view, language would be one of the consequences of a slight modification operated on an ancestral architecture shared with vertebrates (Longa 2013: 1).

Starting from Chomsky's hierarchical classification (1956) into different types of grammars,[50] which generate various linguistic computations, Longa aligns them with three different stages of human linguistic competence (see Longa 2013: 19-21), and demonstrates how the difference between these stages lies not in the fundamental operation of recursive merge, likely present in all of them, but in the amount of memory required:

> the several Types of computational complexity defined by the Chomsky hierarchy are not to do with the computational system itself, but with the amount of memory the computational system is endowed with: "the progression up the scale of complexity is a function of the changes introduced in the memory system, with no other modification of any fundamental property of the computational system being necessary" (Balari & Lorenzo 2013, cited in Longa 2013: 21).

This view is very close to that exposed by Boeckx (2021), who advocates for a higher consideration of continuity in biolinguistics frameworks,[51] and like Longa analyses the Chomskyan hierarchy of grammars identifying it as a starting point to conceive language as a continuum of different levels of complexity, in which the most simple layers likely preceded complex language during the course of human evolution. Furthermore, Longa argues that protolanguages may have existed, and that the development of our competence may have occurred in continuity with previous systems. Its fundamental operations would not have emerged with language but would have already been present in the brain:

> "the system in charge of recursion (not just recursion as an abstract property) subserves many other tasks (not just language) and is most probably a common feature of the brain of vertebrates (not just of humans)" (Balari & Lorenzo, 2013, p. 4). The computational system [...] would be an unspecific device, used not just by language but also by any other motor or cognitive task (or, at least, by many of them) [...] From a cognitive view, that means that many animals, perhaps all the vertebrates, share an unspecific computational system, their differences deriving from the memory associated with that

[50] Chomsky (1956) elaborates a hierarchical classification of formal grammars, in which each type includes the computations of those hierarchically lower. From the simplest to the most complex, we have type 3 grammars, called regular, capable of generating regular languages; type 2 grammars, called context-free, which generate context-free languages; type 1 grammars, called context-sensitive, which generate context-sensitive languages; and finally type 0 grammars, or unrestricted grammars, which Longa excludes from the analysis because they are not found in nature. Human language belongs to type 1.

[51] He also proposed a conceptual passage from minimalism to pluralism, by considering the full complexity of language as a mosaic of different elements which cannot be reduced to a single central trait, believing that most - if not all - of the properties of language are shared with other species.

> common computational system: the more memory the system has at its disposal, the more computational power it is endowed with (Longa 2013: 31).

In this reasoning, the main difference lies in the capacity of memory: in all three types, recursive merge is infinite in potential but limited by memory; in types 2 and 3 (perhaps possessed by hominids), it would not be able to express its full potential and to produce the complex results of human language, which are reserved for type 1. This view connects with the typological classification of merge systems proposed by Rizzi (2016), discussed in section 2.5. Rizzi gives a further contribution from the generative perspective, dividing merge in three levels defined 1-merge, 2-merge and 3-merge (which should not be confused with the Chomskyan hierarchy of grammars described above). In his proposal, the difference between the merge types is not due to the presence or absence of a qualitative distinct operation: merge is always merge, comparable to the simple set merge described by Chomsky (2001) or to the compositional operation described by Szabó (2020) or Gil (2022). What merge does is to combine two meaningful units, or syntactic objects. The crucial aspect are the temporary memory storages, i.e. the places in the brain in which the information of the previous computations is stored: the lack of them grants only the simple 1-merge operation which combines two units. The presence of one storage allows 2-merge, which combines one simple element with a complex one, the latter being recursive. Finally, the presence of two storages leads to the full complexity of recursive combinations of two complex elements, which can contain a (potentially) unlimited set of previous combinations. Rizzi (2016) thus shows that recursive merge can be the consequence of a physical difference in the brain (the presence of one or two temporary memory storages), suggesting that its basic level is a cognitive operation not necessarily unique to humans and, most importantly here, non-recursive. What is recursion then? It could be considered a consequence of the memory storages, or more in general of a high memory or computational power available to humans, which would "unlock" this property, allowing the application of the same simple operation for an unlimited number of times (again, potentially unlimited), thus creating the complex hierarchical structures of human language.

This idea has some implications. First, the key element that divides language complexity into different levels is not merge, but recursion: the difference between humans and other animals might depend on this feature. If we follow Rizzi (2016) and consider the two memory workspaces as two new brain areas, then the difference between humans and other species might be qualitative. However, in my opinion they can also be considered a quantitative modification, in terms of general memory capacity. In this case we would have

a quantitative development (more memory) which leads to a qualitative difference, because 3-merge yields results so complex that they can be considered unique and unprecedented in the animal kingdom. Second, the hierarchical structures that follow could be a direct consequence of the reapplication of merge, and thus not be codified in the genes in any form: they would naturally emerge from the recursive combinations, obeying some general, and not linguistic, constraints. Third, the evolution of merge could have been gradual, from 1-merge to 3-merge, in continuity with hominins and possibly other species, especially if 1-merge is a simple cognitive operation of general domain. This could open to the hypothesis that merge is deeply rooted in the brain of most vertebrates, with very ancient origins, and might be used for other physiological functions. In this case, the question would not be anymore "*How did recursive merge emerge only in humans?*", but it would become: "*Why a shared, simple operation (non-recursive merge) reached the level of complexity observed in humans, and only in humans?*"

Furthermore, in Rizzi's (2021) theory, even the idea of an evolutionary leap might be maintained, although it would be different from that claimed by Berwick & Chomsky (2016): in this case, the sudden cognitive revolution observed in early humans could be traced back to the emergence of the temporary memory storages that allow 3-merge. Rizzi (2021) thus suggest that, if there had been a mutation, it is more probable that was something of this kind. Following this reasoning, it is possible to conceive the evolution of an ability like recursive merge without the need to postulate the emergence of a totally novel combinatorial operation, which is highly unlikely, despite not being impossible. In my opinion, the temporary memory storages do not have to be qualitatively new areas of the brain, and this augmented storage capacity might be considered as a gradual physical increase of computational power more in general. However, even if we consider the temporary memory storages as the consequence of a saltation, it does not seem too "extreme" to hypothesise that a mutation endowed humans with a dedicated memory storage, a set of connections which made the application of an already existing operation (non-recursive merge) more efficient. It would be much more demanding to propose that a mutation endowed humans with the merge operation itself, unprecedented in nature and with no clear evolutive advantage, as Berwick & Chomsky (2016) claim.

Finally, it should be considered that merge and recursion are powerful properties that need one fundamental element: words. Without a lexicon, there is nothing to combine, and thus the possible structures and meanings that can be obtained in language also depend by the size of it. Considering that lexicon is formed by combining discrete units, both on the structural and on the semantic level (see § 4.4), the large size of human repertoire of words

could be a consequence of recursive merge, which should not be seen only as an operation capable to expand the syntactic structure of the sentences, but also as the property which creates the meaningful units that are combined on the syntactic levels. This idea is supported by the approach of Distributed Morphology (DM), a theoretical framework within generative grammar that integrates morphological and syntactic structure, proposing that morphological operations take place post-syntactically. Developed by Halle & Marantz (1993), DM posits that there is no distinct morphological component separate from syntax. Lexical items are inserted after syntactic operations have applied. This model challenges traditional lexicalist approaches by arguing that complex word formation arises from the same syntactic principles governing phrase structure (hence, merge).[52]

After considering all these perspectives (both on the innatist and on the emergentist side), it seems clear that most analyses lean towards the identification of general principles, developed according to evolutionary laws, through a confluence of factors that have given rise to a set of connections, and ultimately to the complex phenomenon of language. This phenomenon might indeed rely on a simple principle like merge (or recursive merge), but it requires an enormous infrastructure for that principle to yield the results we experience every day. Therefore, in my view, among the hypotheses of Hauser *et al*. (2002), the most interesting are precisely the two that are discarded: Hypothesis 1, according to which all the principles are shared with other animals, and Hypothesis 2, according to which they evolved from common rudimental traits.

In light of what has been discussed in this dissertation until now, I believe that the most relevant question is not how or when language formed (as internal or external, due to biology or culture, with monogenesis or polygenesis, etc.), but rather how unique the principles constituting it are. In other words: is the set of mechanisms and factors that led *Homo sapiens* to develop language unique and practically unrepeatable, or could it occur, under the right circumstances and selective pressures, in other species? Are there animals that exhibit a set of characteristics that allow us to identify, in their behaviours, traits attributable to a form – perhaps very primitive – of language faculty? If not, do some animals at least show traces of elements attributable to a remote phase of hominid evolution, and therefore a 'step' in the path that led us from a language-less Last Common Ancestor (LCA) to language-endowed *Homo sapiens*?

[52] Recent developments within DM have focused on the interaction between morphosyntactic features and phonological realisation, further refining the architecture of the grammar (Embick & Noyer 2001; Harley 2014).

Naturally, if the principles and brain operations underlying language were neither species-specific nor strictly linguistic, as argued by Longa (2013), Benítez-Burraco and Boeckx (2015), and others – or if there were no particular principles, as in Tomasello's (2008) view – it would be easier to hypothesise their exaptation from another domain, and justify the "rewiring" – or reconnection, optimisation, depending on how it is described – without the need for an exceptional mutation capable of producing something entirely new. As Fitch (2010) argues, the uniqueness of the human language faculty might "simply" lie in the peculiar way in which the various non-species-specific elements are combined. And it is particularly compelling that a similar idea is proposed by Fitch, who is one of the authors of the FLB and FLN classification, along with Hauser and Chomsky, from which he deviates in this case (see Bickerton 2014: 54). Even Berwick & Chomsky's (2016) description acknowledges that most components are present in many animals, but as we know, they attribute characteristics of uniqueness to the property of recursion, considering it species-specific and linguistic in domain. It is undeniable that Chomsky's theory presupposes uniqueness as a foundational element, and this constitutes a criterion that sets this theory apart from the others presented in this chapter. In my opinion, however, the other elements of Chomsky's proposal are not in contrast with the other theories described in this dissertation. Consequently, to investigate this perspective, in Part II I assess the hypothesis that language arises from a vast complexity of interacting factors and that the principles from which it derives are neither species-specific nor strictly linguistic, and therefore shared with other animals.

Before doing this, however, it is necessary to wrap up all the data and the opinions discussed until now, to clearly identify the traits and the properties that should be searched for in nonhuman animals. I already described some linguistic features and some general traits related to language, but it is important to put order into them and create a systematic classification that could help myself and possibly other scholars in this difficult comparative task. In the following chapters I describe two fundamental types of traits present in humans: the features which are considered purely linguistic, and are thus strictly part of language (although they might also be used in other domains), and the general features that are not linguistic per se, but fundamental to sustain language, and thus should be found in any "linguistic" species, before even hypothesising the presence of a linguistic complexity. In chapter 4 I focus on human linguistic traits, in chapter 5 I focus on human general traits. In Part II, this very classification is applied to nonhuman animals.

4. The properties of language

Language is a complex phenomenon that cannot be easily described. In addition, the several frameworks I presented in chapter 2 reflect different views of language, and have different foci: some aspects are paid more attention to in some theories, and overlooked by others, and vice versa. For this reason, prior to any comparative research, it is important to summarise all these sometimes conflicting ideas, and to elaborate a classification which can take into account all their relevant insights and the properties that are at the basis (or are consequences) of language. Thus, as a first step, I identify and describe the traits that constitute language, and are considered linguistic by experts in the field of theoretical linguistics. This is essential also because sometimes the disciplines of linguistics and ethology struggle to understand each other, and the results coming from the study of nonhuman animals are not accepted by the linguists because they do not fit in their conception of language, which is deeply rooted in the analysis of complex constructions that find no parallel in the animal kingdom (at least at that level of complexity). Hence, identifying traits and definitions of linguistic properties which can be shared by both linguists and ethologists is fundamental to achieve a fruitful convergence.

As it is well-known, the classical levels of analysis are *phonetics* (i.e. physical aspects of the signs), *phonology* (i.e. function of the signs), *morphology* (i.e. formation of words), *syntax* (i.e. formation of sentences), *semantics* (i.e. meaning of word and sentences), and *pragmatics* (i.e. use of language in a real context) (Akmajian *et al*. 2010). Elaborating on this traditional division, I propose a new framework to optimise the comparative study of language, built around 5 levels of analysis: 1) signal, 2) signal learning, 3) structure, 4) meaning, 5) contextual usage. For each level, I describe features pertaining to that specific dimension, following an increasing gradient of complexity. First, since language is a communication system, it must have a series of *signals*, and considering that language needs many of them, they have to be *learned*. These two points share some commonalities with phonetics but are not limited to it. Then, the learned signals must have a *structure*, which allows the simple sound units to be arranged into more complex forms: this level relates to phonology, morphology, and syntax (the latter two are considered together as morphosyntax). Furthermore, these structured learned signals must have a *meaning*, so the form must be associated with a non-linguistic element, or a function: this is the field of semantics. Finally, these meaningful structured learned signals must be used in real life situations, and thus in a *context*. This is related to the field of pragmatics, which involves the important aspects of intentionality, i.e. the necessary will to communicate (without which

there would be no language), signal exchanges, which refer to the actual process of communication, and the pragmatic functions that language fulfils during communication (e.g. sharing information, thoughts, internal states, cooperate, etc.).

4.1 Signal

For language to exist there must be some signals. This category is related to the physical aspects of the signs (sound or gestures), and it is not debated. Language needs a communication channel, e.g. visual or auditory. In humans, the latter is the most prominent one, and thus I mainly focus on this modality (but I also consider gestures). For vocal communication there must be a ***vocal auditory channel***, which allows vocal production and perception (Hockett 1960a). Then, an inevitable property of each vocal expression is that each linguistic signal is "openly" transmitted, so that anyone in the surroundings can hear it, and locate its source. This is defined ***broadcast transmission and directional reception*** (Hockett 1960a). Furthermore, linguistic signals are transitory: they disappear after they are uttered, thus they are subjected to ***rapid fading*** (Hockett 1960a). For the same practical reasons, the speaker can hear the totality of his utterances, receiving ***total feedback*** of his/her speech (Hockett 1960a). Finally, a last property related to the physical nature of the sound (but not to gestures) is ***linearity*** (first proposed by Saussure 1916): the elements composing a message are ordered in time and space in a linear way, one following the other, and thus it is not possible to perceive all the elements of a message at the same moment. This goes for both phonemes (units that compose the words) and words.

4.2 Signal learning

Human language needs many signals; thus it requires the ability to learn new signals and to use pre-existing signals with new functions. Signal learning is a form of social learning referred to new signals of various modalities (e.g. gestural), but the focus is mainly on the vocal ones (Janik & Slater 2000). There are two types of signals: phonemes, i.e. the meaningless units which compose the words, and lexical units (morphemes and words obtained by the combination of phonemes). Both types can be learned, but with a distinction: phonemes are thought to be easily acquired only in youth during the critical period, using procedural/implicit memory, and after that time learning becomes more difficult, if not impossible; lexical units can be learned throughout life, because they are stored in declarative/explicit memory (see Ullman 2001 and § 5.2 below). Having clarified this, signal learning can be decomposed into the following sub-categories: ***Vocal contextual learning***

(only related to lexical units) is the ability to associate an existing signal "with a new context as a result of experience with the usage of signals by other individuals" (Janik & Slater 2000). It is divided into vocal comprehension learning, related to the ability to understand these new associations, and in vocal usage learning, related to the ability to actively use the pre-existing signals in new contexts (e.g. a word like *mouse*, originally referred to the animal, is then used to indicate a computer device). ***Vocal production learning*** (related to both phonemes and lexical units) is the ability to imitate and learn to produce new signals, which are not originally present in the repertoire (they can be new sounds or new combinations of sounds, e.g. foreign words). In this case "signals are modified in form as a result of experience with those of other individuals" (Janik & Slater 2000). However, regarding phonemes, this ability declines after the critical period, to the point that adults rarely acquire new phones or phonemes spontaneously, nor can they perceive them correctly (Lenneberg 1967). Nevertheless, the capacity to learn new sounds (and not just new words) remains undisputed, as even human adults can produce them, albeit with greater effort. Signal learning is connected with ***cultural transmission***, because this ability to learn is not only instinctive, but is encouraged by active teaching and learning in social contexts and interactions.[53] Learning and teaching have thus a great role in language development (Hockett 1958). This leads us to the more specific (and wider) concept of ***learnability*** (Hockett 1966), which indicates that an individual can learn to speak more than one language: not only new sounds or words, but a whole new communicative system, with new rules, grammar, expressions, etc. However, learnability is more complex than the feature described by Hockett: again, a distinction between children and adults should be made, because it is easier to learn a language in youth, during the critical period (Lenneberg 1967). A second distinction relates to the difference between spontaneous acquisition (e.g. without actively studying) and guided acquisition (e.g. through courses) (see Lust 2006 for an overview about the language acquisition debate).

4.3 Structure

The ability to learn allows the mastering of a complex language; however, learning has some limits due to memory and brain structure. To overcome these limits, instead of acquiring completely new signals, humans can enlarge their repertoire by combining pre-existing ones. This structured nature of language is evident both at the level of phonemes –

[53] It should be noted that there are cultures in which adults do not explicitly talk to infants, nor encourage them to talk. Nonetheless, even in these contexts, children proficiently acquire their language (see e.g. Lust 2006, Tomasello 2003).

because even the simplest word is composed by smaller units – and at the level of morphemes and words, which generates hierarchical structures. In this way, the learned units can assume more complex forms, which can be associated with more complex meanings, expanding the expressive potential of language. This is a consequence of a fundamental property of language: duality.

4.3.1 Duality of patterning

Duality of patterning (Hockett 1958, also called double articulation, Martinet 1955) refers to the presence of two levels of articulation in human language: a message can be decomposed in meaningless units (phonemes, first level) and in meaningful units (morphemes and words, second level). This concept is related to the "active" combinatorial ability of language: humans can combine meaningless discrete elements (phonemes) to create meaningful structures (words, morphemes) and then combine these meaningful structures to create more complex meaningful sets (sentences). This leads to two important levels of analysis: phonology and morphosyntax respectively.

4.3.2 Phonology

Phonology relates to the fundamental units that compose language, phonemes: they are not meaningful, because they do not have a semantic meaning (see next section for more details on semantics), but they are more complex than the mere physical sounds that compose a language (phones, studied by phonetics). A phoneme can include more than one sound, or phone, and yet be recognised as a specific unit (think about the different realisations of the /r/). Phonemes have clear boundaries, and they are always distinguishable from each other. The classic way to identify a phoneme are minimal pairs: if the meaning of a word changes with the substitution of a sound, then that sound is a phoneme (e.g. *bat*/*mat*: both /b/ and /m/ are phonemes) (see Graffi & Scalise 2002). Different languages can have different sets of phonemes. The main properties related to phonology are ***discreteness*** and ***combinatoriality***. Discreteness refers to the fact that language is made by discrete units clearly separated, in opposition to the scalar nature of animal calls. Although the phones have a continuous nature and can vary in their production, the phonemes are constant, distinctly categorised in the brain, and always recognisable (Hockett 1958). Thanks to our combinatorial abilities, phonemes can be combined into meaningless structures (e.g. in songs, onomatopoeias, senseless words, babbling, etc.) but also, and most importantly, into meaningful ones

(phonemes → morphemes, Chomsky & Halle 1968, e.g. /k/ /a/ /t/ → *cat*). Combinatoriality is thus related to the first level of articulation.

4.3.3 Morphosyntax

Morphosyntax refers to the meaningful units that compose language. It has two layers that can be considered together as they are supported by similar operations: **morphology** and **syntax**. Morphology is referring to the formation of words (differences between lexical and functional words, processes of word formation like affixation, composition, etc., e.g. *house, cat* → *housecat*, or *invite* → *invitation*), but always at a level superior than the combination of phonemes, i.e. that of morphemes. Syntax refers to the formation of sentences (how words are combined and disposed in the structure of phrases, e.g. *a beautiful housecat*, and sentences, e.g. *I have a beautiful housecat*) (see Graffi & Scalise 2002). As observed by Chomsky (1965), syntax is hierarchical and it is governed by grammatical rules. I will now describe the most important features related to morphosyntax.

Compositionality is related to the second level of articulation, where meaningful units are combined into more complex meaningful structures (e.g. words → compound words/sentences) (Gil 2022, Hurford 2007). A more inclusive definition of compositionality is that "the meaning of a complex expression is determined by the meanings of its constituents and the rules used to combine them" (Zuberbühler 2019: 2), thus including not only the combinatorial ability, but also some aspects about its structure. Syntax can be compositional (*semantic compositionality*: two words merge to create a new structure whose meaning equals or exceeds the sum of the parts) or non-compositional (*semantic combinatoriality*: two words merge to create a new structure with a meaning different from the parts, like in idiomatic expressions), see Engesser & Townsend (2019).

Morphosyntax has the property of ***productivity*** (Hockett 1958): the compositional ability gives humans the ability to produce and understand novel utterances – this is valid not only at the structural, but also at the semantic level (see § 4.3.4 below). In humans, productivity is so extensive that it allows the construction of a potentially infinite number of structures/meanings using finite elements. This infinite generative power is called ***open-endedness***, or *regular creativity* (Chomsky 1968, see also Berruto & Cerruti 1997).

Open-endedness is facilitated by ***recursion*** (see § 1.2), which allows to perform the same finite combinatorial operation for a potentially infinite number of times, applying that operation to its own output (Chomsky 1957, see also Berruto & Cerruti 1997). As we have seen, this operation is called merge (Chomsky 1995), and its recursive application is

considered the core operation of language in generative frameworks (see § 2.1). Despite the proposed (and debated) existence of simpler, non-recursive levels of merge (see the account of Rizzi 2016 in § 2.5), scholars agree that the most complex form of this operation gives rise to complex ***hierarchical syntactic structures*** (Chomsky 1965, see also Berruto & Cerruti 1997). The very structure of language is hierarchical: phonemes combine to create morphemes, which combine into words, which combine into phrases, and then into sentences and combinations of sentences, in a "matryoshka" series of layers each one depending on the previous one. A practical example can be subordination: in *I love the book that you gave me*, the subordinate sentence *that you gave me* hierarchically depends on *book*. However, all the elements composing this string are hierarchically related, e.g. the noun *book* depends on the determiner *the*; the phrase *the book* is the object of the verb *love* and depends on it, etc. Furthermore, there is a complex system of agreement between the elements, e.g. if the subject of the previous sentence were *she* instead of *I*, then the verb would be in third person, *she loves*. This syntactic complexity implies that the meaning of a message does not depend only on word meanings, but also on their relationship and position in the structure, defined as ***structure dependency*** (Chomsky 1965, see also Berruto & Cerruti 1997). Therefore, the position of the elements in a sentence, or ***word order***, has a role in the meaning: all the other elements remaining the same, switching the position of the two units can convey a different message, according to the underlying hierarchical structure in which the units are arranged (Chomsky 1957, see also Berruto & Cerruti 1997). For example, *Abel killed Cain* and *Cain killed Abel* are composed by the same elements and yet have opposite meanings.[54] Although the linear order of the words can reflect to some degree hierarchical relations (due the nature of the sound stream, see linearity, § 1) this is not always true. In human language, syntactic dependencies and relationships can emerge between distant elements, defined ***non-adjacent dependencies*** (e.g. in *The office of the president is big*, "big" is related to "the office", and not to "the president") (Chomsky 1957, see also Berruto & Cerruti 1997).

The properties described in this subsection (along with others that I do not discuss here) form a set of rules which determine the boundaries of a language, with regularities and restrictions that govern the combinations of words and the formation of sentences. Humans have the instinctive ability of ***rule generalisation***: they extract these rules from practical examples, learn them, and then generalise them to create new grammatical structures never heard before (Chomsky 1957, see also Berruto & Cerruti 1997). For example, children can hear the suffix *-able* from the context (e.g. *doable*, *manageable*), understand its meaning of

[54] This observation does not apply to cases in which other syntactic operations are involved (e.g. dislocation, passivisation).

"possible to be done", and apply it to new words to give them that meaning (e.g. *learnable*, *playable*), even inventing new expressions (e.g. *niceable*). The same principle applies to the rules to form plurals, when they ignore irregular forms (e.g. adding *-s* to form *dogs*, *cats*, but also *foots*).

4.4 Meaning

The structured signals obtained through phonological and morphosyntactic processes have meanings. As we saw previously, meaning is pivotal in disentangling the nature of language. The field which studies it is defined **semantics**. Each word or morpheme is associated to (at least) one meaning, from the simple words (created by combinations of phonemes) to the most complex sentences (created by combinations of words): language units are thus pairings of form and meaning (Graffi & Scalise 2002). The following are the most relevant properties related to meaning.

First, a language must have a repertoire of meaningful units shared by a group of individuals, or *words*. This repertoire is defined ***lexicon*** (Graffi & Scalise 2002), and can be composed by thousands of words, which are obtained by combining different elements (see § 4.3 above). It shall be noted that language has two classes of words: lexical and functional. While the lexical have an autonomous semantic content and can appear alone in the discourse (e.g. nouns, adjectives, verbs), functional words serve structural purposes, and do not appear in isolation (e.g. articles, prepositions, auxiliary verbs) (Aitchison 2012). Furthermore, language can attribute properties to the entities, i.e. predicating something about a state of affairs. This is achieved by human children since their two-word stage (around 18–24 months), in which they can already interpret messages predicatively, demonstrating an understanding of basic syntactic structures. For instance, when a child says *"doll broken"*, they likely mean *"the doll is broken"*, showing an early grasp of subject-predicate relationships. Research suggests that children's early word combinations reflect underlying grammatical rules rather than mere word associations (Brown 1973). This supports the idea that even at this stage, their utterances are not just strings of words but meaningful constructions.

In any case, to have a meaningful unit, it is necessary to have the ability to assign a physical signal (signifier) to an extra-linguistic element, concrete or abstract (signified). For example, the word *cat* is a series of meaningless sounds, which constitute its form, but it refers to an existing animal, a real creature of the world, which constitutes its meaning. This fundamental property, defined as ***referentiality*** (Chen 2009, see also Hockett 1960a, who

defines it *semanticity*, and Berruto & Cerruti 1997, who define it *biplanarity*), can have different nuances, depending on what kind of element is referred to the signal, and how closely this element is connected to the external context.

This leads us to the feature of ***arbitrariness***: in language (or at least in most linguistic expressions) there is no natural correspondence between words and meanings (Hockett 1958). *Cat* refers to the feline animal, but nothing in that stream of sounds is really connected to it: that's why the word changes in different languages, becoming *chat* in French*, gatto* in Italian, etc. Hockett considers arbitrariness as a whole, but it can also be conceived as a continuum that goes from expressions strictly connected with the context to "free" expressions, following Peirce (1932 [1897]) distinction in iconicity, indexicality and symbolism. The closest level of resemblance between signal and referent is defined ***iconicity***, in which there is a similarity or analogy between a signifier (sign) and a signified (meaning), like in onomatopoeias (in which the utterance imitates the sound it describes, e.g. splash) or in gestural imitations, and implies almost a lack of arbitrariness (Peirce 1932 [1897], see also Hockett 1958, Deacon 1997, Bickerton 2014). The second level is ***indexicality***, in which a sign refers directly to an element in the context, and cannot be understood without it. For example, deictics like *here, now, today, this, that* are referred to spatial and temporal coordinates; pronouns like *I* or *you*, which indexically refer to the sender or receiver: if you perceive just "*I*" you will never know who is referred to, you will understand it only if you are with the person who produces it. However, if you hear or read *King Charles*, you will know who is the individual even if he is not present) (Peirce 1932 [1897], see also Deacon 1997, Hanks 1999, Bickerton 2014). The highest level of arbitrariness is ***symbolism***, i.e. the ability to represent objects and concepts with signs which do not depend on the context (in a broad sense all words are symbols) (Peirce 1932 [1897], see also Deacon 1997, Croft & Cruse 2004, Planer 2019). It is considered related to the capacity of creating abstract representations of the world shared by groups of individuals, and the first step towards complex culture. According to Berwick & Chomsky (2016), symbolism is one of the hallmarks of our species, and is strictly connected with the human ability to produce language and to have ***mental representations*** (see also Deacon 1997). These mental representations cannot be considered a purely linguistic element, but they are strictly connected with many semantic properties: the presence of concepts in the mind and the ability to label them must be considered as a fundamental requirement for complex language (see Fitch 2010). A further consequence of arbitrariness (less evident in iconicity, more in indexicality and symbolism) is that there is no unequivocal correspondence between signifier and signified. A signifier can have more than one meaning, a signified can have

more words to describe it. For example, the term *cross* refers to an action and an object, and the words *luck* and *fortune* refer to the same concept. This property is defined as ***equivocity*** by Berruto & Cerruti (1997).

Another important effect of symbolism is the property of ***displacement*** (Hockett 1958), one of the most relevant in all language. This is the ability to refer to elements outside of the context (in space or time, i.e. the opposite of indexicality). Considering that symbolic signals are arbitrary, it is possible to signal, for example, *cat*, when a cat is not present, and receivers will likely understand the reference although they do not see the real animal. This builds on the presence of mental representations of the objects of the world, the existence of concepts, and opens to the possibilities of abstraction, imagination, planning, recalling of past events and projections of future happenings. It is believed to be the key to language by some scholars, leading to the development of protolanguage in hominins (see Bickerton 2014). A direct consequence of displacement is the ***detachment from stimuli***, that Berruto & Cerruti (1997) consider a separate property. This is the ability to recall concepts independently from the external stimulus which generates them.

A further aspect of symbolism and arbitrariness is what Hockett (1958) defines ***specialisation***, referring to the fact that there is no particular relationship between the trigger and the direct physical consequences of the messages. For example, hearing an alarm, like "*fire!*" can elicit the same reaction as seeing an actual fire (e.g. running, calling for help), yet "*fire!*" is only a sequence of sounds, and not a physical threat. Since messages are not directly connected to the reality, another consequence is the ability of lying, defined as ***prevarication*** (Hockett 1966). Messages are not always true: they can be false or meaningless (on purpose or not). This high level of abstraction also entails the property of ***reflexiveness***: language can be used to talk about language itself (Hockett 1966).

Finally, all the possibilities described lead to the concept of ***semantic omnipotence*** (Berruto & Cerruti 1997). Language has the theoretical potential to express through language any meaning, even from other codes, systems, and domains. The property of semantic omnipotence shows a connection with "infinite" productivity and open-endedness, which consider the generative potential from a syntactic perspective. In fact, productivity not only creates more and more complex syntactic structures, but also more complex meanings, associated to these structures (thus allowing the semantic omnipotence of expressing any kind of meaning). Another property which encompasses both syntax and semantics is compositionality, which – by combining syntactic elements – combines also meanings, facilitating productivity and leading to semantic omnipotence. A simple form of

compositionality which focus on the semantic aspect instead of the syntactic one is that described by Gil (2022).

4.5 Contextual usage

All the meaningful, structured signals that compose language are used in real communicative contexts, in social interactions which involve the exchange of signals and meanings with other individuals, fulfilling different functions and following communicative rules. This is covered by the field of **pragmatics**, that is the layer of language involved in its active use (like exchanging information, expressing internal states, coordinating, eliciting reactions, changing or influencing the behaviour of others), with the more general goal of having some kind of practical utility (Mey 2006). Pragmatics is thus interested in the higher level of discourse, beyond sentences, and focuses on context and action, with implicatures (messages not explicitly expressed but suggested by the context), speech acts (speech considered as an action, e.g. if I say "*I apologise*" I am not only expressing a message, but also performing the act of apologising), and also non-verbal communication. The most important aspects of pragmatics are the functions that language acquires in the communicative acts, the exchanges of signals, which follow some conversational rules, and the intention to communicate, i.e. the will and purpose of the individual. Here we discuss first the aspects of intentionality and signal exchanges, and then some aspects of the communicative functions of pragmatics in general.

4.5.1 Intentionality

First, to have language it is necessary to have the intention to communicate these meaningful signals. The property of ***intentionality*** thus refers to the purpose of sharing information and express some meanings, and it is a form of action. It is a non-instinctive, intentional usage of language directed at specific individuals (Hurford 2007, Fitch 2010, Townsend *et al*. 2016). To have the intention to communicate with other individuals, and thus somehow changing their state, it is necessary to *understand* that there are other individuals which can receive the information. The ***Theory of Mind*** relates to this understanding, more specifically to the comprehension of being an individual in a community composed by other individuals with different minds and perspectives (Premack & Woodruff 1978, Baron-Cohen 1991, Roth & Dicke 2005, Tomasello 2008, Fitch 2010). It involves the ability of identification in others and is a form of social cognition, related not only to language but also to more general aspects of the mind. After understanding that other

minds exist, the next phase is to understand the necessity (and benefits) of working with them. Hence, the further step that stems from the ToM is the ***shared intentionality***. This concept, proposed by Tomasello (see e.g. Tomasello 2008), refers to ability to cooperate with other individuals/minds to achieve a common goal.

4.5.2 Signal exchanges

All the signals – intentional or not – have to be communicated to other individuals in an external context, and this involves an active exchange of signals between a sender and a receiver. These exchanges follow some conversational rules and principles which are necessary for granting an effective communication. One of the basics of conversation is ***turn-taking***: people generally take turns in communication, because the overlapping of signals compromises the comprehension. First, somebody sends a message and somebody else receives it, then who received replies, and so on: for this reason, in a conversation, senders and receivers frequently switch roles (Drew & Heritage 2006). Another aspect of signal exchanges is that humans can give and receive the same linguistic signals: all messages that can be heard can be also produced. For example, a healthy person can say "*I am ill*", which is not referred to prevarication (lying), but to the possibility to express any kind of message, independently from the state of the individual who expresses it. This is not so obvious: in some animals, certain signals can be produced only by specific individuals – but not all – in the population (e.g. females produce some signals which males can understand but not reproduce). Hockett (1958) calls this property ***interchangeability***. A subsequent elaboration of interchangeability leads to ***parity*** (as intended by Arbib 2005), the fundamental prerequisite for which the interlocutors must agree on the general meaning of a message – e.g. *cat* must have the same meaning for both interlocutors, otherwise a communication about that would be impossible, a concept strictly related to the semantic properties of the signal mentioned above. Finally, it is presupposed that, for communication, the individuals should follow a ***cooperative principle***: there must be active cooperation between the interlocutors, to effectively transmit information. Grice (1975) identifies four maxims: *quantity* (be informative: don't say too much or too few), *quality* (be truthful), *relation* (be relevant: only pertinent information), *manner* (be clear).

4.5.3 Pragmatic functions

Another important feature of signals is that they serve to fulfil some functions, and can change according to the communicative context. First, language has always a degree of

multimodality, involving gestures, movements, proxemics which accompany speech. Pragmatics focuses also on these nonverbal signals which integrate the vocal communication (see Kendon 2004, Arbib 2011). The multimodality implies that the same message can be transmitted in different forms, e.g. orally, written, signed (in signed languages), so it has ***modality-independence*** (Berruto & Cerruti 1997). Finally, there is ***multifunctionality***, i.e. the ability to use language for different functions, which is considered a proper linguistic property by Berruto & Cerruti (1997). These functions include for example representing thought, sharing information and feelings, solving problems, interacting with others. The best description of these functions comes from Jakobson (1960, see also Berruto 1976), that identifies six of them. The *referential* function, related to the context, describes a situation, object, or mental state (e.g. *The earth is not flat*); the *poetic* function, related to the message, focuses on how the linguistic code is used (involves the choice of the stile, like poetry, slogans, e.g. *How art thou, my liege?*); the *emotive* function, related to sender, expresses the internal state of the speaker (e.g. *I am happy!*); the *conative* function, related to the receiver, specifically addresses the hearer (e.g. vocatives, *Come here!*); the *phatic* function, related to the channel, has the role to maintain, verify, or end the communicative act (e.g. *Are you listening?*); the *metalingual* function, related to the code, uses language to reflect on language itself (e.g. *How does language work?*), and it is comparable to the semantic property of reflexiveness (see *Meaning* above).

All these properties, which encompass the most relevant levels, or layers, of language, should be sufficient to describe its multifaceted complexity, and present a way to understand what this phenomenon is, and which of its composing traits can be considered "linguistic". Hence, this systematic classification can be useful as a starting point to search for characteristics related to language in nonhuman species, and to explore the issue of language uniqueness, verifying which features are present in other animals – perhaps in simpler forms – and which traits are indeed uniquely human.

Table 5 summarises the properties of language identified in this chapter.

Human linguistic traits	
Signal *Physical aspects of the sign*	
Vocal-auditory channel *(Hockett 1958)*	Rapid fading *(Hockett 1958)*
Total feedback *(Hockett 1958)*	Linearity *(Saussure 1916)*
Broadcast transmission and directional reception *(Hockett 1958)*	
Signal learning *Ability to learn new signals and use old signals in new contexts*	
Vocal contextual learning *(Janik & Slater 2000)*	Vocal production learning *(Janik & Slater 2000)*
Cultural transmission *(Hockett 1958)*	Learnability *(Hockett 1966)*
Structure *Complex structures created by combining smaller units*	
Duality of patterning (double articulation) *(Hockett 1958, Martinet 1955)*	
Phonology *Combination of meaningless units*	
Discreteness *(Hockett 1958)*	Combinatorial ability *(Chomsky & Halle 1968)*
Morphosyntax *Combination of meaningful units*	
Compositionality *(Zuberbühler 2019)*	Non-compositional syntax *(Townsend et al. 2018)*
Productivity *(Hockett 1958)*	Open-endedness *(Chomsky 1968)*
Recursion *(Chomsky 1957)*	Merge *(Chomsky 1995)*
Recursive merge *(Chomsky 2001)*	Hierarchical structuring *(Chomsky 1965)*
Structure dependency *(Chomsky 1965)*	Importance of word order *(Chomsky 1957)*
Non-adjacent dependencies *(Chomsky 1957)*	Rule generalisation *(Chomsky 1957)*

Meaning *Presence of meaning in language structures (semantics)*	
Lexicon (words) *(Graffi & Scalise 2002)*	Referentiality *(Chen 2009)*
Arbitrariness *(Hockett 1958)*	Equivocity *(Berruto & Cerruti 1997)*
Iconicity *(Peirce 1932 [1897])*	Indexicality *(Peirce 1932 [1897])*
Symbolism *(Peirce 1932 [1897])*	Mental representations *(Fitch 2010)*
Displacement *(Hockett 1958)*	Detachment from stimuli *(Berruto & Cerruti 1997)*
Specialisation *(Hockett 1958)*	Prevarication *(Hockett 1966)*
Reflexiveness *(Hockett 1966))*	Semantic omnipotence *(Berruto & Cerruti 1997)*
Context *Use of language in external communication (pragmatics)*	
Intentionality *Drive to share meanings*	
Theory of Mind *(Premack & Woodruff 1978)*	Shared intentionality *(Tomasello 2008)*
Signal exchanges *Active exchange of linguistic signals*	
Turn-taking *(Drew & Heritage 2006)*	Interchangeability *(Hockett 1958)*
Parity *(Arbib 2005)*	Cooperative principle *(Grice 1975)*
Pragmatic functions *Functions covered by language*	
Multimodality *(Kendon 2004)*	Modality-independence *(Berruto & Cerruti 1997)*
Multifunctionality *(Berruto & Cerruti 1997)*	Jakobson's functions *(Jakobson 1960)*

Table 5. Summary of the properties of language

5. The general traits related to language

In the last chapter, we have seen that many of the properties of language might derive from simpler traits, or can be a consequence of other, more general, systems. Humans and other animals share most of these general systems, and the similarities grow among mammals. Hence, after having identified the linguistic properties that make our communication system so special and complex, it is important to also find the characteristics *behind* them, without which no language could exist. Most of them have been cited in chapter 3 and are defined as crucial by some scholars, who see them as prerequisites or steppingstones towards complex language. These features range from very basic aspects of our biology to subtle aspects of our cognition: some of them are clearly well spread in the animal kingdom, others are almost unique, or are very difficult to demonstrate in other species. First, I briefly cite some fundamental biological traits, and then I focus on three important aspects: brain, cognition, and sociality. About the brain I describe some cerebral and neuronal features that are connected with our intelligence and plasticity, which allow humans to perform many complex operations; concerning cognition, I address some aspects of the mind, which have been already discussed in the previous chapter among the linguistic properties; about sociality, I discuss some traits related to our social structure and interactions, which play a crucial role in our drive to share knowledge and meanings through language.

5.1 Biological traits

Regarding general biological traits (apart from the presence of a *complex vocal auditory channel* which allows to produce and perceive complex vocalisations, discussed above), there are two fundamental characteristics that should be considered. The first one is ***longevity***. Humans are long-lived animals (although not the most long-lived in the animal kingdom), and this implies that our species needs to be adaptable, because the individuals tend to live long enough to witness changes in their environment that are quicker than their lifespans. In past times, before 30,000 years ago, life expectancy was around 30 years (Kyriazis 2020), which can still be considered a good longevity and required adaptability and plasticity. Furthermore, humans have a ***long gestation*** and produce a ***few offspring***, generally one or maximum two children at a time (Liao *et al*. 2005, see also r/K selection theory, Pianka 1970). For this reason, caring for the offspring has a fundamental role in our biological strategy of survival: differently from other species, which invest in quantity – generating numerous offspring of which only a few survive – humans, like other great

mammals, invest in quality, so the survival of every daughter and son has to be ensured with long maternal care and transmission of experience. These two elements, although not being linguistic, represent a fundamental background for language (and for other crucial general properties): a complex language requires a complex way of cultural transmission, adaptability, and plasticity, and these features, in turn, require long-lived animals that invest plenty of energies in the care of their offspring, and in preparing them for a complex, flexible life.

5.2 Brain and cognition

A crucial feature for language and for many other complex functions is, of course, the brain, from which our elaborate cognition stems. I limit to point out some of the most important features that have been connected to language and intelligence, without claiming to be exhaustive in this very large field. A first trait is the dimension of the brain. Considering the different sizes of animals, the sheer mass of the brain is not an effective indicator of intelligence, nor is the brain-body mass ratio. One of the best ways to estimate the intelligence of an animal (although not perfect), is the ***encephalisation quotient*****,** or EQ (Jerison 1973), which "represents the deviation of the regression of the brain-to-body weight ratio" (Cairó 2011: 5). In other words, the bigger the brain in respect of what should be expected by the brain-body mass ratio, the higher the EQ. The average EQ is 1: humans have the highest among animals, reaching 6.54 (Cairó 2011). However, it should be noted that this measurement effectively applies only to mammals (although there are exceptions even among them), and other techniques have been proposed (e.g. forebrain neuron count, see Herculano-Houzel 2017). Another crucial aspect is the shape of the brain. Neubauer *et al*. (2018) observe that the peculiar ***globularity*** of human brain was not present in hominins, and represents a unicum among animals. According to several scholars (e.g. Boeckx 2021), this globularity would imply an unprecedented perinatal development of the brain, and might be responsible for connecting different areas, thus facilitating the formation of a "language-ready brain", i.e. a brain capable to easily acquire language. A further indicator of complexity is the neocortex, which is a part of the brain connected with higher cerebral functions; thus, a ***large neocortex*** with a high degree of gyrification like that of humans suggests a developed intelligence (Diaz & Gleeson 2009). Very important and famous is also the ***FOXP2 gene***, which has been defined the "language gene", although this label has been debunked. Despite not being a gene that codifies language, it is believed that it is connected with its production, and it can be considered a fundamental prerequisite at least for this latter aspect (Nudel &

Newbury 2013, Enard *et al*. 2002). Berwick & Chomsky (2016) describe it as the "cable" which connects the computer (language) to the printer (language production). On the neuronal side, among the most interesting features there are the ***spindle neurons***, also called Von Economo neurons (VEN). These cortical neurons with elongated shape are considered related to complex brain functions like self-control, social sensitivity, processing of emotions and of contradictory information (Allman *et al*. 2002). They might be also related to the need of maintaining connections between cerebral areas in large brains (Allman *et al*. 2002). ***Mirror neurons*** are equally important: they are activated both when an action is performed, and when it is observed while it is performed by another individual (Rizzolatti & Craighero 2004, Gallese *et al*. 1996). Hence, these neurons might be related to imitation, empathy, imagination, and language (Théoret & Pascual-Leone 2002). Mirror neurons might also be connected with ***self-awareness***, i.e. the understanding of having a consciousness. This feature, which might seem to be obvious in humans and won't be discussed here in detail, is the basis of a complex inner perception of the individual and of interactions with others (Jabr 2012).

Memory is a key trait for several cognitive tasks and has a pivotal role in language: without it there would be no language at all (Corballis 2019). Both short-term and long-term memory are fundamental: short-term, or working memory, is required for the sequential production and understanding of language; long-term memory is crucial in its implicit form for knowing the grammatical rules, which are applied in an unconscious way, and in its explicit form as 1) semantic memory, to store the linguistic data, and 2) episodic memory, to store and recall events, opening to the existence of a conceptual world (Corballis 2019, Mayford *et al*. 2012, Ullman 2001). Memory relates to imagination and mental time travel, and allows to remember individuals and social relationships, essential for communication. As we have seen, some scholars (e.g. Rizzi 2016), believe that memory is the key of recursion, and thus leads to the full generative potential of merge and human language. Each account of language and language evolution, or research of language in animals must confront itself with this major aspect of brain and cognition.

All these features of the brain are boosted by ***cerebral plasticity***, which allows the nervous system to create new connections and reorganise its structure following experience, to fulfill new functions, repair injuries and adapt to new external or internal stimuli (Bonfanti & Charvet 2021). This plasticity facilitates learning of new languages and skills by forming and strengthening the brain connections and creating sub-modules, which in turn make these activities easier and less energy-demanding (Paradis 2004). A consequence of this is ***behavioural plasticity***: the changes do not only involve the brain, but also the behaviour,

leading to new adaptations, techniques, and more in general learning of new skills, habits, and solutions to survive and thrive in the environment (Levine *et al.* 2022). Humans are among the most adaptive animals, capable to colonise hostile habitats and develop complex cultures and inventions: this is due to the ability to flexibly change behaviours according to the context, which is reflected into language (and language change). An example of this plasticity is ***tool use***: our species is capable to manipulate the environment, to build instruments to facilitate their tasks, and to transmit this knowledge (Johnson-Frey 2003). Tool use is not linguistic per se, but has an impact on the brain of its users, inducing "functional reorganization of their visuotactile limb representations" (Johnson-Frey 2003: 201), and more importantly, it is involved in casual reasoning (Johnson-Frey 2003): use and production of tools require an understanding of a step by step process, in which an action leads to a result, and it is considered a sign of cognitive complexity. Furthermore, like language, the human ability of tool use is open-ended, because it allows to create and utilise instruments for a potentially infinite set of functions (Baron-Cohen 2020).

Finally, there are the abilities of ***pointing*** and ***pantomime*** (imitation). These forms of primitive "unconventionalized, uncoded communication", as Tomasello (2008: 59) defines them, cannot be considered purely linguistic, but they play a fundamental role in social exchanges and are considered steppingstones towards language (Tomasello 2008). A similar importance can be attributed to the concepts of *Theory of Mind* (Premack & Woodruff 1978) and *shared intentionality* (Tomasello 2008), already discussed as a prerequisite for complex language (see chapter 4 above), but essential for general cognition as well. Equally important and closely related to these features are *mental representations* and the existence of a conceptual world (discussed in § 4.4 above, see also Fitch 2010): in this case it is hard to separate the abstract concepts from their linguistic labels, and thus these traits encompass both linguistic other general domains.

5.3 Sociality

A final, fundamental aspect of human complexity is that of sociality. It can be affirmed that human culture and technology not only depends on our drive to share meanings and information, but also on our need to live together in large social groups. Humans are among the most social animals, and this characteristic is considered by some scholars one of the keys of language and complex communication: according to Freeberg *et al*. (2012), the social complexity hypothesis states that groups with more complex interactions between individuals will develop a more complex communicative system to regulate those

interactions. For reasons of space, I only briefly sketch the main features of human sociality that seem to make it a prerequisite for language.

First, humans live in a multifaceted, ***complex social structure*** with different layers and hierarchies, in which a developed cognition and communication are crucial for its maintenance. In fact, humans have different social structures, which coexist and are sometimes interconnected (e.g. family, religion, state, social classes, etc.) (Lopez & Scott 2000). These structures can have differences across populations, but the tendency to create them is observed worldwide. For the complexity of our societies, humans have been defined a superorganism, in which the individuals share a common fate, and thus they work like cells to achieve a common result which goes beyond the self (Kesebir 2012). It should be noted that, potentially, an individual can survive in isolation. However, our species shows a high ***reliance on cooperation and interaction*** with other individuals. For these features, some scholars believe that humans can be considered a eusocial species, like bees and ants, and not only a prosocial one (Wilson 1971). The criteria for eusociality are: living in common groups, dividing labour, overlapping of generations, and cooperative breeding (Chung 2016). Humans fulfil all these criteria, at least to a certain extent: we live in large, organised groups; we divide the work, so that there are countless specialisations, and nobody is able to master more than a fragment of them; we coexist with at least three different generations; and we cooperate in the care of the offspring, even when unrelated (e.g. nurseries, schools, etc.). These basic traits were also present in more ancient, pre-technological societies, in which for example some individuals took care of the offspring while others went hunting, and similar practices were likely present in hominins (Nowak *et al*. 2010). However, the main difference with eusocial species like ants and bees is that there are no biological differences between individuals of different classes: there are no humans which are born stronger to become soldiers, nor a fertile queen dedicated to reproduction. Hence the idea of humans as a superorganism or eusocial species, although convincing from a conceptual point of view, is grounded in the theory of group selection (according to which natural selection works at the level of the group, and not the genes, see Nowak *et al*. 2010), and it has been highly criticised by supporters of kin selection (according to which natural selection works at the level of the genes, and so "altruism" only works in genetically close individuals, members of the same kin, see e.g. Dawkins 2006). This critique does not imply that humans do not have the features described above, or that they are not fundamental in the complexity of our species. In particular, some scholars have proposed that ***cooperative breeding*** enhanced human cognitive and social evolution, so it is a trait that should be considered

when looking for prerequisites of language in nonhuman species (Burkart *et al*. 2009, Burkart & van Schaik 2010, 2016, but see a critique from Thornton *et al*. 2016).

It is clear that all this complexity is not only the consequence of mere interactions and division of labour, but also of shared knowledge, which has to be earned and transmitted: hence ***cultural transmission*** (see e.g. Mesoudi & Whiten 2008) and ***social learning*** (see e.g. Gariépy *et al*. 2014) are two fundamental features which can be considered important for both language and cognition, and we already discussed them in the previous chapters. These skills lead to the concept of cumulative cultural evolution (see Derex 2021), according to which humans reached the level of complexity that we witness today thanks to the gradual "sedimentation" of conquers and innovations of previous generations, so that the subsequent generations did not have to discover them again, thus leading to a civilisation (with languages, techniques, social structures, etc.) which would have never been reached without this gradual transmission of knowledge.

Finally, a fascinating hypothesis that tries to explain our high degree of sociality is that of *self-domestication* (see e.g. Boeckx 2021), according to which a reduced aggressivity and an increased tolerance among individuals gradually led to modern "ultra-social" humans (Boeckx 2021), also by isolating the more aggressive and non-cooperative individuals and thus gradually selecting the most docile ones. This hypothetical process was accompanied by biological modifications in the cranial and face structure and in some features of the brain, that could have facilitated cultural transmission, sharing of knowledge and language. It would be thus interesting to comparatively search for these features in nonhuman animals (Boeckx 2021).

Table 6 summarises the general traits related to language discussed in this chapter.

Human general traits	
Biological traits *General features related to main aspects of biology*	
Complex vocal auditory channel *(Hockett 1958)*	Longevity *(Kyriazis 2020)*
Long gestation *(Liao et al. 2005)*	Few offspring *(Liao et al. 2005)*
Brain and cognition *Structural aspects of the brain and cognitive features of the mind*	
Encephalisation quotient (high) *(Cairó 2011)*	Globularity *(Boeckx 2021)*
Large neocortex *(Diaz & Gleeson 2009)*	FOXP2 gene *(Nudel & Newbury 2013)*
Spindle neurons *(Allman et al. 2002)*	Mirror neurons *(Gallese et al. 1996)*
Self-awareness *(Jabr 2012)*	Memory *(Corballis 2019)*
Cerebral plasticity *(Bonfanti & Charvet 2021)*	Behavioural plasticity *(Levine et al. 2022)*
Tool use *(Johnson-Frey 2003)*	Pointing *(Tomasello 2008)*
Pantomime/imitation *(Tomasello 2008)*	Theory of Mind *(Premack & Woodruff 1978)*
Shared intentionality *(Tomasello 2008)*	Mental representations *(Fitch 2010)*
Sociality *Aspects related to social groups and interactions*	
Complex social structure *(Lopez & Scott 2000)*	Reliance on cooperation and interaction *(Nowak et al. 2010)*
Cooperative breeding *(Burkart et al. 2009)*	Cultural transmission *(Mesoudi & Whiten 2008)*
Social learning *(Gariépy et al. 2014)*	Self-domestication *(Boeckx 2021)*

Table 6. Summary of the general traits related to language in humans.

6. Back to the core traits: reflections and research questions

In this part, I have offered an overview of both the linguistic properties and the general traits necessary for language, which constitute the backbone of this phenomenon in our species. The identification of the several and variegated traits present in humans leads us to the questions that will be faced in the second and third part of this dissertation.

- Which traits related to language are present in nonhuman animals?

Considering that, as we will see, most of the properties are indeed found, in different degrees, in other animals, another important question that I address is whether there are some species that possess *all* the traits described in these chapters, and, if so, which species they are. This will allow me to identify the animals that are the best candidates for further study and examination: after a wider analysis of the linguistic and general traits of nonhuman animals, I concentrate on the most promising species: great apes, birds, and cetaceans. The cetaceans are the focus of the third part of my dissertation because they are the least studied due to technical limitations, but show some of the most complex communicative, cognitive and social traits in the animal kingdom. After discussing some other interesting species, I focus on the bottlenose dolphin, because they share many similarities with humans, which will be thoroughly debated. The research question will then become:

- Which traits related to language are present in bottlenose dolphins?

A second, important issue is that of uniqueness. I verify the presence of many traits, but also the absence of others. Hence, I ask: are there some characteristics that are not found, in any form, in other species, and can thus be considered unique? Can these characteristics be defined core traits of language, qualitatively distinct from the properties of animals, or are they consequences of a quantitative difference (e.g. more complexity) of the same shared traits? We have seen that, in generative frameworks, recursive merge is considered the key to human complexity, and a unique trait. We have also seen that this strong position represents the main point of dispute in the theories about language evolution, and that – for how it is formulated – impedes a fruitful reconciliation between the opposite poles discussed in chapters 2 and 3. Finally, I observed that other theories propose that merge (or other compositional operations called with a different name) can be non-recursive, non-linguistic,

and that can have a gradual development. Thus, if we want to follow the idea that merge *is* the core trait of language, the question becomes:

- Is non-recursive merge present in nonhuman animals?

After addressing this question in Part II from a general perspective, I focus on cetaceans in Part III, to ask:

- Is non-recursive merge present in dolphins?

These questions, however, are not sufficient to identify the core difference between human language and animal communication. In the previous sections we have seen that recursion might represent the real "evolutionary leap", or at least it might be the property which allows the human generative potential to be expressed. Hence, we should give a special attention to this issue, asking whether recursion is present in nonhuman animals, what is its relationship with merge, and whether this property is related to the size of lexicon (i.e., recursion → bigger lexicon):

- Is recursion uniquely human?
- Is recursion involved in the formation of a large lexicon?

We are now ready to move away from humans and tackle these issues with an interdisciplinary approach that will try to bridge linguistics, ethology, and evolutionary biology.

Part II

Introduction

In the first part of this dissertation, I focused on human language: I identified its fundamental traits and the "core" features that make this phenomenon unique, and I proposed a new classification of linguistic and general traits which summarises the many diverse elements which constitute language and thus have to be present for language to exist. This showed that, although some "core" properties can be hypothesised, this complex phenomenon can hardly be reduced to a single operation or a limited set of skills.[55] I also highlighted the open issues and the several dichotomies about its nature and its origin, showing that this is a topic highly debated in linguistics, for which there are no clear answers: the different theories tend to focus on different features, thus insisting on different approaches to the evolution and acquisition of language. Some of them lean towards the idea of innate skills, others assign a bigger role to experience. I underlined that there are also some intermediate positions, which recognise the importance of both nature and culture (e.g. evo-devo approaches), and which can be a starting point to overcome some of the differences. The same can be affirmed about proposals like that of Rizzi (2016) or Longa (2013), which suggest that even the hallmark of generative frameworks, i.e. recursive merge, can be interpreted as a gradual phenomenon, opening to new comparative perspectives. Some of the approaches that I described hypothesise that the principles constituting the faculty of language, identified in this combinatorial operation (called merge by some scholars, compositionality by others), are neither species-specific nor limited to linguistic domains, or that they evolved starting from shared traits until they became qualitatively unique (corresponding to Hypothesis 1 and 2 presented – but not supported – by Hauser *et al*. 2002). We have seen that several scholars from different frameworks agree with this possibility, according to which, at the root of language, there could be principles of general domain, that would connect different areas of the brain by relying on a complex network of systems (some of which are certainly shared with other animals), all necessary to achieve the final product we call language. In this case, the question is: what evidence can be brought to strengthen these hypotheses?

Obviously, this requires an interdisciplinary stance, and the collaboration of experts in very different fields, from anthropology to zoology, crossing linguistics, ethology, neurosciences, statistics, artificial intelligence, and many others. This kind of research takes us beyond the origin of language, and even beyond the origin of Man: the only way to verify

[55] Even from the Chomskyan perspective, a single operation (recursive merge) would give rise to multiple other properties. Furthermore, the non-linguistic traits associated to the FLB are considered necessary in any theoretical framework.

if the principles underlying language are species-specific and/or limited to linguistic domains is to search for examples of the fundamental operations – or traits attributable to them – outside the linguistic domain, and outside of humans. This is the goal of the second part of this dissertation, which is focused on the research of properties and traits related to language in nonhuman animals (which are empirically observable), adopting a comparative approach. This shows that many characteristics are indeed shared, and that the traits that have to be searched in other species are several, and not the sole "core" properties (although I give them a special emphasis). After all, language is a complex phenomenon composed by different features, and which uses different resources, and none of them should be overlooked. In this task, the classification proposed in Part I (chapters 4 and 5) will come in handy. The comparative method hence proves to be very useful, although not all scholars believe it can help us to better understand the origin of human language (see chapter 2). However, in my opinion, it remains a valuable tool both to confirm certain similarities with the animal kingdom – analogies or homologies – and to exclude them. I therefore believe that the comparative method is the only way to investigate whether there are animals which exhibit a set of characteristics and behaviours that can be considered a rudimental form of language faculty. Or, in other words, it can answer to the questions expressed in the previous part of this dissertation: which traits related to language are present in nonhuman animals? Is non-recursive merge one of these traits?

However, before delving into this topic, I point out two more possible approaches that can explore the proposed human uniqueness and domain-specificity of language. One is to search for the fundamental properties which are considered of linguistic domain (e.g. merge and recursion) in humans, but in *other*, non-linguistic, domains[56], as some have already suggested (cf. Jackendoff & Pinker 2005), and as Hauser *et al*. (2002) invite scholars to do. A second possibility is to search the properties of language in our extinct hominin ancestors and cousins, to verify whether they had a form of protolanguage or some kind of symbolic culture, which could confirm or disprove the idea of a saltation (the former if none of these skills were found, the latter if they were clearly identified). On the one hand, hominins are our closest relatives, so they are the best candidates for this kind of research. On the other hand, the fossils present several limitations to the studies. In the next chapter I describe these alternative approaches, and then I focus on the main topic of my dissertation, analysing more distant – but living – nonhuman animals for the rest of Part II.

[56] This fruitful approach can be applied also to nonhuman animals, as I hint in chapter NR. However, it might be easier to study elusive cognitive skills like recursion in humans, which can give direct answers, than in other species. Although this research does not demonstrate the non-uniqueness of a trait, disputes the idea that this trait is language specific.

1. Two approaches: non-linguistic domains and hominins

In this chapter I briefly discuss two different approaches that display interesting ways to analyse the origin and nature of language, and can be considered complementary to my research on nonhuman animals. This demonstrates once again the interdisciplinarity of this field, and the usefulness of the comparative method, which is also applied to these issues.

1.1 The research of the core language properties in human non-linguistic domains

As I mentioned in the introduction to this chapter, to understand if the fundamental operations of language are exclusively linguistic in nature, one path leads us to search for possible traces of these operations in non-linguistic domains, both in other animals and in humans.

We saw that Hauser *et al.* (2002) encourage the research of fundamental linguistic operations in other domains, alluding to the possibility that the property of recursion may be found, for example, in spatial orientation, number calculation, and social relationships. The authors reiterate this point in their 2005 paper, where they emphasise that "research on non-linguistic cognitive domains (number, navigation, social intelligence, music, and others) is fundamental to the proper eventual delineation of FLN" (Fitch *et al*. 2005: 182). Furthermore, they argue that "it seems clear that the current utility of recursive mental operations is not limited to communication", highlighting its "utility for cognitive functions like interpreting mathematical formulas that are not plausibly adaptations at all", but also in "functions such as planning, problem solving, or social cognition that might themselves be adaptations" (Fitch *et al*. 2005: 186). They thus present an effective example of how recursion might function in the non-linguistic domain of spatial reasoning and navigation:

> consider such concepts as ((((the hole) in the tree) in the glade) by the stream) and ask whether there is an obvious limit to such embedding of place concepts within place concepts (in the forest by the plain between the mountains in the north of the island.). (Fitch *et al*. 2005: 186).

However, while admitting that this competence may respond to "processes of optimal computation already established in animal navigation, like efficient path integration and no backtracking" (Fitch *et al*. 2005: 187), they think that the current use of recursion made by humans does not allow to speculate on its origin: this property might have developed in humans for language purposes and only later been utilised for these other functions. It is

important to emphasise that at present we do not have sufficient knowledge of how neural circuits work, neither in humans nor in other animals. Berwick & Chomsky (2016: 152) candidly admit that the topic of the evolution of language is inevitably speculative because of this lack of knowledge of neural processing, that involves not only humans and other complex animals, but also more simple animals like insects: even in this case, despite having a good idea of the kind of computations required for their navigation and the possibility of doing genetic experiments, the understanding of their actual computations is still limited.

Jackendoff & Pinker (2005), who do not consider recursion the only necessary principle for language, also argue that it is present at least in vision. Here I show the example with the figure they refer to:

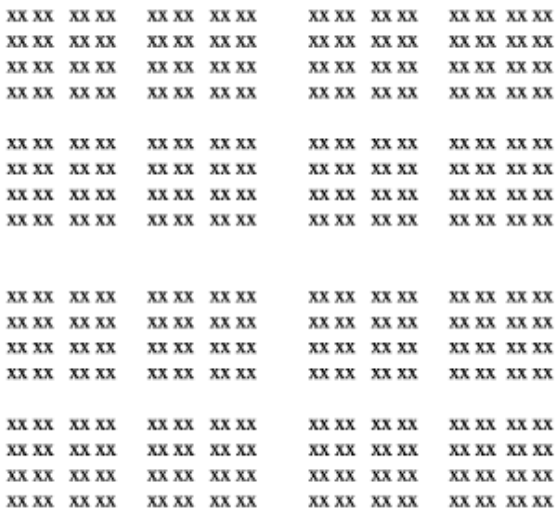

Fig. 2. Example of recursion in visual groupings (Jackendoff & Pinker 2005: 218).

According to the scholars, "this display is perceived as being built recursively out of discrete elements which combine to form larger discrete constituents" (Jackendoff & Pinker 2005: 217). In perception, the pairs of xs can be combined with the adjacent pairs, and then form squares of clusters, and squares of squares, and so on, in a process that they define recursive and potentially infinite: "to use Chomsky's term, we have here a domain of 'discrete infinity' in visual perception, with hierarchical structure of unlimited depth" (Jackendoff & Pinker 2005: 217).

A more extensive investigation is presented by Maurício Dias Martins (2014, see also Dias Martins 2012), who outlines the results of a research program on recursion in the visual and auditory channels, aiming to answer the questions: "Is recursion specific to linguistic domains?" and "Is language necessary to use recursion in non-linguistic domains?"

Fig. 3. Example of the creation of a visual fractal by using a recursive rule (Dias Martins 2014: 178).

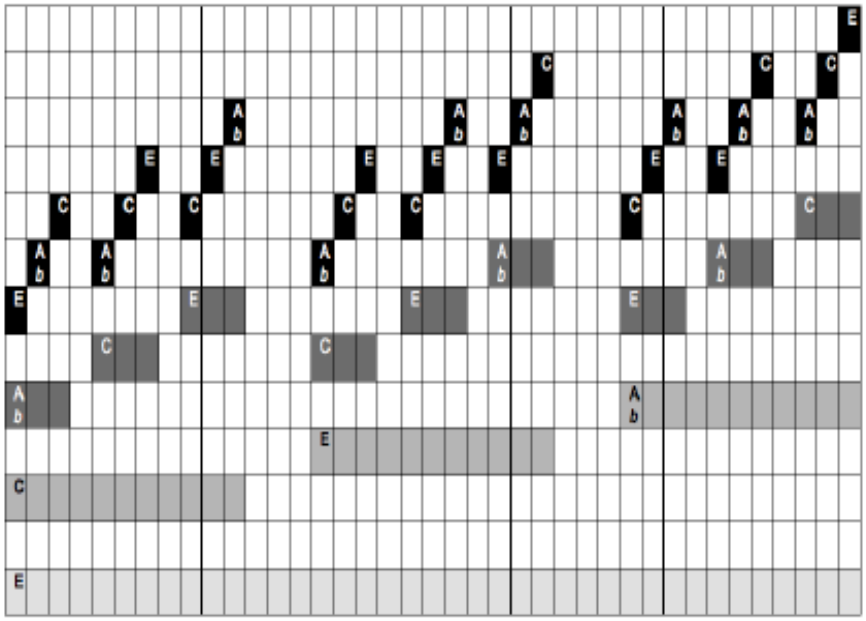

Fig. 4. Example of a musical fractal recursively built (Dias Martins 2014: 179).

For details on the experiments – conducted on samples of children of different ages and aided by brain imaging techniques – I refer to Dias Martins (2014), but we can summarise that, based on the results presented, the question "Is recursion specific to linguistic domains?" is answered by Dias Martins with "no"; and the question "Is language necessary to use recursion in non-linguistic domains?" is answered with "results strongly suggest that humans can build recursive representations independently of language." (Dias Martins 2014: 185). In fact, "the generation of visuospatial hierarchies with recursive rules activates a distributed brain network independent of the classical language areas", and "[during the experiments] Broca's area was not specifically active in the generation of new hierarchical levels" (Dias Martins 2014: 184). As further evidence, Dias Martins adds that verbal resources were not needed to acquire recursive rules in vision, and that the cerebral areas correlated with language processing were not activated while the brain was representing recursive rules on hierarchical levels, suggesting that "grammar comprehension is not specifically correlated with visual recursion" (Dias Martins 2014: 186). Nevertheless, Dias Martins has identified some similarities in recursion in syntactic, visual and spatial domains, like "similar learning constraints (prior acquisition of iteration is necessary)" (Dias

Martins 2014: 186); the activation of the same brain areas related to "categorical/semantic processing", and a correlation between grammar comprehension and recursive visual processing (Dias Martins 2014: 186).

For all these reasons, Dias Martins concludes that "general mechanisms for hierarchical processing may exist, and that recursion is a particular subset of hierarchical processing" (Dias Martins 2014: 186).[57] This is interesting because it outlines recursion as dependent on hierarchy, so that potentially there could be hierarchy without recursion (but not the other way around), while in Chomsky's view the two properties – at least in language – seem connected, with "infinite" hierarchical structures which are a consequence of recursive merge. Furthermore, Dias Martins observes that recursion and hierarchical processing "can operate independently of on-line verbal resources, though they may still depend on the availability of resources shared with language, for example, semantic or categorical knowledge" (Dias Martins 2014: 186). Thus, despite these properties are likely present in other domains and they don't need language to work, they might share some crucial resources with it.

There are also other approaches that view recursion as a property probably specific to our species, but not limited to linguistic domains. I already cited the perspective of Corballis (2011), who believes that recursion is a general mechanism on which language is based. Another perspective is that of Vincenzo Raimondi (2014), according to whom recursion is at the origin of language, but it is not exclusively dedicated to it: the scholar argues that "recursive consensual coordination" is the generative mechanism underlying every human social activity and every form of interaction. Consequently, even the capacity to observe, from which the shared intentionality of Tomasello (2008) derives, would depend on this single process: in Raimondi's view, this domain-general capacity would have emerged biologically as a consequence of human social organisation, in the inevitable relationship between the context and other individuals: hence, collaboration with others and shared activities could be considered species-specific dynamics of human structural associations, and the same would apply to *languaging*[58] (Raimondi 2014: 6). In his perspective, "Although they can be analytically distinguished, human joint activity and language arise from the same process" (Raimondi 2014: 7), and "by recognizing recursive

[57] However, it should be noted that the hierarchy of human language is not limited to elements that contain other elements. According to the theory of c-command (Reinhart 1981), some elements c-command others of the same dimension (or even bigger ones, in the case of heads that c-command whole constituents). There are also other criteria, such as selection (in which a head selects its complement and specifier), that create a hierarchical relationship of dependence. Nonetheless, the study of Dias Martins highlights that some simpler processes, from which this complexity might derive, are present in non-linguistic domains.

[58] *Languaging*, a broader concept of which language is just one aspect, is the process through which meanings are created and organized, and then expressed through language (cf. Raimondi, 2014).

consensual coordination as an invariant organisation of human interactional dynamics, it becomes possible to understand different classes of phenomena, from language acquisition to all kinds of sociocultural practices, as resulting from a single process" (Raimondi 2014: 11).

In this section, I have only briefly sketched a type of research that primarily falls within the realm of neuroscience: hopefully, the scientific advancements of the coming years will help unravel some unresolved questions, presenting new compelling evidence of recursion in other domains. I now shift to an even more uncertain territory: the other extinct members of the *Homo* genus.

1.2. The research of the core language properties in extinct hominins

To understand where language comes from, we must first ask whether the hominins from which we descend, who already possessed unique characteristics in the animal kingdom – a relatively complex culture and social life, the use of fire, the use and creation of stone tools – had an advanced form of communication, from which our competence derives. The most logical approach is to compare *Homo sapiens* with various species of hominins, and identify similarities and differences. Unfortunately, this does not allow us to clarify the issue because there is a lack of direct evidence. For this reason, scholars have focused on indirect evidence, attempting to deduce the presence of language from fossils and available material, and from collateral aspects such as the supposed presence (or absence) of symbolism. However, the issue remains highly debated, with some scholars (e.g. Berwick & Chomsky 2016) believing that hominins had no language faculty, and others admitting the possibility or strong likelihood that other members of the *Homo* genus had forms of protolanguage (see e.g. Fitch 2010). It is then important to assess what the limits of this type of research are and to what extent it can present definitive evidence.

1.2.1 The limitations in the reconstruction of the language of extinct hominins

Since, as we know, all our hominin ancestors are extinct, the closest living relatives (bonobos and chimpanzees) are separated from us by seven million years of independent evolution, and therefore present significant differences compared to humans. To study other *Homo* species, we can only observe what is left: the fossil remains of their bodies and the traces of their creations. We must bear in mind that such findings are usually due to chance, they are fragmentary and deteriorated over time, and represent only a tiny fraction of what actually existed: the absence of certain fossils thus does not necessarily imply that certain

traits never existed, but could also mean that they have not been found or have not been preserved. Pinker and Bloom state regarding this:

> [some biologists] attribute the sudden appearance of fully-formed new kinds of organisms in the fossil record to the fact that speciation typically takes place in small, geographically isolated populations. Thus transitional forms, even if evolving over very long time-spans, are unlikely to appear in the fossil record until they reinvade the ancestral territory; it is only the invasion that is sudden (Pinker & Bloom 1990: 719).

Fossils appear suddenly because species first develop in small communities, gradually and with many intermediate phases, and then – as soon as the various modifications represent a truly significant success – animals multiply and spread. Only then, when they are at the peak of their form and number, do they leave behind enough corpses which increase the chances to become fossils: therefore, we, who can only rely on fossils, often do not see the small intermediate phases, but only the findings related to the periods of splendour of a species, and this can lead us to draw wrong conclusions.

In Figure 5, Ian Tattersall eloquently points out that the history of human evolution was not the linear process of development of a single being, gradually replaced by successive 'versions': in fact, there were many species of hominins – existing even at the same time and often in competition with each other[59] –, some of which managed to prevail and survive longer, while others became extinct. The scenario appears chaotic and diverse, with a great deal of evolutionary experimentation that selected only the most fortunate and capable species. The fact that there is only one species of the genus *Homo*, continues Tattersall (2016: 161), is indeed exceptional, and means that there is something very particular about humans: something that has led us to prevail over other hominins until we completely supplanted them, whether by eliminating the 'competitors' or partly assimilating them[60].

[59] See also Pinker & Bloom (1990).
[60] Indeed, today genetic evidence shows that *Homo sapiens*, during their journey, have assimilated at least partially Neanderthals and Denisovans, as traces of both species are found in our DNA (see Green *et al*. 2010).

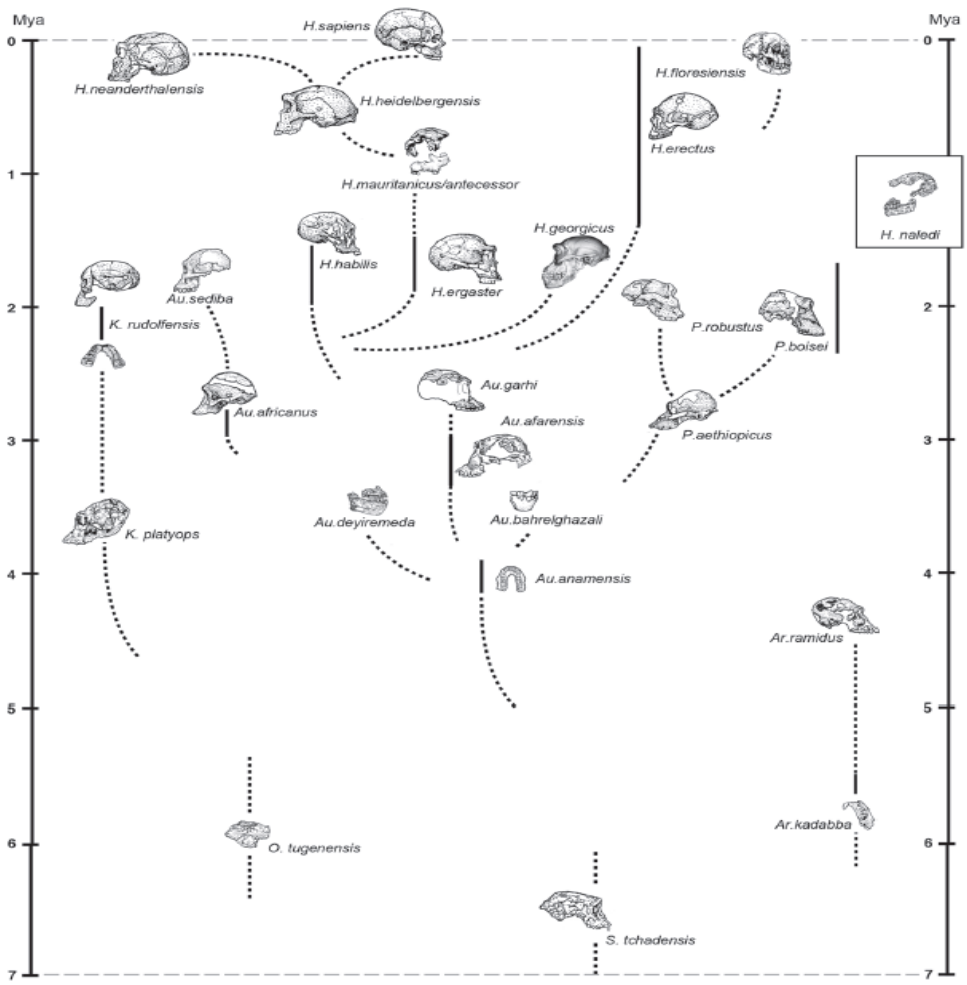

Fig. 5. A tentative genealogical tree of the hominin family (Tattersall 2016: 160).

As if that were not enough, due to the volatile nature of language, the structure of fossils cannot tell us enough about the presence or absence of the phenomenon: Fitch (2000) focuses in particular on the anatomical capacity for speech – probably connected to the position of the larynx, which in humans is lower than in other primates[61] – and considers that the only elements related to speech that fossilise are the skull and the hyoid bone, which are not sufficient to reconstruct the vocal anatomy. In fact, "the vocal tract is a mobile structure that essentially floats in the throat, suspended from the skull by elastic ligament and muscles" (Fitch 2000: 262), so that the fossil record can lead at best to imprecise deductions. Attempts, as Fitch explains, have not been lacking, through various measurements of the anatomical parts that could be involved in speech. For the larynx, attempts have been made to derive its position from the fold at the base of the skull

[61] The issue of the descended larynx has been debated for a long time, both about its actual uniqueness in humans (later disproved) and about its original function: instead of a communicative purpose, size exaggeration has been proposed, which means using the voice to simulate larger body size, a useful 'deception' in both sexual selection and confrontation with predators and rivals (cf. Fitch 2000, 2010).

(basicranial angle), assuming that the two parts were correlated; however, more recent studies have refuted this connection (Fitch 2000: 262). Following the discovery of a Neanderthal hyoid bone, comparable to that of humans, it was then hypothesised that Neanderthals had a descended larynx; however, the position of the hyoid bone is not sufficient to determine the exact position of the larynx. Nevertheless, the finding served to deduce that Neanderthals did not have air sacs, a trait present in primates but absent in humans[62]. Another system "involves using fossils to estimate the size of neural structures involved in spoken language" (Fitch 2000: 262), such as comparing skull endocasts, "to estimate the size of language-related cortical areas" (Fitch 2000: 262), or the size of the hypoglossal canal, whose nerve "contains most of the motor fibers that innervate the tongue and other vocal articulators." (Fitch 2000: 262); or even measuring "the diameter of the thoracic vertebral canal in various extant primates, modern humans and extinct hominids" (Fitch 2000: 262), noting that in *Homo ergaster* and *Homo sapiens*, it is larger than in primates and other hominins: since a larger size would imply greater control of breathing, and breathing is connected to speech, it has been deduced that this capacity was present from *Homo ergaster* onwards, but not in earlier hominids (Fitch 2000: 262). However, Fitch is convinced that none of these approaches has yielded true results, and that "it will be difficult to reconstruct the vocal behaviour of our extinct ancestors with any certainty" (Fitch 2000: 263). The only certainty, in this field, is that around "6 million years since our divergence from chimps, major changes in hominid vocal anatomy and physiology took place, […] together with acquisition of the ability to imitate novel sounds (Fitch 2000: 263).

Although these anatomical interpretations neither exclude nor confirm the presence of speech capacity, it's important to remember that speech doesn't automatically imply the presence of a language faculty; in any case, it remains impossible to prove. The most delicate and debated issue concerns the aforementioned *Homo neanderthalensis*, the hominin most similar to *sapiens*, with whom they share over 95.5% of the genome, from which they separated between 800,000 and 315,000 years ago (depending on the studies), and with whom *sapiens* coexisted for several millennia, until their disappearance (see Paabo 2014). A recent study (Conde-Valverde *et al*. 2021) reconstructed the hearing capacity of Neanderthals through 3D models, derived from fossils of ear bones, then compared with models of the ears of *sapiens* and earlier hominids: based on the results of simulations, the authors argue that Neanderthals were able to perceive sounds in a manner comparable to humans, suggesting that their hearing capacity was perfectly capable of supporting language.

[62] The hyoid bone of chimpanzees has a cavity where air sacs are located, absent in both humans and Neanderthals (Fitch 2000: 262).

This clearly does not mean they could speak or had our 'complete' language: it does, however, once again attest the great compatibility between the two species. Moreover, we have evidence that Neanderthals partly mixed with *sapiens*, as 1-4% of the DNA in the world's human population can be traced back to their species (in Africa to a lesser extent, to the point of being insignificant[63]; see Green *et al*. 2010). Furthermore, as Tattersall asserts, "The Neanderthals were wonderful craftsmen in stone, and left us an incomparable record of very complex lives. They flourished in an age of difficult climates, hunted some fearsomely large animals, and at least occasionally buried their dead" (Tattersall 2016: 162). These arguments lead to believe that it is possible that they had a form of protolanguage, a belief held, for example, by Fitch (2010) and Bickerton (2014), but denied by both Tattersall (2016) and Chomsky, who relies on Tattersall's paleoanthropological studies (see Berwick & Chomsky 2016). Berwick and Chomsky (2016: 147) argue that, even though Neanderthals had slightly larger skulls and brains, the larger dimensions were in the rear of the skull (occipital bun), which would have enhanced visual capacity and tool use but not reasoning, associated with the frontal lobe, which is more developed in *sapiens*.

We have seen that many anatomical details of hominins remain unfathomable, but delving into their mental capacities is even more difficult. Here too is it only possible to rely on indirect evidence and, based on that, try to deduce whether they had a form of language, or rather, protolanguage, as everyone seems to agree that only *Homo sapiens* ever possessed a language faculty as we understand it. In this research, an important clue is represented by archaeological findings, from which some behaviours of hominins can be inferred. As Tattersall observes, despite indicating a high degree of complexity and intelligence, the many findings of "stone tools and butchered animal bones", and their disposition in the hominin settlements, cannot reveal much about cognition, i.e. "about how our precursors subjectively perceived and processed information about the world" (Tattersall 2016: 158). Although a complex lifestyle might suggest the presence of a form of language, this cannot be considered decisive proof in this regard.

One of the most relevant elements associated with language is symbolism, which has been described as a proper linguistic property in Part I, chapter 4. Starting from the idea that "language involves forming and manipulating symbols in the mind, and symbolic reasoning of the human kind is almost inconceivable in its absence" (Tattersall 2016: 159), it is consequently believed that "if we find evidence for symbolic cognition, we can also infer language" (Tattersall 2016: 159). Since "the unusual behaviours associated with symbolic

[63] Indeed, this provides another indirect clue about the possible disappearance of Neanderthals, partly exterminated, and partly absorbed into the sapiens population, upon contact with them after leaving Africa.

thought are not correlated in any useful way with brain mass" (Tattersall 2016: 158), research focuses on findings of symbolic objects, such as ornaments, decorations, or other signs of a culture with ritualistic practices and abstract thinking abilities. Based on this criterion, the earliest definitive evidence of symbolism dates to about 77,000 years ago, and it consists of "two smoothed ochre plaques of slightly different ages, […] that bear similar engraved geometric designs" (Tattersall 2016: 163). These objects were found in Blombos Cave, in southern Africa, and are attributed to anatomically modern *Homo sapiens*: the fact that the plaques were created at different times, depicting similar designs, leads Tattersall (2016: 163) to suppose that they had a symbolic meaning that was transmitted through generations. This is reinforced by another discovery 200 kilometres away, dating back to 60,000 years ago, which features similar engravings. An essential element is indeed the context: a single artifact is not considered sufficient to infer the presence of symbolism, but widespread traces of a well-established symbolic culture, spread over a wider territory and timeframe, can be considered evidence in that direction (Tattersall 2016: 163, see also Berwick & Chomsky 2016). For this reason, older examples (such as a modified shell dated 500,000 years ago, found in Java) have been dismissed as isolated pieces: for Tattersall, they only demonstrate that certain hominins had the cognitive ability to perform such tasks: "To conclude that a Middle Palaeolithic individual was capable of modifying this object is not the same as to demonstrate that the society in which he or she lived depended on the symbolic capacity" (Tattersall 2016: 161). This is certainly true; however, a finding like this one, if genuine, still indicates that this cognitive potential predates *Homo sapiens*, and thus shows that these abilities are not unique: in the perspective of this dissertation, which aims to identify the most rudimental principles beyond language that might represent an alternative to the saltation proposed by Berwick & Chomsky (2016), the existence of similar skills represent an important result. Nonetheless, beyond the debate about the cognitive abilities of extinct hominins, it is evident that after the appearance of *Homo sapiens*, there was an explosion of well-documented innovations[64], apparently unparalleled in hominins, involving various aspects of life. This was followed by the great migration out of Africa (between 60,000 and 70,000 years ago) of "undoubtedly symbolic" humans (Tattersall 2016: 161), who conquered the world, leaving examples of spectacular rock art dating back at least 40,000 years. All of this is only documented in *Homo sapiens*: one of the main arguments in favour of this uniqueness is the way in which culture and innovations among hominins remained stagnant

[64] It should be noted that traces of modern humans from a behavioural perspective, and therefore associated with symbolism, appeared no earlier than 100,000 years ago: this means that anatomically modern *Homo sapiens* went through a period of stagnation of over 150,000 years before this cultural "explosion" occurred (see Tattersall 2016).

for millions of years until the arrival of our species (Tattersall 2016). On the same page is Bickerton who, reflecting on the presence of symbolism and language faculty in hominins, observes that "If language had yielded significant results in developing new and more efficient behaviours, we would expect to see starting around two million years ago a gradual but steady improvement in artifacts and other physical traces of human behaviour" (Bickerton 2014: 90), but no traces of it have been found. For these reasons, it is believed that other hominids did not have any form of symbolism: according to some scholars (see Tattersall 2016, Berwick & Chomsky 2016), this implies that they did not have any form of language, not even qualitatively inferior with respect to ours; according to others, however, this does not exclude the presence of a protolanguage (see e.g. Bickerton 2014, Fitch 2010).

Even in this case, the issue of the Neanderthals is more complex because there are artifacts associated with them that show signs of symbolism. These belong to cultures dating back to the last period of the Neanderthals, following contact with *Homo sapiens*, such as the Châtelperronian (in the area between Spain and France) and its equivalents in some parts of Europe (Longa 2013: 17), along with an artifact recently found in Gibraltar (see Rodriguez-Vidal *et al*. 2014). While some scholars (see d'Errico 2003, cited in Longa 2013) propose that the Neanderthals were undergoing an independent transition towards a more complex culture before the arrival of the *sapiens* in Europe, many argue that the evidence is controversial. Berwick & Chomsky (2016: 146-149) suggest that these artifacts might not belong to the Neanderthals but could be traces of more recent human settlements built on top of Neanderthal sites: Neanderthal fossils would then belong to a more ancient layer. Furthermore, as with other hominins, there is no widespread cultural substrate of symbolism, and the few pieces of evidence found represent a minimal part of the findings from that period, which conversely show no signs of symbolism. As Tattersall states, "Altogether, in a record as geographically, temporally, and materially as expansive as theirs, if the Neanderthals had routinely been symbolic thinkers it seems reasonable to suppose that they would have left more convincing indications of it" (Tattersall 2016: 162). Finally, it is emblematic that the Neanderthals lived for tens of thousands of years in a static manner, and then, with the arrival of symbolic *sapiens*, Neanderthal sites with traces of symbolism appeared (Tattersall 2016: 162). As Longa notes (2013: 17), the idea that the Neanderthals were independently developing symbolism just coinciding with the arrival of *sapiens* is so improbable that even the thesis itself becomes unlikely, and it is more prudent to assume that

the faculty of language was present only in *sapiens* – and it can even be hypothesised that it did not evolve gradually.[65]

In my opinion, although the combination of these arguments is convincing, one cannot ignore the significant similarities between the two species: the absence of symbolism – or conclusive evidence thereof – may suggest that Neanderthals did not have a complex language like humans, but it does not exclude the existence of a protolanguage or another articulated communication system, nor the presence of some rudimental cognitive abilities related to language (e.g. 1-merge), which might have been present although not fully utilised. Furthermore, considering various symbolic findings attributed to Neanderthals (see Rodríguez-Vidal *et al*. 2014; Hoffman *et al*. 2018a and 2018b), it cannot be excluded that these hominids, while not being great innovators, had at least the cognitive abilities to absorb innovations from others through imitation. In this case, the findings might be indeed Neanderthal creations, copied from the *sapiens* who were colonising Europe (Longa 2013: 17). Furthermore, it must be considered that absence of evidence does not automatically mean evidence of absence, and before categorically denying the possibility of symbolism in hominins it should be remembered that our picture is greatly limited by the scarcity of findings, which provide only a partial portrait. Additionally, we must consider that, with the advancement of technology, the picture of human evolution is continually enriched by new discoveries, which sometimes overturn previously established chronologies and set new dates, once deemed unacceptable.[66]

An intriguing approach for measuring the level of complexity of Neanderthal creations is proposed by Longa (2013). Longa does not consider evidence related to anatomical elements or symbolism; by contrast, he seeks for clues of the language faculty by analysing the artifacts from a formal perspective, calculating the type of computation that the various productions of hominins would have required, and therefore what computational brain capacity would have been necessary for them to do what they did. This way, it might be possible through indirect (but somewhat more measurable) means to determine if this computational capacity was also sufficient to support a form of language. To achieve this result, Longa compares geometric engravings belonging to the Neanderthals with engravings belonging to *sapiens*, in both cases made voluntarily.

[65] According to Pinker and Bloom (1990), on the other hand, hominids had a very complex life and a selective pressure towards more efficient linguistic communication from the very beginning.
[66] Consider for example how the date of the appearance of *Homo sapiens* is continuously pushed back due to new genetic studies and fossil discoveries.

Fig. 6. M1-6 ochre fragment, found in Blombos. From Helshinwood *et al.* (2009) in Longa (2013: 29).

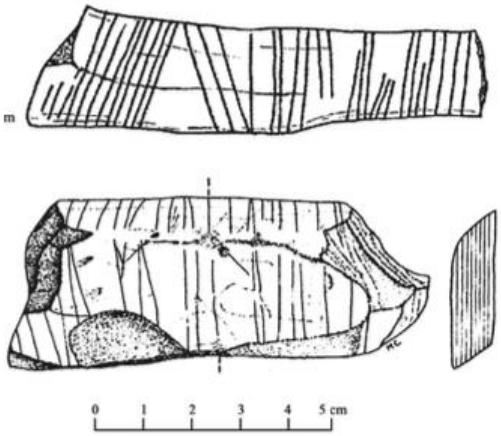

Fig. 7. Parallel incisions attributed to Neanderthals, from Bednarik (2003), on top, and from Crémades *et al.,* (1995), below (in Longa 2013: 23). According to Longa, parallel lines require a lower level of computational complexity compared to 'crossed' ones.

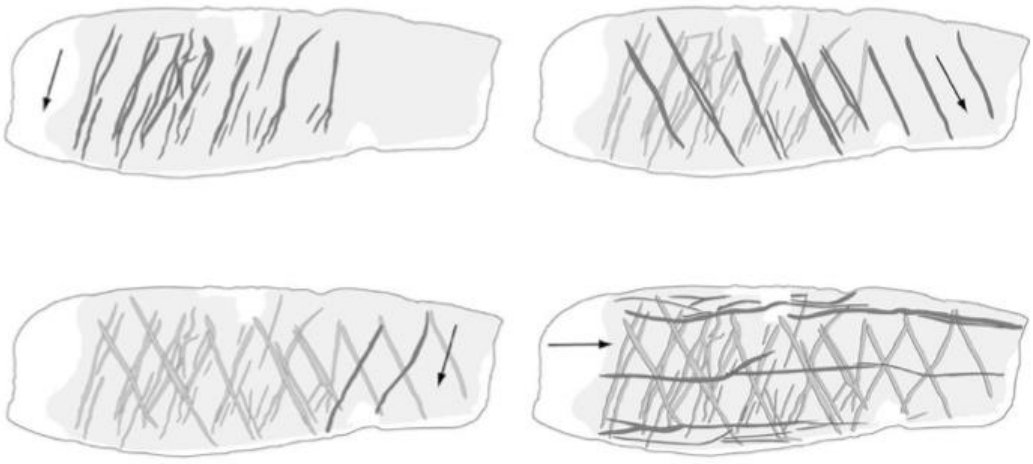

Fig. 8. Crossed incisions attributed to *sapiens*, found in the Blombos Caves. The image, corresponding to the artifact in Figure 6, shows the process necessary to create the complete object (the fourth drawing), which would require a greater amount of memory. From Henshilwood *et al*. (2009), in Longa (2013: 28).

In his opinion, "the arrangement of the lines permits us to infer whether or not a given species was computationally ready for FL." (Longa 2013: 22). For a detailed study, I refer to Longa's article (2013, particularly pp. 22-30), but we can summarise that in his view, engravings based on parallel lines do not require an advanced computational regime, unlike sequences with variations, crossed lines, or different angles. Since "pre-AMH[67] unequivocally intentional geometric engravings are based on parallel lines, not on crossed lines", and the more complex constructions seem "not intentional, or derive from natural causes", Longa concludes that "the computational system of species like Neanderthals or Heidelbergensis was much less powerful than the computational complexity required for FL" (Longa 2013: 24). Nevertheless, Longa believes that these hominins still had a complex communication system, albeit lacking a hierarchical structure, which he himself defines protolanguage (Longa 2013: 24).

Another recent discovery that questions the uniqueness of human complexity is that of *Homo naledi* (Berger *et al*. 2015). These enigmatic hominins have been found in 2013 in a hardly accessible cave of South Africa, and their remains have been dated back to Middle Pleistocene (335,000 – 236,000 years ago) (Dirks *et al*. 2017). The peculiarity of this hominin is that it shows some traits of *Homo* anatomy and some traits of much older ancestors, such as the *Australopithecus*. It is believed that it separated from the lineage which brought to humans very early, at latest around 900,000 years ago (Berger *et al*. 2017): this means that most of its "complex" traits have been reached independently from human ancestors, in a case of mosaic evolution, suggesting that selective pressures led different hominins to develop some similar traits, but not others. It was a small hominin around 145 cm tall, and it was likely more fit for climbing trees than for long-distance travel, despite being able to walk similarly to other *Homo* species (Berger *et al*. 2015). The morphology of the skull, mandible and teeth is similar to that of the *Homo* genus (Irish *et al*. 2018), but the brain size is comparable to that of *Australopithecus*, with an estimated low encephalization quotient of 3.75: this questions the idea that a large brain is necessary for complex behaviours, because it was no bigger than an orange (Berger *et al*. 2017). It is not confirmed that they were capable to use or produce stone tools, but the conformation of the hands, which shows parallels to that of *Homo sapiens*, indicates that they were dexterous enough to do it. Furthermore, stone tool use and production is a possibility because some tools, dated to the same period of *Homo naledi*, have been found in the area, but no remnants of other hominins species corresponding to that time have been discovered (Berger *et al*. 2017).

[67] "pre-Anatomically Modern Humans".

Finally, it has been proposed that *Homo naledi* had some form of cult of the dead, bringing on purpose the corpses into the cave in which they have been found (Dirks e*t al*. 2015). However, this hypothesis has been disputed (Hegeland *et al*. 2018). If the more far-fetched possibilities of burial and tool production were correct, this would change the perception of the anatomical requirements for a complexity that might be considered symbolic, deflating the idea of human uniqueness, also because *Homo naledi* represents a different branch separated from humans by hundreds of thousands of years of evolution. Even if this was not the case, the features of these hominins still lead to reconsider the traditional correlation between brain size and intelligence.

1.2.2 The protolanguage

In the last section, I outlined the issue of extinct hominins, explaining why it is so difficult to determine whether they had the faculty of language, and I traced an overview of the various approaches in this regard. Since I have often mentioned protolanguage, I take this opportunity for a brief description of the phenomenon. I have repeatedly cited those who do not believe it existed, prominently Chomsky among them, but I have not clarified that even among supporters of protolanguage, there are various divergences, both on what type of protolanguage it was and on the period in which it would have emerged.

First, a distinction between the synthetic model and the analytical model is needed. According to the synthetic model, protolanguage would have formed from independent meaningful units – which we could define as proto-words – and then assembled into sentences and broader meanings through a process of synthesis: words would have first 'emerged', and then they would have been assembled into structures[68] (cf. Bickerton 2014). According to the analytical model, on the other hand, sentences, or rather holistic message structures devoid of specific meaning, would have formed first, and then these messages would have gradually been broken down to isolate individual meanings (cf. Fitch 2010). For example, in the message "the lion is coming", according to the synthetic model, concepts and proto-words like 'lion' and 'coming' would have developed first, and then they would have been combined into the complex sentence; according to the analytical model, however, there would have been the indivisible phrase "the lion is coming" first, and only afterward

[68] Although Bickerton does not consider the possibility of a merge-like operation in synthetic lexical protolanguage, recent findings on nonhuman animals (see chapter 2 of this Part) suggest that a computation like 1-merge (cf. Rizzi 2016) might be accessible to some primate species. If a merge-like operation is present in other primates, then it is possible that also hominins had it, and they might have utilised it in their lexical protolanguage.

would the individual units ‘lion’ and ‘coming’ have been extracted from the message (cf. Fitch 2010).

In addition to this distinction, various types of protolanguage have been hypothesised, which can be summarised following Fitch (2010) into three categories: lexical, gestural, and musical (or prosodic). Lexical protolanguage, supported for example by Bickerton (2014), aligns with the synthetic model and would be a system with a broad vocabulary but lacking complex syntax. Fitch summarises that “Words in a lexical protolanguage are not combined into complex syntactic structures, leaving modern syntax as the final step in language evolution”, and notes that “This ‘syntax-final’ model of language evolution is shared by a diverse group of scholars who disagree about almost everything else”. Furthermore, “A lexical protolanguage assumes, as prerequisites, an ability for vocal imitation (necessary to develop a shared spoken vocabulary) and a capacity and drive for referential communication” (Fitch 2010: 401). Bickerton (2014) strongly supports it because it is the only model, in his opinion, capable of justifying the emergence of words: to transition from a holistic protolanguage to a segmentation into words, a processing capacity would have been required that, in his view, could not have existed before syntax.

Another hypothesis is the gestural protolanguage, “in which a visual/manual communication system played a crucial intervening role in the evolution of our current vocally dominated system” (Fitch 2010: 433). As we have already seen (see Part I, § 3.4.1), many scholars consider it highly probable, while some, including Kendon (2011), believe that protolanguage, like human language, was always a combination of gestures and speech: in this way, the gestural and lexical aspects merge, in a very plausible scenario. Regardless of the nuances attributed to it, supporters of gestural protolanguage believe that traces of it are still present in modern humans (for a detailed description, I refer to Fitch 2010: 433-465).

The last one that I mention is the musical[69], or prosodic, protolanguage, favoured by Fitch himself (2010) but also, among others, by Charles Darwin (1871), Otto Jespersen (1922), Steven Brown (2017). In this scenario – which coincides with the analytical model – the capacity for vocal imitation would have developed in hominins for reasons other than the expression of meanings: humans would be “singing” animals, and this ability would have evolved for ‘aesthetic’ reasons, “to fulfill many of the same functions for which modern humans use music today (courtship, bonding, territorial advertisement and defense,

[69] The relationship between music and language could constitute an independent dissertation, and for this reason I completely refrain from delving into the matter, although it could be one of the keys to understanding the origin of language.

competitive displays, etc.)" (Fitch 2010: 494). Building on Darwin's description, Fitch (2010: 494) defines it prosodic protolanguage, rather than musical, stating that modern music cannot reflect what the musical 'language' of the time was, from which certain aspects that are typical of music today, like notes and rhythm, are excluded. According to Fitch, the musicality – or prosody – laid precisely in the sung cadence of the syllables, which produced larger units which were hierarchically structured.

Among the various models of musical protolanguage, Brown's (2017) is of particular interest, because he sees the development of this phenomenon primarily aimed at communicating the emotions of the rich inner life of early humans, which involved "ritualized group-level vocalizations" (Brown 2017:1), and involved both "affective and intonational aspects of prosody" and "holistic and combinatorial mechanisms of phrase generation" (Brown 2017:1). Starting from this stage, there would have been a subsequent division between music and language.

In my opinion, since the various hypothesised protolanguages highlight different aspects of linguistic evolution, they all offer valid insights, but it is difficult to determine which one actually existed (or came first). Naturally, it is equally challenging to establish whether a protolanguage existed at all. With this clarified, we can proceed and extend this analysis beyond protolanguage and beyond extinct hominins, venturing into the vast and precarious terrain of comparison with other animals.

2. The research of language properties in nonhuman animals

After discussing extinct hominins and considering the limits of reconstructions, the research on the fundamental operations of language shifts towards more distant, yet still existing relatives. Thanks to the comparative method, it is possible to study the communication systems of other animals and compare them with human language to identify similarities and differences. An approach that I consider effective, at least as a methodological tool, is the division of language into different components, to separately analyse the various traits (as we have seen in the systematic classification of linguistic and general properties of human language presented in Part I, chapters 4 and 5). This decomposition allows focusing the investigation on elements that are easier to isolate: in this way, one can search and identify homologies and analogies with other living beings. Homologies are those characteristics inherited from a common ancestor and thus shared by the various species descending from it. Many of the similarities that we have with other primates, for example, depend on the long journey that our ancestors took together before we separated, and are therefore homologies. Naturally, this can be extended back to mammals, vertebrates, and so on, but in this case, the similarities become more general and less distinctive (e.g. reproductive system of mammals, bones in vertebrates), despite the existence of "deep homologies" (cf. Fitch 2010) in the structure of living beings, to which certain fundamental mechanisms can be traced back. Sometimes, however, surprising similarities are found in animals phylogenetically distant from each other, which are absent in evolutionarily closer species: in this case, these are analogies. Analogies are those common traits that develop independently, due to convergent evolution, in species facing the same selective pressures, which gradually shape them to fit a specific ecological niche. An example is the eye of vertebrates compared to the eye of cephalopods: through different paths, both have independently developed a very similar system for seeing[70].

Although the presence of homologies and analogies with other animals is an established fact, we must also consider the significant differences that separate our language from other communication systems. It is therefore legitimate to ask: to what extent can the comparative study of other animals help us to understand something about human language? Can this approach shed light on the nature and origin of this phenomenon, or is human language so qualitatively different that any comparison is useless, and it is a totally different field? This is the object of a debate. Hauser *et al*. (2002) encouraged the comparative approach and the division of language into different components or building blocks, which

[70] The eye of cephalopods is even more efficient than that of humans (cf. Dawkins 1986).

can allow for a better identification of the traits present in other animals, recognising homologies or analogies. For them, this approach is a fundamental requirement and must be adopted before any claims of human uniqueness: in their view, the starting hypothesis must be that, until proven otherwise, any human trait is shared with other animals[71]. This does not mean that they believe that all traits are indeed shared (as we know, they hypothesise that FLN is exclusively human), but they think that the first step in the investigation is to exclude certain similarities through comparison. Particularly enthusiastic of the comparative method is Fitch (2000, 2010), who states that it "provides a principled way to use empirical data from living animals to deduce the behavioural abilities of extinct common ancestors, together with clues to their adaptive function" (Fitch 2000: 263). This allows the identification of homologies, while analogies "can provide clues to the types of problems that particular morphological or behavioural mechanisms are 'designed' to solve" (Fitch 2000: 263). Tomasello is also a supporter of the comparative method, which represents the starting point of his investigations: in his view, the comparison with animals can lead to the identification of the cognitive mechanisms underlying language, as well as other social and communicative behaviours related to shared intentionality (see Tomasello 2008). Bickerton, on the other hand, is more sceptical; he criticises the division into components and the great attention given to uniqueness when it refers to humans: "With regard to the 'uniqueness issue', one needs to ask why that is treated as 'an overarching concern' when human uniqueness is under consideration but not when dealing with uniqueness in other species" (Bickerton 2014: 50). For Bickerton, in this case, humans receive a special treatment because "Nobody in the history of biology has treated any other species-specific trait as something that could be decomposed into a list of allegedly component features" (Bickerton 2014: 50) to be sought in other species and different contexts. For him, this approach is thus fundamentally irrelevant, as every adaptation is unique and follows its own path, and it is useless to divide it into components as if they were abstract and somewhat arbitrary parts: for Bickerton, what has evolved is never the single trait that makes up the final product (in this case, language), but the composite whole that serves a certain function. Even if some traits are shared, they are nothing but functional parts of the whole, and it is the whole that must be considered. According to Bickerton, the comparative method is at best an excellent tool for identifying analogies: for him, the similarities should be investigated in the whole niche, i.e. in the selective pressures generated by necessity, rather than in the presence of a specific shared trait. A different conclusion is adopted by Bolhuis *et al.* (2014, see also

[71] However, Berwick & Chomsky (2016) reverse this perspective by considering that the starting hypothesis about recursive merge is that is uniquely human, because it is not found in other species.

Berwick & Chomsky 2016). For them, the division into components is not useless, nor is the comparative method: rather, these approaches have demonstrated that, from a linguistic point of view, there are no similarities between the fundamental operations of human language and animal communication systems. Therefore, according to them, the comparative method has served to understand that any comparison related to syntax is useless, due to the extreme qualitative distance between the phenomena, and thus is unable to provide us with information about the central competence of language.

In my opinion, to verify whether the principles underlying language (i.e. the computational operations performed by the brain, identifiable with recursive merge) are neither species-specific nor language-specific, the comparative method can be very useful. Finding traces of these operations in other animals, in fact, would confirm the idea that such processes are not exclusive to humans, and therefore 1) they would respond to general ancestral principles that concern the deep architecture of the brain and would be homologies shared with other animals, or 2) if they were analogies, this would demonstrate that they can still develop in other beings through convergent evolution, given the right set of circumstances and the right selective pressures. On the other hand, if it were possible to exclude beyond any reasonable doubt the presence of these operations in other animals – or if their total absence in the linguistic domain were determined – there would be strong evidence supporting the hypothesis that the faculty of language is a species-specific characteristic of humans and is exclusive to the linguistic domain. The comparative method would prove useful in either case. Moreover, it is also important to research other characteristics, not necessarily connected to the fundamental operations attributable to the FLN of Hauser *et al*. (2002), but rather to their FLB, and even to non-linguistic elements considered external to the faculty of language. Anatomical, cognitive, social, and behavioural aspects are, in my view, determining 'ingredients' for the existence of language, and identifying even some of these elements in other species could suggest that, although it likely took a multifaceted set of fortunate circumstances to arrive at our faculty of language, some of these circumstances have occurred and continue to occur in the animal kingdom.

Once the importance of the comparative research on animals is assessed, it becomes clear that this study presents unprecedented problems: while humans can be questioned and express judgements about their linguistic abilities (the property of reflexiveness described in Part I, chapter 4), animals do not allow for language to be directly investigated, thus the only way to explore the issue is to observe their behaviour and deduce some linguistic properties from it (Berthet *et al*. 2023). There are different approaches to embark in this complex task, divided into two main possibilities with their pros and cons: the study of captive or wild

animals. Captive animals can be trained, and thus offer the possibility to perform more invasive experiments during which they are constantly monitored. The studies and the training can last weeks or even years, and lead to a closer interaction with the animals: it is thus possible to test complex cognitive abilities (e.g. self-recognition, numerical processing, detection and generalization of rules, see Reiss & Marino 2001, de la Mora & Toro 2013), and artificial languages can be taught (e.g. associations of arbitrary signals and meanings, meaningless finite-state languages, etc., see Savage-Rumbaugh 1986, Ravignani *et al*. 2013). However, this close interaction might also influence the results of the experiments, causing abnormal behaviour (Lambrechts *et al*. 1999) and non-genuine responses (e.g. when an animal seems to generalise a rule, but instead is imitating the human examiners which subjected him or her to several trials). Furthermore, the studies on captive animals cannot describe what happens in the actual environment of their species, because the specimens live in very different conditions from their natural ones (and many individuals are born in captivity): for example, the fact that a bonobo (see e.g. Savage-Rumbaugh 1986 and the following chapters of this Part II) is able to understand or perform a rudimental form of language does not imply that the species displays this ability in its environment. Nonetheless, in the perspective of this dissertation, which aims to unravel the most rudimentary principles and operations *behind* language, the discovery that an animal is capable to perform a complex cognitive operation related to language would still be a very important result, because it would show that – provided there is a necessary training – the brain of that animal can support that operation, which therefore would not be unique to humans. This would not mean that other animals can be compared to humans to the full extent of their competence, nor that they make active and constant use of it, but it would still strengthen the hypothesis of a gradual evolution of language starting from shared traits. The presence of a latent trait (e.g. the ability to understand some syntactic rules) in a nonhuman animal (especially in primates) could explain the emergence of that trait also in humans, not through a very favourable mutation as Berwick & Chomsky (2016) suggest, but through other phenomena like exaptation or selective pressures which led an 'underlying' trait to become used in new situations and more developed. For this reason, the results of captive animals can be fundamental for this kind of research, although they have to be taken with caution.

On the other hand, the studies on wild animals offer a different perspective: the observations reflect the real conditions of the species, and the behaviours are not influenced by artificial stimuli or training, thus they offer a valid portrait of a certain animal or population in its own environment. However, the data are more difficult to be interpreted and scientists cannot interact with the animals to dig deeper. Furthermore, the animals'

reactions to the examiners can compromise the studies, because wild individuals are not used to humans and can perceive them as a threat or as a distraction. The ideal condition, as Lambrechts *et al*. (1999) consider, would be to perform parallel long-term studies with wild and captive populations, to produce testable hypotheses based on the observation of wild animals and then test them with captive individuals. However, this approach is difficult and rarely performed. An approach that can be defined "intermediate" between studies with wild and captive animals is that of research involving free-ranging animals habituated to humans, like for example the great apes of certain African nature reserves, which are not captive but live in territories supervised by humans: this makes them more "traceable" in the forest and less hostile or scared by the examiners.

A fundamental practical way to study vocalisations and behaviours – in all three cases – is that of playback experiments, in which animal calls are recorded and then played back in different ways to verify the reaction of the animals in different contexts (e.g. without the predator threat to which an alarm call is referred, or with a dummy that represents it). A practical example of this approach is the study performed by Suzuki & Matsumoto (2022), whose results will be discussed in the next section. To analyse the perception of call combinations in Japanese tits (*Parus minor*), the scholars 1) recorded two of their meaningful calls (ABC and D) in the wild, 2) they edited them producing different sequences (the natural ABC-D and the unnatural D-ABC), and 3) they played them back to wild Japanese tits in their environment. To do so, they installed two speakers on the trees, and they reproduced the sounds in both orders (the correct and the incorrect one) and coming from different speakers (e.g. in one case the ABC-D and D-ABC sequences were reproduced by the same speaker, simulating a single individual, in another case one speaker reproduced the call ABC, and the other reproduced the call D, simulating two different individuals). Furthermore, they put a fake predator (a snake model) to represent the threat associated to one of those calls. This method allowed the scholars to inspect the response of the animals in a way that is more interactive than the sole observation, but less invasive than experimenting on captive trained animals.

In any case, all these approaches can be applied to search for the fundamental properties of language in nonhuman animals. However, as Berthet *et al*. (2023) point out, the first step in every comparative research of language skills is to avoid anthropomorphism, by excluding the simpler alternative hypotheses which might explain the animal abilities without the need of postulating a linguistic complexity. Following the so-called Morgan's Canon – which resembles the concept of Occam's razor – if a behaviour can be interpreted by taking into account simple cognitive abilities spread in the animal kingdom, then there is

no need to postulate the presence of more complex human-like skills to explain it (Berthet *et al*. 2023). While keeping this in mind, several scholars made attempts to apply linguistic terms and approaches to animal communication. This is a complex task also because of the difficulties represented by connecting biology and linguistics, which are very different fields that hardly communicate between each other. This is mainly due to the differences in the type of research performed, in methodologies[72], and also in the terminology, which might create confusion with discrepancies in the definitions (e.g. the term "syntax" is used with different meanings in biology and linguistics). A recent tentative to bridge the two disciplines in the field of 'animal linguistics' has been done by Berthet *et al*. (2023), who present a description of the fundamental linguistic levels of semantics, pragmatics, and syntax which can be applied to both animal communication and human language. By focusing on the main properties of the signals, their associations with semantic and pragmatic meanings, and their combinations, they present a quick guide for the research of language properties rich with examples of animal calls and behaviours, and with descriptions of methods to study 'animal linguistics' (see Berthet *et al.* 2023 for further information).

An analysis more centred on the "syntactic" concepts of combinatoriality and compositionality is that proposed by Engesser & Townsend (2019), who review the presence of these abilities in several species (see chapter 3 below). In their paper, the scholars describe different types of combinations: among them there are combinations comparable to some constructions of human language (e.g. combinations of meaningless "phoneme-like" units to form songs or meaningful calls; combinations of meaningful "morpheme-like" units to form more complex calls), but also very peculiar phenomena like temporal structures, i.e. sequences obtained by the repetition of the same unit with different temporal arrangements (conceptually comparable to a sort of morse code)[73]. What emerges is that combinatorial and compositional abilities seem spread in the animal kingdom, and they can assume several shapes, some of which are very distant from those observed in humans, also due to the biological differences of the species. This suggests that evolution might find many ways to expand a repertoire of signals and convey more information: hence, complexity should be searched also in phenomena that are not intuitive and may have been traditionally overlooked.

[72] Berthet et *al*. (2023) cite a study of Prat (2019) which used the typical tools of animal communication research to analyse human languages: apparently, "it failed to highlight semantics, syntactic structure or vocal learning in human language" (Berthet *et al*. 2023: 93). This gives an emblematic example of the incommunicability of the two fields.

[73] All these elements will be discussed in the following chapter.

Compositionality and the merge operation are conceptually related also in the perspective of animals, like in human language. Despite the complexity of human linguistic constructions is not observed in any other species, the framework of Rizzi (2016) discussed in Part I also comes in handy for animals. While there is no evidence of recursive 2-merge and 3-merge outside humans, it has been proposed that non-recursive 1-merge is accessible at least to some bird and primate species. According to Rizzi (2016, 2021), given how merge is described by Chomsky (1995, 2001), it is very complicate to deny that some animal call combinations are not a simple realization of this operation. Gil (2022) reaches a similar conclusion, with his framework of bare and constructional compositionality (see Part I, chapter 2) which considers to be applicable to several species of primates and birds, in different degrees of complexity (see § 3.3 below). However, following Morgan's Canon, there are also alternative, non-compositional interpretations which do not need the presence of merge to explain the properties of these animal calls, and involve simpler explanations like pragmatic principles and associative learning (see e.g. Schlenker *et al*. 2022, 2023): they have to be considered as well, although they do not exclude the possibility of 1-merge and compositionality.

The most relevant animal calls and their corresponding interpretations will be now discussed. However, like for humans, it is not possible to limit the analysis to the sole merge operation, which is inserted into a wider scenario of interacting properties and phenomena. Hence, in the next two chapters I adopt the same systematic classification presented in Part I for humans: first, I describe all the properties related to language in animals, focusing on those considered "purely linguistic" (including merge and compositionality), and verifying which ones seem absent and which ones seem present (and in what forms); then I focus on the general non-linguistic traits important to sustain language, with the same approach. After this analysis, I get back to the concepts of merge and compositionality in light of the empirical data of nonhuman animals.

3. The properties of language in nonhuman animals

The first attempt to compare human and animal linguistic properties was made by Hockett (1958), who assigns some of his language features (see Part I, chapter 1) also to nonhuman animals, while others are considered uniquely human. In a chart (Hockett 1958: 574) the author compares the first seven features of language with the communication systems of four species (bee dancing, stickleback courtship, herring gull care of offspring and gibbon calls), indicating whether these features are present or not. He finds interchangeability, specialisation and arbitrariness (the latter only scarcely developed) present in bee dancing and gibbon calls, while he assigns productivity and displacement only to bee dancing. Duality and cultural transmission are not observed outside humans, but Hockett does not exclude that these first seven "key properties" are present – and yet to be found – in other species. In the second formulation (Hockett 1960a, [1960b] 1977), the features of broadcast transmission and directional reception, rapid fading, and total feedback, are consequences of the physical nature of the vocal auditory channel, thus are present in each animal which produces and hears sounds. Semanticity is assigned to bee dancing and gibbon calls, and discreteness is attributed only to humans, in opposition with the continuous nature of other communication systems. The last three properties (prevarication, reflexiveness, learnability) are less focused on the comparison with other species and more on the nature of language, so they seem to be assigned only to humans (Hockett 1966, 1968).

Starting from Hockett's work, in the previous part of this dissertation I identified the main linguistic properties that characterise language present in humans. Now I turn to nonhuman animals, to address the question: which linguistic properties are present in them? I thus discuss *signal*, *signal learning*, *structure*, *meaning* and *contextual usage*. I focus both on the traits which are observed in the wild, and on the properties that emerge in captivity, where animals can be subjected to more invasive experiments which can highlight the extent of their cognitive capacities underlying language.

3.1 Signal

Like in humans, the aspects related to the nature of the signal are not disputed, so I will not extensively discuss it. Hockett's (1960a) properties of ***vocal auditory channel***, ***broadcast transmission and directional reception***, ***rapid fading***, and ***total feedback*** are present in all animals with vocal communication (and similar properties are present in visual communication too). ***Linearity*** (Berruto & Cerruti 1997) is equally present in animal vocal communication; however, two special cases need discussing here. First, the visual modality

is used extensively in animals, which, like in human sign language, do not follow the linearity principle. Furthermore, some animal species can produce two different sounds at the same time, violating the linearity principle and leading to a potential level of complexity absent in humans. This phenomenon of biphonation, observed for example in cetaceans (e.g. Jones *et al*. 2020), allows some species to convey two different messages at the same time, although it has been poorly explored.[74]

3.2 Signal learning

Signal learning, albeit being fundamental for humans, is not documented in many species. Instead, the majority of animal species have been traditionally described as relying mostly on genetically fixed signals to communicate. Nevertheless, researchers in animal communication recently implemented a novel framework considering signal learning on a continuum rather than a whole or nothing trait (Watson *et al.* 2022) and several social species have been shown to possess different degrees of such capacity. In humans, I differentiated between learning of meaningless signals (phonemes) and meaningful signals (morphemes and words); however, in nonhuman animals this distinction is more blurred, because many calls are not composed by a set of discrete elements, so that – conceptually – in most cases "phonemes" and lexical units coincide. ***Vocal contextual learning*** (referred to morphemes and words) is present in "a broad group of species including parrots, chickens, budgerigars, mynahs, cats, dogs, rats, lemurs, sea lions and dolphins […] bats, seals and elephants" (Vernes *et al*. 2021: 3, see also Janik & Slater 2000), but also in some nonhuman primates like chimpanzees (*Pan troglodytes*), vervet monkeys (*Cercopithecus aethiops*), gibbons (*Hylobates* spp.) (Zuberbühler *et al.* 2023, Watson *et al*. 2015, Fischer & Hammerschmidt 2019). All these animals have both the ability of comprehension and usage learning, so they can understand a new association between a signal and a function, and they can also actively use it. This ability is more easily studied in captivity, in which animals can be trained "to produce vocalizations in response of a conditioned stimulus" (Vernes *et al*. 2021: 3): among the animals proficient in this there are crows (*Corvus* spp.), rhesus monkeys (*Macaca mulatta*), gray seals (*Halichoerus grypus*) or bottlenose dolphins (Vernes *et al*. 2021). An example of vocal contextual learning in captivity is that of pig-tailed macaques (*Macaca nemestrina*), which have been trained to produce their natural 'coo' social call as a request

[74] A process which might be conceptually related to biphonation has been studied on human language, i.e. in the simultaneous production of signs (from a sign language) and words (from an oral language) by bilingual children (see e.g. Branchini & Donati 2016). In this case, two different meanings could be produced at the same time, with no linearity between the two modalities.

for food to human examiners (Coudé *et al.* 2011), thus learning to associate the existing signal to a new function (comprehension) and to use it (usage learning). In the wild, infant vervet monkeys produce alarm calls when they encounter any aerial disturbance, but growing up they restrict their usage only to aerial predators (Seyfarth & Cheney 1980). In this case, maturational effects are not excluded; however, Hauser (1989) shows that the infants in groups with more interactions with other groups mastered the 'correct' use of the 'wrr' call earlier than others, thus suggesting a role of social learning (both comprehension and usage) (see also Janik & Slater 2000). Another interesting possibility is that usage learning is involved in deception, e.g. when fork-tailed drongos (*Dicrurus adsimilis*) produce an alarm call in the absence of predators, to gain an advantage about food (Vernes *et al*. 2021).

Vocal production learning (referred both to "whole" word-like calls and discrete meaningless phoneme-like sounds) is much rarer, and refers to what the literature in animal communication calls *true* vocal learning. It is observed in three avian taxa (parrots, hummingbirds, songbirds) and in four mammal taxa (cetaceans, pinnipeds, bats, elephants) (see Brenowitz & Beecher 2023, Janik & Knörnschild 2021, Vernes *et al*. 2021, Janik & Slater 2000). A historic example of such vocal production learning comes from Alex, a grey parrot (*Psittacus Erithacus*) that was able to learn human words (Pepperberg 1999). Strikingly, nonhuman primates seem to be absent in the list of species abled to learn to produce new vocal signals. Historically, researchers in the 70s aimed to teach apes human language, and although their learning ability was significant in the visual domain (e.g. Washoe the chimpanzee, see Gardner *et al.* 1989) or when using lexigrams (e.g. Kanzi the bonobo, see Savage-Rumbaugh 1986), the results were much less convincing in the vocal domain (Viki the chimpanzee, see Hayes 1951). From an evolutionary perspective, as nonhuman primates and great apes specifically are our closest living relatives, these results appeared to demonstrate that the vocal production learning observed in animals is the result of convergent evolution. However, according to Watson *et al*. (2022), some captive chimpanzees showed some aspects of vocal production learning, by producing acoustically novel 'raspberry' sounds and 'grunts' which were not observed in the wild, that could be socially transmitted to the offspring by the mothers, and had the function to attract the attention of caregivers (see also Lameira 2017, Taglialatela *et al*. 2012). Similar abilities have been observed also in captive orangutans (*Pongo* spp.), and – according to Wich *et al.* (2012) – perhaps even in the wild (with the existence of certain call types present in a population but not in others). In any case this ability would be much more limited than that of humans, but also of other animals like songbirds or dolphins. However, if we consider

also gestural learning, then trained great apes show a higher level of proficiency (e.g. learning signals from American Sign Language or lexigrams, see Watson *et al.* 2022).

Cultural transmission is a feature that goes beyond language; however, here it is referred to the ability to learn communicative signals through social interactions. The ability to learn is present in all animals with vocal production learning (see above), but Hockett's ***learnability*** (1966) refers to the ability to "learn more than one language", thus not only some signals, but an entire repertoire. To my knowledge, as of today this definition shows some parallels only to wild orcas, in which some individuals have been observed learning a new "dialect" (a repertoire of signals that changes from group to group) to be included in a new pod (Filatova *et al.* 2010, see also Part III, § 2.2 and 3.2). However, trained animals like Alex the grey parrot studied by Irene Pepperberg (1999), Kanzi the bonobo (*Pan paniscus*) (Savage-Rumbaugh 1986), and the bottlenose dolphins Ake and Phoenix (Herman *et al*. 1984) learned new communication systems with rules that were novel to them (see Part III, § 3.3): despite not being clear cases of learnability, they suggest that this property seems to be within the reach of the cognitive abilities of nonhuman animals.

3.3 Structure

Many nonhuman animals produce calls that are not structured, meaning that they are continuous signals not divisible in simpler discrete units. However, there are several examples of structured signals, which in some cases are clearly formed by the combination of smaller units. Where the smaller units are meaningless, these combinations can be considered a form of phonology, and when they are meaningful, the combinations resemble rudimentary forms of morphosyntax. Although these comparison remains highly controversial, especially among linguists, such a comparative approach is fundamental. Most scholars have argued that the uniqueness of human language lies on the extensive complexity of its structure (see e.g. Berwick & Chomsky 2016), and the comparative approach is the only method that will enable us to test whether such structure is indeed unique to humans, or whether we can find similar constructions in other species.

3.3.1 Duality of patterning

Duality of patterning (or double articulation), already described as the existence of two levels in each linguistic expression, one of meaningless signalling units and one of meaningful units, is not found in nonhuman animals. Although there are animals capable to combine meaningless units, and animals capable to combine meaningful units (see §§ 3.3.2

and 3.3.3 below), a crucial difference with humans is that these two elements seem to never coexist. There is no clear proof of the presence of both levels in the same species.

3.3.2 Phonology

In humans, phonology relates to the level of meaningless, although functionally distinctive, units. In animals we can identify two different types of phonology: phonocoding, or phonological syntax (see e.g. Engesser & Townsend 2019), in which meaningless units are combined to form songs with a "holistic" meaning (e.g. courtship, territorial display), and "proper" phonology, in which meaningless units are combined in different dispositions to form different meaningful calls. The former is widely documented in animals (e.g. songbirds, whales), but despite showing an ability to create and process complex structures, it is not considered by linguists as a form of language, because of the lack of complex meanings (see Fitch 2010, Berwick & Chomsky 2016). The meaningless units combined into meaningful calls are the closest to human phonemes, which allow to create several words through their different arrangement. As of today, the best example of phonology is represented by chestnut-crowned babblers (*Pomatostomus ruficeps*) (Engesser *et al.* 2015), an Australian cooperative bird that produces two functionally relevant calls by rearranging the meaningless units A and B, obtaining AB and BAB: AB sequences are produced while flying, and BAB sequences are produced while feeding the nestlings. According to Engesser *et al.* (2015: 1), adding or omitting the B element "generates a phoneme-like contrast that is sufficient to distinguish the meaning between the two calls". There are two properties connected with phonology: discreteness and the combinatorial ability which allows to produce songs and other structures.

Discreteness is not openly described with this term in any nonhuman species. However, the A and B units of chestnut-crowned babblers (see above) can be considered discrete, being able to convey a different meaning with a different disposition, and thus being clearly identifiable (Engesser *et al.* 2015). Furthermore, it is true that the repertoires of many species are graded, but nonetheless the understanding of their signals can be discrete (known as categorical perception, like in human speech) as shown by playback experiments on primates, birds, or even frogs (see Green *et al.* 2020 for a general review). For example, Japanese macaques (*Macaca fuscata*) produce several 'coo' vocalisations which are acoustically similar: among them there are coos produced by juveniles who are separated from their group, and coos produced by courting females, whose main difference is the peak frequency. Although 'coo' calls appear acoustically continuous, individuals can discriminate

between the two variants, indicating categorical perception (Green *et al.* 2020). Also, complex vocalisations like birdsongs (see e.g. Garland & McGregor 2020, Kershenbaum *et al*. 2012), independently from their meaning, appear to be composed by discrete units, to the extent that the animals can produce them in different arrangements. It should also be noted that discreteness, despite being traditionally associated with phonology, can relate also to meaningful units. However, from the perspective of linguistics, to have discreteness is not sufficient to have a word that is segmentable into smaller units, nor that these units are distinguishable from the others. Thus, a contrast with another element, that represents a binary value (e.g., voiced/unvoiced) is required. Following this conceptual approach, a similar contrast might be proposed, for example, in the opposition of Campbell's monkey's *krak-oo* and *hok-oo* calls (see § 3.3 below), in which *krak* and *hok* might resemble a minimal pair, if we accept the interpretation by Schlenker *et al.* (2016) that *krak* and *hok* instantiate the contrast "terrestrial" vs "aerial" threats (hence we could associate them to a feature [+/-terrestrial]).

The ***combinatorial ability*** of meaningless units is generally attributed to animals, and it is effectively described by Engesser & Townsend (2019), which identify several ways in which meaningless elements can be combined: the already cited *phonocoding*, *multi element calls* (which are observed in chestnut-crowned babblers, but also in other birds of the Paridae family that combine meaningless signals to produce functionally relevant sequences), *temporal structures* (related to the variation of the temporal arrangement of the same sound in a sequence, observed for example in Mexican free-tailed bats, *Tadarida brasiliensis*), and *stochastic/proportional structures* (in which meaningful structures are obtained by the proportional distribution of the meaningless units in the sequence, proposed for black fronted titi monkeys, *Callicebus nigrifrons*, see Berthet *et al*. 2019) (see Engesser & Townsend 2019 for a general review). I classified these types of vocalisations in the phonology section because, following Martinet (1955), the first level of articulation (referred to phonology) relates to the possibility to decompose a word in smaller, meaningless units, and in reverse, to create a word by combining those units. This is possible in all the cases described: although not each of them can be considered proper phonology (e.g. temporal structures), they still denote the presence of this fundamental combinatorial ability related to language, which is the basis for phonological operations.

3.3.3 Morphosyntax

Morphosyntax is related to the combination of meaningful structures. Syntax, in particular, is considered to be the hallmark of human language and the stronghold of its uniqueness (Berwick & Chomsky 2016). Only recently some studies have begun to openly propose its presence in nonhuman animals (see e.g. Zuberbühler 2019, Abe & Watanabe 2011); however, this approach has been disputed (see Schlenker *et al*. 2022), and strongly criticised by generativists (see e.g. Berwick & Chomsky 2016, Bolhuis *et al.* 2018). Despite the criticism, there are valid reasons to believe that syntax, at least in its most rudimentary form, is indeed present in some nonhuman species (see Leroux *et al*. 2023, Suzuki & Matsumoto 2022). The research on syntax in nonhuman animals is probably the most relevant approach to test the uniqueness of the core traits of language, because if its principles were to be undeniably found in other species, then there would be almost nothing left to be considered uniquely human, and the FLN would likely be empty.

The key property of morphosyntax is ***compositionality***, i.e. the ability to combine meaningful units to create more complex meaningful structures, in which the meaning of the structure is derived from the meaning of the parts. It is known that nonhuman animals combine calls, i.e. meaning-bearing units, in several ways (see e.g. Engesser & Townsend 2019): among these types of combinations there are *segmental concatenations* (identity-encoding segments followed by other elements related to the animal behaviour or internal state, e.g. "individual X is scared" in Diana monkeys, *Cercopithecus diana*), *meaning-modifying segmental structures* (segments which do not appear in isolation and combine with calls, modifying their meaning like affixes, e.g. Campbell's monkeys' *-oo* described below), *semantic compositionality*, (combinations of two meaningful units which can be produced also in isolation, e.g. alert + recruitment in Japanese tits, see below) (see Engesser & Townsend 2019 for a review). One possibility is that these concatenations are simply random and non-deliberate, and interpreted by the receivers as "online readouts", i.e. combinations of different information produced by the caller (see Engesser & Townsend 2019). However, there is growing evidence that these concatenations are not random, nor guided only by pragmatics, as indicated by statistical methods like collocation analyses, which indicate that some co-occurrences are more frequent than others (see e.g. Leroux *et al.* 2021). Furthermore, some studies show that certain combinations are produced in specific contexts, thus being potentially meaningful: for example, Leroux *et al.* (2021) show that chimpanzees combine pant-hoot (utterances which convey information about the identity of the caller) with food calls only in feeding situations, when a beta male finds some food and

communicates its presence to the alpha males to avoid retaliation. It has been observed that if an alpha male finds the food and notices that it was not informed about that, it becomes aggressive towards the males which are present. Hence, this 'ID + food call' combination would be a way for the beta males to protect themselves, and it should be intended as "I am the individual X and there is food" (Leroux *et al.* 2021). Finally, some studies with playback experiments also confirm that receivers process these combinations in a compositional way: for example, Suzuki & Matsumoto (2022) showed that Japanese tits perceive two of their calls as combined in a more complex meaning only when they are produced in the correct order by the same individual, thus eliciting a behaviour which is not performed otherwise (see below).

All these elements thus suggest that compositionality is present in nonhuman animals. Effective examples come from captive animals, with nonhuman primates and birds that seem able to perform or understand combinations of meaningful signals. Among primates, Kanzi the bonobo (Greenfield & Savage-Rumbaugh 1990) was able to understand and combine meaningful lexigrams from a keyboard (e.g. "Liz hide", "water hide"), and Chantek the orangutan (Miles 1990) combined signals from the American Sign Language (e.g. "beard pull", "you pull") (examples taken from Gil 2022). Among birds I can mention the famous parrot Alex (Pepperberg 1999) that, in addition to an impressive ability to learn and produce words in the right context (see chapter 4), also combined some of them (e.g. "what colour").

In the wild, the already cited Campbell's monkeys have a call, *krak*, which relates to the presence of a leopard (Ouattara *et al.* 2009) or, in another interpretation, to a general disturbance (Schlenker *et al*. 2016), and another call, *hok*, related to the presence of an eagle (Ouattara *et al*. 2009) or an aerial disturbance (Schlenker *et al*. 2016). Both calls can be combined with the element *-oo*, which is not emitted in isolation, and seems to modify the calls adding a meaning of "more general", in a first interpretation, or "non urgent" in a subsequent interpretation (Ouattara *et al*. 2009, Schlenker *et al*. 2016). If the first interpretation is correct, *-oo* would be a suffix comparable to *-ish*, so that the call *leopard* (*krak*) would become *leopardish*, "something similar to a leopard" (and *eagle*, *hok*, would become *eagleish*) (Schlenker *et al.* 2023). In the second interpretation, *krak* would mean "there is a general disturbance", and *krakoo* "there is a non-urgent general disturbance" (Schlenker *et al*. 2016). This compositional phenomenon shows an interesting parallel with human morphology, in which words can be created by adding affixes in a similar process (see Part I, § 4.3).

Another interesting example of compositionality in the wild comes from the alert recruitment sequences observed in chimpanzees, pied babblers and Japanese tits.

Chimpanzees produce *alarm-huus* calls when the animals are surprised, while *waa-barks* relate to the recruitment of conspecifics during hunting or aggressions (Leroux *et al*. 2023): in dangerous situations, especially in the presence of snakes, they tend to produce the two calls together (*alarm-huus* + *waa-bark*), which elicit a stronger collective reaction to the threat. This has been verified with playback experiments associated with snake presentations, and it has been defined one of the strongest cases for compositional syntax in nonhuman primates (Leroux *et al*. 2023). Similar abilities have been proposed also for the other great apes: flanged-red orangutan (*Pongo pygmaeus wurmbii*), gorilla (*Gorilla gorilla*) and bonobo (Spillman *et al.* 2010, Hedwig *et al.* 2015, Schamberg *et al*. 2017, see also Leroux *et al*. 2023). Concerning birds, pied babblers (*Turdoides bicolor*) (Engesser *et al*. 2016) and Japanese tits (Suzuki & Matsumoto 2022) have an *alert* call, referred to the presence of a threat, which puts the conspecifics in alarm and makes them look for a predator, and a *recruitment* call, which has the effect to attract them to the source of the signal. When these two calls are produced together (*alert* + *recruitment*), they elicit a *mobbing sequence*, a behaviour in which the animals go to the source of the signal with the intent to mob a predator. Hence, the combination could be intended as "recruitment for an alert", or a threat (Engesser *et al*. 2016, Suzuki & Matsumoto 2022), with a meaning that can be deduced by its parts, similarly to what Leroux *et al*. (2023) claim. Suzuki & Matsumoto (2022) further examine the properties of this phenomenon in Japanese tits with playback experiments in which the alert call (ABC) and the recruitment call (D) are reproduced in different dispositions: what emerges is that the mobbing sequence is elicited only if the combination has the right order (ABC-D), while the inverse option (D-ABC) does not elicit the behaviour. Furthermore, the effect is observed only if the two calls come from the same source: if *alert* is produced by an individual, and *recruitment* by another, it does not trigger the reaction. Suzuki and Matsumoto (2022) define this property *core merge*, and give the human example of "come talk": the message "come to talk" will be understood only if both words come from the same person; otherwise, it will be intended that somebody wants the receiver to come, and somebody else wants the receiver to talk, in two separate actions. This study shows the presence of a combination of meanings which, like in human syntax, also needs some specific rules to be understood (in this case, the word order). In a recent study, Engesser *et al.* (2024) perform a similar playback experiment with Southern pied babblers (although they do not use two speakers, so that all the signals came from the same source). Interestingly, babblers were able to produce the mobbing sequence even when the alert and recruitment calls were reproduced in the wrong order, non-existent in nature, i.e. recruitment-alert. According to Engesser *et al.* (2024), this demonstrates that babblers could decompose the

combination into its constituting parts, alert and recruitment, and then interpret it correctly, showing a degree of flexibility in their perception, which they define open compositionality. The scholars give the example in English of *the wolf, beware of*, which can be understood by humans even if it is not grammatical, because we know the semantic meaning of the single words that compose the sentence, so that we are capable to repair it. According to Engesser *et al*. (2024), babblers would perform a similar cognitive operation, and this would be the proof that the alert-recruitment sequence is interpreted in a compositional way. It should be noted that there is a contradiction between the interpretations offered by Suzuki & Matsumoto (2022) and Engesser *et al*. (2024). According to Suzuki & Matsumoto (2022), the fact that Japanese tits only recognise the alert-recruitment sequence when it occurs in the correct order demonstrates that there is a "syntactic" rule that governs the combination. By contrast, according to Engesser *et al.* (2024) the ability of pied babblers to also interpret the combinations in the reversed order (recruitment-alert) demonstrates that babblers decompose the combination into its constituent parts and reinterpret it. In the two studies, the rigidness of the word order in the understanding of the sequence is used as an indicator of compositionality, but in opposite ways: in Suzuki & Matsumoto (2022), compositionality is confirmed by the presence of a rigid order, in Engesser *et al*. (2024), it is confirmed by its absence. This discrepancy should be further investigated in future.

Another formulation of compositionality, more focused on the semantic nature of the elements combined, is that of Gil (2022), which divides the binary combinations of meaningful units in two levels. The first is *bare compositionality*, and it relates to simple, unordered combinations, where the meaning is the same independently from which element comes first: he describes it in captive bonobos (Kanzi) and orangutan (Chantek), chickadees and bees. The second is *constructional compositionality*, in which the combination is between a lexical unit and what he defines as a grammatical or a configurational element, the latter being features like order, repetition, intensity, duration, which can convey additional information, although they are not traditionally considered purely linguistic (see Gil 2022). He finds this type of compositionality in Diana monkeys, putty-nosed monkeys, chickadees, Campbell's monkeys, colobus monkeys (*Colobus guereza* and *Colobus polykomos*), chimpanzees, pied babblers, and Japanese tits. The classification of Gil (2022) is not in contrast with the concept of semantic compositionality (and the animals described are the same), but gives a slightly different perspective. Furthermore, the author identifies examples of each type of compositionality in human languages, drawing interesting parallels and showing that even the simplest levels are present in human communication. However, Gil's (2022) definition of grammatical items appears to be more inclusive than that adopted

in most linguistic theories. According to his account, elements such as diminutive suffixes should be regarded as grammatical, as exemplified by the Hebrew word *teyale* ('tea.DIM'), which conveys the meaning of 'tea' with an affective nuance. This example closely resembles several constructions in Italian, such as *piccolo/piccolino* and *bello/bellino*, in which a diminutive suffix (-ino) encodes a form of endearment. Nonetheless, such elements lie at the intersection between grammar and the lexicon: they are lexical in nature insofar as they carry semantic content (e.g., *-ino* functions as a diminutive) and give rise to new lexical items (e.g., *piccolo* → *piccolino*); at the same time, they exhibit systematic and productive behaviour, and may acquire grammaticalised features (e.g., the formations are regular and predictable). Tipically, such elements are classified as lexical. However, within grammaticalisation frameworks (e.g., Haspelmath 1999), they may be viewed as peripheral or 'borderline' elements, occupying a transitional space between lexical and grammatical categories. Accordingly, while Gil (2022) classifies the *-oo* element in Campbell's monkey vocal repertoire as grammatical, in light of the observations presented here, it may be more appropriately described as a "mixed" element, exhibiting both lexical and functional properties.

It must be considered that Schlenker *et al*. (2016) also propose different analyses of the calls described in this section. While a possibility is that the calls are indeed combined in a rudimental form of syntax, there could be other pragmatic principles that determine the position of the elements: according to the Urgency principle, the call with the most urgent information (e.g. a threat) should come first; according to the Informativity principle, the most informative call should always be used unless it is false. The calls might thus not be compositional, neither in the intention of the caller nor in the comprehension of the receiver, and it would be a case of *trivial compositionality* (see Schlenker *et al*. 2023). The same scholars present two deflationary theories that should always be considered as alternative options to compositionality: *one-word* calls, in which the two elements (e.g. *krak* and *-oo*) are part of the same call, and thus do not have individual meanings, and *separate utterances*, in which there are two distinct calls but they are interpreted in a linear way (first one, then the other), without a composition (see Schlenker *et al*. 2023).

Finally, a case of combination which can be considered syntactic but does not fall under compositionality is that of non-compositional syntax, or *semantic combinatoriality* (Engesser & Townsend 2019). In these constructions, the combined units are meaningful calls, but their combinations do not produce a meaning which equals the sum of their parts. The best example is that of putty-nosed monkeys (*Cercopithecus nictitans*), that have a call, *pyow*, related to a general disturbance, and a call, *hack*, related to an eagle alert. However,

the sequences of *pyow-hack* elicit group movement (Arnold & Zuberbühler 2006, 2008, 2012). This phenomenon has been compared to human idiomatic expressions, like *pain in the neck* (see Engesser & Townsend 2019, Townsend *et al*. 2018).

The "syntactic" abilities described above indicate that nonhuman animals can combine meaningful units into more complex structures. However, they show a low degree of ***productivity***: the combinations that we analysed only refer to a few, likely hardwired, calls. To have productivity, an animal should be able to spontaneously create new signals, or at least to use its own repertoire to create new combinations to express new meanings, showing a "creative" use of its communication system; and conspecifics should be able to understand and use these new structures. Although theoretically possible, to the best of my knowledge this has never been observed. Even the most proficient captive animals showed at best a very limited productivity (e.g. Kanzi seemed to spontaneously combine two or three signs, but nothing more, see Fitch 2010). However, although the combinations seem to be restricted to only two (already existing) calls (Miyagawa & Clarke 2019), this vision is being challenged by data showing longer combinations (Girard-Buttoz *et al.* 2022), and future research should investigate whether such structures are meaningful to receivers. Furthermore, the raspberry sounds spontaneously produced by captive chimpanzees (see § 3.2), beside indicating a certain degree of vocal production learning, might be considered a case of "simple" productivity (new signals with a new function, learnable by other individuals). Among the animals described by Hockett, he assigns this property only to the waggle dance of bees (Hockett 1958), because with this system (made by a movement which indicates the direction, and a movement which indicates the distance), they can communicate the location and distance of any kind of food. The apparent lack of productivity in the communication in non-human animals results in a system that is limited in the number of messages it can convey, as opposed to the ***open-ended*** pivotal characteristic of human language (see Beecher 2021). This does not mean that the principles beyond open-endedness are necessarily unique, because it could rely on vocal production learning and simple compositional abilities, but the difference between humans and animals in this matter is one of the sharpest that I have found.

Another crucial trait of language is ***recursion***. Despite being considered a linguistic property, some evidence shows that in humans it is present also in non-linguistic domains, which are likely more ancient than language, such as auditory processes and visual perception (Martins 2014, Jackendoff & Pinker 2005, Givón 1995). According to Hauser *et al*. (2002) and Fitch *et al*. (2005), similar pseudo recursive abilities might exist also in nonhuman animals, in hearing, vision, but also orientation, the latter especially in migrating

animals (expressed by the ability to add an element to another one, e.g. fly to the nest inside the tree behind the river inside the forest behind the road, etc.). In addition, the presence of a basic recursive processing has been recently claimed for trained monkeys (Ferrigno *et al.* 2020) and crows (Liao *et al*. 2022). However, recursion seems absent in animal communication (according to Hauser *et al.* 2002, it is confined to non-linguistic domains), while humans are capable to embed sentences and create constructions with a potentially infinite level of subordination. On this subject, the most interesting result would be to find a "limited" recursion in animal communication, e.g. a hierarchical structure organised in three or more levels. An imaginary example of what should be searched is the following. Let us assume that Japanese tits had also signals dedicated to the predator type, e.g. eagle and snake, and that they combined these signals with their (already existing) alert call, producing different alarmed reactions (e.g. looking up vs looking down). In this case they would have *alert* + *eagle* and *alert* + *snake*, and it would be a case of semantic compositionality (like the documented *alert* + *recruitment*). Then, let us imagine that the animals further combined these two compounds with a recruitment call, creating *alert-eagle* + *recruitment* and *alert-snake* + *recruitment*, which could be interpreted as "recruitment for an eagle threat" (generating a mobbing sequence towards the air) and "recruitment for a snake threat" (generating a mobbing sequence towards the ground). In this case, there would be the extremes to hypothesise an embedded phrase with some kind of recursion, or a non-recursive cascade effect (limited, but crucially more complex than a combination of two units), which in generative frameworks would be represented hierarchically.[75] This example is a gross generalisation, but finding something like that would challenge the idea that recursion *per se* is uniquely human: considering that real infinite recursion is not possible, then the difference between two hierarchical levels or more could be considered quantitative instead of qualitative. Research should investigate this topic further to verify whether similar processes are present in animal communication.

The hypothetical scenario described above appears to be in contrast with the generative concept of *recursive merge* (see Part I, § 2.1), which considers the compositional ability inseparable from "infinite" recursion, seen as a uniquely human, "all or nothing" trait, which cannot exist in intermediate forms (Berwick & Chomsky 2019). For how this operation is conceived, it is not possible to find it outside humans. However, the "syntactic" abilities of nonhuman animals discussed above, which are limited to two elements, suggest that compositionality can indeed exist without recursion. This is further supported by the

[75] However, it might also be considered a cascade effect of more 1-merge systems (see Rizzi 2016), although the difference with 'proper' recursive systems is blurred.

classification presented by Rizzi (2016) (see Part I, § 2.5), which shows that, even from a generative perspective, the operation merge can be seen as having different levels of complexity. The simplest, 1-merge, seems to be equivalent with the definition of compositionality. Rizzi (2016) observes that 1-merge can describe the compositional skills of Campbell's and Diana monkeys analysed in Schlenker *et al*. (2016), but the same reasoning can apply to more recent studies on chimpanzees (Leroux *et al*. 2021, 2023), pied babblers (Engesser *et al*. 2024) and Japanese tits (Suzuki *et al.* 2016, Suzuki & Matsumoto 2022), and to the sign combinations produced by captive apes (Greenfield & Savage-Rumbaugh 1990, Miles 1990). According to Rizzi (2016), for how merge is conceived in the Minimalist Program it is difficult to deny that the call combinations of primates and birds that we discussed are not a case of merge: "[Merge] takes two elements a, b already constructed and creates a new one consisting of the two; in the simplest case {a, b}". (Chomsky 2001: 6). Hence, its simplest form (1-merge) might be shared with other species.

All the features that we discussed in this section are related to syntax. In humans, a fundamental aspect of syntax is the presence of ***hierarchical structures*** which convey complex meanings, while in animal communication hierarchy has been proposed only for "holistic" messages like songs of birds and humpback whales (Suzuki *et al*. 2006) which are closer to phonology (see § 3.3.2). Albeit hierarchical syntax is not found in other species, we saw in this section that simpler instances of syntactic combinations are probably present in nonhuman animals, even phylogenetically distant from humans like birds (so that this could be a case of convergent evolution). Furthermore, Townsend *et al*. (2018) suggest that also humans have several forms of simple, non-hierarchical, and nonproductive syntax, that can be effectively compared with the operations of other animals. It should also be considered that, like recursion, also hierarchical structuring seems to be present in other domains, like in the social organisation of several species (e.g wolves, ants, see also Tibbets *et al.* 2022). These types of hierarchy cannot be compared to the hierarchical syntax of human language, but they suggest that simpler mechanisms that follow similar basic principles might exist in nature. The hierarchical structuring of language reveals the property of ***structure dependency***, which is not explicitly claimed in the literature about animal communication as a whole. However, some aspects related to it can be found in nonhuman species.

A crucial feature is ***word order***, which does not necessarily reflect a linear understanding of the units, but is due to the linear nature of the sound stream. If we consider the definition that a different disposition of the same elements can convey a different message (e.g. *The dog bites the cat* vs *The cat bites the dog*), this ability is mostly found in captive animals. The most interesting example is again Kanzi, which according to Fitch (2010: 167) "uses

word order in spoken sentences to disambiguate meanings (e.g. correctly enacting "make the doggie bite the snake" and "make the snake bite the doggie" with dolls)." In cases like this one, the understanding of different word orders might suggest the presence of an underlying syntactic knowledge. In the wild, there are many species in which the disposition of the elements in a sequence plays a role for the meaning, like bottlenose dolphins (Marino *et al*. 2007, see Part III, § 3-4 below), Japanese tits (Suzuki *et al.* 2016, Suzuki & Matsumoto 2022, widely discussed above) and putty-nosed monkeys (according to Gil 2022). Similarly, most of the combined units described above as compositionality follow a rigid order (e.g. Campbell's monkeys, Diana monkeys, pied babblers) (see Gil 2022 who defines them constructional compositionality).

A more complex aspect of word relationships are ***non-adjacent dependencies*** (NAD). Traditionally, this phenomenon has been studied by performing experiments on trained specimens, to verify if they are able to understand relationships between distant elements: this ability opens to the possibility that they can process deeper hierarchies like humans. This approach is preferred because, in the current state of the art, there is no clear evidence of animal call combinations made by more than two units, therefore NAD would be absent by nature. However, as we discussed, new studies are suggesting the presence of combinations of more than two units (Girard-Buttoz *et al.* 2022); hence, further research of NAD should investigate them in animal repertoires. As of today, evidence comes only from captive individuals: the tests are made with artificial grammars much simpler than human language, and focus on the ability to perceive non-adjacent auditory (or visual) stimuli, and not meaningful units. Among the animals which gave promising results there are some bird species like Japanese tits and zebra finches (*Taenopygia* spp.) (see Abe & Watanabe 2011, Chen & ten Cate 2017) and some nonhuman primates like squirrel monkeys (*Saimiri sciureus*) (Ravignani *et al*. 2013), tamarins (*Saguinus oedipus*) (Newport *et al.* 2017), common marmosets (*Callithrix jacchus*) and chimpanzees (Watson *et al*. 2020, see also Wilson *et al*. 2020). In the experiment of Ravignani *et al*. (2013), for example, experimenters taught some squirrel monkeys a finite-state language ABnA, by using two acoustic signals which were familiar to the monkeys: one high (A) and one low (B). The rule to be learned was that A always came first, followed by a variable number of B, and then A again (e.g. ABBBA, ABA, ABBA, etc.). The sequence had no semantic content, but it was repeatedly reproduced and associated with a prize between one section and the other. After the training, the monkeys were exposed to familiar and unfamiliar stimuli, and the results suggest that instead of focusing on the pitch of the sounds, they focused on the relationship between the elements, and were able to generalise the rules, applying them to novel constructions. This

led Ravignani *et al*. (2013) to suggest that the ability to detect non-adjacent dependencies is older than language, and it has to be tracked back to an ancient common ancestor of humans and monkeys, a conclusion shared also by Watson *et al*. (2020).

What I have just described is an example of ***rule generalisation***. If a trained animal manages to recognise as correct the same pattern which it has learned, but in a sequence never heard before, then there is reason to believe that it has generalised and reapplied the rules following a key mechanism of human language and cognition. This ability has been claimed to be present in nonhuman primates (an animal which detects a non-adjacent dependency is also capable of generalisation, see Watson *et al*. 2020), in some birds like budgerigars (*Melopsittacus undulatus*) (see Petkov & ten Cate 2020), Bengalese finches (*Lonchura striata domestica*) (proposed by Abe & Watanabe 2011, criticised by Berwick & Chomsky 2016 and Maes *et al*. 2015), and in mice (de la Mora & Toro 2013). Some of the studies are focused on simple artificial grammars: de la Mora & Toro (2013), for example, trained mice with a "language" with invented vocal words composed by three vowels and three consonants, in two different orders: AAB (the first two with an identical vowel and the third different, e.g. *tabaki*) and ABC (all the vowels different, e.g. *tebaki*). The mice had to push a button that gave food, but only if it was pushed after hearing the correct sequence AAB. After the training, the mice were exposed to new words with vowels and consonants never heard before, and they still were able to detect the correct sequence AAB and receive the prize (de la Mora & Toro 2013). These examples, together with the other features discussed in this section show that there are many complex abilities related to language which might be present in nonhuman cognition, but they are difficult to identify. Now I focus on another fundamental layer of language, that of meaning.

3.4 Meaning

The concept of meaning has been reconsidered in animals: initially it was believed that the reactions of other species were merely determined by automatic reflexes with no form of understanding, in the framework of behaviourism (Skinner 1976). Now it is accepted that animal cognition is more complex, and involves some aspects originally believed to be uniquely human (see Fitch 2010). Many of the aspects of meaning (also defined *function*) have already been discussed in the section related to the structure, and in particular morphosyntax, because the two features are strictly connected, and they are part of the complex nature of language. Here we focus more precisely on the semantic properties of these signals. It should be noted that until recently, researchers in animal communication

used the function of a signal as a proxy for its meaning (Berthet *et al.* 2023). However, this can be inaccurate, because the function does not necessarily correspond to the meaning: for example, the function of an alarm call is to warn conspecifics of danger, but the meaning might be "watch out!" or "there is a predator" (Berthet *et al.* 2023).

Like humans, other species have a ***lexicon***, i.e. a repertoire of signals that are used in specific contexts and could be attributed different sets of meanings. Following this definition, all communicating animals have a specific lexicon. However, the repertoires of nonhuman animals seem much more limited in contrast to the thousands of words forming the human lexicon. Furthermore, it is not easy to analyse the possible distinction between lexical and functional words: most of the calls should be considered "lexical", being related to a meaning (e.g. the *krak* call of Campbell's monkeys, referred to a terrestrial threat), and there are no clear examples of grammatical items. [76] In any case, meaningful signals can express the key property of ***referentiality*** (although not all meaningful signals are referential). Referentiality is evident in trained animals like the already discussed captive bonobo, orangutan, and parrots, which show the ability to understand new pairings of form and meanings, and seem to associate human words with external events. In the wild, vervet monkeys produce different alarm calls in the presence of different predators, i.e. leopards, eagles and snakes. In their seminal study, Seyfarth *et al*. (1980a, 1980b) played back those calls when the predators were not present, showing that the receivers still performed the corresponding behaviours (hiding on trees for leopard calls, looking at the sky for eagle calls, and looking down for snake calls). This indicates that there was a clear relationship between the signal (signifier) and its meaning (signified), although it was not certain whether the meaning of the calls was "run" or "snake/eagle/leopard". For this reason, Seyfarth *et al*. (1980a) introduced the term "functionally referential" to describe them. However, a subsequent work on Diana monkeys demonstrated experimentally that the calls of these animals, who have the same kind of alarm system of vervet monkeys, are mentally represented as "leopard" vs "eagle" (Zuberbühler 1999, 2000), and thus can be considered referential following the definition of Hockett (1960a). The property of referentiality has been explicitly proposed for some birds (reviewed by Gill & Bierema 2013, Lapergola *et al*. 2023); for several primates like vervet monkeys (Seyfarth *et al.* 1980a, 1980b), Diana monkeys (Zuberbühler 2000), and chimpanzees (Watson *et al*. 2015, Slocombe &

[76] Another aspect that should be explored is the presence of predicative structures. For an existential interpretation of signals – which might also be considered predicative – such as Japanese tits' alert and recruitment, see Schlenker *et al.* 2024).

Zuberbühler 2005, see also Townsend & Manser 2012); cetaceans (see Zamorano-Abramson *et al*. 2023 for a general review), and some rodent species (see Macedonia & Evans 1993).

A concept strictly connected with referentiality is that of ***arbitrariness***, which relates to the lack of a natural correspondence between a signal and its signified (Hockett 1958). Watson *et al*. (2022) highlight that arbitrariness is not absolute (present or absent), but can be seen as a continuum which ranges from its absence (the presence of fixed, instinctive, and non-modifiable signals, i.e. signals "stuck" with their meanings) to its full potential (fully arbitrary signals). They define this novel framework *optionality*, or the "capacity to associate alternative functions with a signal, or alternative signals with a function" (Watson *et al*. 2022). Optionality also encompasses aspects of signal learning (both usage and production, see § 3.2).

At the lowest level of arbitrariness there is ***iconicity***, that presents the strongest bond between the sign and its meaning, which have a natural connection. Hence, to be iconic, a signal should represent the element it refers to, e.g. an alarm call that reproduces the sound of a predator, or a gesture that mimics the act of eating to communicate about food. While this is considered the simplest level, it does not seem to be common in nonhuman animals. Fitch (2010) argues that cases of imitation and pantomime could be considered iconic, and might have been a first step towards the development of a gestural protolanguage. However, although some animals have the ability to imitate conspecifics and humans (e.g. parrots, Pepperberg 1999), they don't seem to use this skill to produce iconic signals. Nonhuman primates confirm this trend: according to Tomasello *et al*. (2019), they do not do iconic gestures, but some trained apes seemed capable to understand them (Bohn *et al*. 2016). A possible example of iconicity comes from the bees' waggle dance, which according to Bickerton (2014) is in part iconic (in the movement related to the distance of the resource), and in part indexical (in the movement related to the direction).

Indexicality is much more spread, being related to a sign that needs the context to be understood. Fitch (2010) describes as indexical all the animal communication systems with context-specific calls (e.g., vervet monkeys). Another property considered indexical, both in animals and humans, is the ability to point and understand *pointing*, which has been observed in several mammals but not in birds (but see Kaplan 2011). Production and comprehension of pointing have been described in great apes (Tauzin et *al*. 2020) and bottlenose dolphins (Xitco *et al*. 2001), while the sole comprehension is described in elephants (Smet & Byrne 2014), monkeys, bats, and several species of domesticated animals as dogs, cats, goats, pigs, ferrets (see Krause *et al*. 2018 for a complete review). A recent study (Wilke *et al*. 2022) documented a wild female chimpanzee performing a referential gesture, i.e. showing an item

to her mother (a leaf in the context of grooming): according to the authors, this attention-directing behaviour might have no other goal than the desire of sharing, in a way qualitatively similar to humans. Like pantomime, pointing is considered one of the first steps in the evolution of language by some scholars (e.g. Tomasello 2008).

The highest level of arbitrariness, ***symbolism***, is traditionally considered uniquely human (Berwick & Chomsky 2016). The presence of it in hominins is highly debated (see e.g. Longa 2013, Tattersall 2016). This property has been recently proposed for sperm whales, whose identity codas might contain a symbolic reference to their cultural groups (Hersh *et al*. 2022), although this would not be comparable to the much more complex human shared symbolic constructions. The idea of culture in animals is debated, and mostly revolves around the question of how much complex a system should be to be defined "cultural" (see e.g. Whitehead & Rendell 2015). Like many other concepts (e.g. open-endedness, recursion), symbolism cannot be found to its full extent in nonhuman animals; however, not all animal signals need the context to be understood, and thus cannot be considered fully indexical. For example, trained animals can learn arbitrary signals referred to entities, actions, and objects (Pepperberg 1999, Savage-Rumbaugh 1986, Herman *et al*. 1984, see also Bruck *et al*. 2022), which could be considered symbols, without the need of having a symbolic culture. In the wild, the already discussed alarm calls (e.g. in vervet monkeys) trigger their associated behaviour even in the absence of predators (Seyfarth *et al*. 1980a, 1980b). All these animals are likely capable to understand the signal even if they don't see the actual referent: this suggests that some animals have a degree of arbitrariness that goes beyond indexicality, although it might not be considered symbolism (certainly not according to Berwick & Chomsky 2016, who explicitly deny it).

This issue is related to the problem of ***mental representations***, which have been proposed for several animals, although they are difficult to demonstrate (see Premack 1984, Hauser *et al*. 2002, Emery & Clayton 2004); they have been discussed more in the field of philosophy (see Saidel 2009, see also Thornton *et al*. 2012). The most careful approach is to admit this possibility, without claiming their presence. It seems reasonable that, for example, when a chimpanzee hears a food call, the representation of food is triggered in their mind. The same reasoning could be applied to several other species. A first convincing evidence for mental representations in nonhuman animals comes from a prime-probe experiment performed on Diana monkeys (Zuberbühler *et al.* 1999). The researchers used playback experiments, presenting the monkeys with recorded predator vocalisations and male alarm calls. The stimuli consisted of two categories: real predator sounds, such as leopard growls and eagle shrieks, and alarm calls that male Diana monkeys naturally produce when encountering these

predators. Previous studies had shown that when a male emits an alarm call for a particular predator, females respond with their own matching alarm call, suggesting that they recognise the meaning behind these signals. To test how Diana monkeys process alarm calls, the researchers played two sounds in sequence, separated by five minutes of silence. The experiment was structured into three conditions. In the baseline condition, the same sound was played twice (e.g., an eagle shriek followed by another eagle shriek). If monkeys were simply responding to the acoustic stimulus itself, their reaction should diminish upon hearing the second call, as they had already processed the information. In the test condition**,** a male's alarm call was followed by the corresponding predator's vocalisation (e.g., an eagle alarm call followed by eagle shrieks). If monkeys interpreted the alarm call as a meaningful warning about the predator, they should become habituated and react less to the second sound. Finally, in the control condition, a male's alarm call was followed by the vocalisation of a *different* predator (e.g., a leopard alarm call followed by eagle shrieks). If the monkeys were responding only to acoustic features, their reaction should decrease. However, if they understood the meaning of the alarm call, they would still react strongly to the unexpected predator sound. The results provided compelling evidence that Diana monkeys process alarm calls in a conceptual way. In the baseline condition, the monkeys showed habituation, responding less to the repeated predator sound, which confirmed that they recognised it as the same information. In the test condition, they also exhibited reduced responses when an alarm call preceded the matching predator sound, suggesting that they had already anticipated its presence. However, in the control condition, their response remained strong, indicating surprise when the predator sound did not match the prior alarm call. These findings led the authors to conclude that Diana monkeys possibly had a mental representation or concept of the predator, associated to both the alarm call and the actual predator vocalisation. Other studies suggest that animals have concepts (see Bruck *et al*. 2022, Zentall *et al*. 2014, Fitch 2010). A recent research suggests that even dogs (*Canis familiaris*) have mental representations associated with words, and this ability to understand referential signal has been measured with brain scans (Boros *et al*. 2024). This might support the idea of Bruck et al. (2022: 4) that "labels commonly occur in animal minds", although their clear identification presents several problems.

A natural consequence of these levels of arbitrariness is ***equivocity*** (Berruto & Cerruti 1997): it is found in the animals that have a signal which is related to more than one function, and/or more signals related to the same function. It is a characteristic present, for example, in the species capable of vocal usage learning that associate an existing call to a new function (e.g. some nonhuman primates or bird species, see Vernes *et al*. 2021, Watson *et al*. 2022).

A demonstration of equivocity is the already cited experiment of pig-tailed macaques (Coudé *et al.* 2011), which learned to produce their social signal 'coo' also to request food: one signal, two functions.

Arbitrariness, symbolism and mental representations relate to the property of ***displacement***, or displaced reference, which allows to communicate about elements outside of the context. It is the opposite of indexicality, and is not clearly demonstrated in any nonhuman species: on the one hand, several animals seem to understand the absence of a labeled object, maintaining a representation of it when it is not present. On the other hand, the active production of displaced reference is more controversial, and at best very rare: according to Bickerton (2014), it is present in bees and ants, which can communicate to conspecifics the presence of distant food, and possibly in common ravens (*Corvus corax*) (see also Heinrich & Marzluff 1991). It has been also proposed for wild orangutans (Lameira & Call 2018), captive great apes (Lyn *et al*. 2014, Bohn *et al*. 2015), and grey parrots (Pepperberg 1999). However, further studies are needed to claim the presence of this property.

Related to displacement is the ***detachment from stimuli*** (Berruto & Cerruti 1997): if the signal produced by an animal is not triggered by internal or external stimuli (e.g. hunger or danger), then it would be "detached". This property could be researched in behaviours like requests of play, or attention not connected with primary needs. Like mental representations, it is very difficult to be demonstrated, and to my knowledge there are no studies that specifically address this issue.

A property more widespread in animals is ***specialisation***: it applies to all animal calls which elicit a behaviour without the presence of their referent: for example, in vervet monkeys, hearing an alarm call triggers the same behaviour triggered by seeing the predator it refers to, albeit the call itself is not an actual threat (Seyfarth *et al*. 1980). Specialization is a matter of degree: when there is a strict relationship between trigger and reaction (e.g. seeing the predator and running), the system is not specialised; the more this connection gets subtle (e.g. the sound waves of a call that alert an animal), the more the system is specialised. This property is different from displacement because it is still related to the context.

Another important property is ***prevarication*** (Hockett 1966). Deception is widespread in the animal kingdom (e.g. an animal pretending to be dead, or camouflage), but "intentional" deception is much more rare, especially when it relates to vocal communication: in the wilderness it has been proposed for orangutans, which seem to use leaves on the mouth to modify their whistles to reach an effect of size exaggeration (Hardus *et al*. 2009), and for fork-tailed drongos, which mimic the alarm calls of other species, producing false alarms

that scare the targets away allowing the deceivers to steal the food (Flower *et al*. 2015). Furthermore, drongos change their fake alarm calls to avoid habituation (Flower *et al*. 2015). In captive animals there are several cases of tactical deception among great apes, which might for example pretend to be tired to avoid some examination (see Karg *et al*. 2015, Whiten & Byrne 1988, Byrne & Whiten 1992) and birds, in particularly referring to the grey parrot Alex, which was reported to have intentionally given wrong answers to known questions in front of strangers, and then give the correct answers in their absence (Pepperberg 2009). A deceptive behaviour has been observed even in dogs (Heberlein *et al*. 2017). The most complex cases of deception presuppose the understanding of the perspective of other individuals, and thus – possibly – a Theory of Mind (see Part I, § 4.5.1 above, and § 3.5 below).

The last properties discussed here are ***reflexiveness*** (Hockett 1966) and ***semantic omnipotence*** (Berruto & Cerruti 1997), which both seem absent in nonhuman animals: there is no evidence to date showing that other species use their communication system in a reflexive way, to talk about the communication system itself; similarly no animal communication system appears to grant the potential to express any kind of meaning (also because the possible meanings are quite limited). Finally, despite the abilities of productivity and compositionality have been described in the section "*Structure*" above, they also relate to the meanings: the two layers of form and function cannot be separated, so the operations which combine syntactic objects and allow to produce more complex structures can in turn also combine semantic elements and produce more complex meanings.

3.5 Contextual usage

The final layer of language that I discuss is that of context. Like for humans, I focus on the main aspects of pragmatics, i.e intentionality, signal exchanges and communicative functions.

3.5.1 Intentionality

As we have seen, ***intentionality*** is the fundamental requirement for a complex communication. Most of the signals produced by animals are not intentional, because they are genetically determined responses triggered by internal or external stimuli (Skinner 1976). Although difficult to empirically demonstrate (Hauser *et al*. 2002), intentionality has been discussed in several species (see Fitch 2010, Engesser & Townsend 2019, Slocombe & Zuberbühler 2007, Zuberbühler 2008). Dennett (1983) identifies three levels of

intentionality: zero-order (no real intentions, reactions determined by instinctive reflexes), first-order (the individual has beliefs and desires, thus intentions, and even the will to modify the behaviour of others, but lacks the understanding of other minds), and second-order (the individual also understands the mental states of other individuals). Zero-order intentionality is present in nonhuman animals, and consist in the lack of intentionality. First order intentionality is present, for example, in the case of an animal that produces an alarm call to make other conspecifics escape from a threat, with the goal of eliciting their movement (Schel *et al*. 2013). It is believed to exist in great apes, both in gestures and – following more recent observations – in vocalisations, to a certain extent (Graham *et al*. 2019, Zuberbühler 2018, Schel *et al*. 2013). A way to understand whether a signal is intentional is to verify 1) if it is directed toward a receiver, 2) if the signalling individual checks that the receiver is giving attention to it, 3) if it waits for a response, and then, in the case of lack of response, 4) if it continues to produce more signals (Graham *et al*. 2019, see also Townsend *et al*. 2016). All great ape species showed to fulfil these criteria, but Graham *et al*. (2019) suggest that first-order intentionality might be present also in other animals with complex audience effect, i.e. that modify their signalling depending on the presence of receivers, and in some species also depending on *who* the receivers are (e.g. males or females). This expands the list of potential candidates to animals like ravens, dogs, squirrels, and even some fish species (Graham *et al*. 2019).

It is possible that other animals do not fulfil all the criteria for first-order intentionality, but still show some traits which put them above the total lack of intentions, or zero-order intentionality: for example, Coudé *et al*. (2011) taught pig-tailed macaques to associate their natural social call 'coo' to a request of food to humans, and noticed that the calls produced in the new context elicited a brain activity *before* the call, which did not happen in their natural, instinctive use, "indicating pre-planning or voluntary control over the production" (Watson *et al*. 2022: 9). However, this level of intentionality is not enough: according to Grice (1975), for a real communication it is necessary that both sender and receiver have a representation of each other's mind, so a form of second-order intentionality should be required. To have second-order intentionality, an animal would have to produce, for example, an alarm call to inform another animal of the presence of a predator, thus changing its state of knowledge, and not only the behaviour (Schel *et al*. 2013). This has not been confirmed in any nonhuman species, although Crockford *et al.* (2017), by replicating the experiments of Schel *et al*. (2013), hypothesise that chimpanzees show a certain extent of it (i.e producing alarm calls with the goal of informing other individuals of the presence of a snake).

Related to intentionality is the ***Theory of Mind*** (Premack & Woodruff 1978, Baron-Cohen 1991), whose presence in nonhuman species is an open issue (see Krupenye & Call 2019, Fitch 2010, for a general overview): signs of it are observed in chimpanzees (Call & Tomasello 1998), bonobos (Krupenye *et al*. 2016), grey parrots (Pepperberg 1999), corvids (Bugnyar *et al*. 2016). It must be observed that it can be conceived as a gradual continuum, so that some species might have only some aspects of it: even in humans the understanding of other minds and beliefs develops after the first years of life (Fitch 2010). The most complex form of the Theory of Mind, which coincides with second-order intentionality, is ***shared intentionality***, i.e. the ability not only to understand the minds and the will of other individuals, but to use this knowledge to cooperate with them in an effective way to achieve a common goal (Tomasello 2008). Although many species are highly social and cooperative, this ability is considered uniquely human (see O'Madagain & Tomasello 2021), and according to Tomasello (2008) evolved only in our species through interaction and social cognition, representing the most important skill behind language. Some clues of it might be present in chimpanzees (Duguid & Melis 2020) and bonobos (Krupenye & Call 2019). In other animals it has not been found in any form, or not even explored.

3.5.2 Signal exchanges

Another key aspect of pragmatics are the signal exchanges, which are present in nonhuman animals without the need of complex intentionality. An aspect of signal exchanges is the ***turn-taking*** behaviour, with an alternation of sender and receiver roles, which is present in many species from early in life. It is believed to be an ancient mechanism underlying the evolution of language (Pika *et al.* 2018). Among mammals it is observed in nonhuman primates, cetaceans, bats, elephants, naked-mole rats (*Heterocephalus glaber*) and meerkats; it is also present in several bird species and even in anurans and insects (see Pika *et al*. 2018 for a general overview, see also Vernes *et al*. 2021). One well-known example of turn-taking behaviour is observed in certain bird species during duet singing, which has the function to reinforce pair bonding and territorial defense (Pika *et al*. 2018). In nonhuman primates, there is evidence that antiphonal calling (i.e. an individual emits a call and another responds) is influenced by learning and incentivised by parents instead of overlapping (Pika *et al*. 2018). Pougnault *et al.* (2021) reviewed turn-taking in great apes, observing that the type of vocal interactions was related to the type of social interactions: more competitive species like chimpanzees had a higher degree of isolated calling and overlapping, while turn taking was more spread in tolerant species like bonobos and gorillas,

and likely enhanced group cohesion. Similarly, turn-taking of vocalisations in cetaceans facilitates coordination and social interactions (Pika *et al*. 2018). According to Ravignani *et al*. (2019), turn-taking can be considered a subset of chorusing, i.e. the synchronous coordinated production of the signals, and it need not necessarily be learned nor be intentionally cooperative: this approach can give a wider perspective on this behaviour and facilitate comparison between different species.

Another property related to pragmatics is ***interchangeability*** **(**Hockett 1958), according to which the same signal can be sent and received by any member of the species. This is the case of many animals (e.g. several primates, birds, cetaceans, etc.); however, there are also many species that lack this ability: a classic example are the animals that produce songs for mate attraction, generally restricted to males (e.g. songbirds, humpback whales), and not reproducible by females. Conversely, among meaningful, compositional vocalisations, only female Diana monkeys produce their complex calls, but not males (see e.g. Engesser & Townsend 2019).

According to Arbib (2005), ***parity*** means that a message must have the same meaning for both sender and receiver. We find this property present in most of the signals described until now: for example, the combination *alert* + *recruitment* (see § 3.3.3) has the same meaning for all the Japanese tits. Parity is not debated, but it is an important step towards more complex imitations and pantomimes that, in Arbib's (2005) theory, led to language; thus, it is a fundamental prerequisite, judged a key trait of communication also by Grice (1975).

Finally, a prerequisite for an effective communication is the ***cooperative principle***, i.e. the cooperation between sender and receiver, which agree to some rules. While this might seem to be present in nonhuman animals, albeit in an unintentional form (as it is shown by turn-taking, interchangeability and parity), the cooperative principle of Grice (1975) requires an active contribution of cooperation which appears absent in animals. According to Fitch (2010), in nonhuman species the receivers do most of the work, inferring the meaning from the senders which do not really make an effort to be informative. Animal communication is asymmetric, while human communication is symmetric, made by a constant adjustment of both interlocutors (Fitch 2010). The Gricean idea of communication is thus rooted in the Theory of Mind and in second-order intentionality, and for how it is conceived it cannot be found in nonhuman species (Fitch 2010). However, it can be interesting to briefly consider the conversational maxims of Grice (1975) from the general perspective of animal communication: *quantity* (be informative: don't say too much or too few) is not always respected, as animals tend to produce repetitive signals in variable quantity; *quality* (be

truthful) is usually respected, but we have seen that some species in some contexts can give false information; *relation* (be relevant: only pertinent information) is hardly respected, in the measure that animals produce several sounds and calls which might contain redundant or conflicting information (or no information at all); *manner* (be clear) is respected in some cases, to the extent that some signals have a clear meaning and elicit a clear response, although this cannot be affirmed for all signals. The Gricean maxims are a concept built on the complexity of human language, and thus is difficult to apply them to animal communication.

3.5.3 Pragmatic functions

Language is used to fulfil some pragmatic functions. Communication of nonhuman animals assumes different functions and takes many shapes according to the context. This is expressed by ***multimodality***. Most animals, like humans, are not limited to vocal communication, but produce signs in several ways, from movements to chemical signals (and many animals do not have a vocal communication at all, relying on other features). The coexistence of different signals is also documented: multimodal communication has been described, for example, in apes and monkeys (Leavens *et al*. 2004), and in some cetacean species, which combine vocalisations, kinesthetics, and touch (see Zamorano-Abramson *et al*. 2023 for a review). The multimodality emerges also in the ability of cross-modal recognition and imitation: interestingly, the only animals considered capable of cross-modal imitation of vocal and gestural signals are some bird species (budgerigars and grey parrots) and cetaceans (Zamorano-Abramson *et al*. 2023).

Related to multimodality is the ability of ***modality-independence*** (Berruto & Cerruti 1997), i.e. the possibility to transmit the same message through different channels (e.g. to have a vocal signal and a gestural signal which can be used interchangeably with the same meaning). This does not seem to be accessible to nonhuman animals, neither in the wilderness nor in captivity. At best, this ability could be present in their cognition, and might be connected to cross-modal recognition.

Multifunctionality, instead, is a property present in several animal communication systems: it is observed in species that use their repertoire of signals to express different functions, e.g. sexual selection, social interaction, expression of internal states, identity, etc. For example, songbirds have vocalisations used for courtship, and others for territorial defence; some primate utterances can communicate emotions like fear, and others can indicate identity, or threats; similarly, bottlenose dolphins use their vocalisations to

coordinate with conspecifics, to call other individuals, to stun the prey, for orientation, etc. (see e.g. Jones *et al.* 2020). In fact, it would be probably more peculiar if a communication system conveyed only a single type of information. Among the six main functions of human language (Jakobson 1960, see also Berruto 1976), some of them seem to be present in nonhuman species: the *referential* function is present in all the animals with referential signals (see § 3.4 above); the *poetic* function is absent, unless we consider the complexity of the courtship songs of birds as a form of "poetic" enrichment, but this definition would be too far-fetched; the *emotive* function is present in several species, in the measure that utterances can communicate internal states; the *conative* function does not seem to be present (but see Part III, § 5, for a discussion of this function in dolphins); the *phatic* function can be considered present in those animals who use calls to maintain contact and social bonds, even if these signals are not meaningful; the *metalingual* function, close to the linguistic property of reflexiveness, seems to be absent. Overall, some aspects of the multifunctionality of human language and animal communication converge, although the complexity of the possibilities which derive from human language appears unparalleled.

Table 7 summarises all the linguistic properties described in this chapter.

<table>
<tr><th colspan="2">Animal linguistic traits</th></tr>
<tr><td colspan="2">Signal
Physical aspects of the sign</td></tr>
<tr><td>Vocal-auditory channel
All animals with vocal communication</td><td>Rapid fading
All animals with vocal communication</td></tr>
<tr><td>Total feedback
All animals with vocal communication</td><td>Linearity
All animals with vocal communication</td></tr>
<tr><td colspan="2">Broadcast transmission and directional reception
All animals with vocal communication</td></tr>
<tr><td colspan="2">Signal learning
Ability to learn new signals and use old signals in new contexts</td></tr>
<tr><td>Vocal contextual learning
• Birds (e.g. parrots, chickens, budgerigars, mynahs, ravens)
• Nonhuman primates (e.g. vervet monkeys, chimpanzees, gibbons, macaques, lemurs)
• Marine mammals (e.g. cetaceans, pinnipeds, sea lions)
• Land mammals (e.g. elephants, bats, dogs, cats, rats)</td><td>Vocal production learning
• Birds: parrots, hummingbirds, songbirds
• Nonhuman primates (to a lesser extent: captive chimpanzees)
• Marine mammals: cetaceans, pinnipeds
• Land mammals: bats, elephants</td></tr>
</table>

<table>
<tr><td>Cultural transmission
(depends on the definition of culture)
• Birds (e.g. songbirds)
• Nonhuman primates (e.g. chimpanzees, orangutans)
• Cetaceans (e.g. orcas, dolphins, humpback whales)</td><td>Learnability
(learning of a new “language”)
• Wild orcas
• Captive birds (e.g. Alex the parrot)
• Captive primates (e.g. Kanzi the bonobo)
• Captive bottlenose dolphins</td></tr>
<tr><td colspan="2">Structure
Complex structures created by combining smaller units</td></tr>
<tr><td colspan="2">Duality of patterning (double articulation)
No proof in animals, but partial clues in nonhuman primates, birds, cetaceans</td></tr>
<tr><td colspan="2">Phonology
Combination of meaningless units</td></tr>
<tr><td>Discreteness

Controversial: not openly described, but discrete signals are present at least in:

• Birds
• Nonhuman primates</td><td>Combinatorial ability
• Phonocoding: songbirds, hyraxes bats, gibbons, whales
• Multi element calls: birds (titmice, chickadees, tits, babblers)
• Temporal structures: Colobus monkeys, Mexican free-tailed bats, pied babblers (cetaceans?)
• Stochastic/proportional structures: titi monkeys (bonobos?)</td></tr>
<tr><td colspan="2">Morphosyntax
Combination of meaningful units</td></tr>
<tr><td>Compositionality
• Segmental concatenations: Diana monkeys, Campbell’s monkeys, meerkats, mongooses, dingoes
• Affixation: Campbell’s monkeys, pied babblers
• Semantic compositionality: chimpanzees, Japanese tits, pied babblers, captive birds (parrot), primates (bonobo, orangutan)</td><td>Non-compositional syntax

• Semantic combinatoriality: putty-nosed monkeys</td></tr>
<tr><td>Productivity
No animal: bees (?), captive bonobo (?)</td><td>Open-endedness
No animal (only without meaning)</td></tr>
<tr><td>Recursion
No animal (possibly in non-linguistic domains, e.g. vision, orientation)</td><td>Merge
If accepted as non-recursive, it is comparable with compositionality</td></tr>
<tr><td>Recursive merge
No animal</td><td>Hierarchical structuring
No animal in its full complexity (but birds and primates show simpler forms)</td></tr>
<tr><td>Structure dependency

No animal in its full complexity (but birds and primates show simpler forms)</td><td>Importance of word order
• Nonhuman primates (e.g. Diana, putty-nosed, Campbell’s monkeys)
• Birds: Japanese tits, pied babblers
• Cetaceans: bottlenose dolphins</td></tr>
</table>

Non-adjacent dependencies • *Birds: Bengalese and zebra finches* • *Nonhuman primates: chimpanzees, common marmosets, tamarins, squirrel monkeys*	Rule generalisation • *Birds: Bengalese and zebra finches, budgerigars* • *Nonhuman primates (e.g. tamarins, chimpanzees, squirrel monkeys)* • *Cetaceans: bottlenose dolphins* • *Other mammals: rats*
Meaning *Presence of meaning in language structures (semantics)*	
Lexicon (words) *All animals with a call repertoire*	Referentiality *All animals with functionally referential calls (e.g. alarm calls)* • *Birds, primates, cetaceans, rodents* • *Evident in trained animals: parrot, bonobo, orangutan*
Arbitrariness • *Low: all animals with referentiality* • *High: no animal (see optionality)*	Equivocity • *Animals with vocal usage learning (e.g. birds, nonhuman primates)*
Iconicity • *Almost no animal (only bees)* • *Understanding: some trained apes*	Indexicality • *All animals whose vocalisations reveal sex, size, etc.* • *Species with context-specific calls* • *Species able to point: great apes, bottlenose dolphins* • *Species which understand pointing (e.g. primates, cetaceans, domesticated dogs, cats, pigs, etc.)*
Symbolism • *No animal comparable to humans* • *Proposed for sperm whales* • *However: some animals might have a level of arbitrariness higher than indexicality but lower than symbolism*	Mental representations *Controversial: likely but not demonstrated* *Possibly many species, proposed for:* • *Birds (e.g. ravens)* • *Nonhuman primates* • *Dogs*
Displacement *Controversial. Possible in:* • *Insects: bees, ants* • *Birds: grey parrots, ravens* • *Great apes: orangutans, chimpanzees*	Detachment from stimuli *Difficult to demonstrate.* *Possible in species with displacement*
Specialization *All animals with referential calls*	Prevarication • *Birds (e.g. fork-tailed drongo)* • *Great apes (e.g. orangutan)* • *Captive animals (e.g. grey parrot, great apes, domesticated dogs)*
Reflexiveness *No animal*	Semantic omnipotence *No animal*
Context *Use of language in external communication (pragmatics)*	
Intentionality *Drive to share meanings* • *First-order intentionality: great apes, possibly many species with referentiality*	
Theory of Mind *Controversial: difficult to demonstrate* • *Nonhuman primates (?)* • *Birds: parrots, corvids (?)*	Shared intentionality *No animal*

Signal exchanges *Active exchange of linguistic signals*	
Turn-taking *Many animal species (e.g. birds, nonhuman primates, cetaceans, elephants, naked mole rats, bats, insects)*	Interchangeability *All animals which can produce all the signals of their species. Not universal (e.g. not in species in which only males sing).*
Parity *Common: all animals which share the meanings of their calls*	Cooperative principle *No proof in any species: animal communication is asymmetric*
Pragmatic functions *Functions covered by language*	
Multimodality • *Nonhuman primates*	Modality-independence *Apparently no animal*
Multifunctionality *Common: most animal calls have different functions*	Jakobson's functions • *Referential: common* • *Poetic: no animal* • *Emotive: common* • *Conative: bottlenose dolphins* • *Phatic: common* • *Metalingual: no animal*

Table 7. Linguistic properties in nonhuman animals.

4. The general traits related to language in nonhuman animals

After this review of the linguistic properties of humans and nonhuman animals, it seems evident that many linguistic traits are present in other species in several forms and in all the levels described (signal, signal learning, structure, meaning and contextual usage). The properties which, to date, appear to be unique in humans are open-endedness, semantic omnipotence, recursion in communication, complex symbolism, reflexiveness, and shared intentionality. However, they might be conceived as consequences of simpler principles which are present in nonhuman animals in more rudimental forms. Humans show to have more complex syntactic constructions, more abstract meanings and more complex interactions, but these unparalleled skills might be grounded in shared, simpler traits. This does not imply that all properties of language developed in continuity, but at least should suggest that the "seeds" of language are not unique, because they can develop in other species through convergent evolution (e.g. in birds, bees, cetaceans). This review also shows that some individual properties (e.g. the ability to detect non-adjacent dependencies) can be present in animals without having the "full package" of language: some aspects of syntax, for example, can emerge even in animals who do not produce complex syntactic structures, and there can be complex structures without complex meanings (e.g. birdsongs). For these reasons, it is also important to focus on those traits that – as I already clarified – are not linguistic per se, but are still fundamental to support language, and could be considered prerequisites on which the more complex linguistic properties lay. Hence, before searching for linguistic properties in animals, the presence of these features should be assessed. Following the same classification adopted for humans, I focus on the main traits related to biology, brain and cognition, and sociality, identifying which species display these important characteristics. This review cannot be fully comprehensive of every single general trait which indicates biological, cerebral, cognitive, or social complexity that can be connected to language, but it can still be useful to understand which key elements should be searched, and to highlight the species which show most of them.

4.1 Biological traits

The feature of ***longevity***, crucial for humans, is shared by several complex species, although not all long-living animals have complex communication (e.g. turtles, crocodiles). However, most of the species which show linguistic traits, described in the previous chapter, also have a long lifespan: this is the case for nonhuman primates, in which potential life-expectancy ranges from the 15 years of lemurs to the 60 years of chimpanzees (Groves &

Napier 2024); of cetaceans (see Part III), in which several species live more than 40 years, with females surpassing even the 60 years of age (e.g. killer whales, bottlenose dolphins), and with cases of whales like the bowhead whale (*Balaena mysticetus*) that can live more than 200 years (Berta *et al*. 2006); elephants (up to 65 years, see Lee *et al*. 2016); and some bird species, e.g. parrots (up to 60 years in captivity), corvids (up to 20 years in the wild, with an average of 7-8 years), including also the birds that show the most interesting signs of compositionality, as Japanese tits or pied babblers (both up to 15 years in the wild, although average lifespan is much lower) (see Lovette & Fitzpatrick 2016). These animals seem to confirm a correlation between long lives and complex communication, especially because, as far as we know, there is no species which shows linguistic properties and has a very short life-expectancy (with the notable exception of bees and ants, which are eusocial insects, though). Furthermore, there is a correlation between size and longevity: bigger species tend to live more.

With the obvious exception of birds, the long-lived mammals described above have also a ***long gestation*** period, which generates a ***few offspring*** at a time. Following the r/K selection theory about reproductive strategies (Pianka 1970),[77] this suggest that they are K selected: this definition is referred to the species that invest a high amount of energy in the production and care of a few calves, favouring quality over quantity. In these species the mortality rate of the offspring is reduced with respect to the r selected type (like, for example, birds, but also many mammals), in which there is a high mortality rate compensated by the high quantity of the brood. However, this does not imply that birds do not invest in caring of the offspring (see § 4.3 below). A few examples of long gestation and reproduction of few calves at a time are nonhuman primates (ranging from an average of 132 days of lemurs to the 255 days of gorillas, see Groves & Napier 2024), cetaceans (ranging from the 7 months of porpoises to the 17 of sperm whales, see Berta *et al*. 2006), and elephants (from 18 to 22 months, see Hildebrandt *et al.* 2007). Another interesting feature of some of these species is that they keep living after the end of their fertility period, meaning that even after that, their lives still have a function which goes beyond mere reproduction, especially in females (Lee *et al*. 2016). This function might be searched in social relationships and transmission of knowledge: for example, in orcas, social groups (named pods) are usually guided by an old female leader, which can reach the age of 90 years (Berta *et al*. 2006, see also Part III, § 2.2).

[77] In the r/K selection theory, r is referred to "rate", and K to "capacity limit" (from the German word *Kapazitätsgrenze*).

4.2 Brain and cognition

Although many traits related to cognition are difficult to demonstrate, as we saw in the previous chapters, there are many features of the brain that can be measured and can offer comparative insights about the capabilities of nonhuman species. It is interesting to notice that many animals with complex properties related to language also show similarities in these neural features, although these resemblances should be taken with caution. Furthermore, the traits which are not observed in other species might be useful clues of human uniqueness, to be further explored. Among the animals with the highest ***encephalisation quotient*** (i.e. brain size related to body mass, see Part I, § 5.2) after humans (6.54) there are cetaceans, especially odontocetes, led by dolphins: bottlenose dolphins have an EQ of 4.14 (see Marino 2002, 2004, Hof *et al*. 2005), or even, according to Roth & Dicke 2005, 5.3 (see also Part III below). Both measures are much higher than that of nonhuman primates (chimpanzees are ranked 2.2-2.5, gorillas 1.5-1.8, while capuchin monkey range between 2.4-4.8, see Roth & Dicke 2005), while elephants have only 1.8 (Roth & Dicke 2005, see also Shoshani *et al.* 2006). In any case, each measure above 1 indicates a brain which is bigger than it should be expected. However, the are some evident discrepancies between EQ and intelligence: for example, capuchin monkeys have a higher EQ than great apes, and this does not reflect their observed capabilities. Other animals which do not respect this measure are mysticetes (with an EQ below 1, see Part III, § 1.3), because their body has an increased dimension due to physiological reasons related to life underwater, despite showing relevant signs of intelligence (Marino 2002). Furthermore, it has been observed that EQ might not be an effective measure of intelligence because brain size also depends on factors like alimentation (e.g. diet rich of meat leads to bigger brains). On the other hand, differences in body size in individuals of the same species determine different EQs (Cairó 2011). In birds, some corvid and parrot species have an EQ around 1, which is lower than that of mammals but the highest among birds (see Emery 2006, 2004). About this, Roth & Dicke (2005) observe that bird cell size is smaller than that of mammals, so that a small bird brain might have more connections than one of a mammal of the same size. Finally, Roth & Dicke (2005) also consider the number of cortical neurons, showing that in this different classification humans, elephants and killer whales are the best ranked. Despite its flaws, EQ remains a useful tool to explore animal intelligence, although the results cannot be taken as an objective measure.

Another feature of human brain, considered by Boeckx (2021) one of the hallmarks of human complexity, is ***globularity***, for which to the best of my knowledge there are no specific studies in the animal kingdom. It is confirmed that hominins and other living

nonhuman primates do not have this shape (Boeckx 2021). The only reference about this trait comes from Marino (2004), who writes that cetacean brain has a globular shape, but she does not elaborate further (see Part III, § 1.3). This issue should be explored more in depth in the future.

Another potential indicator of intelligence is a ***large neocortex***. Cetaceans are the animals with the highest level of gyrification, superior to humans (Marino 2002, 2004, Hof *et al*. 2005). However, their neocortex is thinner, so that, despite having a larger surface in respect of that of our species, it has a different structure: Marino (2002) defines it a case of "convergent complexity", because it is part of a brain capable of performing complex operations, but this complexity has been reached through convergent evolution (see Part III, § 1.3 for a detailed discussion). Other animals with a large neocortex are elephants (Shoshani *et al*. 2006), pinnipeds (Hauksema *et al*. 2021) and nonhuman primates, although the latter have a lower extension than humans (Haug 1987). Finally, birds have a different forebrain structure than mammals, but according to Emery (2006) the pallium can be compared with the cortex, thus leading to possible comparisons which will not be discussed here. It should also be noted that humans and song-learning birds share some similarities in the brain circuits related to vocal production, showing connections which are absent in both nonhuman primates and non-singing birds, representing another interesting case of convergent evolution (see Pfenning *et al.* 2014).

Connected with vocal production is the ***FOXP2 gene***, widely studied in humans. Orthologs of this gene are found in many vertebrate species, several of which are capable of vocal communication (e.g. nonhuman primates, cetaceans, elephants, pinnipeds, mice, birds), and its protein is highly conserved (Nudel & Newbury 2013, Enard *et al*. 2002). This indicates that it cannot account for the unique properties of language. However, the amino-acid substitutions observed in humans are not found in other animals, thus indicating a possible further specialization which makes this gene different in our species. Research of similar substitutions in nonhuman animals with complex vocal communication (e.g. birds, cetaceans) did not lead to significant results, e.g. showing that the FOXP2 gene in songbirds is no different than that of mice (Nudel & Newbury 2013).

Spindle neurons (VEN), however, are less spread in the animal kingdom. They have been found in great apes (Allman *et al*. 2002, Nimchinsky *et al*. 1999), in cetaceans, in which they are more abundant than in humans (Butti *et al*. 2009, see also Part III, § 1.3), in elephants (Hakeem *et al*. 2009), and macaques (Evrard *et al*. 2012): all animals with a large brain, a certain extent of vocal production learning, and several other properties related to language. However, the finding of VEN in the pygmy hippopotamus (one of the closest

living relatives of cetaceans) and other mammals (e.g. cow, pig, sheep, deer, horse, zebra, rhinoceros, hyrax, walrus, manatee) led Raghanti *et al*. (2015) to propose that these neurons are not correlated with large brains, and they might have an ancient evolutionary origin, being then lost by some modern species and conserved by others. Nonetheless, Shubha & Sudhi (2017) found spindle neurons in the Indian parrot (*Psittacula krameria*), which represents the first case of a non-mammal with this feature (in a capable vocal learner), thus indicating both that a large brain is not always necessary for VENs, and that they can likely develop through convergent evolution. Finally, Banovac *et al.* (2021) observe that human spindle neurons have different characteristics than those of other species, and suggest that further studies are needed to evaluate their structure in animals with complex cognition and communication. ***Mirror neurons*** seem even more rare: starting from the study of Gallese *et al*. (1996), which found them in macaques, they have been identified in other nonhuman primates (Rizzolatti & Craighero 2004), and, as of today, they are not observed in other species. However, other animals have "mirroring systems" which seem to have the same functions, like some bird species (Prather *et al*. 2008), or rats and mice, which are capable to mirror some emotions felt by conspecifics – pain – but not others, like fear (Carrillo *et al*. 2019).

A debated topic connected with imitation skills and with a possible role of mirror neurons is that of ***self-awareness***, intended as the ability not only to have a consciousness with will and desires, but also to be aware of being an individual with a consciousness (Jabr 2012). Intelligent species like great apes, corvids and cetaceans have been considered candidates for this property. One effective technique to test self-awareness is that of the mirror test created by Gallup (1970). Basically, an animal is marked in a part of his body, and then exposed to his reflection: if the animal inspects or tries to clean the mark, then there are reasons to believe that it recognises its image.[78] This test has been passed by a few species: interestingly, most of them show several linguistic properties, and are social animals, indicating once again that there is a connection between complexity in communication, sociality and cognition. However, although some individuals pass the test, showing that it is possible for certain species, it shall be noted that several others fail it. That said, the animals which passed the test and showed the ability of *mirror self-recognition* are great apes (Anderson & Gallup 2015), bottlenose dolphins (Reiss & Marino 2001), orcas (Delfour & Marten 2001), one Asian elephant (Plotnik *et al*. 2006), and magpies (*Pica pica*, see Prior *et al*. 2008). Surprisingly, other animals which seemed to pass the test are ants

[78] See Part III, § 3, for a detailed description of the experiment in dolphins.

(Tricot & Cammaerts 2015), giant manta rays (Ari & D'Agostino 2016, but see the critical response of Stewart *et al*. 2017, according to which the mantas were engaging in social behaviours with the reflection), and cleaner fish (*Labroides dimidiatus*, see Kohda *et al*. 2023).

As I have already discussed, another crucial requirement for language is ***memory***, necessary to store linguistic data, to apply the grammatical rules and to remember the events and information which should be communicated and the individuals which should receive that communication. All animals have memory, but in some species it is very limited. Regarding *short-term memory*, working memory is a fundamental component of cognition related to "decision making, planning, goal directed behaviour", and thus "refers to the ability to memorize information over short periods and, importantly, to process this information" (Fongaro & Rose 2020: 1). It is not a passive skill, because it requires "active maintenance of information" (Fongaro & Rose 2020: 1), and for this reason it is also limited in humans and dependent from attention. This kind of memory has only been observed in great apes and corvids (Fongaro & Rose 2020). Other forms of short-term memory are spread in the animal kingdom, but do not require the level of active processing that is crucial for language. Although denoting a higher level of complexity, *long term memory* is also spread in several species, like great apes (Lewis *et al*. 2023, Vanduffel 2018), cetaceans (Marino *et al*. 2007, Bruck 2013, see also Part III, § 2-3 below), elephants (Chusyd *et al*. 2021), bats (Dixon *et al*. 2022), rats (Panoz-Brown *et al*. 2018), some birds (e.g. jays, see Jo *et al*. 2023; robins, see Shaw & Harvin 2020), and even octopuses (Carls-Diamante 2022). Procedural memory, i.e. the ability to remember learned skills (how to do something), is also observed in all the animals described above. Social memory, i.e. the ability to remember individuals and social relations, is also spread in social species like great apes (Lewis *et al*. 2023), cetaceans (Marino *et al*. 2007, Bruck 2013), elephants (Chusyd *et al*. 2021), rats (Cum *et al*. 2024), and also in some birds (see e.g. Emery 2004). A more complex level of long-term memory is episodic-like memory, which is related to mental time travel and a stronger sense of self. This has been observed in great apes (Martin-Ordas *et al*. 2010), bottlenose dolphins (Davies *et al*. 2022) and likely other cetaceans, elephants (Chusyd *et al*. 2021), rats (Babb & Crystal 2006), food-catching birds like corvids and jays (Clayton *et al*. 2007, Feeney *et al*. 2009, Zinkiyskay *et al*. 2009), parrots (Torres Ortiz *et al*. 2022), cuttlefish (Jozet-Alves *et al*. 2013), and bees (Pahl *et al*. 2007); see also Dere *et al*. (2006), Allen & Fortin (2013) for a general overview. Furthermore, some studies highlighted incidental memory in cats (Takagi *et al*. 2017) and dogs (Fujita *et al*. 2012). Although it is a prerequisite for complex language, the presence of memory does not imply it.

Cerebral plasticity is also present in several forms in many species, in a continuum that goes from a low to a high degree of flexibility (see Fitch 2010). The latter is present in large-brained mammals like nonhuman primates, elephants (see La Rosa & Bonfanti 2018), cetaceans (Marino *et al*. 2007), but also in birds (Nottebohm 2005). Similarly, ***behavioural plasticity*** is a very general concept (see Ghalambor *et al*. 2010), and it is found to different extents: for example, the animals which show complex behaviours that change depending on the population (e.g. a group has a shared hunting technique or food preference which is not present in other groups) have a high degree of plasticity. This is evident in nonhuman primates (see e.g. Fröhlich *et al*. 2022) or in cetaceans (see e.g. Marino *et al*. 2007, Whitehead & Rendell 2015, and Part III below), and in all the captive animals (e.g. bonobos, parrots, dolphins) which were able to learn new skills and thus adapt their behaviour to new contexts.

Tool use is a sign of cerebral and behavioural plasticity. Once believed to be uniquely human (and evolved only in hominins), now it is demonstrated in several species, although it cannot be considered common. The fact that this skill is spread in many taxa of vertebrates and invertebrates indicates that it can be reached through convergent evolution. Furthermore, there is a distinction between the ability to use tools, more widespread, and the rarer ability to produce them. Bentley-Condit & Smith (2010) present a detailed review of tool use in animal kingdom, which range from insects to great apes, also identifying the cases in which only one individual has been observed, and the data is thus anecdotal (see also Seed & Byrne 2010 and Hunt *et al*. 2013). Among the animals discussed in our paper, and excluding the anecdotal cases, tool use is present in great apes, monkeys, bottlenose dolphins (see Krützen *et al*. 2005), elephants, passerine birds (e.g. corvids, tits), parrots, finches, octopuses, ants (see Bentley-Condit & Smith 2010), and tool production is present in great apes (chimpanzees, gorilla), macaques, and passerine birds (e.g. New Caledonian crow, *Corvus moneduloides*) (see Bentley-Condit & Smith 2010). Generally, tool use in animals is restricted to one function; however, passerine birds and nonhuman primates have shown a more flexible use of this skill (see Bentley-Condit & Smith 2010).

The abilities of ***pointing*** and ***pantomime***, or complex imitation, have been already discussed in the previous chapters, because they are strictly connected with language properties, albeit they are not purely linguistic. Pointing has been defined a type of indexical signal, observed in great apes, dolphins and (considering the sole comprehension) also in several other mammals (see § 3.4). Pantomime, or imitation, encompasses several linguistic and non-linguistic areas, such as vocal production learning, cultural transmission, parity, Theory of Mind (see §§ 3.2 and 3.5), and it also relates to mirror neurons. Here I just remind

that this skill is present in nonhuman primates (Tomasello 2008, Tomasello & Call 1997), cetaceans (Zamorano-Abramson *et al*. 2023), elephants (Kaufmann *et al*. 2023), some bird species (see Zentall 2006) and octopuses (Gómez-Moreno 2019). Equally important concepts which we discussed above (see §§ 3.4 and 3.5) are *Theory of Mind*, *shared intentionality*, and *mental representations*. These three cognitive features are so intertwined with language that I treated them as linguistic properties, albeit embracing many aspects of general cognition. Although shared intentionality seems to be uniquely human in the measure that no animal shows our degree of mind-reading and cooperation, traces of Theory of Mind and mental representations on which it is rooted have been found in other species, as great apes, corvids, and cetaceans (see § 3.4).

4.3 Sociality

The final aspect that I discuss is sociality. Many species are highly social, and not all of them can be considered "linguistic". However, even distant, small-brained species like bees and ants developed properties related to language which facilitate the transmission of information (e.g. iconicity, indexicality, displacement). This suggests that sociality seems to be another prerequisite for language, in line with the social complexity hypothesis (Freeberg *et al*. 2012). In fact, a ***complex social structure*** is observed in all the species with complex communication. Not only eusocial insects like bees and ants, but also bigger animals like some mammals and birds live in complex societies characterised by a large group size and a strong cohesion (see e.g. Kappeler 2019, Lukas & Clutton-Brock 2018). I cannot describe all the features of all animal societies, but an interesting fact can be noticed: many animals with complex communication live in fission-fusion societies, i.e. large flexible groups that frequently split (fission) and form smaller unions to perform different activities (e.g. foraging, taking care of the calves, etc.), and then go back to the larger associations of individuals (fusion). This kind of society is observed also in humans, for example in hunter-gatherers, and it has been a precursor of our modern organisation; furthermore, it was likely the type of society present in hominins (Grove *et al*. 2011). Among the species with fission-fusion societies we find great apes (chimpanzees, Chapman *et al*. 1995, orangutans, van Schaik 1999), bottlenose dolphins (King *et al*. 2022, see Part III, § 3), elephants (Fishlock *et al*. 2013, Wittemyer *et al*. 2005) bats (Ortega 2016, Kashima *et al*. 2013), parrots, corvids (Emery 2006, 2004). It does not seem a coincidence that these animals have also developed a complex communication system, because this type of fluid society implies a need for exchange of information, coordination, and recognition of other individuals. Like humans,

these species ***rely on cooperation and interaction***, so that they have a natural inclination in forming groups and having a prosocial behaviour. Although none of these animals, except bees and ants, can be defined eusocial, nonhuman primates, cetaceans and certain bird species show some traits of eusociality (see Part I, § 5.3), i.e. they live in common groups, they divide their tasks (e.g. hunting), and in some cases, like cetaceans, there is also an overlapping of generations. Finally, ***cooperative breeding*** is observed in many insects (e.g. bees and ants), in almost 8% of birds (see Cockburn 2006) and in 5% of mammal species like canids, mongooses, new world monkeys, rodents (Lukas & Clutton-Brock 2012), but not in great apes. In its proper definition, cooperative breeding is referred to those species which biologically inhibit the fertility of some individuals (only some of them reproduce themselves, while the others are dedicated only to the care of the offspring). However, if the concept of cooperative breeding is extended also to the species that help in the care of non-relatives (alloparental care), then the list becomes larger, including humans (see Part I, § 5.3), great apes (Reddy & Mitani 2019), and cetaceans (Whitehead & Rendell 2015). For example, young chimpanzees have been adopted by an adult female after their mother had died (Reddy & Mitani 2019). A similar behaviour has been observed also in bonobos (Tokuyama *et al*. 2021).

Two aspects already discussed as linguistic and cognitive traits are cultural transmission and social learning, which refer to the ability to transmit information by teaching and learning from conspecifics in social contexts. The two concepts are related, because there could be no cultural transmission without learning and social interactions. Putting aside the linguistic aspects, and the fact that the definition of culture is debated, phenomena of ***cultural transmission*** and teaching are observed in many species (see Whiten 2021) as nonhuman primates (Whiten & van de Waal 2018, Mesoudi & Whiten 2008), cetaceans (Whitehead & Rendell 2015, Krützen *et al*. 2005), birds (Garland & McGregor 2020, Williams *et al*. 2013 – interestingly also in pied babblers, which show the property of semantic compositionality, see Raihani *et al*. 2008), even some fish (Helfman & Schultz 1984), bees, ants, and several other species (see Hoppitt *et al*. 2008, Hillix & Rumbaugh 2004, Caro & Hauser 1992). ***Social learning*** is also spread in many taxa (see Reader 2016, Janik & Slater 2000 for a general overview), but the best examples are in nonhuman primates (Whiten & van de Waal 2018), cetaceans (King *et al*. 2022, Marino *et al*. 2007), elephants (Greco *et al*. 2012), bats (Ortega 2016), and birds (Wild *et al*. 2021, Slagswold & Wiebe 2012). Finally, clues for a possible *self-domestication* (which is only a hypothesis at the moment) have been found in bonobos, elephants, and to a lesser extent in dogs and cats (Raviv *et al*. 2023), although this peculiar feature has still to be explored in other species.

Table 8 summarises the general traits related to language present in nonhuman animals discussed in this chapter.

<table>
<tr><th colspan="2">Animal general traits</th></tr>
<tr><td colspan="2">Biological traits
General features related to main aspects of biology</td></tr>
<tr><td>Complex vocal auditory channel

All animals with vocal communication</td><td>Longevity
Nonhuman primates (15-60 years)
Cetaceans (40-200 years)
Elephants (65 years)
Birds (parrots 60 years, corvids 7- 20 years, tits & babblers 15 years)</td></tr>
<tr><td>Long gestation
Nonhuman primates (4-8,5 months)
Cetaceans (7-17 months)
Elephants (18-22 months)</td><td>Few offspring
Nonhuman primates
Cetaceans
Elephants</td></tr>
<tr><td colspan="2">Brain and cognition
Structural aspects of the brain and cognitive features of the mind</td></tr>
<tr><td>Encephalization quotient (high)
Bottlenose dolphins: 4.14-5.3
Chimpanzees: 2.2-2.5
Elephants: 1.8
Grey parrots: 1</td><td>Globularity

No animal
(some observations about cetaceans)</td></tr>
<tr><td>Large neocortex
Cetaceans (larger than humans)
Elephants (larger than humans)
Pinnipeds
Nonhuman primates
Birds: different structure (pallium)</td><td>FOXP2 gene

Many vertebrate species (e.g. nonhuman primates, cetaceans, elephants, pinnipeds, mice, birds), but in humans is different.</td></tr>
<tr><td>Spindle neurons
Macaques
Great apes
Cetaceans (more than in humans)
Elephants
Other mammals with smaller brain
Indian parrot</td><td>Mirror neurons
Only in nonhuman primates (monkeys and apes)
Similar systems might be present at least in mice and rats, song-learning birds, cetaceans.</td></tr>
<tr><td>Self-awareness
Animals who passed the mirror test:
Great apes
Cetaceans (dolphins and orcas)
Asian elephant (only one)
Magpies
Possible: ants, mantas, cleaner fish</td><td>Memory
Many species have forms of short and long-term memory, but few have:
Working memory: great apes, corvids
Episodic-like memory: great apes, cetaceans, elephants, rats, parrots, corvids, jays, bees, cuttlefish</td></tr>
</table>

<table>
<tr><td>Cerebral plasticity
Many animals in different degrees. Highest level: nonhuman primates, cetaceans, elephants, some bird species</td><td>Behavioural plasticity
Many animals in different degrees. Highest level: nonhuman primates, cetaceans, corvids, captive apes and birds</td></tr>
<tr><td>Tool use
Several species from insects to great apes.
Use: e.g. nonhuman primates, bottlenose dolphins, elephants, passerine birds, parrots, finches
Manufacture: e.g. great apes, macaques, passerine birds</td><td>Pointing
Production: great apes, bottlenose dolphins, dogs (?)
Comprehension: several species, e.g. nonhuman primates cetaceans, domesticated dogs, cats, pigs, etc.)</td></tr>
<tr><td>Pantomime/imitation
Nonhuman primates (scarce)
Cetaceans
Elephants
Some bird species (e.g. parrots)</td><td>Theory of Mind
Controversial: difficult to demonstrate
Nonhuman primates (?)
Birds: parrots, corvids (?)
Cetaceans (?)</td></tr>
<tr><td>Shared intentionality

No animal</td><td>Mental representations
Controversial: likely but not demonstrated.
Possibly many species (e.g. nonhuman primates, cetaceans, dogs, ravens)</td></tr>
<tr><td colspan="2">Sociality
Aspects related to social groups and interactions</td></tr>
<tr><td>Complex social structure
Many animals in different degrees
Eusociality: insects (e.g. ants, bees)
Fission-fusion societies: great apes, dolphins, elephants, bats, parrots, corvids</td><td>Reliance on cooperation and interaction

Many species, e.g. those with complex social structure (ants, bees, nonhuman primates, cetaceans, elephants, bats, etc.)</td></tr>
<tr><td>Cooperative breeding
In strong sense (with infertility of some classes of individuals): insects, 8% of bird species, 5% of mammals
In weak sense (alloparental care): great apes, cetaceans</td><td>Cultural transmission
(depends on the definition of culture)
Nonhuman primates
Cetaceans
Some bird species (e.g. babblers)
Some fish species
Insects (e.g. bees, ants)</td></tr>
<tr><td>Social learning
Spread in many taxa with social interactions, e.g. nonhuman primates, cetaceans, elephants, bats, birds</td><td>Self-domestication
Not demonstrated: clues of it observed in bonobos, elephants, dogs, cats.</td></tr>
</table>

Table 8. Summary of general traits related to language in nonhuman animals.

5. Overview: the most promising "linguistic" animals

In chapters 3 and 4 I outlined the main properties and features related to language present in nonhuman animals. What emerged is a composite scenario in which many traits are present in a diverse range of species, but no nonhuman animal displays all of them. Furthermore, some trained specimens showed an understanding of linguistic processes traditionally considered uniquely human, which can point to the presence of latent abilities, not fully exploited in their natural environment (see e.g. Pepperberg 1999, Greenfield & Savage-Rumbaugh 1990, Herman *et al.*1984). As Fitch (2010) suggests, all these features and properties might be spread in the animal kingdom, but they would coexist only in humans. Considering the theories of the nature of language discussed in Part I, chapter 2, the view of language as a mosaic seems more compelling than the view of a system with a central core. If this idea is correct, then the human uniqueness should not be searched in a genetic mutation, but in more variegated and gradual processes, related to general, non-linguistic dynamics and features. However, even in the case that Berwick & Chomsky (2016) are right in considering the merge operation as the feature which is closest to what could be defined a central core, then it should also be observed that, first, merge alone is not sufficient to produce language, and, second, that this operation seems to be present in other species, in simpler, non-recursive forms. As illustrated in section 3.3, the species which display the most convincing evidence of these compositional combinations are Campbell's monkeys, chimpanzees, Japanese tits and Southern pied babblers. Here I briefly isolate these animals from the wider list presented in the previous chapters, and I summarise their call combinations, showing their compatibility with non-recursive 1-merge proposed by Rizzi (2016).

Campbell's monkeys seem to combine their calls *krak* (terrestrial threat) and *hok* (non-terrestrial threat) with the "meaning-modifying suffix" *-oo* (weak disturbance), obtaining *krakoo* (weak terrestrial threat) and *hokoo* (weak non-terrestrial threat) (Ouattara *et al.* 2009). In this case, the non-compositional interpretations proposed by Schlenker *et al.* (2016, 2022, 2023) are hardly compatible with these vocalisations: *krak* and *-oo* cannot be considered *separate utterances* interpreted in a linear way because *-oo* is not found in isolation, and thus it is not an independent call (the same reasoning applies for *hok* and *-oo*). The *one-word* interpretation, according to which *krakoo* and *hokoo* are single calls with their own non-compositional meaning, only acoustically complex, cannot be excluded. However, it is highly unlikely, because in this case there would be no syntactic nor semantic relationship between the calls *krak*, *hok*, *krakoo* and *hokoo*: the resemblances in meaning

and structure should be just a coincidence. Hence, the compositional interpretation, according to which *krak* and *hok* combine with *-oo* to create a more complex meaning, remains a reasonable possibility.

Regarding the *hoo – waa* (*alert* + *recruitment*) combinations of **chimpanzees**, it is unlikely that their calls *hoo* (related to alert situations) and *waa* (related to recruitment) are *one-word*, because they are both also produced in isolation (Leroux *et al.* 2023). It is also unlikely that they are *separate utterances* interpreted linearly because – although they are indeed two calls – the behaviour observed with *hoo waa* combinations is more complex than the simple sequencing of *hoo* and *waa*. The animals join to scare the threat away (generally a snake) only when the two calls are combined (Leroux *et al.* 2023). A similar argumentation stands for **Japanese tits** alert (ABC) and recruitment (D) combination (ABC-D, eliciting the mobbing sequence): it is not *one-word* because the two calls are used in isolation, and despite being *separate utterances*, their interpretation is not linear because the mobbing sequence (scaring the predator away) is produced only when the calls are emitted in the correct order (alert-recruitment) from the same source. The experiment by Suzuki & Matsumoto (2022) clearly shows the difference between the alert-recruitment sequence produced by two speakers (no compositional interpretation, no mobbing sequence), and the sequence produced by one speaker (compositional interpretation, mobbing sequence). Schlenker *et al*. (2024) propose a further non-compositional alternative, defined *bird implicatures*: from this perspective, alert, recruitment, and alert-recruitment are three separate calls with separate meanings, in competition among themselves because of the Informativity principle. In brief, according to Schlenker *et al*. (2024), the alert-recruitment call is the most informative one, and thus should always be used in any situation in which it is true, i.e., when there is a predator to mob. If it is not used, this means that there is no predator to mob. Hence, when a bird hears alert, understands that there is no recruitment to do, and conversely, when it hears recruitment, it understands that there is no alert (otherwise, the more informative alert-recruitment call would have been used). The complete meaning of the calls then would be "alert but not recruitment" in one case, and "recruitment but not alert" in the other. This has been proposed to explain why the birds in Suzuki & Matsumoto experiment did not produce the mobbing sequence when the calls came from two different sources: basically, hearing the first call (alert) automatically excluded the other one (Schlenker *et al.* 2024). However, this explanation requires a system of implicatures which appears complex, and it is not necessarily more simple than the hypothesis of a compositional, 1-merge like interpretation. Finally, in **Southern pied babblers** (that have a similar alert-recruitment combination that elicits a mobbing sequence), alert-recruitment is not *one-word* because the calls exist in

isolation, and they are hardly *separate utterances* because the combination produces a behaviour which is more complex than the simple juxtaposition of the two calls (like for chimpanzees and Japanese tits). Furthermore, Engesser *et al.* (2024) showed that the animals can decompose the incorrect sequence recruitment-alert into its constituting parts and reinterpret it correctly, indicating that they have a semantic knowledge of both calls, and that they combine them to obtain a compositional meaning.

For all the cases just discussed, the compositional interpretation, according to which these animals combine two meaningful units with a non-recursive operation like 1-merge, cannot be discarded, and in my opinion it is a convincing hypothesis. It is consistent also with the results of captive animals like Kanzi the bonobo (Savage-Rumbaugh 1986) and Alex the grey parrot (Pepperberg 1999), who were capable to combine lexigrams (Kanzi) and English words (Alex) in very simple compositions (see § 3.3). In any case, the properties of nonhuman primates and birds suggest that, regardless of whether merge is a central core or not, this operation could be non-uniquely human.

However, it is striking to observe that all the animals discussed, even the ones with the most interesting cases of compositionality, only display combinations of two elements, which reflect at best the skills of human infants in their two-word stage. In animal communication systems there are no clear signs of recursion, nor of cascade 1-merge. Furthermore, the lexicon of all the species described is restricted to a few calls. There seems to be a correlation between recursion and lexicon size: humans have recursion and a wide, complex lexicon, while other animals do not display recursion and have a very limited set of calls. Hence, it could be argued that recursive merge (i.e., the combination of the compositional 1-merge operation with recursion, leading to 3-merge) might be indeed responsible for the expansion of human lexicon. This correlation should be further explored.

Among the several species that I described, some of them appeared more frequently in this review. Although the general and linguistic traits are sparsely distributed in the animal kingdom (suggesting that they can evolve independently, without the need of "complete" language), some animals (nonhuman primates, birds, cetaceans) showed to have most of these traits: hence, it would be interesting to further analyse them in search of the properties that they seem to lack, to verify if they are indeed not present, or if they have just not been found yet. The classification proposed in this dissertation can thus be useful to identify the most promising animals to be studied from the perspective of language. What emerges is that cetaceans display several traits related to language, although they have not been thoroughly studied from this perspective. Furthermore, they show many similarities with the nonhuman primates and birds that present the most convincing cases of compositionality.

For this reason, in the third part of this dissertation I focus on cetaceans, and in particular on bottlenose dolphins, with the goal of identifying their "linguistic" abilities and to verify whether some of their complex call combinations can be considered an example of 1-merge. Furthermore, I investigate their communicative and cognitive skills to check if cascade 1-merge systems (i.e., combinations of more than two elements) or recursion are present in this species, and if their repertoire of calls (lexicon) has a size comparable to that of the other animals discussed in this part of the dissertation.

Part III

Introduction

In the second part of this dissertation I described and discussed the issue of language properties in nonhuman animals in general, after having highlighted the complex scenario of the nature and evolution of human language in Part I. From the systematic classification I proposed in the previous chapters, it emerged that several nonhuman species display a set of general skills and basic "linguistic" properties that play a pivotal role in human language, although no species (except for humans) has all of them. It also emerged that cetaceans present a high number of features related to language (e.g. a complex brain, a complex social structure, a complex system of vocalisations). Cetaceans are interesting for the comparative research of language properties because they display both the complex vocal production and imitation traditionally observed in birds, and the complex cognitive dynamics traditionally studied in nonhuman primates. These two important features related to language have been separately explored in animal language research, because the species that possess complex vocal skills usually do not have complex brains and social dynamics, and vice versa: they coexist only in humans and, apparently, in cetaceans. Cetaceans can thus be considered one of the most promising groups of animals for this kind of research. Furthermore, experiments with captive individuals showed that they are capable to perform complex cognitive operations, which suggest that they have a set of latent abilities that might not be used in their environment. This is important because the presence of a latent ability can justify the emergence of a certain phenomenon (e.g. complex communication, compositionality) without the need of a mutation: in these cases, selective pressures might lead a pre-existing trait to be used for a new function. The observation of similar phenomena in nonhuman animals can shed light also on the evolution of language in humans, and potentially in hominins: as in the case of other species, if hominins had a latent ability like, for example, 1-merge, even if they did not use it in communication, they might have exapted it as a result of selective pressures. This would support some theories of language evolution which advocate for gradual development (see emergentist frameworks in Chapters 2-3 of Part II), rather than those which propose a saltation (see innatist frameworks in Chapters 2-3 of Part II).

As I will show, cetaceans present more than just latent abilities. In particular, many of their behavioural, cognitive and social traits have similarities with those of humans, and might indicate a level of complexity even higher than that of the nonhuman primates and birds described in Part II. The main problems with cetaceans are the practical difficulties encountered in the studies: being marine animals, they spend most of their time underwater

(often in very isolated regions of the world), and thus it is complicated to observe and record them. In the case of vocalisations, it is difficult to identify the individual who emits the calls, as well as the reactions of the other animals, so it is not always possible to assign a function or meaning to each signal. This will be discussed (together with possible ways to get around the problem) in the following chapters. However, the features of complexity which can be observed suggest that cetaceans perform several operations related to language, and thus they are worth to be studied from this perspective. The lack of evidence regarding core linguistic properties as compositionality and 1-merge might be due to these technical issues, and not to their absence.

Therefore, in this Part III I try to assess the extent of the linguistic abilities of cetaceans, and in particular of bottlenose dolphins, focusing on the presence of the complex features highlighted in the systematic classifications of linguistic and general traits presented in Part I for humans and in Part II for nonhuman animals. First, I present an overview of cetacean biology and evolution, to highlight the peculiarities which make them different from many other species in terms of plasticity and adaptability. Then, I concentrate on behaviour, cognition and sociality of humpback whales, killer whales, and sperm whales, with a special attention to the properties of their complex vocalisations. The remaining part of the dissertation is focused on bottlenose dolphins: I describe and discuss their behavioural, cognitive, social, and communicative features, and I analyse their vocalisations to verify whether they display combinations of meaningful units comparable to those observed in nonhuman primates and birds, which might be considered a case of 1-merge. I also discuss some statistical and computational approaches which can help to shed light on the issue. Furthermore, I investigate the presence of recursion (in linguistic and non-linguistic domains) and the size of lexicon. Then, I apply the systematic classification proposed in Part I and II to bottlenose dolphins and other cetaceans, to gain a clear picture of the skills which are present in them and the ones which are not.
This helps to answer the following questions:

- Which traits related to language are present in bottlenose dolphins?
- Is non-recursive merge present in dolphins?
- Is recursion present in dolphins?

Finally, I extend the analysis to the wider picture of the animal kingdom, comparing dolphin skills and properties with those of the other species discussed in the previous parts of the dissertation, including humans and extinct hominins. This contributes to a reassessment of

the theories of language evolution discussed in Part I, in light of the data coming from ethology. I thus address the questions:

- Are dolphin traits similar to those present in other species with complex communication (including humans and hominins)?
- Did human language evolve starting from similar traits (or pressures)?
- Which theory can better describe both the evolution of human language and the complex communicative phenomena observed in other species?

1. Biology of cetaceans

From a classification standpoint, cetaceans constitute an infraorder, which in turn includes two extant parvorders, Mysticeti and Odontoceti, and one extinct parvorder, the Archaeoceti, from which they are believed to have descended. It appears that the common ancestor of primates and cetaceans lived 95 million years ago and that cetaceans have evolved independently from other mammals for the past 55 million years (Marino *et al.* 2007: 966). Mysticeti and Odontoceti are thought to have diverged approximately 35 million years ago (Berta *et al.* 2006: 62).

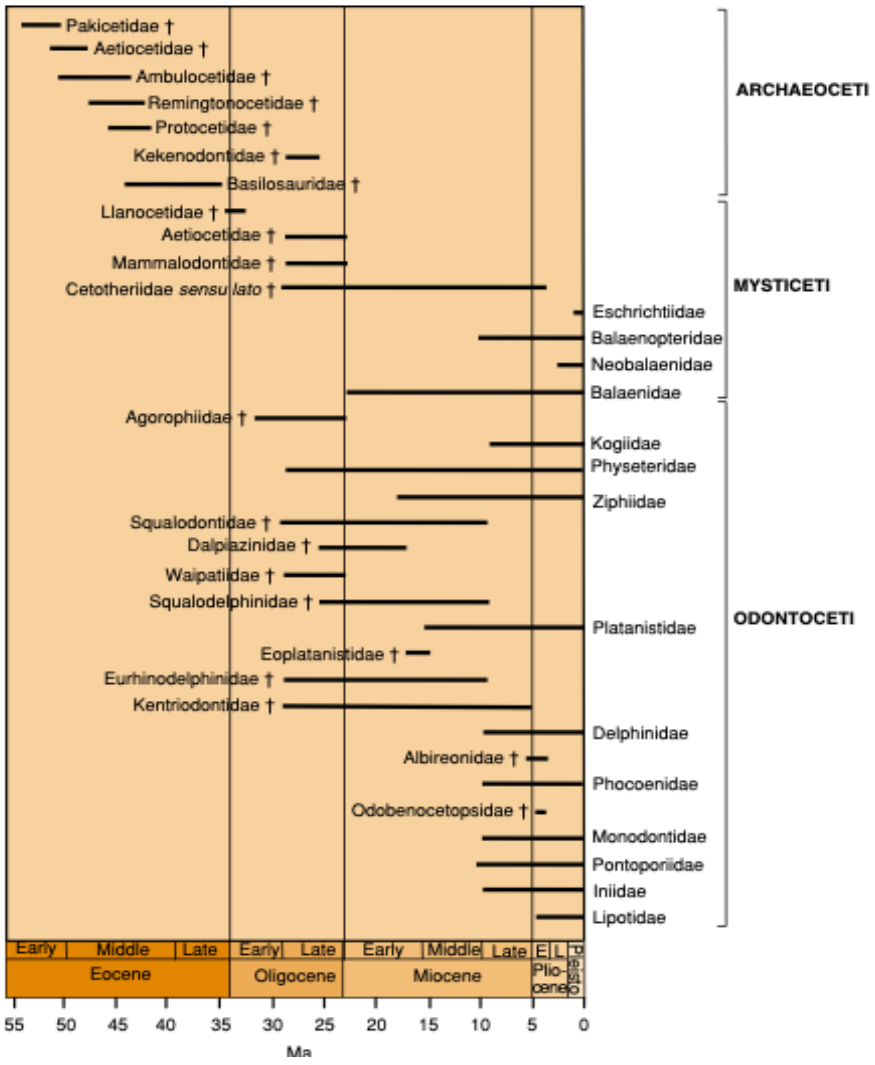

Fig. 9. The diagram (Berta *et al.* 2006: 52) illustrates the emergence and evolution of the various families up to the present day, each of which includes multiple species.

The main current difference between the two parvorders lies in the presence of teeth in odontocetes, which are replaced in mysticetes by a unique adaptation in the animal kingdom:

baleen. Baleen consists of keratinous structures arranged in the mouth like a comb, functioning as a filter for plankton, krill, and small fish that form the diet of mysticetes. These animals filter large quantities of water, trapping organisms in their baleen before ingesting them (Berta *et al.* 2006: 325-332). For this reason, while mysticetes rely on a more "passive" form of feeding, odontocetes are active hunters and predators, primarily feeding on cephalopods and pinnipeds. Odontocetes (including delphinids and, notably, orcas) also tend to be smaller in size, with the exception of the sperm whale, which can rival the largest mysticetes, such as the blue whale, right whale, and humpback whale. In total, there are an estimated 85 species, with 41 subspecies, 40 genera, and 14 families. Odontocetes are more numerous, with about 70 species, whereas only 14 mysticete species have been identified (Reeves *et al.* 2003: 5-9). Despite their differences, all cetacean species share a fusiform, hydrodynamic body shape, an adaptation to aquatic life. A key characteristic of cetaceans, as with other marine mammals, is that they were originally terrestrial animals that gradually occupied a new ecosystem, undergoing radical transformations in their anatomy. Figure 10, which compares reconstructions of different cetacean ancestors and their modes of movement, is particularly effective in illustrating the anatomical changes they have undergone.

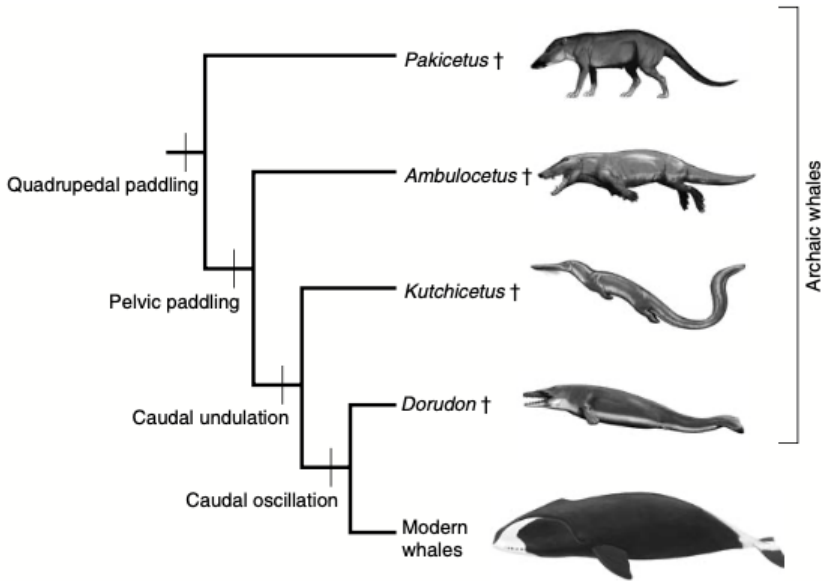

Fig. 10. Reconstruction of the movement characteristics of various extinct archaeocetes, illustrated by C. Buell, alongside a modern mysticete, the bowhead whale, illustrated by P. Folkens (Berta *et al.* 2006: 15).

Contrary to what one might assume, other marine mammals (pinnipeds and sirenians) followed independent evolutionary paths. According to current studies, the closest relatives of cetaceans are hippopotamuses. Consequently, some researchers classify cetaceans within

the order *Artiodactyla*, the same order that includes not only cetaceans and Hippopotamidae but also most ungulates, such as deer and cattle (Berta *et al.* 2006: 183). However, it is important to note that classifications are constantly evolving due to new discoveries that sometimes challenge previous views, ongoing debates among scholars, and the fact that anatomical similarities – which often determine the classification of a species – are sometimes merely examples of convergent evolution. In this regard, Bickerton (2014) has proposed an alternative tree of life that does not consider phylogenetic relationships but instead groups of organisms based on their functional adaptations to occupy a niche. In this model, animals would no longer be categorised by descent but rather by how nature has shaped them and the roles they play in the ecosystem. With this clarification, Figure 11 presents a possible reconstruction of the relationship between cetaceans and artiodactyls, developed by O'Leary *et al.* (2004) by integrating morphological and molecular data.

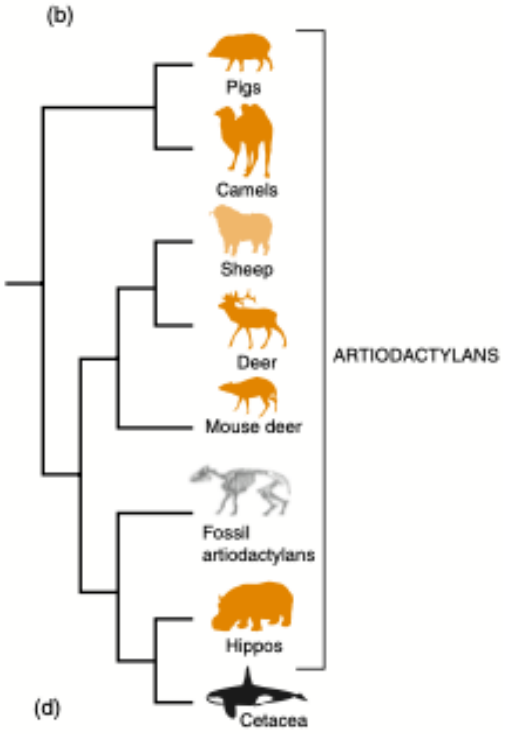

Fig. 11. Hypothetical reconstruction of the relationships between cetaceans and other ungulates, developed by O'Leary *et al.* (2004), in Berta *et al.* (2006: 56).

Cetaceans are distributed across all the world's oceans: some species, such as orcas, are found in almost every sea, while others are restricted to specific regions or latitudes – for example, the bowhead whale (*Balaena mysticetus*) lives only in boreal waters, and the pygmy right whale (*Caperea marginata*) only in the southern hemisphere. Other cetaceans, such as the members of the *Platanistoidea* family, inhabit certain freshwater rivers and lakes, although in increasingly reduced numbers (Reeves *et al.* 2003). Several mysticetes migrate

seasonally between feeding and breeding grounds. Humpback whales, for instance, spend summers in polar waters rich in krill and migrate in winter to calmer equatorial regions, where they mate and give birth, safe from orcas (see Corkeron & Connor 1999). Odontocetes also undertake migrations: some smaller species move from coastal regions to deeper waters in search of food, while sperm whales, the largest odontocete species, can cross entire oceans. Nevertheless, it remains unclear how cetaceans navigate and follow their migration routes. According to Kenney *et al.* (2001), who summarise several hypotheses, they might use the position of the sun (drawing a distant parallel with bees), the Earth's geomagnetic field, ocean currents, or navigate by detecting ultra-low-frequency sounds.[79] Before the arrival of humans, cetaceans had no natural predators and occupied the top of the food chain. However, they are now severely threatened. When studying them, it is essential to consider that they are currently facing unprecedented challenges and are experiencing a phase of decline.

In this chapter I describe the general features of these animals (anatomy, physiology, brain structure, systems of vocal production and perception), treating them as a whole block, while highlighting some differences. In the following chapters I focus on the most interesting species from a cognitive and "linguistic" perspective (humpback whales, killer whales, sperm whales, bottlenose dolphins).

1.1 Anatomy of cetaceans

As previously mentioned, all cetaceans share a fusiform body shape, making them resemble fish due to a phenomenon of convergent evolution. Their skin, which has lost its hair (present only during early development) and sweat glands, retains the same layers as other mammals (epidermis, dermis, and hypodermis); however, these layers are up to twenty times thicker, containing high amounts of fat and oil, which provide insulation for the rest of the body (Berta *et al.* 2006: 133). As a result, the skeletal structure is concealed and not discernible from the external appearance, as illustrated in the following figures.

[79] As we saw in Part I, section 3.1, Hauser *et al.* (2002) acknowledge that the complex navigation of migratory species might require recursion; although this has not been thoroughly studied or demonstrated, it could be the case for cetaceans.

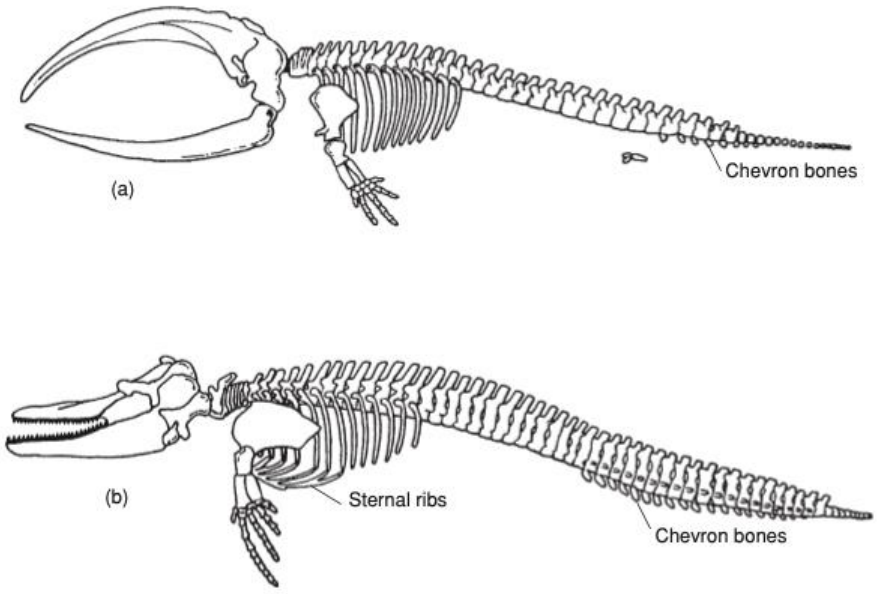

Fig. 12. Indicative representation of the skeleton of (a) mysticetes and (b) odontocetes (Berta *et al.* 2006: 184, modified from Harrison & Bryden 1988). Note the structure of the forelimbs, which still have digits, and the disappearance of the hind limbs, of which only vestigial traces remain, separate from the skeleton, as evident in (a).

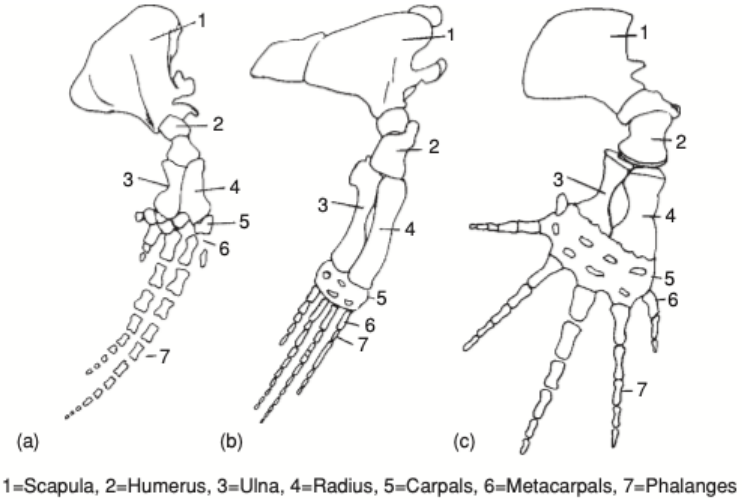

Fig. 13. Cetacean forelimbs: pilot whale (a), blue whale (b), right whale (c), from Evans 1987, in Berta (2006: 189). Note the correspondence with human arm bones.

As shown in Fig. 12, cetaceans have an elongated, beak-like skull, known as a *rostrum*, and during their evolution, their forelimbs have developed into pectoral fins, functioning like paddles. The bones of these fins are entirely covered by fat and skin, yet the digits remain present. The hind limbs, on the other hand, have disappeared entirely, but vestigial traces remain within the body, separate from the rest of the skeleton and externally invisible. The dorsal fin, present in all cetaceans, is not visible in Fig. 12 because it is composed of

connective tissue and evolved independently of the skeleton. Finally, the tail has transformed into a caudal fin, revealing another key difference from fish: since the vertebral column maintains the same orientation as in other mammals, its movement is vertical (up and down), whereas in fish, it is horizontal (side to side) (Berta *et al.* 2006: 178-194). These features further clarify the terrestrial origin of cetaceans, which can be appreciated in detail in the forelimb structure shown in Fig. 13. Among other noteworthy anatomical characteristics, a crucial feature is the *blowhole*, through which cetaceans breathe. It is an opening located at the top of the head, used to expel the characteristic spouts of vapor from previously inhaled air before taking in new air for submersion. The blowhole corresponds to the nostrils of other mammals, which have been reduced to a single opening in odontocetes, while mysticetes have retained two (Berta *et al.* 2006: 244). Besides its respiratory function, the blowhole plays a role more akin to lips than a nose, as it is also used for vocal production, enabling both communication and echolocation (see § 1.4). Other anatomical adaptations to hydrodynamic needs include the loss of external ear flaps and the absence of external genitalia. Although their reproductive organs do not differ significantly from those of other mammals, they are housed within genital slits (Berta *et al.* 2006: 366).

Despite these morphological similarities, cetaceans vary greatly in size. The smallest cetacean, the vaquita (*Phocoena sinus*) of the Gulf of California, measures about 1.5 metres and weighs around 30 kilograms (Brownell *et al.* 1987: 26), whereas the largest, the blue whale (*Balaenoptera musculus*), can exceed 30 metres in length and weigh over 150 tons (McClain *et al.* 2015: 49). The blue whale is the largest living animal and the largest to have ever existed, surpassing even the most massive dinosaurs (Berta *et al.* 2006: 65). It is understandable that the largest animals are aquatic: their skeletons are lighter because water supports most of their weight, allowing for the development of massive structures without the risk of structural collapse. However, this is not the reason for the gigantism observed in mysticetes. Large bodies require substantial amounts of energy for movement, making mysticetes exceptional within the animal kingdom. Slater, Goldbogen, and Pyenson (2017) suggest that their rapid increase in size, which began around 5 million years ago, resulted from changes in ocean currents following the natural closure of the Isthmus of Panama. This event shifted the krill populations that mysticetes feed on toward polar regions. The resulting abundance and, most importantly, the density of krill concentrated in specific areas (which allowed whales to capture enormous quantities of food in a single gulp, optimising energy consumption), likely drove mysticetes – already equipped with their unique and efficient filtration system – to change their habits, transitioning from relatively sedentary animals to long-distance migrants. The extreme increase in size was thus an unintended consequence

of their successful adaptation to a new ecological niche for which they were already evolutionarily prepared (Slater *et al.* 2017). This evolutionary mechanism bears some conceptual resemblance to the rapid development of humans, as it demonstrates that, under the right conditions, nature can produce radical changes in a relatively short time, leading to novel and highly specialised adaptations. The more unusual and specialised an adaptation, the more unstable it becomes, as it can lead to great evolutionary success but also vulnerability – any shift in the ecosystem could render it unsuitable, leading to the species' decline or extinction. To sustain such complex anatomical structures, an equally complex physiology is required (see § 1.2 below). Some species also exhibit sexual dimorphism: in all mysticetes, females are larger than males, whereas in some odontocete species, the opposite is true (e.g., the common bottlenose dolphin, *Tursiops truncatus*). Additionally, in certain species, such as beaked whales and the well-known narwhals (*Monodon monoceros*), males possess an overdeveloped tooth that resembles a horn, used in battles between rival individuals (Berta *et al.* 2006: 389-390).

1.2 Physiology of cetaceans

Here I briefly describe the physiology of cetaceans, highlighting some general characteristics of the respiratory, circulatory, and reproductive systems, as well as their sensory organs. As it is evident, these animals have adapted to life in water: visibility in water is limited compared to air, so vision is not their primary sense. However, their eyes can see both at the surface and underwater, and the spherical shape of their lenses is particularly suited to capturing light even in the ocean depths (Berta *et al.* 2006: 152-155). Sound waves, on the other hand, propagate more easily, making hearing the most developed sense, which I examine in more detail in section 1.4. Hearing is essential not only for communication but also for orientation, at least in odontocetes, which possess the melon organ to help direct and propagate sound (see § 1.4). The least developed sense is smell, which is almost absent in mysticetes and entirely absent in odontocetes (despite being highly developed in other aquatic animals, such as sharks). It can be said that cetaceans have sacrificed their sense of smell in favour of vocal production and perception (see § 1.4), as the anatomical structure of their nostrils has transformed into a blowhole, and the brain regions dedicated to hearing have become highly developed at the expense of those related to smell (Marino 2004: 5). Taste and touch are present, though based on current knowledge, they are not considered crucial (Berta *et al.* 2006: 155-157).

Like other mammals, cetaceans breathe air, so they must periodically surface to expel carbon dioxide and inhale oxygen. Unlike other mammals, however, breathing – performed through the aforementioned blowhole (§ 1.1) – is not an automatic process but requires voluntary control. Additionally, their lungs have evolved for greater efficiency in oxygen absorption: their highly vascularised alveoli can assimilate up to 90% of the inhaled oxygen into the bloodstream, whereas humans only assimilate about 20%. This allows cetaceans to stay underwater for long periods with a single breath: the duration varies by species, ranging from about ten minutes for bottlenose dolphins to over two hours for sperm whales. Sperm whales also reach the greatest depths among cetaceans, with estimates suggesting they can dive as deep as 3,000 metres. To survive at such high pressures, their lungs are capable of collapsing (complete collapse is estimated to occur at around 100 metres in depth) preventing nitrogen accumulation that could cause embolisms.

The circulatory system is not drastically different from that of other mammals, but it has also adapted to aquatic life: during dives, the heart rate slows down (bradycardia), reducing cardiac frequency. This helps maintain stable arterial pressure regardless of external pressure, protecting the organs. Additionally, to facilitate oxygen transport during dives, cetacean blood contains high levels of haemoglobin, nearly double that of terrestrial mammals. Due to space constraints, I have highlighted only the most representative aspects, but for a detailed discussion, see Berta *et al.* (2006: 237-262), from which the above information has been extracted. For further details on the circulatory system and the physiological challenges of aquatic life, I recommend Williams & Worthy (2002: 73-93).

Another notable system is the reproductive system. I have already mentioned that the genitalia are concealed in pockets for hydrodynamic efficiency, that sexual dimorphism is present in certain cases, and that mating is seasonal and linked to migration in most species (Berta *et al.* 2006: 363-370). As placental mammals, cetaceans develop their offspring internally and give birth to live young. In odontocetes, gestation ranges from seven months for Dall's porpoise (*Phocoenoides dalli*) to 17 months for sperm whales, with longer gestation periods for larger animals (as seen in terrestrial mammals). However, this does not apply to mysticetes, which, despite being larger, have shorter and more uniform gestation periods, ranging from ten months for right whales (*Eubalaena* genus) to thirteen months for gray whales (*Eschrichtius robustus*). These gestation times are partly dictated by seasonality and migration: many cetaceans give birth in the same location and season in which they conceived the previous year. A longer (or shorter) pregnancy, depending on the species, would disrupt their migration rhythms (Berta *et al.* 2006: 372-377). In both suborders, each pregnancy typically results in a single offspring, and in cases of twin births, one of the two

calves usually does not survive. What is relevant to the topic of this thesis is that, similar to humans (Part I, § 5.1) and to some other nonhuman animals (Part II, § 4.1), this reproductive system implies significant parental investment: cetaceans produce fewer offspring, but each has a higher chance of survival. Nursing lasts from approximately four months for small odontocetes to nearly two years for sperm whales, meaning that young cetaceans spend the early part of their lives in close contact with their mothers, forming strong bonds (Berta *et al.* 2006: 395-406). A key aspect, which will be discussed later (§§ 2-3), is that odontocetes (and mysticetes as well) learn many behaviours from their mothers that cannot be attributed solely to instinct. As Berta *et al.* (2006: 377) state: "odontocetes typically produce single offspring that are physically well developed but socially undeveloped. Extended periods of postnatal social learning are crucial elements of the normal development of odontocetes."

Another related element, which is important for my analysis, is longevity. Due to their marine nature, it is difficult to determine cetacean lifespans with precision, as current tracking techniques do not allow for definitive measurement. Some methods estimate age by examining tooth layers and the scars that form in the uteri of mysticetes (see Berta *et al.* 2002: 438-439), but these are still approximations[80]. Nevertheless, it is believed that most species live between 20 and 30 years, with notable exceptions among both odontocetes and mysticetes. Some mysticetes, such as blue and fin whales, can surpass 100 years of age, while bowhead whales (*Balaena mysticetus*) have been estimated to live over 200 years in some cases (George *et al.* 1999).

As we have seen in Part II, in complex animals, long lifespans are generally associated with advanced intelligence and cognitive abilities such as memory and learning. This further aligns cetaceans with humans. I now examine some aspects of their brain structure, which reveal additional similarities and intriguing analogies.

1.3 Brain structure of cetaceans

Cetaceans have some of the largest brains in the animal kingdom. The sperm whale, with an average brain weight of 7.8 kg in adult males, possesses the largest brain overall, followed by the blue whale (around 7 kg), the orca (between 5.4 and 6.8 kg), and the elephant, the largest land animal (between 4 and 5 kg). Among smaller animals, the bottlenose dolphin (*Tursiops truncatus*) has a brain weight of approximately 1.5–1.7 kg,

[80]As an indirect example, an NBC News article (https://www.nbcnews.com/id/wbna19195624) reports that in 2007, a bowhead whale was captured with the head of an explosive harpoon embedded in its neck. The harpoon was manufactured in the late 1800s, meaning the whale not only survived the encounter but had lived for over a century beyond it.

slightly surpassing humans (1.2–1.3 kg) and significantly exceeding chimpanzees (around 400 g).

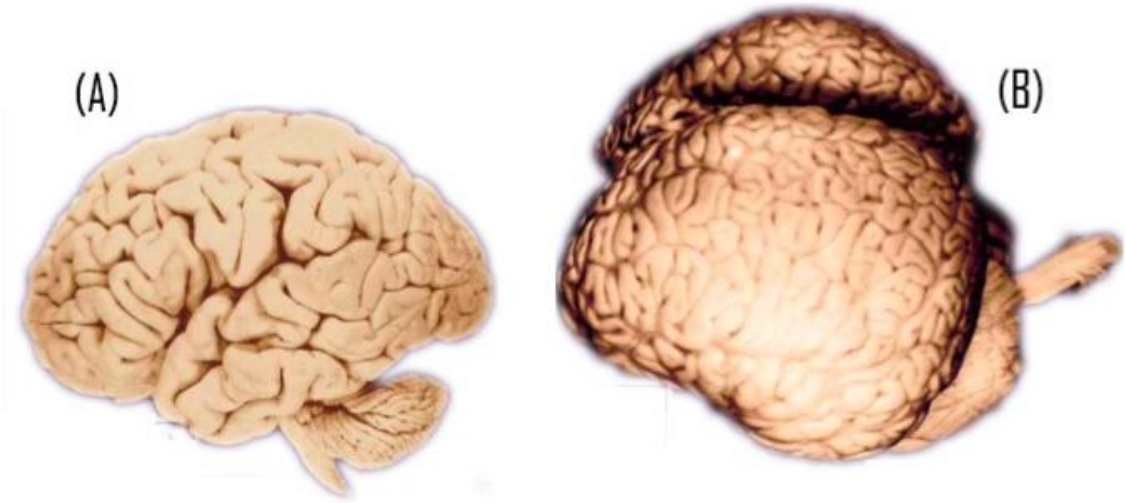

Fig. 14. Comparison between (A) a human brain and (B) a bottlenose dolphin brain (Marino, 2002: 24).

While brain mass can be an indicator of intelligence, various studies have revealed significant cognitive abilities in animals with much smaller brains, such as birds (see Part II, § 3-4), suggesting that brain size alone is not a determining factor (see Marino 2004). Other parameters are therefore considered, such as the ratio between brain size and body size (meaning absolute size alone is not the key factor) and, more reliably, the encephalization quotient (EQ). As discussed in Part I (§ 5.2) and Part II (§ 4.2), the encephalization quotient measures the actual brain size of an animal relative to what would be expected based on its body size. The EQ value thus indicates how much a given species' brain deviates from the typical body-to-brain proportions: an EQ of 1 represents proportionality, an EQ below 1 indicates a relatively larger body compared to the brain, and an EQ above 1 suggests an unusually large brain in relation to body size (Marino 2004: 2). Marino (2004: 4) presents a table of EQ values[81] for various odontocetes, ranging from 0.58 to 4.56, with an average of 2.56. The lowest value, 0.58, belongs to the sperm whale (despite having the largest brain in absolute terms), and it is the only one significantly below 1. However, according to Lori Marino (2004: 4), in this case, EQ is not highly meaningful because the sperm whale has evolved a disproportionately large body for adaptive reasons. The highest value (4.56) belongs to the Tucuxi dolphin (*Sotalia fluviatilis*), though several dolphin species exceed an EQ of 4, including the well-known and extensively studied bottlenose dolphin (4.14)[82], while

[81] There are several ways to calculate EQ: Marino (2004) bases her calculations on Jerison's (1973) formula
[82] Note that according to Roth & Dicke 2005, bottlenose dolphins have an EQ of 5.3 (see also Part II, § 4.1).

orcas have 2.57. For comparison, “EQs of modern nonhuman anthropoid primates range from 1.02 to 3.21 with a mean of 2.0.” (Marino 2004: 5), while that of modern humans is 7.0[83] (Marino 2004: 5). Marino thus notes that “the modern odontocete values reveal that some odontocetes have achieved a level of encephalization second only to modern humans and significantly higher than any of the modern nonhuman anthropoid primates” (Marino 2004: 5). The brains of mysticetes are also very large (as previously mentioned, the blue whale’s brain exceeds 7 kg), but their EQ values are all well below 1: “mysticetes have undergone extraordinary increases in body size without allometric increases in brain size” (Marino 2004: 5). Therefore, according to Marino, as with the sperm whale, “EQ is not particularly meaningful in mysticetes because the general rules about brain and body relationships that underlie EQ do not hold for this group” (Marino 2004: 5).

Moreover, “the large size, high cortical convolutional index and highly derived morphology of mysticete brains establish that these brains have indeed undergone substantial enlargement and elaboration during the course of their evolution” (Marino 2004: 5). Another important detail concerns brain development: most mammals are born with a brain that already weighs about 90% of its adult size; in humans, however, this figure is only about 28% of the final brain size. Cetaceans fall between humans and the mammalian average: for example, in bottlenose dolphins, the brain at birth is 42.5% of its adult size. Among terrestrial mammals, similar values are seen in chimpanzees (54%) and elephants (35%) (Poole 1999). These data suggest that a large brain relative to body size, which undergoes significant postnatal development, is better suited for managing complex cognitive processes.

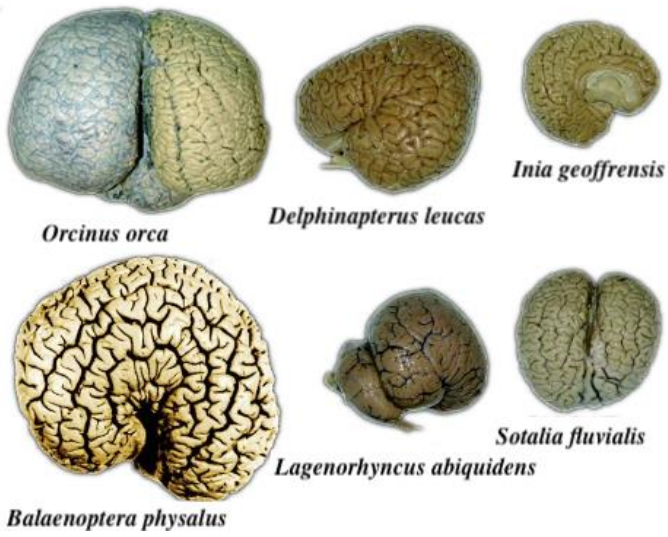

Fig. 15. Brains of different cetacean species, from Hof *et al.* (2005:1144). From left to right: orca, beluga, Amazon river dolphin, fin whale, lagenorhynchus, and sotalia.

[83] However, according to Cairó (2011), human EQ is 6.54.

Beyond their size, the structure of cetacean brains is particularly noteworthy: the hemispheres are highly developed, and the brain surface exhibits many convolutions, in odontocetes even more so than in humans (Berta *et al.* 2006: 148). Marino (2004: 3) notes that throughout evolution, the cetacean brain has acquired a globular shape – an extremely important feature that, according to Benítez-Burraco & Boeckx (2015), may have led to the unique neural connections seen in humans (see Part I, § 3.4). However, brain expansion has primarily affected the parietal, temporal, and occipital regions, with less elaboration of the frontal areas. Compared to archaeocetes, which had much smaller brains, the most evident difference in modern cetaceans is the loss of the sense of smell: the brain regions associated with olfaction, which were clearly present in archaeocetes, have regressed to the point of disappearing entirely in odontocetes and almost completely in mysticetes. What has developed instead, as mentioned earlier, are the structures dedicated to hearing. To put this into perspective, their neural areas related to auditory perception are about ten times larger than those in humans, whereas their visual processing areas are much smaller – approximately ten times more reduced (Marino 2004: 5).

Another less developed aspect compared to terrestrial mammals is the limbic system, particularly the hippocampus. According to Marino (2004: 5), this is due to the reduction of the sense of smell. However, she highlights that the amygdala – believed to be linked to emotion processing, memory, and comparison of past and present experiences – is highly developed, much like in humans. This suggests that its preservation is due to the retention of the limbic system's "non-olfactory" functions. Marino (2004: 6) also notes that near the small limbic system lies an extremely developed limbic cortical lobe: "this juxtaposition of a vastly reduced archicortex and a highly elaborated periarchicortical/entorhinal zone leads to intriguing questions about whether there was a transfer of hippocampuslike functions from the olfactory-based hippocampal domain to other cortical regions" (Marino 2004: 6). Among the other less developed structures is the corpus callosum, which connects the two hemispheres. On the other hand, particularly enlarged regions include the rhombencephalon and the cerebellum, the latter being even larger than in humans (Marino 2004: 7). The most significant issue, however, concerns the neocortex, whose surface area is more extensive than in any other animal (including humans), although it is thinner:

> Most notable among the regions of enlargement in the cetacean brain is the neocortex. The telencephalon is arranged into three concentric tiers of tissue comprising limbic, paralimbic, and supralimbic regions. The high degree of cortical gyrification in many cetacean brains and resulting neocortical surface area of approximately 3745 cm2 is

> unsurpassed among mammals, including humans (2275 cm2; Elias & Schwartz, 1969; Ridgway & Brownson, 1984). Therefore, from both the point of view of total brain size and neocorticalization cetacean brains are highly elaborated. However, the cetacean neocortex is relatively thin with a width between 1.3 and 1.8 mm, as compared with the 3.0 mm thick human neocortex (Haug, 1969; Kesarev, 1971; Ridgway & Brownson, 1984). This combination of extreme surface area and narrow cortical width in cetaceans betrays a unique underlying neocortical organizational scheme that has been the focal point of longstanding and unresolved controversy about the computational capacities of the dolphin brain (Marino 2004: 7).

The neocortex is also organised differently from that of other mammals. For specific details, I refer to Marino (2004: 8-10), but in summary, since the ancient divergence between the ancestors of modern mammals and those of cetaceans, the latter's brain "has apparently exploited a highly conserved neocortical organizational scheme to evolve an extremely elaborated brain capable of complex cognitive processing." (Marino 2004: 8). The result is a brain that, despite originating from a very ancient common ancestor, has followed a different evolutionary path than terrestrial mammals and primates. However, through convergent evolution, it has developed in a similarly complex manner, utilising evolutionary possibilities unexplored by other animals yet arriving at a comparable outcome. For this reason, Marino (2004: 8) argues that "the large cetacean neocortex represents a striking alternative to the forms of elaboration one sees in other large brains, such as those of primates." Consequently, she adds that "the cetacean brain is uniquely valuable for revealing the wide range of structural and functional possibilities that the mammalian neocortex can express" (Marino 2004: 8). Brain connectivity also appears different: cetaceans exhibit what is known as "cortical adjacency", meaning that motor and sensory regions, as well as visual and auditory structures, are adjacent to one another but are not connected by any intervening cortex. In these animals, they are particularly clustered in the frontal region of the brain (Marino 2004: 8). Additionally, the layers that compose the cortex have different thicknesses compared to those of other mammals. According to Marino *et al.* (2007: 969), this could indicate that "thalamocortical projections of cetaceans rely on a very different wiring scheme than in terrestrial species" and that the neuron groupings in mysticetes (which are themselves different from those in odontocetes) might represent "a strategy to optimize intrahemispheric connectivity in their very large brains". What I want to emphasise is that cetaceans possess a brain and neural connectivity system that is very different from that of humans, and yet they may achieve similar results. According to Marino *et al.* (2007: 967), "there is certainly no evidence that the 'cetacean scheme' is incapable of supporting complex processing

similar to that in primates and other mammals" and furthermore, "there is no reason to expect that cetacean and primate prefrontal cortical analogs would be, in fact, located in the same region of the brain". For these reasons, Marino *et al.* (2007: 967) state that "the expansion of the insular and cingulate cortices in cetaceans is consistent with high-level cognitive functions – such as attention, judgment, intuition, and social awareness – known to be associated with these regions in primates". Consequently, Marino (2004: 10) argues that the development of the cetacean neocortex, with its unique connections, "is a special way of achieving computational complexity that represents an alternative to the evolutionary route taken by other large mammals, including primates". She then cites a compelling explanation from Morgane *et al.* (1990):

> […] we are dealing with a very unique situation of a massive cerebral cortex, in many cases considerably larger than in the human brain, but one in which the fundamental cortical structural plan is similar to that of smaller, more conservatively organized brains such as seen in basal insectivores and bats. Computationally, such a cortex is of special interest in that the total number of functional cortical modules may be very great but the fundamental organization is one of relative simplicity compared to those of more progressive terrestrial mammalian forms. (Morgane *et al.* 1990, cit. in Marino 2004: 10).

The great complexity of the cetacean brain is also suggested by the presence of spindle neurons (VEN), which until the 2000s were believed to be exclusive to hominins. Today, they have been identified not only in cetaceans (both mysticetes and odontocetes) but also in elephants and the pygmy hippopotamus – the closest living relative of cetaceans (Raghanti *et al.* 2015), along with other species (see Part II, § 4.2). Spindle neurons have an elongated shape, with one end consisting of a single apical axon (the branch that transmits signals) and the other end featuring a single dendrite (the branch that receives signals), oriented in the opposite direction (Watson *et al.* 2006). These neurons appear to facilitate interregional brain connections and rapid information processing. It is believed that they can evolve through convergent evolution in large brains that need to manage complex connections (Butti *et al.* 2009). However, their presence in the pygmy hippopotamus demonstrates that they can also exist in smaller animals. In humans, these neurons are thought to be linked to intelligence and involved in social behaviour, emotions, judgment, and theory of mind (see Part I, § 5.2). In cetaceans, they are found in brain areas homologous to those in humans – albeit in even greater numbers – suggesting that they may serve a similar function (Hof & Van Der Gucht 2007).

Other neurons believed to be involved in cognitive processes, self-awareness, and imitation are the so-called mirror neurons, which have been identified only in humans, other primates, and songbirds. To the best of my knowledge, the presence of these neurons has not been confirmed in cetaceans. However, several researchers argue that this is likely due to a lack of studies rather than to their effective absence (see Jeon & Lee 2018; Kilner & Lemon 2013). Cetaceans (particularly odontocetes) exhibit complex imitative abilities (see below). Given that spindle neurons have been discovered in various unrelated species and that structures analogous to mirror neurons have evolved convergently in non-primate species, such as songbirds (Prather *et al.* 2008), it is reasonable that this possibility should be kept open, as suggested by Vittorio Gallese, the scientist who first identified these neurons (Marshall 2014). This has been observed also by Herman (2012) who believes that cetaceans must have some kind of analogous system, because they show the greatest imitative skills among mammals and other signs of complex cognition. Another trait which has been considered connected with language production is the FOXP2 gene (see Part I, § 5.2, Part II, § 4.2), present in many cetacean species (Webb & Zhang 2005).[84]

All these features indicate that the brains of modern cetaceans are highly complex. It is intriguing to consider what led to such development, especially since archaeocetes did not exhibit the distinctive characteristics of their descendants. One theory, proposed by Marino *et al.* (2007: 966), suggests that cetacean brains evolved to support the cognitive demands of complex social life, "characterized by communication and collaboration as well as competition among group members." Indeed, various cetacean species – especially odontocetes – demonstrate these traits (see § 2, 3 and 4 below). "In such a society, individuals can benefit from the recognition of others and knowledge of their relationships and from flexibility in adapting to or implementing new behaviours as social or ecological context shifts" (Marino *et al.* 2007: 966). If this hypothesis proves correct, it would have significant implications for the emergence of complex cognition (and possibly language) in humans, suggesting that brain size in hominins may have also increased due to social selective pressures, aligning with Tomasello's (2008) perspective. Another idea is that odontocete brains evolved primarily to support echolocation (Marino *et al.* 2007: 966), although this explanation does not account for mysticetes, which exhibit complex cognition despite lacking this ability. Manger (2006) presents a contrary view, arguing that cetacean brain development resulted from adaptation to a drop in ocean temperatures around 33 million years ago, during the transition between the Eocene and Oligocene, and therefore

[84] Strangely, there are few specific references in the literature about FOXP2 in cetaceans, although several papers claim their present without delving into the topic.

had nothing to do with cognition. Marino *et al.* (2007: 966) counter this argument, stating that while the cooling did occur, and unusual brain development appears to have begun during that period, cetaceans at the time already had sufficient size and thermoregulatory adaptations to withstand low temperatures. Additionally, organisms typically respond to colder climates by increasing body size, not decreasing it. Thus, according to Marino *et al.* (2007: 966), the increase in relative brain size was not linked to climate change but rather to cognitive and social demands, which led to a significant decrease in body mass associated with a (less pronounced) increase in brain size. Regardless of which hypothesis is correct, the evidence indicates that cetaceans exhibit complex cognitive abilities, which are discussed in the following chapters. However, before delving into topics such as cognition, behaviour and communication, it is necessary to detail one last crucial aspect about cetacean biology, strictly related to all their complex skills: how they perceive and produce sounds.

1.4 Vocal production and perception in cetaceans

Berta *et al.* (2006) describes the fundamental characteristics of the sound production and perception system in cetaceans, considering the differences between mysticetes and odontocetes. Regarding perception, all cetaceans are believed to have a well-developed hearing. Odontocetes can perceive frequencies ranging from 30 Hz to 200 kHz, whereas for mysticetes,[85] estimates suggest a range from 10 Hz to 36 kHz (Mooney *et al.* 2012). For comparison, human hearing ranges from 20 Hz to 20 kHz. Without delving into a detailed anatomical description of the ear (for which I refer to Ketten 1992), it is important to note that the auditory system of mammals evolved to perceive sounds transmitted through air and, consequently, underwent significant modifications in cetaceans,[86] which emit and perceive their signals underwater.

In odontocetes, echolocation (see below) requires a hearing system capable of simultaneously perceiving both the strong sounds produced by the animal while diving and the faint echoes of previously emitted sounds. As with many other aspects, experimental evidence is primarily based on studies conducted on small captive species, meaning our understanding remains incomplete. However, it is known that odontocetes have a unique way of perceiving sound among mammals: sound waves are intercepted by the mandible, which directly transmits vibrations to the middle ear. The external auditory canal is still present and retains the typical mammalian structure, but it is very narrow and less sensitive

[85] For more details on how mysticete auditory capabilities are assessed, see Mooney *et al.* (2012: 230-231).
[86] In section 1.3 I have discussed how significant developments have also occurred in brain structures related to auditory perception.

to sound than the mandible.[87] The two posterior ends of the mandible are very thin and protruding, containing two fat-filled sacs that are directly connected to the auditory bulla of the middle ear. When sound waves strike the ends of the mandible, the vibrations travel through the fat, which – like the melon and spermaceti organs (see below) – helps direct the sound toward the middle ear (Berta *et al.* 2006: 280). Figure 16 illustrates the anatomical structures involved in sound perception.

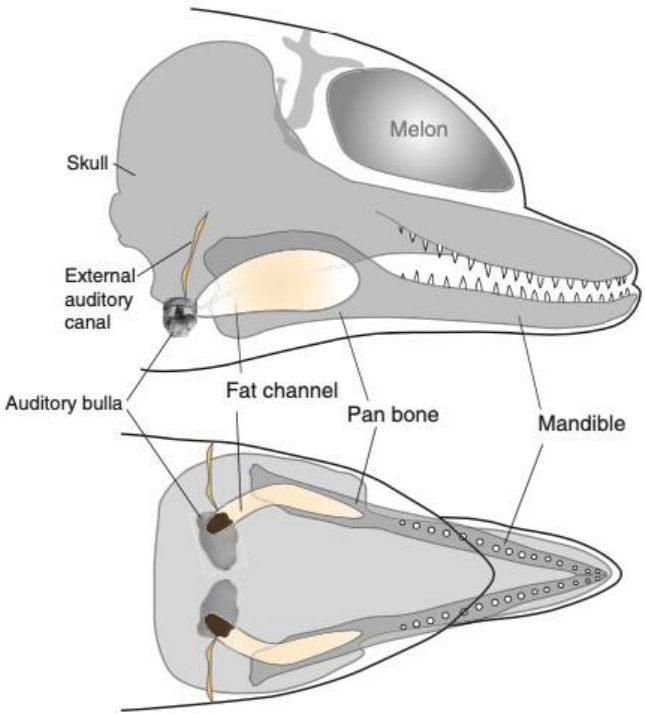

Fig. 16. Illustration of a dolphin's head showing the position of the fat sacs within the mandible and their proximity to the auditory bulla, taken from Berta *et al.* (2006: 282). The underdeveloped external auditory canal is also visible.

The odontocete ear has two distinct components: the tympanic bones and the periotic bones, which are separated from the rest of the skull by air-filled cavities containing oil and insulating mucus. As a result, the ears remain suspended and isolated from the skull – and even from each other – allowing them to function independently in sound reception (Berta *et al.* 2006: 282). The tympanic membrane is reduced to a calcified ligament connected to the ossicles: the malleus, incus, and stapes. The exact mechanism of sound transmission from the middle ear to the inner ear remains unclear, but the preferred model according to Berta *et al.* (2006) is that of Hemilä *et al.* (2001). In this model, vibrations reaching the tympanic

[87] In an experiment by Brill *et al.* (1988), the mandibles of some bottlenose dolphins were covered with rubber shields that dampened sound. These individuals, when inhibited in this way, showed reduced echolocation abilities.

bone are transmitted through the malleus, then via the incus and stapes to the inner ear, whose cochlea does not differ significantly from that of humans (Berta *et al.* 2006: 283).

The hearing apparatus of mysticetes is far less studied and remains poorly understood. However, there are some similarities to odontocete structures, as well as several key differences. First, the external auditory canal of mysticetes is completely closed, and its functionality is unclear. However, an extension of the tympanic membrane occupies part of the canal, ending in a waxy, horn-shaped protrusion that can reach up to one metre in length. The tympanic and periotic bones differ from those in odontocetes, yet they share some characteristics, including the reduced tympanic membrane, which is connected to the malleus. Since the mandible does not exhibit the same sound reception properties as in odontocetes, it is believed that sound is still received through the obstructed auditory canal (Berta *et al.* 2006: 280). A more recent study (Yamato *et al.* 2012) found that right whales possess "well-formed fat body adjacent to the mandibular ramus and lateral to the tympano-periotic complex [...] This fat body inserts into the tympano-periotic complex at the juncture between the tympanic and periotic bones and is in contact with the ossicles" (Mooney *et al.* 2012: 231). Such findings suggest that even the less-studied mysticetes may have a sound reception system that utilises fat-filled structures (Mooney *et al.* 2012: 231).

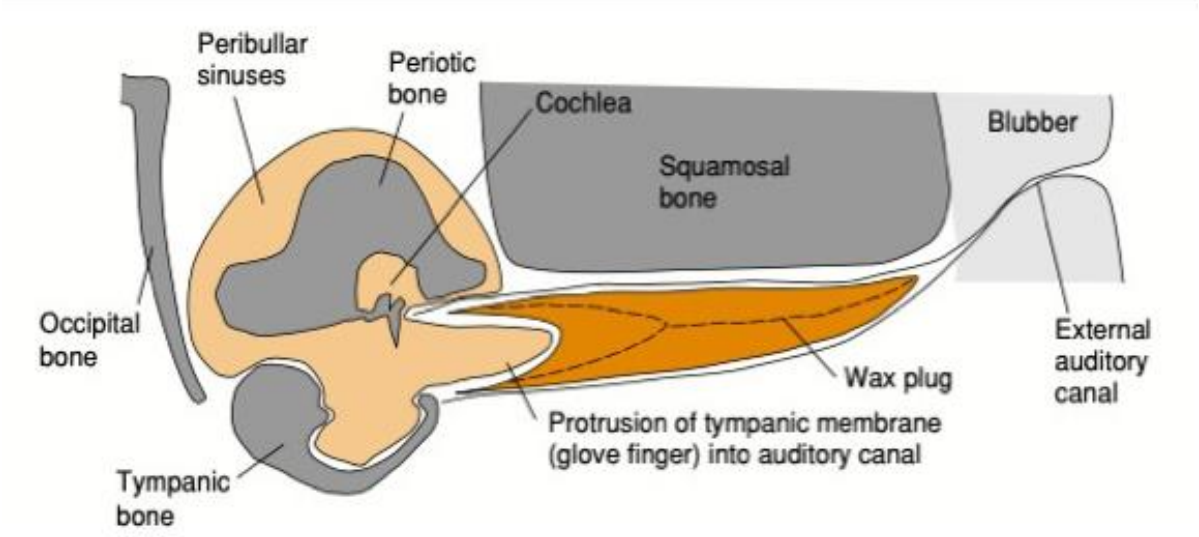

Fig. 17. Section of the ear region of a mysticete (Berta *et al.* 2006: 281).

Archaeocetes, the common ancestors of mysticetes and odontocetes, are believed to have used high frequencies. Thus, it has been proposed that the ability of mysticetes to hear very low frequencies developed after their divergence. According to Ketten (1992), this adaptation may have resulted from an increase in body size, possibly due to ocean cooling and the previously described concentration of new food sources in polar regions (see § 1.2).

A later study (Park *et al.* 2017) challenges this perspective, arguing that archaeocetes used low frequencies, contrary to previous assumptions. In this reconstruction, mysticetes retained this characteristic, which would therefore predate both their increase in size and the emergence of baleen; thus, it would have been odontocetes that deviated more significantly from their ancestors by developing a novel high-frequency sound system.

Regarding vocal production, cetaceans produce sounds ranging from 30 Hz to 200 kHz in odontocetes and from 10 Hz to 31 kHz in mysticetes, meaning that mysticetes can generate much lower-frequency sounds (Berta *et al.* 2006: 289-293). In humans, by comparison, the average frequency of speech ranges from approximately 100 Hz in adult males to around 300 Hz in children. The differences, of course, do not end here. In humans, sounds are produced by expelling air from the lungs, which, upon passing through the larynx, is shaped by the vocal cords and further modified by the throat, mouth, tongue, and lips. Cetaceans, by contrast, lack vocal cords, and their mouths are not involved in sound production; in fact, air does not even need to be expelled from the body. Their mechanisms are therefore fundamentally different from ours and also differ between mysticetes and odontocetes. Odontocetes produce two distinct types of sounds: clicks, which are used for echolocation, and various whistles and other sounds associated with communication.[88] Within odontocetes, it is important to differentiate between the vocal structures of smaller species (such as dolphins, orcas, and belugas) and those of sperm whales, which are significantly larger. Both groups possess an organ in their heads called the melon, which is believed to be involved in sound transmission. However, the sperm whale's melon is highly developed and has a different composition, and this species also uniquely possesses the spermaceti organ (Berta *et al.* 2006: 277). Despite these anatomical differences, the function of these structures appears to be the same. Both adaptations are unique in the animal kingdom and exhibit significant homologies, suggesting that while differences exist, they represent evolutionary modifications of the same fundamental mechanism (Berta *et al.* 2006: 277). For this reason, they can be described together, while also highlighting their divergences. Figures 18 and 19 illustrate the vocal production structures of a bottlenose dolphin, as a representative of smaller odontocetes, and a sperm whale, respectively.

[88] See § 2.2, 2.3, 3, 4 for a detailed analysis of the vocalisations produces by several odontocete species.

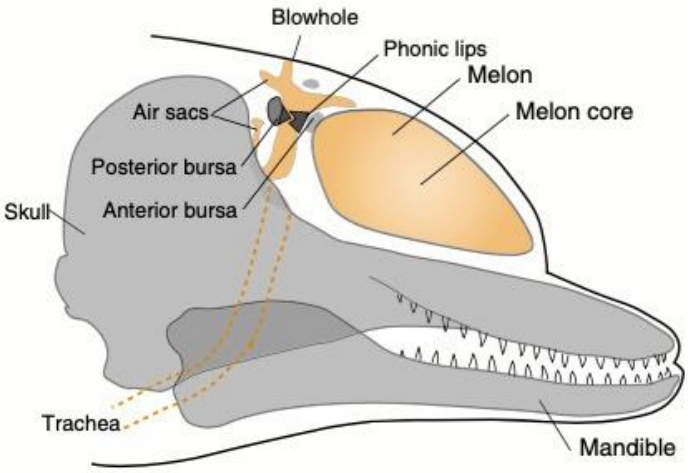

Fig. 18. Semi-transparent illustration of a dolphin's head, showing the structures associated with sound production, modified from Cranford *et al.* 1996 (in Berta *et al.* 2006: 276). Air from the trachea moves up toward the blowhole and passes through the phonic lips, which cause the anterior and posterior bursae to vibrate. The resulting sound is directed toward the melon, which amplifies and precisely channels it in the desired direction. The air is not expelled but can be reused for the same process. Returning echoes first strike the mandible, which transmits the vibrations to the ear.

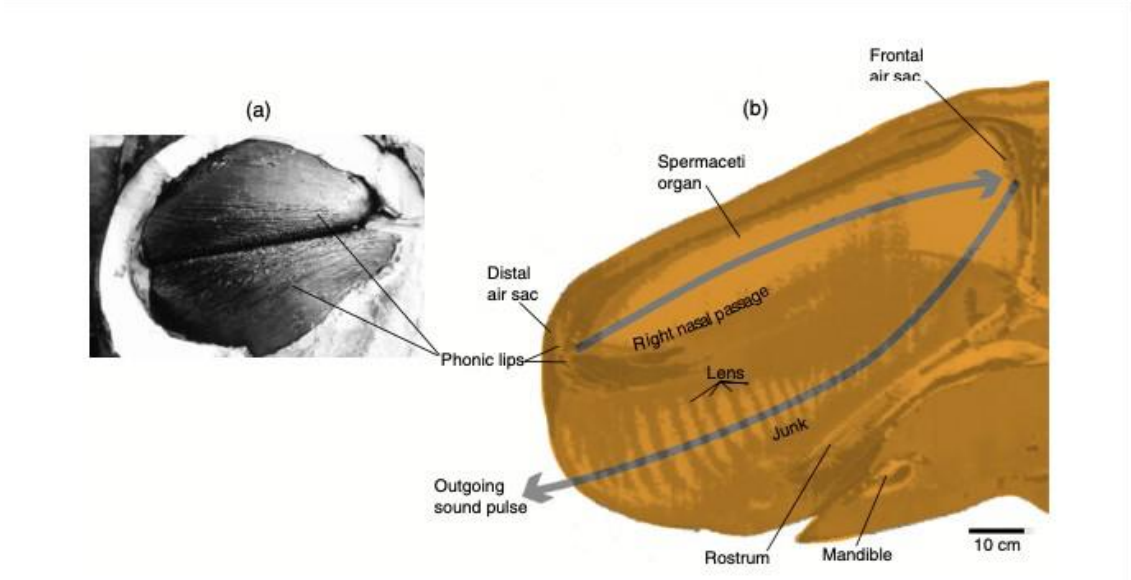

Fig. 19. (a) Phonic lips of a sperm whale, with the upper air sac partially removed. (b) Section of the head of a newborn sperm whale, highlighting the structures involved in phonation, from Cranford, 1999 (in Berta *et al.* 2006: 279). The arrows indicate the presumed sound pathway: air passes through the phonic lips, generating powerful clicks that are directed inward. These clicks travel through the spermaceti organ and "bounce" off the frontal air sac, from which they are redirected into the large melon. The sperm whale's melon is divided into cartilage compartments that separate layers of fat. These structures amplify and direct the signal with extreme intensity and precision, producing what is considered the most powerful sound in the animal kingdom.[89]

[89] For a reconstruction of click production and reception, see www.youtube.com/watch?v=CJP8jC1SikQ.

As illustrated in the figures, odontocetes produce sound by passing air through the phonic lips,[90] a structure located in the upper nasal cavity beneath the blowhole. More precisely, the phonic lip complex is positioned below the ventral plane of the vestibular air sacs and consists of two pairs of fat-filled dorsal sacs – one anterior and one posterior – into which the actual phonic lips are embedded.[91] These lips are two slit-shaped muscular tissues. Additionally, there are a strong cartilage component and a robust ligament connected to the blowhole above. These elements are interconnected and "all suspended within a complex array of muscles and air spaces" (Berta *et al.* 2006: 277). When air passes through the phonic lips, which can contract or expand similarly to human vocal cords to regulate airflow, vibrations are produced in the dorsal fat-filled sacs. These vibrations are further modulated by the entire structure of muscles and cartilage, which, much like the human larynx, can unconsciously produce sounds with great precision. Moreover, the air is not expelled through the blowhole; instead, it continues to circulate and can be reused in the process. This apparatus is believed to be responsible not only for the production of clicks used in echolocation – through the rapid opening and closing of the phonic lips – but also for the generation of whistles and other communication sounds (Berta *et al.* 2006: 277, see also § 4 below). These sounds are then directed to the melon, the organ visible in Figures 18 and 19 (in the latter it is labelled as junk,[92] for sperm whales). The melon is composed of low-density lipids, which function as an acoustic lens: sound waves are focused into a beam that is precisely directed outward, into the direction the head is pointing (Berta *et al.* 2006: 277). This beam is used for echolocation, as illustrated in Figure 20.

[90] Originally, they were called "monkey lips" due to their form in sperm whales. However, since they have different shapes in other odontocetes, the term "phonic lips" was preferred. Nevertheless, the acronym MLDB (monkey lips – dorsal bursa) is still used to refer to the entire structure (Berta *et al.* 2006: 276).

[91] Sperm whales have only one set of phonic lips, while other odontocetes have two, allowing them to produce two different sounds simultaneously (Berta *et al.*, 2006: 277).

[92] The term junk, meaning trash, was coined by whalers, who discarded that part because it is full of cartilages that hindered the extraction of spermaceti, which was present in lesser quantity and quality (Berta *et al.* 2006: 182).

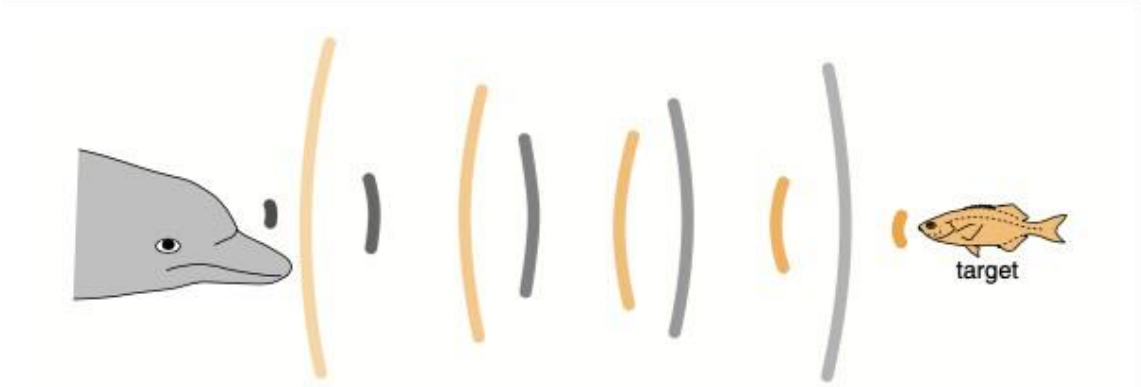

Fig. 20. Illustration of how echolocation functions as a biological sonar (Berta *et al.* 2006: 290). The dolphin produces a series of clicks, known as "trains" (gray lines), which reflect off surfaces and return as echoes (orange lines). These echoes reach the dolphin's mandible, transmitting vibrations to the ear and allowing it to locate prey.[93]

Although the mechanism of sperm whales, illustrated in Fig. 19, is more complex, the principle of echolocation is the same:[94] all odontocetes produce rapid sequences of clicks by opening and closing the phonic lips, direct them through a fat-filled organ, use them for orientation and hunting, in a manner similar to bats, and receive the echoes produced by the bio-sonar through the mandible, which propagates the vibrations to the ear (Berta *et al.* 2006: 288-292). Echolocation deserves to be described in greater detail: "The sound production system of all odontocetes generates trains or pulses of broad-frequency clicks of very short duration [...]. As each click strikes a target, a portion of its sound energy is reflected back to the source [...]. Each click lasts from 10 to 100 μs and may be repeated as many as 600 times each second" (Berta *et al.* 2006: 289). To give some examples of range, the clicks of bottlenose dolphins range from 30 to 150 kHz, going well beyond human acoustic capability, those of porpoises range from 40 Hz to 150 kHz, and those of Commerson's dolphins (*Cephalorhynchus commersonii*) even reach 200 kHz. The signals can be of low, medium, or high frequency, ranging approximately from 30 Hz to 200 kHz. It is believed that low-frequency sounds are used for environmental detection and communication, and high-frequency ones for the identification and classification of objects and prey: low frequencies travel farther, and therefore could serve to identify distant elements in an indistinct manner; high frequencies, directed towards closer objects, could provide greater sharpness at the expense of range (Berta *et al.* 2006: 289). Odontocetes, while producing these trains of

[93] For a video of echolocation, see en.wikipedia.org/wiki/Whale_vocalisation#/media/File:Delfinekko.gif
[94] It should be noted that echolocation has been experimentally confirmed only in a few species of captive delphinids, due to the obvious impossibility of conducting experiments with larger species; nevertheless, all odontocetes in nature produce clicks and present homologous anatomical structures, so it is logical to assume that it is an ability present in the entire parvorder, although with some variations (Berta *et al.* 2006: 288).

clicks, continuously adjust the repetition rate, so that the returning echoes reach the mandible during the silent intervals between one sound and another. The time elapsed between the production of a click and its return in the form of an echo indicates the distance from the object that has been struck by the sound, and the variation of the echoes of successive clicks allows the estimation of its speed and direction. The mechanism, highly sophisticated, is naturally unconscious, and is integrated with further adjustments: since odontocetes (except sperm whales) have two pairs of phonic lips, while one pair is engaged in producing these clicks, the other can independently produce frequency-modulated whistles, ranging from 2 to 30 kHz. Meanwhile, the animal continues to modify the frequency of the 'main' clicks to adapt them to the surrounding environment, avoiding background noises that could disturb the signal. Based on certain measurements carried out in experiments, the maximum echolocation range detected in bottlenose dolphins is one hundred metres,[95] but it is believed that sperm whale clicks can cover much greater distances (Berta *et al.* 2006: 290-291). Sperm whales produce two types of clicks: the 'traditional' ones show long and variable intervals between one sound and another, while those defined as 'creaks' have a long duration, between 5 and 6 seconds, and very short intervals; moreover, they are repeated in rapid succession. It is believed that 'traditional' clicks are used to scan the seabed in search of prey, while the 'creaks', which are always produced while diving in foraging areas, are linked to rotational movements that sperm whales would perform before capturing their targets: according to the calculations of Madsen *et al.* (2002) the 'traditional' clicks would be capable of identifying targets up to 16 km away, and the 'creaks' up to 6 km (Berta *et al.* 2006: 292).

An alternative use of vocal production is the emission of so-called bangs, very loud sounds intended to damage, stun, or temporarily disorient prey, allowing them to be captured before they attempt to escape. This strategy, likely generated by the same system used for clicks, is a response to an adaptation of certain types of prey, fish from the Clupeidae family (such as herrings). These fish have learned to detect the frequencies of the clicks used in echolocation and associate them with the danger of a predator: when they perceive this type of sound wave, they flee (Berta *et al.* 2006: 299). This phenomenon, which presents a fascinating evolutionary convergence with the ability developed by moths to evade bats, raises questions about the origin of echolocation. Clupeid fish are older than odontocetes: they have existed for about 130 million years, and thus, it is believed that this ability is well established. Since odontocetes are thought to have evolved echolocation around 33 million

[95] Ivanov (2004) argues that the range can exceed 650 metres (Berta *et al.* 2006: 291).

years ago, these fish could not have developed the 'anti-echolocation' ability to defend themselves from them: consequently, it is reasonable to assume that echolocation, which likely exerted selective pressure leading fish to develop their countermeasure, must have already been present in some aquatic life form preceding odontocetes (Berta *et al.* 2006: 301).

Unlike odontocetes, very little can be said about the vocal production of mysticetes. According to Berta *et al.* (2006: 279), "no anatomical studies have shown equivalent structural specializations for sound generation or transmission in mysticetes". What we do know is that mysticetes have a larynx but lack vocal cords, as well as the complex apparatus of phonic lips and the melon: "the cranial sinuses of mysticetes are thought to be involved in phonation although no precise or general mechanism has so far been demonstrated" (Berta *et al.* 2006: 279). More recent studies, referenced by Cazau *et al.* (2013), have identified anatomical parts that seem to be involved in vocal production, located in the laryngeal region: the U folds (Reidenberg & Laitman 2007), and the laryngeal sac (Reidenberg & Laitman 2008). The U-shaped cords present many similarities with human vocal cords and are embedded in a complex of muscles and cartilages capable of movement. The laryngeal sac is described as "a soft extensible oval balloon with an extensive surrounding musculature" (Cazau *et al.* 2013: 2269). On the opposite side of the laryngeal sac, the nasal region consists of an area of muscles and soft tissues leading to two nasal cavities, which end at the blowhole. It is believed that sound is produced by the recirculation of air: pushed by the lungs into the trachea, when it reaches the larynx, it passes through three respiratory valves in the region (one of which is the U-shaped cords). Then, the air is not expelled through the blowhole but returns to the lungs, and so on: this means that sound is produced in both directions (as air moves from the lungs to the blowhole and back again). Cazau *et al.* (2013), to whom I refer for details, summarise that, considering the combinations of opening and closing of the three valves, three different configurations are possible, meaning three different paths the air can take, producing different sounds. This was deduced by studying real anatomical samples of humpback whales and reconstructing their structures; then, the three configurations were 'tested' to verify if they were compatible with recorded humpback whale vocalisations in nature. Since this mechanism has only been inferred but not observed, it is not known whether all these configurations are actually used, or if they are truly related to the frequency changes in humpback whale songs. It is also unclear whether they are voluntarily controlled or used specifically to shape sounds: they could, for instance, be used to facilitate breathing, which is certainly the primary function of the system. In this case, mysticete sounds would be nothing more than an involuntary noise produced by air

recirculation, a kind of meaningless 'snoring' (Mercado & Handel 2012). However, Cazau *et al.* (2013) believe that the production is intentional and has a communicative function, as I show in section 2.1.

It is important to emphasize once again that animals of this size are not easily studied: if there were no "small" odontocete species, such as bottlenose dolphins (from which most knowledge derives), their mechanisms of vocal production and perception would likely be poorly understood as well. This also applies to echolocation : Berta *et al.* (2006: 292) report that broadband click trains have been recorded in the presence of gray whales, rorquals, and right whales, both in the Pacific and Atlantic Oceans. These recordings were made by a network of hydrophones originally dedicated to intercepting Soviet submarines, which the U.S. Navy made available to scientists to listen to the deep sea, allowing mysticetes to be tracked instead of submarines. Since these sounds have very low frequencies and very high wavelengths and are alternated with long periods of silence, it is believed that they are used for long-distance communication rather than for echolocation.[96] An undemonstrated hypothesis in favour of echolocation has been proposed by Clark & Ellison (2004): according to them, "these tone pulses might be used by large mysticetes to detect very large scale oceanic features, such as continental shelves, islands, and possibly sharp differences in water density associated with divergences or upwelling of cold water" (Berta *et al.* 2006: 294). At present, however, these theories remain in the realm of speculation. In the next chapter, I discuss how these features are used in communication.

1.4.1 The problems with the definition of "dialect"

Before proceeding with a description of the communicative skills of cetaceans, a terminological clarification is required. As we will see in sections 2.2 and 2.3, in the study of nonhuman animal communication, the concept of "dialect" has been applied to describe culturally transmitted vocal variations within certain species, most notably in killer whales. In this context, "dialect" refers to a group-specific repertoire of discrete, repeated vocalisations (commonly known as calls), which are acquired socially and persist across generations within stable matrilineal groups (Ford 1991, Yurk *et al.* 2002, see also § 2.2). These acoustic patterns are not attributed to genetic or ecological differences but instead reflect social learning and group identity (Rendell & Whitehead 2001). However, while this usage of the term draws a useful parallel to human linguistic variation, it is not directly

[96] If such large waves (which can reach 100 metres in rorquals) were used for echolocation, they would not be able to detect smaller objects (Berta *et al.* 2006: 292).

comparable to the definition of “dialect” in linguistics, due to fundamental differences in cognitive, structural, and symbolic complexity. In linguistics, a dialect is defined as a regional or social variety of a language, distinguished by systematic differences in phonology, vocabulary, and grammar (Berruto 2012, Chambers & Trudgill 1998).[97] Human dialects are not merely sets of sounds but fully developed linguistic systems embedded within broader sociohistorical and cultural contexts. They reflect both identity and social dynamics and are often shaped by political, geographical, and historical factors. Despite these differences, there are some interesting points of convergence between the two phenomena. Both human and orca dialects are socially learned and serve as markers of group membership. They are maintained through intergenerational transmission and can function to reinforce social cohesion (Janik & Slater 2000, Deecke *et al*. 2000). However, human dialects involve complex syntax and semantic content, and are part of an intentional and symbolic communicative system. In contrast, orca dialects, while structured and group-specific, lack evidence of syntactic organisation or symbolic reference (see § 2.2 below). Thus, although the term “dialect” is applied in both domains, it denotes phenomena that, while functionally analogous in some respects, are fundamentally distinct in form and complexity. For this reason, a more appropriate definition for orcas’ (and other species’) sets of calls would be “in-group repertoires”. However, being “dialect” the term that has always been used for this phenomenon in ethology I will maintain it, but with the use of quotation marks to distinguish it from human dialects.

[97] Note that the term "dialect" refers to two different types of language varieties (Coseriu 1980). So-called primary dialects emerge directly from an ancestral language they share with the standard language (e.g. the Italian dialects, that emerged from Latin). Thus, they evolve in parallel with a standard language. Secondary dialects, instead, develop later from a standard language (e.g., the diatopic varieties that emerge from standard Italian).

2. Humpback whales, killer whales, and sperm whales

In the previous chapter we have seen that cetaceans possess highly developed brains which, although different from the human brain, appear capable of sustaining complex cognition – a hypothesis supported by empirical demonstrations of intelligence observed in the wild and through experiments in captivity, which will be discussed in the following chapters of this dissertation. Although the idea that cetaceans are highly intelligent is widely accepted (Marino *et al.* 2007), their actual level of cognition remains debated, particularly in the case of delphinids. Marino (2004: 11) highlights that most research on cetacean cognition and behaviour has focused on bottlenose dolphins, with less attention given to orcas, belugas (*Delphinapterus leucas*), and other small odontocetes, which are easier to study compared to large mysticetes. Because of this, the overall picture is incomplete, and "cognitive and behavioral differences across cetacean species" should be expected. "However, the sizeable literature on the bottlenose dolphin can serve as an indication or 'existence proof' of cognitive and behavioral abilities in cetaceans in general" (Marino 2004: 11). Among these abilities there are a developed memory, complex skills in imitation (both vocal and behavioural), innovation, tool use, and even the understanding of abstract rules by captive animals (Marino *et al.* 2007, Whitehead & Rendell 2015). The learning ability of cetaceans is well-documented and is believed to give rise to 'cultures': "field studies have documented impressive cultural learning of "dialects", foraging sites, and foraging and feeding strategies in cetaceans" (Marino *et al.* 2007: 970). Understandably, these abilities have been identified mainly in the most studied species (bottlenose dolphin, orca, sperm whale, humpback whale), which nevertheless provide an effective overview, as they include both mysticetes and odontocetes, representing animals with varying degrees of distant or close relatedness (Marino *et al.* 2007: 970). Some species even exhibit 'cultural diversity', with groups belonging to different cultures (social groups with their own "dialects", kinship structures, and 'customs') coexisting in the same regions (Marino *et al.* 2007: 970). There are reasons to believe that many cetacean behaviours are transmitted horizontally, in the sense that they are learned by adults through imitation of other adults. In addition to the vertical transmission from mother to offspring, other elements appear to depend on a broader social context (Whitehead & Rendell 2015). In these cases, culture is defined as "information or behaviour – shared within a community – which is acquired from conspecifics through some form of social learning" (Whitehead & Rendell 2015: 29). This differs from genetically transmitted animal behaviours, which do not involve active learning since they are already part of the evolutionary 'package'. However, it is important to note that Whitehead & Rendell's (2015; see also Rendell & Whitehead 2001) definition of culture has been widely

criticised for being too inclusive. Some argue that if cetacean behaviour is to be considered 'culture', then a new term must be found to describe human capabilities, which exhibit vastly more extensive and developed properties, forming a kind of 'hyper-culture' (Barkow 2001). Others have 'accused' Rendell and Whitehead of deliberately using this definition to include cetaceans, arguing that social learning is widespread among various species, without needing to invoke a unique form in dolphins and whales or comparisons with humans (Tyack 2001).

Independently of this debate, the developed social skills of cetaceans – and their ability to learn in ways that generate forms of culture – "are arguably dependent on a complex and flexible communication system, encompassing vocal, visual, tactual, and possibly chemical signals" (Marino *et al.* 2007: 970). Here, it is important to underline the distinction between mysticetes and odontocetes, which have been separated by approximately 33 million years of independent evolution. Odontocetes produce rapid sequences of sounds known as clicks, which are also used in echolocation, along with an array of whistles and low-frequency noises (see § 1.4 above). Mysticetes, on the other hand, produce entirely different vocalisations, and they do not appear to use echolocation (see § 2.1 below). However, the mechanisms they use for orientation remain unknown. For both suborders, in addition to vocalisations, other senses such as vision and touch must be considered, as they are believed to play important roles in communication and social relationships. Regarding touch, cetaceans are thought to have heightened sensitivity in certain areas of their bodies – Tyack (2000) compares the sensitivity of their fins, head, and genital areas to that of human lips – and various species interact by rubbing against and caressing each other[98] (see Tyack 2000). As for vision, it also plays a role in non-vocal communication, as several species use gestures – such as head movements – to signal hostility. Coloration can also convey information: Tyack (2000: 277) notes that each humpback whale has slightly different pigmentation, allowing individuals to recognise each other. Similarly, since spotted dolphins (genus *Stenella*) change colour as they age, this feature helps other individuals estimate the age of a conspecific.

Keeping in mind these peculiarities, I now focus more in detail on the species which offer more insights about the complex cognition, sociality and communication of cetaceans: for mysticetes I describe humpback whales (§ 2.1), for odontocetes I describe orcas (§ 2.2), sperm whales (§ 2.3) and bottlenose dolphins, which are discussed in greater detail (§§ 3-4). This analysis will also allow to highlight the "linguistic" traits of these animals, with a

[98] According to some researchers, this activity strengthens social bonds and is comparable to primate grooming (see Sakai *et al.* 2006).

special attention to the properties that I identified as pivotal for human language in the previous sections: compositionality (1-merge), recursion, and size of lexicon.

2.1 Humpback whales

Mysticetes are more difficult to study than odontocetes due to their big size. However, these animals are often observed alone, and it is believed that they have less structured interactions than odontocetes. The mother-offspring bond – including the care and teachings it entails – is nonetheless strong and enduring, entailing a phase in which the calf learns vocalisations and behaviours from the mother. This kind of bond proves to be a common characteristic of cetaceans, fundamental to their social system (Whitehead & Rendell 2015). However, there is insufficient information to describe the characteristics of mysticetes with certainty. For example, Lomac-MacNair & Smultea (2016) state that most behaviours of blue whales, even the most basic ones, have never been analysed. The same applies to minke whales (Risch *et al.* 2019) and other species, making it difficult to provide a comprehensive description of their habits.

Most of the knowledge about mysticetes comes from humpback whales, the best studied species. They appear to have a fission-fusion society based on reciprocal altruism,[99] and it is thought that this may apply to other mysticetes as well (Valsecchi *et al.* 2002). This complex and flexible social structure varies depending on their location, reproductive state, and seasonal migration patterns. Humpback whales are typically observed alone or in small, temporary groups that form for specific purposes, such as cooperative feeding, mating, or migrating (Whitehead & Rendell 2015). These animals undertake long migrations between high-latitude feeding areas and low-latitude breeding grounds. During migrations, groups of individuals form, which can in turn merge into larger groups in breeding and feeding areas (Valsecchi *et al.* 2002). Some whales aggregate in specific stopover areas, such as Hervey Bay, where females with calves tend to linger longer than other groups, likely for calf development and protection from predators (Brooks *et al.* 2024). The composition of groups along migratory routes changes frequently. During the breeding season, males engage in competitive groups, where multiple individuals compete for access to a female. According to Connor *et al.* (1998), the fission-fusion societies of humpback whales might be compared to those of chimpanzees.

[99] 'Reciprocal altruism' refers to the fact that animals help each other in times of need but do not have 'family' structures that maintain bonds beyond immediate necessity.

A crucial feature for humpback whales, as for other cetaceans, is cooperation, whose best example is the coordinated hunting technique of *bubble-net feeding* (Wiley *et al.* 2011). In this strategy – executed either by an individual alone or in groups of up to twenty – humpback whales dive a few meters below a school of fish and release bubbles from their blowholes to encircle the prey. This *bubble net* disorients the fish, forcing them into a tighter formation, at which point the whales surface with their mouths wide open, capturing large amounts of prey in a single gulp (Wiley *et al.* 2011). The whales perform this movement in a spiralling pattern while releasing bubbles, sometimes incorporating a tail slap, which appears to make the technique even more effective (Wiley *et al.* 2011).[100]

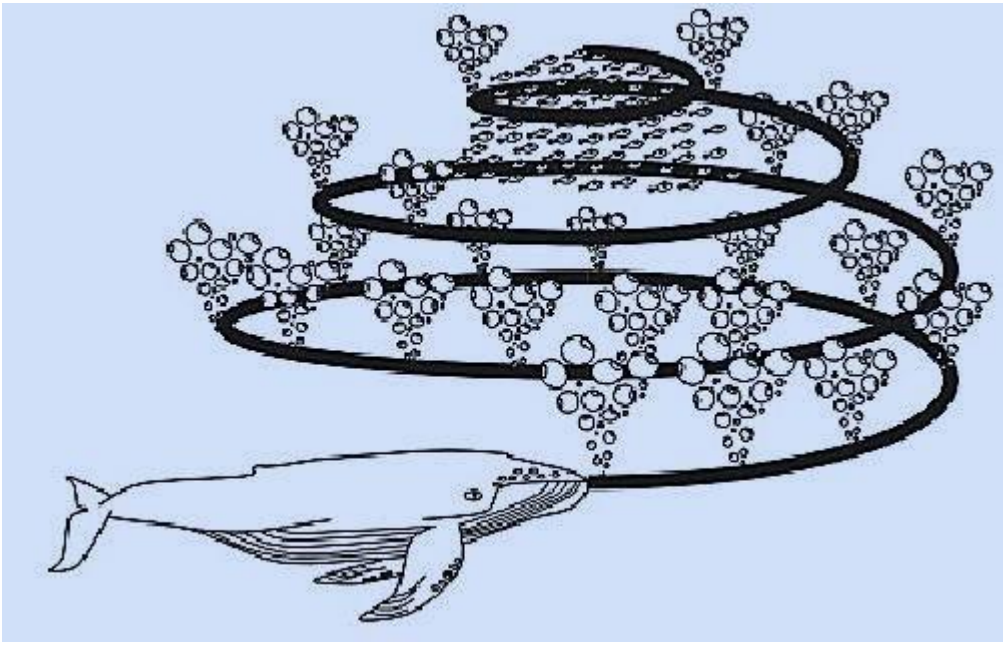

Fig. 21. Illustration of the spiralling movement in bubble-net feeding (El-Amary *et al.* 2018).

Despite its complexity, this technique might reflect general cognitive abilities rather than social learning. However, Whitehead & Rendell (2015) argue that the final tail slap is a more recent innovation, first observed in a single individual before spreading to others within and beyond the original group. If this is true, it would constitute an example of horizontal learning, shaped more by culture than by genetics – though this remains undetermined. Similar techniques have been observed in other whale species, such as the *lunge feeding* of rorqual whales.[101] The most interesting example of horizontal learning, nonetheless, come from the vocalisations that are discussed below.

[100] For a video of bubble-net feeding, see www.youtube.com/watch?v=JNhldKgPRg0.
[101] For a video of lunge feeding, see www.youtube.com/watch?v=cbxSBDopVyw.

2.1.1 Vocalisations of mysticetes

Mysticete vocalisations are more challenging to analyse than odontocetes due to their size and migratory nature. Moreover, their communication occurs over vast distances, making it difficult to observe both the sender and the receiver of a message simultaneously. Although modern technology allows researchers to record their sounds using hydrophones, this is rarely sufficient to determine the context in which they are produced. Some of these sounds have remained a mystery for years and were initially classified as noises of unknown origin. One example is the so-called *bio-duck* sound,[102] first detected by U.S. submarines off Antarctica during the Cold War and later consistently recorded by military hydrophones made available to science. It was not even clear whether this was a biologically produced sound until Risch *et al.* (2014) managed to attribute it to the Antarctic minke whale (*Balaenoptera bonaerensis*). Another case is the *biotwang*,[103] a sound recorded near the Mariana Trench in an entirely unexpected location and context. Its origin remains unconfirmed, but it is believed to be comparable to the peculiar vocalisations of minke whales, which are still poorly understood (Nieukirk *et al.* 2016). This underscores that the knowledge presented in this section is incomplete, subject to future changes, and could be completely overturned in some cases. Nonetheless, Laela Sayigh (2013) provides a useful review of studies on mysticete communication, categorising them by family and presenting the current state of knowledge. What stands out is the limited understanding of both communication processes and their precise functions, highlighting the need for extensive research. However, I can outline the fundamental known aspects, as well as the differences and similarities between species. First, I distinguish between two types of vocalisations: simple calls, used by males and females of all ages, and more complex, structured calls, which are believed to be produced only by males in specific situations and are referred to as *songs*. While calls are common across all species, songs are produced only by some, mainly those in the *Balaenopteridae* family. I first describe the functions of the calls and then focus more closely on the structure of songs.[104]

Despite the variety of 'simple' calls, they are thought to serve a similar social function across different species, related to maintaining contact with conspecifics. Some experiments with humpback whales have shown that these sounds – played back by researchers in breeding areas – attracted other individuals from up to 9 km away (Tyack

[102] For a recording of the bio-duck, see it.wikipedia.org/wiki/File:Bio-duck.flac.
[103] For a recording of the biotwang and the vocalisations of minke whales, known as Star Wars sounds (see below), see www.youtube.com/watch?v=HxOHV65DpSs.
[104] For acoustic recordings of various mysticetes, see en.wikipedia.org/wiki/List_of_whale_vocalisations.

1983). Staying within the *Balaenopteridae* family, certain calls of blue whales, fin whales, and minke whales are believed to play a role in conveying information about food or predators (Oleson *et al.* 2007; Sirovic *et al.* 2013). Given the variability in calls, it is also believed that they help individuals coordinate collective activities, such as defending food resources (Sayigh 2013). There is ongoing debate about how far mysticete calls can travel. Sayigh (2013) summarises the extant hypotheses suggesting that, like sperm whales (see § 2.3 below), baleen whales might produce vocalisations so powerful and low-frequency that they can travel across oceans (see Sirovic *et al.* 2007). However, while this is theoretically possible, Simon *et al.* (2010) propose a more moderate estimate, suggesting a range of about 90 km. In the *Balaenidae* family, which includes right whales and bowhead whales, vocalisations are also associated with contact between individuals. Various studies have shown that these calls are produced by approaching animals and cease once they reach each other. Würsig & Clark (1990), for example, described this behaviour in a bowhead whale mother and her calf. In right whales, which, unlike bowhead whales, do not produce songs, other types of calls with rising frequencies – called *upcalls* – have been identified (Clark 1983). Upcalls appear to function as responses to normal calls: when an isolated individual produces a call, other group members may respond with upcalls, which stop once they reunite. Additionally, female calls can attract males (Sayigh 2013). Parks *et al.* (2011) found that right whale vocalisations vary depending on behaviour, occurring more frequently during feeding and movement periods and less frequently during surface activities. The last two families, each containing only one species, have also demonstrated the ability to produce vocalisations but not songs. The gray whale (*Eschrichtiidae* family) produces repetitive sequences of sounds (*pulses* and *clicks*) interspersed with long periods of silence, which are believed to help maintain contact or call conspecifics. The lesser-studied pygmy right whale (*Neobalaenidae* family) also produces vocalisations, but these remain poorly understood and insufficiently documented (see Sayigh 2013).

More complex vocalisations, or *songs*, deserve closer attention, especially regarding their structure. While varying across species, a song – a sequence of sounds repeated many times over several hours – is believed to be composed of discrete *units*, which combine to form *phrases*, which in turn form *themes* (Payne & McVay 1971). These songs have been identified in all species of the *Balaenopteridae* family (various rorquals and humpback whales) and, within the *Balaenidae* family, in bowhead whales (but not in right whales). The complexity of songs varies by species. Humpback whales[105] (see below) exhibit the most

[105] Humpback whales (Megaptera) also belong to a different genus than rorquals (Balaenoptera), of which they are the only representatives.

intricate and well-studied structures, whereas all rorquals produce simpler sequences, with blue whale songs appearing slightly more sophisticated (Sayigh 2013). However, the minke whale produces irregular, poorly understood sequences referred to as *Star Wars* sounds due to their 'electronic' quality (Sayigh 2013). Bowhead whale songs, on the other hand, have an intermediate complexity between those of rorquals and humpback whales, consisting of one or two themes that can be repeated for over ten hours (Würsig & Clark 1990). Songs are thought to serve a sexual function, as they are produced exclusively by solitary, sexually mature males, primarily during the breeding season and in breeding areas. This is considered valid for all species (Sayigh 2013). Beyond attracting females, songs might serve to mark territory and intimidate rivals (Sayigh 2013). Additionally, it has been suggested that when males collectively sing, their songs may help synchronise female ovulation (Mobley & Herman 1985). However, various recordings show that songs can also be produced in contexts not strictly related to reproduction, such as in feeding areas, during migrations, or by males accompanied by females. While their role is still believed to be sexual, it is unclear whether they are used exclusively for interactions between males and females or also among individuals of the same sex (Sayigh 2013). In this regard, they have often been compared to the vocalisations of songbirds (see Fitch 2010). It is important to note that, although it is widely accepted that only males produce songs, this has only been confirmed for a small number of humpbacks in a breeding area. Most natural recordings do not allow researchers to determine the sex of the singing individual, though it is assumed to be male (see Berta *et al.* 2006).

Like orca "dialects" (see § 2.2 below), songs are shared by a specific group and differ from those of geographically distant populations. The main difference with respect to orcas is that individuals in a given breeding area always reproduce the same song, without significant modifications from one individual to another.[106] For example, the songs of humpback whales in the Atlantic Ocean significantly differ from those in the Pacific Ocean (Berta *et al.* 2006: 297). However, there is an important level of variation: the structure of each humpback whale's song gradually changes over months and years, showing modifications in the frequency and duration of the discrete units that compose it. Additionally, certain phrases and themes may be altered or removed, while new ones can be added (Payne *et al.* 1983). The sequence can also change: the same units, phrases, or themes can be reversed or combined in a new order to create new compositions (Berta *et al.* 2006:

[106] However, some studies argue that individual variation has been greatly underestimated and that this "personalisation" from one specimen to another play an important yet-to-be-defined role (see Cholewiak *et al.*, 2012).

297). This variation generally occurs in the winter months, when humpback whales migrate from food-rich polar waters to temperate breeding areas.[107] As a result, the song of a group gradually changes during that period, and by the end of the season, when migration resumes, it is substantially different from how it was at the beginning. Nevertheless, all individuals in acoustic contact reproduce a very similar song (though not in perfect synchrony), and they appear to adjust to each other's modifications. As a result, while collective variation is tangible, individual variation is minimal or insignificant (Berta *et al.* 2006: 297). For this reason, it is believed that songs are actively learned through vocal learning and imitation, culturally rather than genetically, in a refined process of constant adaptation (Sayigh 2013). Until a few years ago, it was thought that humpback whale songs changed constantly but gradually, so that an entire song would transform completely (replacing all its units, phrases, and themes in the process) in about fifteen years. This idea was challenged by observations off the coast of Australia, where a much faster rate of change was noted, as also observed by Fitch (2010). For some unclear reason, certain songs introduced by western Australian populations gained enormous popularity among eastern Australian populations, which adopted and spread them until they completely replaced their original song in just two years (Noad *et al.* 2000). It is now believed that the degree of variation is greater than previously thought, both in the song of an individual (which is less stable than expected) and among the songs of different individuals within the same group, which are now considered less homogeneous (Cholewiak *et al.* 2012). In other species, the rate of variation may be greater, as in the case of the bowhead whale, where according to Tervo *et al.* (2012), songs are unrecognizable from one season to the next; or lower, as suggested by McDonald *et al.* (2012) for the blue whale, whose song appears to remain stable for decades. The authors argue that blue whale songs are not influenced by vocal learning.[108] A more recent study (Fournet *et al.* 2018) claims to have identified an almost unexplored characteristic in humpback whales: according to the authors, while the songs produced in breeding areas vary according to the mechanisms described above, the songs produced in feeding areas remain essentially unchanged for decades (the study identified a stability period of 36 years). Naturally, further research would be needed, but if confirmed, this would imply that songs expressed in contexts other than breeding serve a different function that has yet to be explored.

[107] The reason humpback whales are the most studied mysticetes – besides the fact that they communicate at frequencies audible to humans and recordable with instruments available as early as the 1960s – is that one of their breeding areas is located near Hawaii, making it easily accessible to researchers.
[108] However, it is natural to wonder how it is possible that vocal learning does not play a significant role in that species while being crucial in almost all the others.

Another aspect to consider is intentionality: to what extent are these vocalisations produced deliberately? What is the actual role of vocal learning? While the information presented suggests that songs have at least a sexual function, it has been proposed that mysticete sounds may be a secondary effect of their movement, especially of the passage of air through their respiratory channels (Mercado & Handel 2012). According to this hypothesis, changes in notes could result from internal physical factors due to bidirectional phonation (two different notes would be produced depending on the direction of the airflow), while changes in themes could be caused by external factors, such as water pressure, determined by the depth at which the animal is located. Cazau *et al.* (2013) respond that in animals so well adapted to aquatic life, it would be more logical to expect an internal pressure regulation system that does not compromise respiration. They do not dispute this primary function but hypothesise that, starting from the constraints of the physical structure of the vocal apparatus, mysticetes have ended up using the sounds produced by this apparatus for communication, in a phenomenon that can be defined as *exaptation*. Cazau *et al.* (2013) therefore support vocal and cultural learning of vocalisations, emphasising that humpback whale calves initially produce imprecise sounds accompanied by bursts of bubbles from the blowhole – likely involuntary – and gradually refine their repertoire over time. According to Cazau *et al.* (2013), young whales first emit sounds constrained by respiration (in the scenario Mercado & Handel 2012 attribute to adults as well), and later learn to separate them, using the musculature of the laryngeal area with greater precision: "all these features manifest a vocal learning in progress, during which the calf may acquire the ability of mastering the secondary function of vocal production from the primary one of respiration" (Cazau *et al.* 2013: 2272).

2.1.2 The songs of humpback whales

In this subsection, I describe the structure of humpback whale songs[109] in greater detail, both because they are the most documented and because they are the most complex. Since their structure has often been described as hierarchical and even syntactic (see Suzuki *et al.* 2006; Allen *et al.* 2019), this discussion is particularly relevant to my thesis: do their combinatorial mechanisms (if they exist) and the resulting order indicate a form of compositionality? Do their discrete units carry meaning, and can they therefore be considered a lexicon? Of course, before asserting the presence of syntax, it should be verified how the authors of these studies – who are generally not linguists – use the term 'syntax', as

[109] To listen to humpback whale songs, see https://www.youtube.com/watch?v=sjkxUA041nM

they may be applying it too broadly. However, I also believe it is worth considering the reverse perspective: could the definition established by linguists be too restrictive and overly human-centric, thereby excluding from the analysis any phenomenon that does not perfectly conform to human combinatory rules and structural patterns? Additionally, it should be recognised that the tests used with humans to identify linguistic properties are not easily applicable to animals whose characteristics differ so much from our own. As a result, there is currently no effective method for determining whether these properties – or the underlying principles – are present in other species. However, while caution prevails among linguists, and the belief that no animal possesses a mechanism comparable to language is deeply ingrained, it is also true that non-linguists have often leaned toward the idea of animal language without providing reliable evidence. With this clarification, I can now proceed with the description provided by Cholewiak *et al.* (2012), who summarise with great precision four characteristics universally observed in humpback whale songs:

> (1) populations or groups of males that are in acoustic contact at some point in time and space sing "similar" songs, which is to say that their songs are comprised of visually and aurally recognizable similarities in patterns that are characteristic of that time/ region; (2) the overall hierarchical structure is observed globally, thus a heritable species-level characteristic, although the details of song patterns differ between populations of males that are acoustically isolated during all seasons; (3) song patterns change over time as a result of individual males modifying the spectral and temporal features of song units, as well as their order and repetition; and (4) males in acoustic contact incorporate similar changes into their own songs, maintaining continuity within populations despite progressive temporal changes (Cholewiak *et al.* 2012: 314).

To describe these songs, as noted by Cholewiak *et al.* (2012), a model is used, and like any model, it is inevitably an approximation. In the first attempt, developed by Payne & McVay (1971), a song is divided into themes, phrases, units, and subunits. In humpback whales, the same song is repeated multiple times with minor variations, forming what they define as a session. Sessions can last several hours, with the longest recorded duration being 22 hours. A song, on the other hand, can last anywhere from 6 to 35 minutes, according to observations (Winn & Winn 1978). To facilitate visualization, I present the Payne & McVay (1971) model.

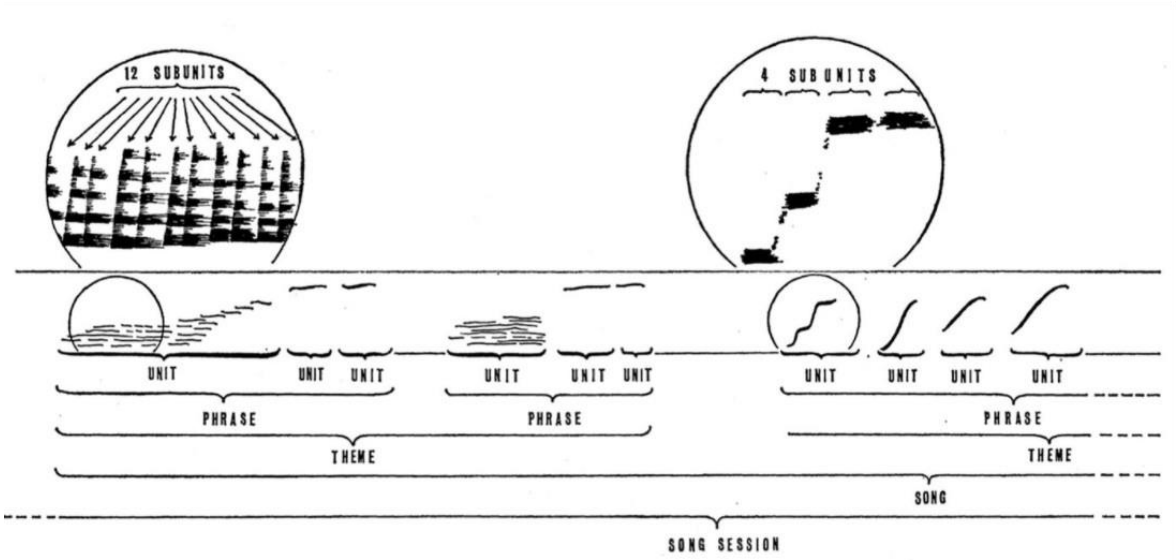

Fig. 22. Illustration by Payne & McVay (1971) showing their breakdown of whale songs: a session is composed of repetitions of the same song, and the song consists of themes, phrases, units, and subunits.

As of today, the subunits (which form each discrete sound unit) are analysed from the perspective of automatic classification. Researchers are investigating whether they could be useful elements for computational models describing whale songs (see below). A unit, on the other hand, is defined as "the shortest sound that seems continuous when evaluated at real speed", analogous to the "note" in the literature on songbirds (Cholewiak *et al.* 2012: 316). According to Cholewiak *et al.* (2012), these two definitions are effective and should be maintained, as the segmentation of these parts appears unequivocal. Things become more complex with phrases, which are further divided into sub-phrases (Payne *et al.* 1983). A sub-phrase is "a sequence of one or more units that is sometimes repeated in a series" (Cholewiak *et al.* 2012: 316). A group of sub-phrases forms a phrase. Similar phrases are generally repeated multiple times in varying amounts before different combinations are introduced. However, even consecutive phrases exhibit differences: they may have more or fewer units, altered frequency, and other variations that do not compromise their structure, allowing the sequence to remain identifiable. According to Cholewiak *et al.* (2012), the core of the variation that generates song diversity is precisely at the phrase level. Payne & McVay (1971), as cited in Cholewiak *et al.* (2012: 317), describe phrases as "inexact replicas" of one another. Moreover, Cholewiak *et al.*, diverging from Payne & McVay, propose a similarity between bird song and this specific level of humpback whale song. Phrases, which Cholewiak *et al.* (2012) consider the most important element to analyse, present a challenge: they are not easily segmentable because the time intervals between phrases are the same as those between individual units (unlike in the songs of most songbirds). Since clear separations cannot be drawn, every division depends on the interpretation of the transcriber

and is therefore arbitrary. This, of course, represents a significant limitation: while units can be clearly identified, phrases change based on human classification parameters, making it impossible to confirm whether this structural level truly exists. However, Cholewiak *et al.* (2012: 319) argue that "despite challenges in labeling and delineating phrases, multiple authors have recognised that phrase duration is one of the most stable features of humpback whale song". This supports the view of the phrase "as an element that is important to humpback whales, and delineation of phrases is not an arbitrary construct of researchers attempting to force organization upon the display".

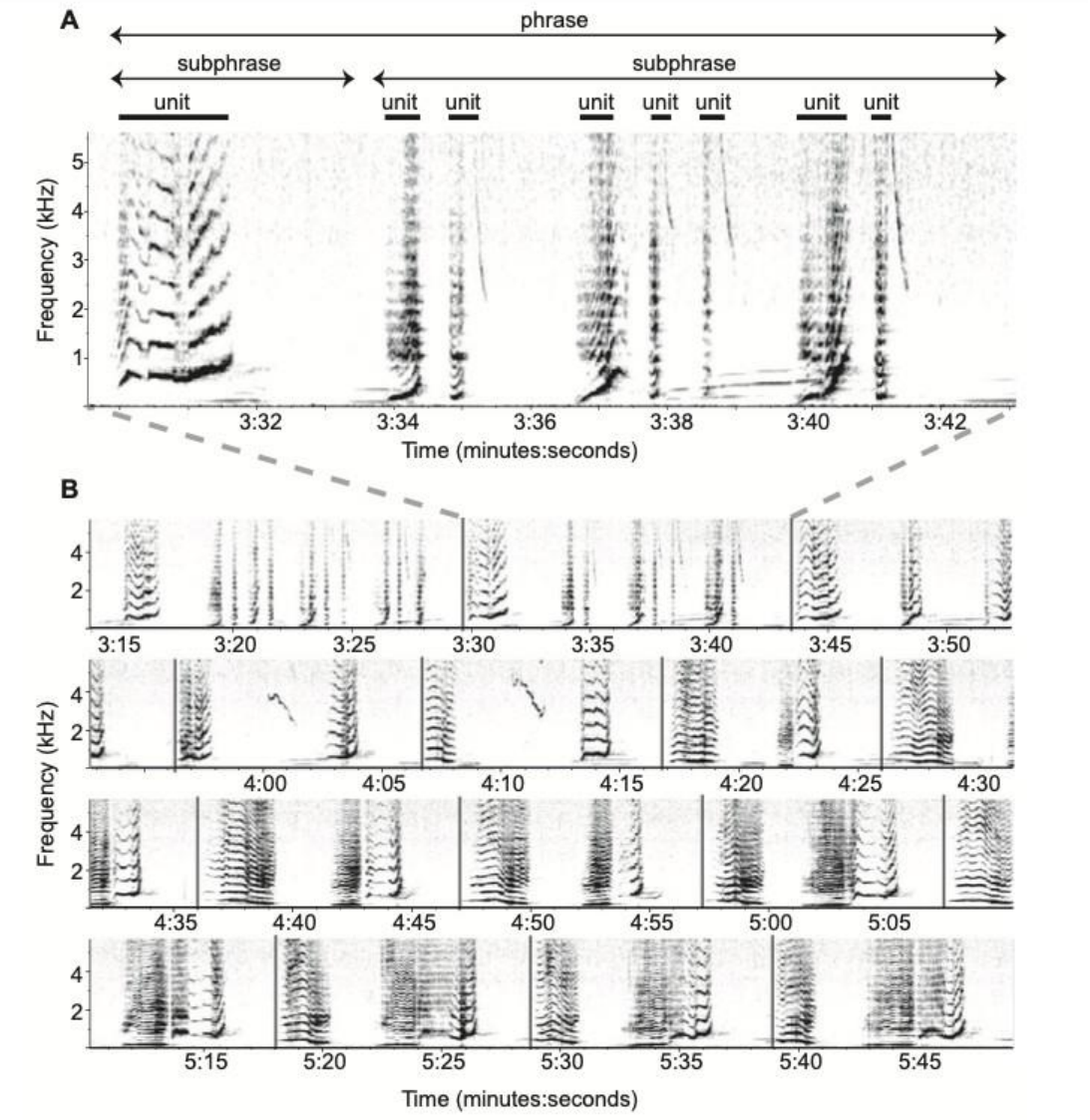

Fig. 23. Spectrographic representation of a sequence from a humpback whale song, recorded at Isla Socorro, Mexico, on March 27, 2006. The zoomed-in section (A) shows a phrase divided into sub-phrases and units. Note the difference in composition between the two sub-phrases. In (B), a broader sequence is displayed: phrases are separated by vertical lines (Cholewiak *et al.* 2012: 318).

To address these classification issues, Cholewiak *et al.* (2012) propose parameters that I will not discuss in detail, but which serve a dual purpose: on one hand, to standardise the procedures of different researchers so that the results are more easily comparable; on the other, to develop solutions that allow for the "absorption" of variations and reduce the occurrence of incomplete phrases, facilitating segmentation. For example, transitional phrases (Payne & Payne 1985) are phrases specifically created to "group" those elements that vary and to preserve the structure. It is more than reasonable to ask, as Mercado & Perazio (2021) do, what correspondence such an expedient might have with nature.

Continuing with the structure of the song, a sequence of similar phrases constitutes a theme. Since, as we have said, phrases gradually vary until they change significantly, the first phrase that is considerably different from the previous sequence marks the beginning of a new theme (Cholewiak *et al.* 2012). The length of themes can also vary, both within the same individual and in comparison to other specimens, as humpback whales tend to repeat an undetermined number of phrases. Frumhoff (1983) classified recurring themes, present in over 90% of song recordings in an entire season, as "fundamental". Payne & Payne (1985) identified three types of themes based on their organization: static themes, in which sequences contain almost identical phrases; shifting themes, in which phrases evolve gradually until they change almost completely; and unpatterned themes, in which the high unpredictable diversity of units prevents division into phrases, leading them to be considered as a single very long phrase. Songs, finally, in line with the original conception of Payne & McVay (1971), can be considered as sequences of themes, which are repeated an indefinite number of times. Frumhoff (1983), expanding the definition, describes the song as "a series of at least three themes, organized in a predictable sequence, repeated in the same order two or more times" (Cholewiak *et al.* 2012: 322). Since humpback whales sing in succession, without pauses between one cycle and another, the segmentation of songs is arbitrary, as in the case of phrases and themes: in fact, there are no ways to determine which is the first theme and therefore the beginning of the song.

Despite the efforts of researchers, what emerges is that the tools used to describe humpback whale songs are not sufficient to determine their structure with certainty. The cyclical form, which does not seem to have a defined beginning and end, presents inevitable difficulties for this type of classification. Moreover, while it was previously believed that songs were regular and substantially homogeneous both among specimens and in the different repetitions of the same individual, today the greater detected variability also challenges the "fixed" aspect of the structure. The main issues, therefore, seem to lie in the method of classification and subdivision of structures higher than units. A possible solution

comes from computational procedures: Suzuki *et al.* (2006) attempted to resolve the problem of the arbitrariness of human-made segmentations by developing a program to classify song elements, assigning each sound an abstract symbol. Then, they applied information theory to verify how much information could theoretically be produced with that system and what type of structure would be necessary to support it. Suzuki *et al.* thus focused on employing "an automated classifier to group the units and applying objective information theory techniques to study the song structure" (Suzuki *et al.* 2006: 1849). Information theory is an approach developed in the field of computer science, which allows for the calculation of the structural properties of a given code and its informational potential, defined as entropy (see Suzuki *et al.* 2006). The classifier is defined as "a method of extracting individual units from a continuous sound recording and converting them into discrete symbols that represent the particular types of sound, preserving the sequence order", whose tasks are to absorb "minor variations between sounds by lumping them together as a single symbol", and to assign a new symbol "when it encounters a sound element that is deemed sufficiently different from all previously encountered units" (Suzuki *et al.* 2006: 1850). One might observe that, to replace an arbitrary division made by humans, humans have created an equally arbitrary automatic rule. The approach of Suzuki *et al.* (2006) seeks "to determine whether the song contains mathematical properties consistent with hierarchical structure": in fact, vocal units are transformed into abstract symbols, and "acoustic features, duration, and other details of units and gaps between units" are not considered (Suzuki *et al.* 2006: 1850). They also clarify their use of the terms syntax and grammar: "in this paper, the structural constraints in the symbol sequence are referred to as a syntax or grammar, with no relation to semantics that may or may not exist" (Suzuki *et al.* 2006: 1850). Through their analysis, they seek to answer three questions:

> First, what is the quantitative measure of the information conveyed by the songs? Second, can any of the common stochastic models used in animal communication studies reasonably approximate the statistical features of the songs? Third, does a hierarchical model provide the best match to the structure of the songs? (Suzuki *et al.* 2006: 1850).

For highly technical details on entropy calculation, I refer to the original text, but I can summarise that, based on their analysis, they argue that 1) the amount of conveyed information is low; 2) traditional models used for animals are insufficient to describe the songs; 3) in their view, a hierarchical model is the most appropriate. Suzuki *et al.* are

convinced that humpback whale songs have a hierarchical structure, exhibiting properties generally considered unique to humans (see Berwick & Chomsky 2015; Bickerton 2014):

> The correlation data demonstrate that the songs possess strong long-distance dependencies of the sort discussed in Hauser *et al.* (2002) as a hallmark of phrase structure grammar. In closing, there is substantial quantitative evidence consistent with the sequence of units in the humpback songs being organized in a hierarchical structure, but equally strong evidence that this sequence is carrying relatively little information (Suzuki *et al.* 2006: 1864).

Other authors have been sceptical in this regard: Hurford (2011), in response to Suzuki *et al.* (2006), argues that whale songs do not require a hierarchical structure and can be described using a finite-state grammar.[110] To demonstrate this, he constructs a finite-state model of a song identified by Payne & McVay (1971).

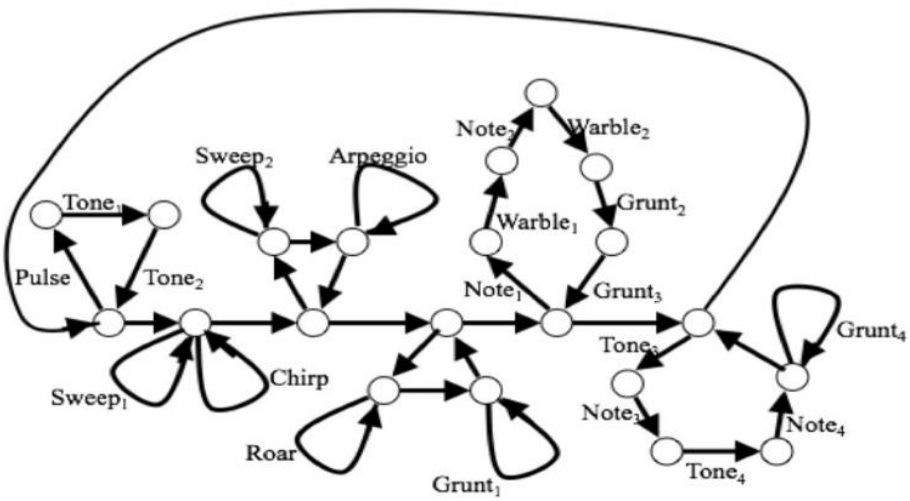

Fig. 24. Finite-state model by Hurford (2011: 67) to describe the song of a humpback whale as reported by Payne & McVay (1971).

Eduardo Mercado III (see Mercado & Handel 2012; Mercado & Perazio 2021) challenges both the hierarchical structure and the information theory approach, which he considers misleading: the abstract structures that are built do not take into account the reality of physical aspects, obscuring acoustic characteristics. According to Mercado & Perazio (2021: 41), new measurements have shown that over the years, humpback whales have modified the acoustic properties of their phrases, revealing "a potentially universal feature

[110] A finite-state grammar is a simplified model of generative grammar that creates sentences through linear rather than hierarchical relationships. It is therefore not suitable for describing human language but is sufficient for expressing other types of messages.

of songs not noted in past subjective analyses." These aspects are disregarded by information theory and by models that transform sounds into symbols without considering their actual characteristics: according to Mercado & Perazio (2021: 41), this approach can "lead to errant conclusions about how other species produce and use vocal sequences." Taking a clear stance, they state that:

> applying information theoretic analyses to humpback whale songs is guaranteed to distort both descriptions of song cycles and their variations over time, erasing complexity where it exists and introducing information-related complexity as a methodological artifact. (Mercado & Perazio 2021: 45).

Mercado is also unconvinced about the hierarchical structure, arguing that despite what is suggested or implied, there is actually no evidence to support it (Mercado & Handel 2012). An argument in favour of the existence of discrete and segmentable units is presented by Garland *et al.* (2017): by studying humpback whale songs during a transitional phase between one song and another, they observed that, in rare cases, individuals reproduced parts of both songs, replacing or "mixing" phrases in a phenomenon of hybridization. According to them, this would serve as evidence of imitation and cultural learning, and this event would also help better understand song segmentation by identifying as compact units the blocks taken from one song and transferred to another, in a kind of "constituency test". For Mercado & Perazio (2021), this actually means the opposite: they note that the presence of such modifications should lead to a much greater diversity than what is observed. New studies, in fact, suggest that "detailed acoustic similarities exist between humpback whale songs produced across acoustically isolated populations" (Mercado & Perazio 2021: 41). Furthermore, variations in songs have been observed to be more pronounced than previously thought (Cholewiak *et al.* 2012). According to Mercado & Perazio (2021), cultural transmission and social learning do not play a significant role in these modifications: if they did, the songs of different, separated populations should show almost no similarities. Consequently, they argue that there are physical and structural constraints that limit expressive possibilities, at least due to respiration and diving depth. However, these universal principles governing song composition have not been identified from a physical, cognitive, or syntactic perspective. Considering all this, what could be done to adequately describe the hierarchy of songs? For Mercado (Mercado & Handel 2012; Mercado & Perazio 2021), this is the wrong question: there would be no hierarchy to determine, but rather a "heterarchy" – a horizontal, non-hierarchical relationship of elements arranged in a sequence, dictated by physiological principles. Therefore, what should be studied are the

acoustic variations in songs in relation to humpback whale behaviours and their breathing cycles. As we previously mentioned (see § 2.1.1), Mercado believes that bidirectional phonation is responsible for differences in notes, while changes in pressure and depth account for differences between themes (Mercado & Handel 2012; Mercado & Perazio 2021).

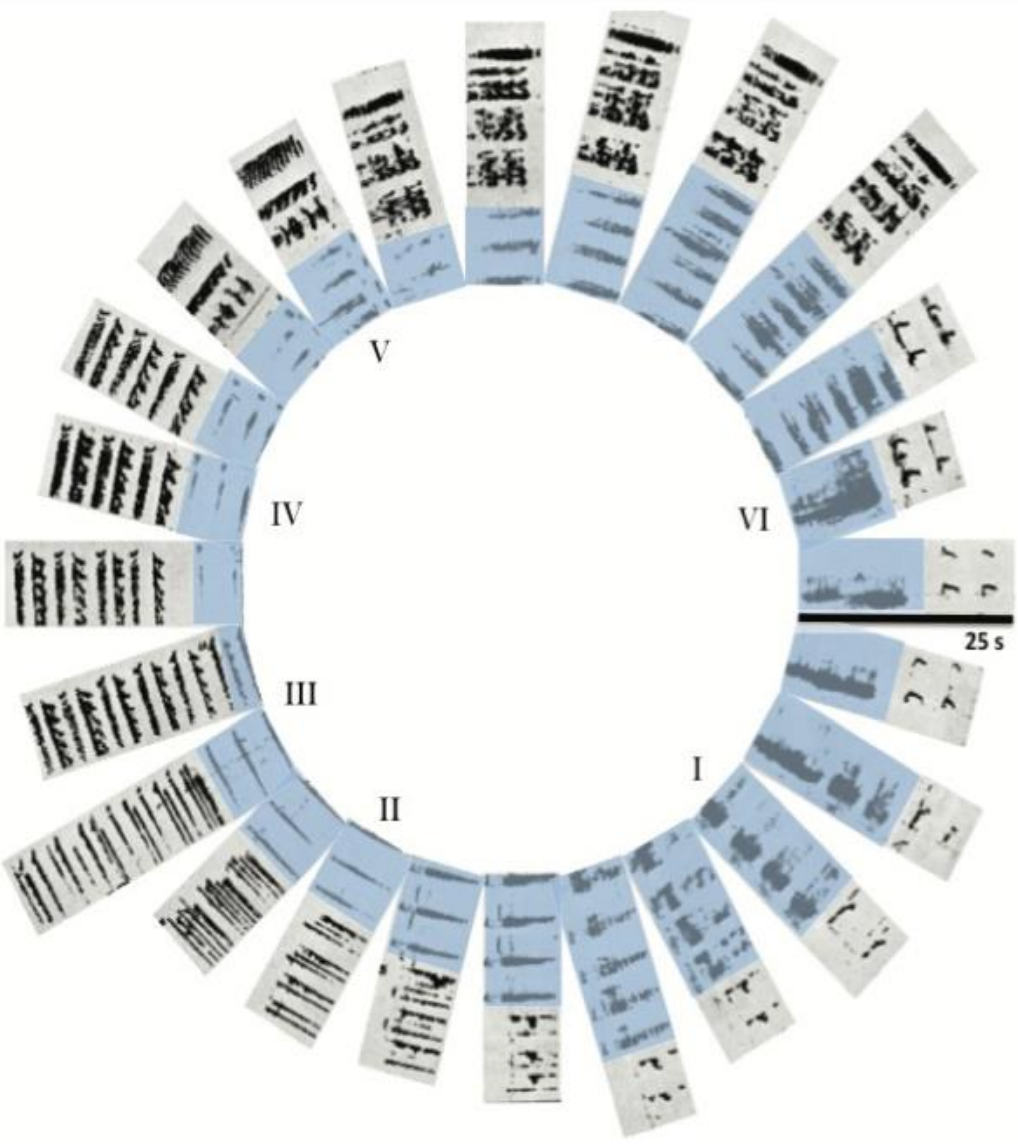

Fig. 25. Circular representation of 25 spectrograms from Winn *et al.* (1970), reinterpreted in this model by Mercado and Perazio (2021: 37).

Figure 25 illustrates an attempt to represent whale songs using a non-hierarchical, circular model aligned with the physiological rhythms of swimming and breathing. The spectrograms are arranged in a circle and are read one at a time, from the inside outward, in a clockwise direction. According to Mercado & Perazio (2021: 37), "this representation reveals the

temporal regularity of unit production as well as acoustic continuity and progressive morphing across consecutive phrases throughout the song cycle." At the points marked by Roman numerals – indicating the beginning of the themes identified by Winn *et al.* (1970), from which Mercado and Perazio derived the spectrograms – it is possible to observe the difference between adjacent phrases. Within a theme, however, the variation progresses gradually. Based on this 'physiological' reasoning, Mercado & Perazio (2021) also reject the idea that whale songs serve a sexual function. However, they are unable to explain why the songs vary in this manner, though they suggest it may be related to sound perception – perhaps the need to optimise vocalisation clarity in response to surrounding noise. They argue that "the complexity of humpback whale songs lies primarily in how listening whales are processing and interpreting what they hear, not in how aesthetically impressive song qualities are to human observers attempting to make sense of them" (Mercado & Perazio 2021: 46). Nonetheless, they acknowledge that vocal imitation and social interactions play a role: "social interactions almost certainly are contributing to how variations in songs progress", and "vocal imitation abilities, which humpback whales seem likely to possess, may contribute to the flexibility and speed with which singers can modify their songs" (Mercado & Perazio 2021: 46). Consequently, while they believe their findings raise doubts about the importance of these phenomena, they admit they cannot fully explain why whale songs undergo constant change (Mercado & Perazio 2021).

Despite Mercado's well-reasoned objections, other researchers continue to view information theory and the search for rules and structures through computational models as the most promising path forward. Approaches as machine learning and deep learning[111] could have significant applications in this field. From this perspective, while Mercado's insights certainly deserve further exploration, computational approaches may prove equally useful. One interesting attempt in this direction was made by Malcolm Kennedy (2018). Although his proposals have not received significant validation, I consider it worthwhile to present his alternative and relatively unexplored approach. Kennedy analyses the internal structure of humpback whale phrases, attempting to explain elements that appear extraneous when segmenting themes and are included in transitional phrases (see Cholewiak *et al.* 2012). His idea is that certain repetition processes within phrases cannot be explained without hierarchical tree structures featuring multiple levels of nesting – essentially, the typical architecture of human syntax. To test this hypothesis, Kennedy employs computational tools to create probabilistic context-free grammars and uses them to describe

[111] See Ntalampiras (2017) for a discussion of the potential application of machine learning to humpback whale vocalisations.

a humpback whale song recorded by Paul Knapp in 1992 in the British Virgin Islands. Without delving into the specifics of his approach, I can summarise that his computational models successfully described individual phrases using a binary hierarchical tree structure. Below, I present some examples of these structures, which offer useful ways to visualise the supposed relationships between the 'constituents' of whale songs.

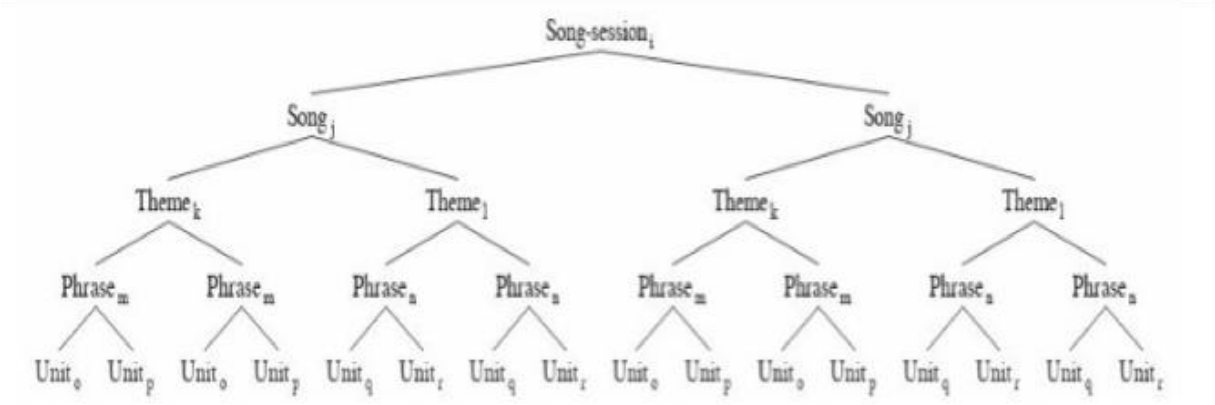

Fig. 26. Alternative graphical representation of the classic structure of whale songs, comparable to the scheme of Payne & McVay (1971). The hierarchy is depicted using a binary tree architecture (Kennedy 2018).

Below, I present a tree structure automatically generated by Kennedy's program for one of the phrases he analysed. The letters A, B, C, D, and E represent discrete units, with each corresponding to a specific sound produced by the humpback whale in this particular phrase.

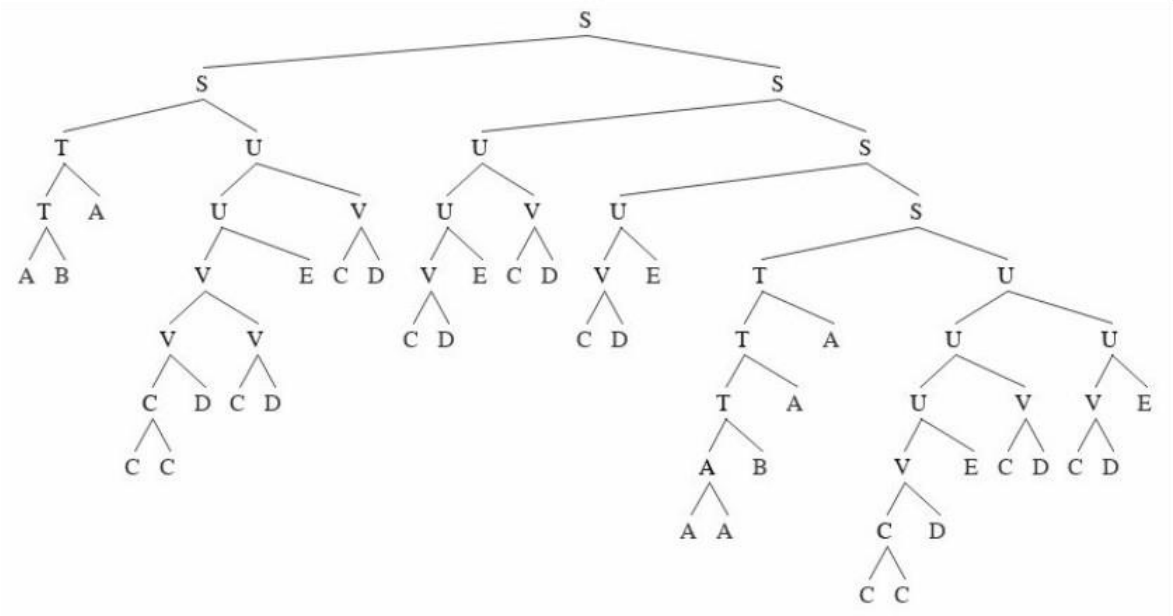

Fig. 27. A binary tree automatically generated by Kennedy's program (2018), describing the first phrase of the song while maintaining a binary structure.

Based on these correspondences, Kennedy hypothesises that humpback whales may possess the property of recursive merge. Consequently, their songs would exhibit a syntactic structure in a strictly linguistic sense, although, in his view, they would not convey any semantic content. Under this perspective, humpback whale songs would share the same properties as human music: a hierarchical, syntactic structure that can only be described through tree structures but devoid of meaning. However, even though several studies discussed in this thesis suggest that other animals may possess some fundamental operations underlying language – whether in linguistic or non-linguistic domains (see Part II) – the data presented by Kennedy do not seem sufficient to confirm this with certainty. Moreover, as Kennedy himself acknowledges, the binary structure collapses when representing the various phrases that compose a theme. To describe a theme, the structure requires a number of branches corresponding to the number of phrases; in Kennedy's example, five. Nonetheless, Kennedy attributes this limitation to the reduced memory capacity of humpback whales. Due to the structure of their syntax, which is based on repetition, and the constraints imposed by externalization in a linear form, he argues that whales are unable to process and embed higher-order sequences.

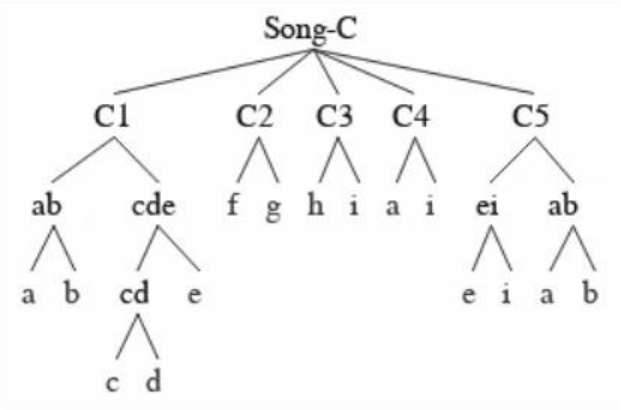

Fig. 28. Tree representation showing the incompatibility of the upper level with a binary structure (Kennedy 2018).

This approach, like all formal reconstructions, does not take into account the acoustic aspects that Mercado & Perazio (2021) have so strongly emphasised. Based on the points discussed, I suggest considering it only as an example and a possible avenue for future research.

A different approach has been pursued by Pace *et al.* (2010), which introduced a novel method to classify humpback whale calls by segmenting them into subunits rather than analysing entire units. Songs recorded in Madagascar were analysed with an automatic method to identify vocalisations, followed by manual segmentation of subunits based on

spectral characteristics. Subunits were then classified using a clustering algorithm. Results showed that subunit-based classification was more accurate than unit-based methods in 83.5% of cases, as subunits exhibited lower acoustic variability and shorter duration, making them more suitable for automatic recognition. Additionally, certain subunits recurred across different song sequences, implying that humpback whales may construct their vocal repertoire using a finite set of acoustically distinct subunits. This challenges the conventional approach to whale song analysis and suggests that subunits could provide a more fundamental and stable framework for classification, potentially allowing cross-population and longitudinal comparisons. Future research is required to determine whether these subunits remain consistent over time and across different geographical regions.

Despite the studies reviewed, for now it is neither clear nor demonstrable whether humpback whale songs contain semantic content, nor is it easy to determine the extent to which their structures are comparable to human language. Furthermore, being songs that seem linked to a sexual function, their complexity is probably related only to structure, and not to meaning, as in the case of birdsong (Fitch 2010). As of today, there is no evidence confirming or refuting the presence of a hierarchical structure in humpback whale songs, nor of the presence of the merge operation and recursion in their brain structure. However, the features described support the idea that cetaceans are promising animals for this kind of research. The next section focuses on the vocalisations and the properties related to the language of odontocetes.

2.2 Killer whales

As we have seen in chapter 1, odontocetes are easier to study than mysticetes. One of the most famous and studied animals is the killer whale, also called orca, which has complex repertoires of vocalisations and a peculiar society, different from that of other cetaceans. According to Connor *et al.* (1998), their social organisation is unprecedented, and currently it has no equivalent in the animal kingdom.[112] These animals live in matrilineal – and possibly matriarchal – stable and closed groups, which can be classified into pods, clans, clan associations, and 'ecotypes' (so that more pods form a clan, more clans form an association, and more associations form an ecotype, see Whitehead & Rendell 2015). Pods, which vary in size (typically not exceeding 20 individuals), always include a reproductive-age female, her offspring, other non-reproductive females, and at least one male (Berta *et al.* 2006: 295). Orcas tend to mate within their pod or, at most, within their clan, avoiding

[112] For an in-depth discussion of cetacean societies, see Whitehead and Rendell (2015).

‘outsiders’. This appears to contradict the biological principle that animals generally mate outside their group to ensure genetic diversity. Since different groups do not mix, in addition to biological kinship (with physical similarities, e.g., in size), the result is a great diversity in habits. Clans are distinctly different in “dialects” (see §§ 1.4.1 and 2.2.1), diet, and feeding strategies.

The most studied orca communities live in the North Pacific, and they are divided into residents (larger, more numerous, and territorial, feeding on salmon), transients (smaller groups, slightly larger individuals, feeding on pinnipeds), and offshores (deep-sea dwellers feeding mainly on fish, especially sharks), see Whitehead & Rendell 2015. In the Atlantic, at least two communities are believed to exist: one primarily feeding on other cetaceans and another with a more general diet. In Antarctic waters, there are at least five different ecotypes specialised in hunting fin whales, seals, penguins, and in two cases, fish (Whitehead & Rendell 2015). According to Whitehead & Rendell (2015), these are not ‘morphotypes’ but ‘ecotypes’ because, despite having the same body structure, they use it in different ways to occupy specific ecological niches. Unlike other species where differences arise due to geographical barriers separating populations and resources, orcas are found in all oceans and face no such physical barriers. Yet, each group remains isolated despite contact with others. In Antarctica, for example, different orca ecotypes can inhabit the same waters, yet each group hunts only its customary prey, ignoring others (Whitehead & Rendell 2015). This separation is also evident in their genetics, as different ecotypes have not interbred for many generations (although matings may occur between pods of the same clan). Lance Barrett-Lennard, cited by Whitehead & Rendell (2015), argues that orcas succeeded in becoming one of the most efficient predators in the animal kingdom thanks to their being ‘picky eaters’, ‘xenophobic’, and cultural. ‘Picky eaters’ refers to their precise and habitual food preferences. For example, Antarctic orcas that prey on seals prefer a specific, larger species and ignore others. Before attacking, they often surface to check the ice floe for the ‘correct’ prey (Whitehead & Rendell 2015). Even in captivity, these dietary habits persist, sometimes making adaptation to new food difficult, and, in documented cases, leading to death (Whitehead & Rendell 2015). Their ‘xenophobia’ is reflected in their conservative nature, affecting mating, diet, hunting techniques, “dialects”, and social interactions. Pods of different clans tend to ignore or avoid each other and, in some cases, even confront each other or reject more ‘nomadic’ groups. The third aspect, culture, ties into the other two. Whitehead & Rendell (2015) argue that this conservatism is not genetic but arises from how knowledge is transmitted and maintained within groups. Orcas live and hunt alongside the same individuals for life, sharing vocalisation systems, habits, and strategies reinforced by

strong bonds. Despite their high imitative ability, they favour established practices and are resistant to innovation (Whitehead & Rendell 2015). According to the researchers, the distinctiveness of ecotypes and clan-specific strategies originated from individual innovations that proved effective and were imitated until they became the new 'norm'. Once established, due to the species' conservative nature, these strategies saw little further innovation (Whitehead & Rendell 2015).

In these animals, the transmission of knowledge plays a fundamental role. One of the most representative examples involves certain hunting techniques. As I have described, orcas exhibit a great diversity of specializations, characteristic of each ecotype. For example, orca populations in the Antarctic Peninsula that hunt seals have developed a technique to dislodge their prey from floating ice sheets where they seek refuge. This behaviour was first described by Smith *et al.* (1981). Orcas have since been observed and filmed multiple times performing this activity: first, they break the ice sheets to reduce the seal's available space, then they generate waves that overturn the ice platform or sweep away the prey. These waves are created by body movements and tail slaps, sometimes executed by groups of up to three individuals simultaneously (Visser *et al.* 2007). Figure 29 illustrates the technique more clearly.

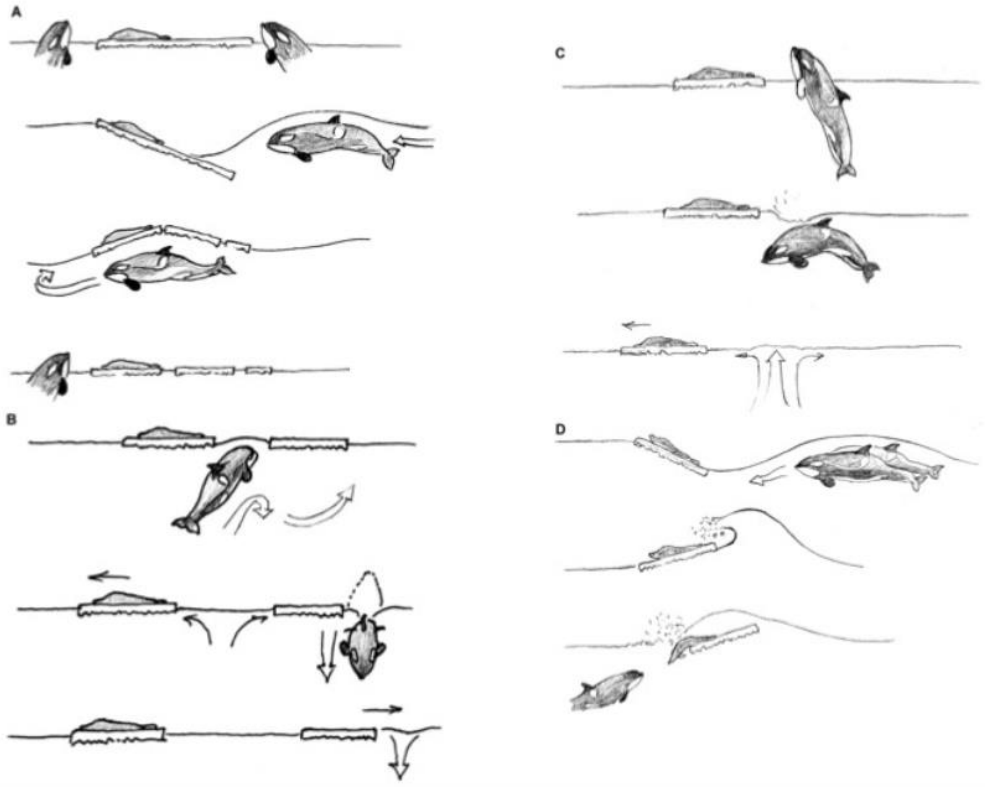

Fig. 29. Illustration of the orca hunting technique[113] (Visser *et al.* 2007: 230-231).

[113] For a video of this hunting strategy, see https://www.youtube.com/watch?v=VyfOp_keW0A.

This technique is practiced exclusively by orcas living around the Antarctic Peninsula. Other pods of the same ecotype in different Antarctic regions or in the Arctic that also hunt seals have not developed this method (Whitehead & Rendell, 2015). Another unusual strategy, known as *intentional stranding*, has been observed in certain orca pods both in Patagonia, Argentina (Lopez & Lopez 1985), and in the Crozet Archipelago in the southern Indian Ocean (Guinet 1991). This technique involves swimming toward the shore, beaching themselves to catch a sea lion calf, and then dragging themselves back into the water, jumping and rolling with their prey in their jaws[114] (Guinet 1991). This behaviour is highly risky, as individuals risk to remain stranded, and in some cases this has happened. Guinet (1991) observed that only adults practiced this technique, while their young stayed with other pod members. However, the juveniles were gradually involved in these activities, sometimes even in the absence of prey, as a form of play. Guinet (1991) noted that play might serve as a learning mechanism, helping young orcas master hunting strategies. Guinet & Bouvier (1995), studying the orca populations of Possession Island in the Crozet Archipelago, observed that young orcas practiced stranding by themselves only around the ages of 4-5, and even then, only without prey. It was only around age 6 that they began attempting to catch elephant seals, though they still needed their mothers' assistance to return to the water with their *catch*. Furthermore, they noted that "learning hunting techniques needs a high degree of skill and requires high parental investment to reduce the associated risk. Furthermore, social transfer, through apprenticeship, is probably one of the mechanisms that enables the high degree of adaptability observed in killer whales" (Guinet & Bouvier 1995: 27). A final example of the orcas' strategic intelligence was recorded at SeaWorld in San Diego: a captive orca, after receiving a fish, deliberately placed it at the pool's edge as a bait, where birds were gathered. When one of the birds attempted to take the fish, the orca swiftly grabbed and devoured it.[115] Despite being an anecdotal observation, this behaviour indicates the flexibility and adaptability of killer whales, which are capable to elaborate new solutions to face the challenges of their environment. The complexity of their intelligence and cognition is also indicated by the fact that orcas are among the few animals which passed the mirror self-recognition test (Delfour & Marten 2001).[116]

[114] For a recording of intentional stranding, see www.youtube.com/watch?v=LqIC8D5xyU4.
[115] For a video of the event, see https://www.youtube.com/watch?v=14wWxaMR2Mg.
[116] This type of test is described in detail in chapter 3 for bottlenose dolphins.

2.2.1 The "dialects" of the orcas

Communication in odontocetes has been studied more extensively than in mysticetes, primarily because some species are easier to analyse, due to their smaller size. Most data, in fact, comes from bottlenose dolphins (see § 3-4 below) and orcas (family Delphinidae).[117] Marino *et al.* (2007) summarise their capabilities as follows:

> Extensive fieldwork has shown that cetacean call types exhibit enormous variation (Ford 1991, Yurk *et al.* 2002), evolve over time (Deecke *et al.* 2000), and are used differently across social groups (Rendell & Whitehead 2003). In some cases, the variation is so pronounced that other species have learned to use it in judging predation risk (Deecke *et al.* 2002). (Marino *et al.* 2007: 0971).

As I have already mentioned, the communication system is flexible and combines vocal production with visual signals, such as changes in posture, tactile signals, for example fin rubbing, and other movement-related behaviours. According to Marino *et al.* (2007: 970), "the temporal sequencing of these latter nonvocal communicative events can be highly structured, demonstrating a complex and diverse nonvocal communication system." Herzing (2015) defines it as a multimodal system (as Kendon defines human language, see Part I, § 2) and describes the three categories of sounds produced by odontocetes, generated through the mechanisms described in section 1.4: frequency-modulated whistles, considered entirely related to communication; echolocation clicks, used for orientation and – in the case of lower frequencies – probably also in communication; and burst pulse sounds, sequences of strong clicks with variable repetition rates, which remain difficult to categorise and correlate to social functions. Herzing (2015) also points out that these categories might be less defined than commonly assumed, and rather than being neatly separated sounds in terms of characteristics and function, they may form a more flexible continuum. This applies to most odontocete species.

More specifically, orcas are characterised by so-called "dialects" (Ford 1991, see also § 1.4.1), where the repertoire of sounds emitted by individuals varies depending on the social group to which they belong.[118] As I have discussed above, odontocete species like orcas live in structured societies that appear to have a form of culture and knowledge transmission, which manifests both in the teaching and learning of behaviours and hunting and defence

[117] It should be noted that other families also exhibit vocalisations and imitation abilities, such as Monodontidae (belugas and narwhals). Others, like Ziphiidae and Phocoenidae, show lower tendencies for vocalisation and social behaviour, but they are understudied animals. For further details, see Sayigh (2013).
[118] A similar system of "dialects" is observed also in sperm whales, see § 2.3 below.

techniques, as well as in the acquisition and imitation of sounds that make up an individual's vocal repertoire. Orcas produce series of calls (usually consisting of a repertoire of about ten sounds that are repeated), which are shared among members of the same pod but differ from those of other groups. However, similarities exist, becoming vaguer as the geographical distance between populations increases.[119] Each group, therefore, uses a specific repertoire of discrete calls (i.e., well-defined and not easily confused), which can remain virtually unchanged for decades and show similarities to the systems used by geographically adjacent populations (Berta *et al.* 2006: 295). Comparative analyses of the various "dialects" used by different orca populations, combined with genetic analyses of certain individuals, have shown that the greater the genetic divergence, the more distinct their calls become. This suggests that groups using similar "dialects" are more closely related (Filatova *et al.* 2010).

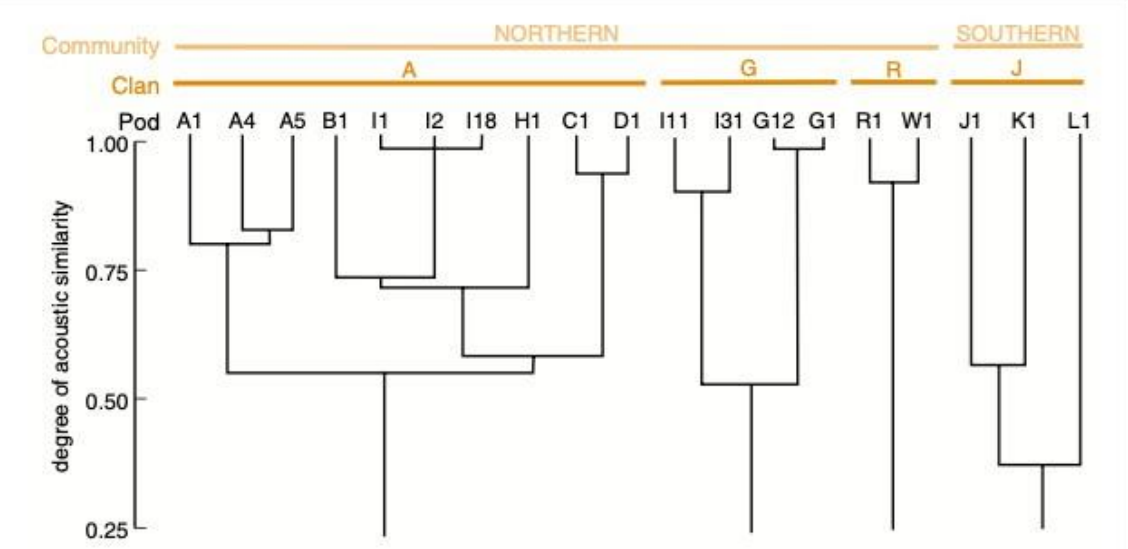

Fig. 30. Tree diagram of the linguistic and genealogical relationships among the orca groups studied by Ford (1991) (image modified by Berta *et al.* 2006: 296). The numerous pods are classified into four different clans, with corresponding vocalisations that are geographically distributed in a homogeneous manner (the pods of each clan are found in adjacent regions). It is unclear whether the clans themselves originated from a single ancestral community.

Ford's (1991) studies identified 16 different pods in the waters of British Columbia, forming four distinct clans based on similarities in their calls. All pods within a clan shared various calls, though some remained exclusive to specific pods, and no calls were shared between different clans. This hierarchical structure led Ford to propose that the traditions of each group developed culturally over generations, in a system where members of the same clan, united by linguistic similarities, descended from a common ancestral group. This group

[119] To listen to orca vocalisations, see: www.youtube.com/watch?v=29neQ5mLT5E.

eventually diversified into numerous pods with progressively differentiated calls. The repertoire of calls, transmitted matrilineally, would have evolved alongside the dispersal of descendant groups (Berta *et al.* 2006: 295). Reconstructing the "linguistic families" of orcas suggests that all pod "dialects" within a clan originate from a common and shared ancestral system, which gradually diversified over time, much like Schleicher's (1861) genealogical tree theory. According to this model, variations and innovations are primarily transmitted vertically to descendants. The concept of "dialect" provides a useful framework for understanding this phenomenon (Ford 1991), although, as I clarified in section 1.4.1, it cannot be compared with human dialects, because of their fundamental differences (see § 1.4.1). Furthermore, it conceptually parallels the work of Cavalli-Sforza (1997) and Ruhlen (1994), who compared the genealogical tree of humans with that of languages – originally proposed by Schleicher – to suggest the past existence of a protolanguage.[120] Similarly, one could symbolically refer to a search for an "original language" of orcas, tracing the progressive diversification of their vocal and communicative repertoire back to a shared origin. Since Ford's initial studies, research on this phenomenon has expanded, analysing populations of various species in different regions and accumulating more specific data on "dialect" variation. However, the precise content of the calls remains uncertain. It is believed that they serve a social function related to long-distance contact maintenance, communication and coordination among conspecifics, and recognition of group members after dispersion (see Deecke *et al.* 2000). "Dialects" allow recognition of an individual's membership within a particular community, and imitation and vocal learning play a fundamental role in their understanding and formation. It is believed that vocal repertoire acquisition primarily occurs through imitating the mother, resembling the way humans acquire language. However, orcas also exhibit "horizontal" learning systems, where adults influence each other's calls (Filatova *et al.* 2010). In addition to the divergences between different groups, variations in vocal traditions have been observed within individual pods, including innovations, modifications, additions, or removals of calls that accumulate over time, contributing to "dialect" diversification (Berta *et al.* 2006: 295). From this perspective, it can be observed that despite not being the primary transmission method, the presence of horizontal influences – rather than exclusively vertical ones – in "dialect" variation bears some resemblance to human language. This is somewhat reminiscent of Schuchardt's wave theory (cf. Schmidt 1872), although innovations primarily spread within the same group or

[120] In this case, the term protolanguage is intended as an ancient language spoken by all humans (with recursive merge and all the properties of modern language), from which all languages spoken today descended, and not a primitive phase *before* complex language, with more rudimental properties, as the one described in Part II, § 1.

among groups within the same clan (and thus genetically related) rather than necessarily by geographical proximity. The fact that even adult orcas can learn new signals and modify their "dialect" allows for greater variation, including the adoption of new calls and the potential influence of "borrowed" calls from external groups. However, despite documented instances of this behaviour, Filatova *et al.* (2010) argue that it remains uncertain whether such occurrences are common (as in humans) or exceptional. It has been observed that outsiders can imitate a pod's "dialect", and in some cases, they learn it and are integrated into the community (Filatova *et al.* 2010). Additionally, some odontocetes have been observed mimicking the call systems of another species. In rare cases, these animals were even accepted into the group, demonstrating exceptional interspecies communication and cooperation.[121]

What could be the reason for the existence of "dialects" and their variation? Logically, if vocalisations were determined solely by instinct or by anatomical characteristics of the vocal apparatus, they would not exhibit these kinds of changes and would likely be much more similar across individuals. It has been suggested that orca repertoires – given their strong imitative abilities[122] – gradually diverge between pods due to random errors in imitation and innovation, but other dynamics may also be at play (Filatova *et al.* 2010). Fitch (2000), who does not dedicate much attention to cetaceans and often discusses them alongside songbirds,[123] describes an intriguing hypothesis that we might call the *password hypothesis*: if the primary function of vocalisations is social, then vocal learning and imitation serve to establish and maintain bonds within the group. In such a scenario, it is plausible that these abilities would develop because individuals more skilled at reproducing vocalisations would have greater evolutionary success. This would also allow external individuals to learn a community's "dialect" and be accepted into it, fostering a form of social dynamism – something documented in orcas. However, a group (as well as a specific "dialect") does not only serve to unite but also to distinguish itself and defend against rivals. Just as a well-intentioned outsider might learn a group's "dialect" through imitation, the same could hold true for adversaries. This would naturally drive all groups to

[121] I was unable to find scientific literature on this topic, but there are documented cases. In one instance, a bottlenose dolphin was observed adopting a pilot whale calf, which learned to imitate dolphin vocalisations and integrated into the group, forming a strong bond with its adoptive mother (see: www.smithsonianmag.com/smart-news/researchers-document-first-known-instance-dolphin-mom-adopting-whale-calf-180972769/).

[122] Abramson *et al.* (2018) demonstrated that orcas are capable of imitating human sounds, leading them to classify these animals as "open-ended vocal learners", meaning their vocal learning abilities are "open" and not restricted to a fixed, limited repertoire.

[123] Similarly, in *The Origin of Language* (Fitch 2010), Fitch does not pay much attention to studies on cetacean cognition and their "linguistic" abilities, nor does he attempt to refute them.

gradually modify their repertoire. A code functions both to include and to exclude; therefore, variation could be explained in this direction, as part of an "arms race" where groups simultaneously attempt to decipher the "dialects" of others while protecting their own (Fitch 2000: 264). According to Fitch (2000), imitation, linguistic variation, and complex communication may have arisen from these social pressures, highlighting intriguing parallels between humans and cetaceans, as also suggested by Louis Herman (2012, see below). I now present Fitch's argument in full:

> These cetaceans, like chimpanzees and contemporary humans, live in social groups characterized by within-group cooperation and competition between groups. Such social systems put a premium on reliable indicators of group membership, vocal or otherwise. They also encourage an ability in newcomers who have emigrated into a group to learn the shared vocal indicators or group membership or 'passwords', and so could select for vocal learning. Once basic learning capabilities have evolved, however, they would also allow interlopers to master the shibboleth of an invaded territory quickly. Such code-breaking would in turn select for increased discrimination on the part of group members or increasingly complex and hard-to-master 'passwords' (or both). If the value of being a group member is high, and vocal indicators of group membership are important, the stage would be set for a runaway selection leading to increased vocal learning and finer perceptual tuning. (Fitch 2000: 264)

The password hypothesis is also "compatible with the observation that modern humans use accent to differentiate readily between individuals raised in their natal environment and newcomers" (Fitch 2000: 265). Furthermore, this hypothesis shows a parallel with the sociolinguistic studies regarding the difference between *we-codes* and *they-codes* (Fishman 1972). A *we-code* is defined as the language (or language variety) associated to the "in-group", i.e. a specific speech community determined by sociolinguistic factors (e.g. age or socio-economial group); a they-code is used in the "out-group", i.e. in the communication with individuals that do not belong to that community. For example, in a Latin-American community in the USA, Spanish can be used as a we-code related to the in-group, while English represents the they-code related to the out-group. From this perspective, orcas' "dialects" might be considered we-codes.[124] Finally, the password hypothesis "makes a number of testable predictions (e.g. individuals should be more likely to aid a stranger that

[124] However, orcas seem to lack a they-code (comparable to English in the example above), although some discrete units are observed across different populations.

shares their dialect than other strangers)" (Fitch 2000: 265), and it aligns with data from other species, such as cetaceans.

Despite the complexity displayed in these vocalisations, the meaning of the calls is still poorly understood, so it is not easy to investigate the presence of compositionality, and of other properties related to language. However, on the structural level, the complexity of orcas' vocalisations has been analysed. Shapiro *et al.* (2011) compared call-based and subunit-based methods for categorising Norwegian killer whale vocalisations, testing whether calls are composed of smaller acoustic units. The study tested three classification methods: (1) whole-call classification, where each vocalisation was treated as a distinct, indivisible unit; (2) unique subunits per call type, segmenting calls into smaller acoustic components specific to each type; and (3) shared subunits across calls, grouping recurring sound patterns into a common set of elements used to form multiple call types. Data were collected using digital archival tags and processed using speech recognition techniques, including pitch-tracking algorithms and Gaussian mixture models (GMMs), which are commonly employed in human language analysis to identify and classify phonemic structures. The study found that the classification accuracy of all three methods – whole-call, unique subunits, and shared subunits – was statistically similar, indicating that killer whale calls could be reliably categorised based on either complete vocalisations or smaller acoustic units. However, the shared-subunit approach required fewer elements to represent the full repertoire, suggesting that killer whales may construct their calls by combining a limited set of fundamental subunits rather than memorising entire calls. Moreover, nearly 75% of stereotyped calls contained at least one shared subunit, reinforcing the idea that these elements function as building blocks of communication. Some of these subunits were also detected in variable (non-stereotyped) calls and in the vocalisations of killer whales from other populations, hinting at a structured and potentially universal vocal system. This pattern parallels human speech, where phonemes are combined to form words, implying that killer whale communication may be governed by combinatorial rules rather than a fixed repertoire of calls. According to Shapiro *et al.* (2011: 383), their data "suggest a rule-based system in which subunits are only arranged in certain orders and combinations". Further research is needed to confirm whether these subunits form the basis of call production and perception.

2.3 Sperm whales

Sperm whales are a well-known species which represents an exception among odontocetes, due to their big size that makes them superficially look like mysticetes.

However, these animals share several features with other odontocetes. First, they have a stable and long-lasting social structure, though different from that of orcas. Here too, the society is matrilineal and possibly hierarchical, but it appears more fluid: groups generally consist of two or three females with their offspring and display strong bonds that can last a lifetime. However, males leave the group after their early years, though they periodically reunite with their relatives and never lose their sense of belonging, maintaining their "dialect" (see below) and habits (Whitehead & Rendell 2015). Despite these bonds, sperm whales spend most of their time on their own (mothers with their calves). Female groups remain in the same area and gather daily at the surface only for brief periods. Males, on the other hand, travel thousands of kilometres and grow to almost twice the size of their female counterparts (Connor *et al.* 1998). One aspect that highlights the bonds among females is cooperative calf care: since sperm whales dive to great depths, calves are initially unable to follow them. For this reason, when the mother dives in search of food, the calf stays at the surface, entrusted to another female in the group. When it is the other female's turn to dive, the mother returns the favour by caring for her calf. Observations indicate that when there are no calves, sperm whales do not adopt this system of shifts (Connor *et al.* 1998). In the Pacific Ocean, these groups divide into clans with distinct "dialects" (see § 2.3.1 below), whereas in the Atlantic Ocean, this division does not seem to occur. Whitehead & Rendell (2015) suggest that this difference may be due to pressure from orcas, which have learned to hunt sperm whales in the Pacific but not in the Atlantic. I have already described the conservatism of orcas: according to Whitehead and Rendell (2015), sperm whales may have structured themselves accordingly, forming tightly knit clans and compact groups to defend against their only possible predators aside from humans. This is supported by Whitehead & Rendell's (2015) observations, which show that Pacific sperm whales assume a coordinated defensive position in the presence of orcas, whereas Atlantic sperm whales remain completely indifferent when orcas pass by. According to Connor *et al.* (1998), the stable societies of sperm whales resemble those of elephants.[125]

2.3.1 The clicks of sperm whales

From the perspective of communication, sperm whales exhibit traits similar to both the "dialects" of orcas (see § 2.2.1 above) and the signature whistles of bottlenose dolphins (see § 4 below). One key distinction from other species is that these animals produce only

[125] It is worth noting that elephants and sperm whales are among the most intelligent animals on the planet, possessing a complex and developed brain structure with spindle neurons and possibly mirror neurons (see Marino 2004).

clicks, meaning that both echolocation and communication rely on this single type of sound. Sperm whales arrange these clicks in rhythmic sequences with temporal arrangements known as *codas*, which convey information in a Morse-like way, and are used for social bonding and interaction (Whitehead & Rendell 2015). Codas can last up to two seconds and consist of 3 to 30 sounds (Berta *et al.* 2006: 296).[126] Some of them are shared among conspecifics, while others are unique and distinctive for each individual (Oliveira *et al.* 2016). It has been suggested that the latter, much like dolphin signature whistles (see § 4), may serve for individual recognition. They may also help group members maintain contact and signal their location, particularly when sperm whales disperse over vast distances in search of food, sometimes separating by many kilometres (Berta *et al.* 2006: 297). Unlike smaller odontocetes, sperm whales cover greater distances, and individuals can remain separated for extended periods; thus, this system would have a clear function in tracking distant companions. In addition to distinctive *codas*, there are also sequences used by multiple individuals, suggesting an additional communicative function unrelated to identity, much like the whistles of other odontocetes. This is not surprising given the evolutionary relationship between these species (Weilgart & Whitehead 1997). The distinctive repertoires of different groups can persist for years and, similarly to orca vocal systems, they show geographical variation, though to a lesser extent. The similarities are significant enough that, according to Weilgart & Whitehead (1997), sperm whale vocalisations can also be considered "dialects", following the definition presented in section 1.4.1. Based on vocal repertoire analysis, Rendell & Whitehead (2003) identified 18 social units among sperm whales, which could be grouped into six acoustic clans. One difference from orcas is that members of these groups can be found thousands of kilometres apart. Whitehead *et al.* (1998), comparing genetic and vocal data, discovered that members of the same acoustic clan are also genetically related. They hypothesised that "dialects", along with other behaviours and hunting techniques, are learned from (or taught by) the mother during early life, likely through cultural transmission, and persist even after the geographical dispersion of individuals. According to Hersh *et al.* (2022), some shared codas might convey the identity of the group, and thus might contain a symbolic reference to their specific cultural clans. Although this would not be comparable to the complexity of human shared symbolic constructions, it represents one of the few claims of 'symbolic culture' in nonhuman animals:

> Our study provides quantitative evidence that identity codas act as symbolic markers of sperm whale cultural (i.e., clan) identity in the Pacific Ocean, with whales adjusting

[126] To listen to sperm whale clicks, see: www.youtube.com/watch?v=HS9OIBifQVw.

> their vocal behavior in ways that could reaffirm their cultural identity in areas of high spatial overlap. The distributional scales and trends seen in Pacific sperm whales resemble those found in human ethnic groups. Although we cannot know for certain how whales perceive codas or label the identity of conspecifics, the observed patterns could arise from the selection for identity signals that enable social assortment and cooperation in sperm whales, as has been shown for humans and suggested for other taxa. (Hersh *et al.* 2022: 6).

Further studies investigating the linguistic components of coda systems would be necessary to determine the presence of syntax, structure-dependent rules, compositionality, and recursion in sperm whale communication. A recent step in this direction is the CETI project[127] (*Cetacean Translation Initiative*), an interdisciplinary initiative aimed at studying cetacean vocalisations through a linguistic approach, attempting to 'decipher' their code. Since artificial intelligence and machine learning have yielded remarkable results in the systematic analysis of human languages, CETI plans to apply these computational tools to the analysis of animal calls to identify structural rules, correspondences, and – possibily – a lexicon and syntax. Over the next years, this ambitious project will specifically analyse the vocalisations of sperm whales, chosen among cetaceans for "their highly-developed neuroanatomical features, cognitive abilities, social structures, and discrete click-based encoding", which "make for an excellent starting point for advanced machine learning tools that can be applied to other animals in the future" (Andreas *et al.* 2021: 1). Naturally, the analyses and results produced by computers depend on the input, so the CETI research team is first focused on gathering and organising an extensive dataset. Sperm whales will only serve as the initial experimental subjects for this approach, and, in my view, this method could lead to significant progress, as such an analysis has never been conducted before – nor was it previously feasible due to inadequate technology. The CETI project team – comprising scientists, cryptographers, engineers, robotics and AI experts, computer scientists, linguists, and marine biologists – is highly optimistic:

> We posit that machine learning will be the cornerstone of future collection, processing, and analysis of multimodal streams of data in animal communication studies, including bioacoustic, behavioral, biological, and environmental data. (Andreas *et al.* 2021: 1).

This project is still in its early stages, but it already stands out for its unprecedented, modern, and original effort to clarify one of the oldest questions in science and philosophy: *do*

[127] Notably, there is a clever resemblance to the SETI project (Search for Extra-Terrestrial Intelligence).

animals have language? However, until results emerge, caution is required: as of today, the meaning of sperm whale coda is poorly understood, and it is not clear whether codas combine to form more complex, meaningful units. In the next chapters I focus on an animal that, for its size, diffusion, habits and communicative skills, has been more thoroughly studied, and offers more insights about the complex "linguistic" properties of cetaceans: the bottlenose dolphin.

3. The bottlenose dolphin

Bottlenose dolphins represent a promising species for the study of communication complexity and the search for linguistic properties in non-human animals. Their sophisticated social structures and complex signalling systems suggest advanced cognitive and communicative abilities (see below). Widely distributed in many seas, including the Mediterranean, dolphins are among the most extensively studied cetaceans, benefiting from long-term datasets collected across different environments. Their relative tolerance to human presence facilitates observational research, while laboratory studies have demonstrated remarkable cognitive skills (see below), reinforcing their potential as a model species.

However, significant challenges remain. Despite extensive research, most dolphin signals remain poorly understood, and their precise functions within social and ecological contexts are difficult to determine. Studying dolphin communication in the wild is further complicated by the aquatic environment, which limits direct observation and individual signal attribution. Underwater behaviour often remains hidden, and vocalisations are typically recorded at the group level, making it challenging to establish clear associations between specific signals and behaviours. Despite these limitations, the wealth of available data and the species' cognitive potential make dolphins an invaluable subject for advancing our understanding of complex communication systems beyond the human domain, as we will see in the following of this thesis. First, I describe some features related to their social structure and their behaviour in the wild, highlighting their complexity and adaptability (§ 3.1). Then, I discuss experiments on captive animals which have shown the ability to understand or perform complex cognitive tasks, some of them related to linguistic properties (§ 3.2). Finally, I describe and analyse dolphin vocalisations in the wild, verifying whether their properties can be considered a form of compositionality or non-recursive merge, as it has been proposed for other nonhuman species (§ 4).

3.1 Society and behaviour of bottlenose dolphins

Bottlenose dolphins, like other delphinids, exhibit a fission-fusion society (Marino *et al.* 2007). This means that they live in large, fluid groups that are constantly changing, forming temporary associations called schools, which last for short periods and generally serve specific functions (e.g., hunting). These schools form and dissolve multiple times a day, with individuals recombining into new compositions. However, more lasting bonds, called bands, also exist. Studies have shown that females sometimes reunite with their mothers for childbirth and child-rearing, forming intergenerational groups. Males, on the

other hand, often form pairs that can live together for years[128] (Whitehead & Rendell 2015). Additionally, males have been observed creating stable alliances in which they cooperate to control fertile females. These alliances can further form alliances of alliances, joining other groups to compete against rival alliances (Whitehead & Rendell 2015). This indicates a hierarchical structure which might even suggest a form of embedding (a group which includes another group). Alliances are organised in three levels: first order alliances involve two dolphins, which can be part of second order alliances (composed by more first order alliances), which can in turn be part of third order alliances (composed by more second order alliances), forming larger and larger groups (King *et al.* 2021, Connor *et al.* 2001). However, although dolphins of both sexes can cooperate in various circumstances, associations between males and females tend to be short-lived. A key ability that enables this continuous formation of new groups is spontaneous imitation, which is also believed to be at least partly responsible for the coordination observed in dolphin movements, whose synchronization (e.g., in breathing cycles when swimming in formation) is extremely precise, down to the millisecond (Jaakkola *et al.* 2018).[129] Dolphins thus rely on cooperation and coordination; furthermore, in their fluid society, they have the need to identify the specific individuals they are going to join in their activity (see § 4 below).

As we have seen, dolphins – like other cetaceans – are long-lived animals with a very complex brain, and they invest energy in the care of the calves (§ 1.2 and 1.3). These features imply a high degree of behavioural and cerebral plasticity, which becomes evident in the differences observed among populations. These differences involve both the repertoire of signals (see e.g. La Manna *et al.* 2020 and § 4 below) and behaviours (Whitehead & Rendell 2015). For example, some foraging strategies or complex techniques are observed only in certain groups, thus indicating that experience plays a role in their formation (Whitehead & Rendell 2015). An effective example is the *mud plume feeding*, a hunting technique which is similar to humpback whales' *bubble net feeding* (see § 2.1). However, unlike the orca and humpback whale strategies, this *behaviour*[130] – where dolphins stir up sediment with their tails to trap fish – is limited to specific communities in the waters south of Florida (Lewis 2003), suggesting it is more of a cultural practice than an instinctual one.

[128] Unions lasting up to twenty years have been documented (see King *et al.* 2021).
[129] Other odontocete families that exhibit fission-fusion societies include beaked whales (Ziphiidae), although they remain understudied (Fedutin *et al.* 2015), and porpoises (Phocoenidae), which show less inclination toward collaboration (Whitehead & Rendell 2015). By contrast, according to a recent study (O'Corry-Crowe *et al.* 2020) belugas demonstrate a more stable yet highly dynamic society, compared to delphinids. They form long-term groups that do not appear to be constrained by sex, age, or kinship. Non-related belugas have been observed traveling together for years, and associations can be either matrilineal or patrilineal. However, further research is needed to clarify these unusual relationships.
[130] For a video of mud plume feeding, see www.youtube.com/watch?v=bzfqPQm-ThU.

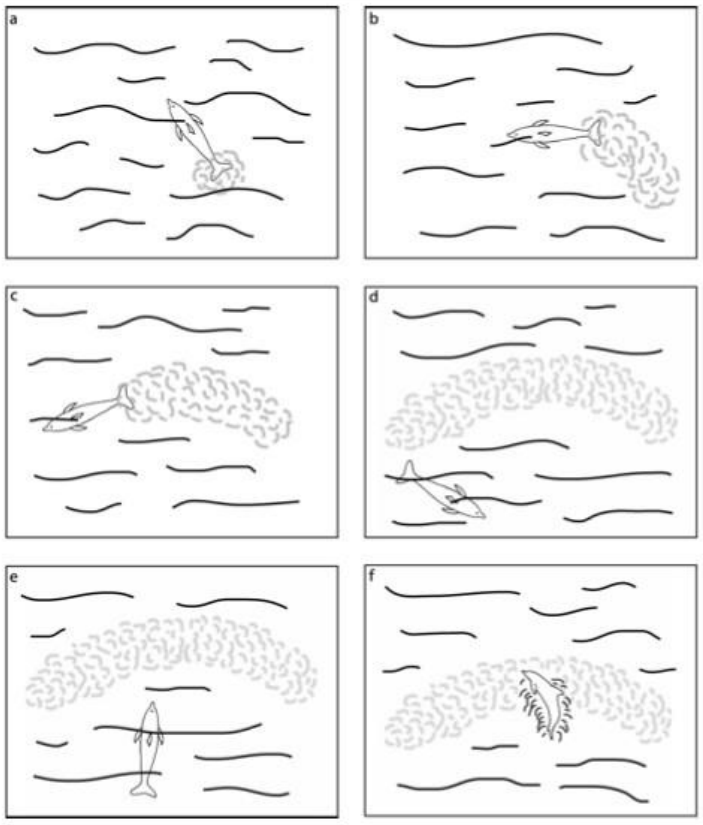

Fig. 31. Illustration of mud plume feeding, where bottlenose dolphins stir up sediment to trap fish, similar to the bubble technique used by humpback whales (Lewis 2003: 94).

Another notable example of variation of techniques among populations is tool use, which is rarely documented outside of humans. Over several years, researchers observed that a small group of bottlenose dolphins in Shark Bay, Western Australia, frequently swam with sponges wrapped around their beaks. Since dolphins in this area forage at depths of around ten meters, the function of this behaviour was initially unclear. Later, researchers confirmed one of Smolker *et al.*'s (1997) hypotheses: the sponges serve as protective coverings, shielding the dolphins' beaks as they dig into the seabed in search of food (Krützen *et al.* 2005). According to Krützen *et al.* (2005: 8939), this phenomenon "classifies as the first case of an existing material culture in a marine mammal species". By analysing mitochondrial DNA, Krützen *et al.* (2005) concluded that this technique is socially transmitted, passed vertically within a single maternal lineage – specifically from mothers to female offspring. Furthermore, genetic evidence suggests that all '*spongers'* trace back to a common ancestor – dubbed the *Mitochondrial Sponging Eve* by the study's authors – who first developed and culturally transmitted the technique, leading to a lineage of *spongers*. Numerous other traits also merit further exploration, such as interspecies collaboration

(Herzing & Johnson 1997). One striking example occurs in Laguna, Brazil, where dolphins spontaneously *assist* human fishermen by herding fish toward their nets, then catching the escapees for themselves – a mutually beneficial arrangement[131] (Machado *et al.* 2019). This phenomenon is not exclusive to Brazil; similar cases have been documented elsewhere and even mentioned by Pliny the Elder (Frohoff & Oriel 2017). Other behaviours worth investigating include surface activities such as breaching, tail slaps, spontaneous acrobatics, and other movements that lack an apparent function. Additionally, dolphins exhibit traits generally associated to humans like curiosity, an instinct to help (Kuczaj *et al.* 2015), and a strong inclination toward plays (Kuczaj & Eskelinen 2014).

Another feature which might reflect complex social dynamics are dolphin reactions in response to dead conspecifics, displaying what researchers interpret as epimeletic (helping) or possibly grief-related behaviours (Bearzi *et al.* 2017). Various documented cases suggest that dolphins may engage in physical interactions with carcasses, prolonging contact with deceased individuals, which raises intriguing questions about their cognitive and emotional capacities. Research has recorded instances of dolphins attempting to keep dead calves afloat, sometimes carrying them for extended periods (see e.g. Pedrazzi *et al.* 2022, Dudzinski *et al.* 2003). Such behaviours, often attributed to maternal instincts, are thought to reflect a form of social attachment or an inability to immediately recognise death. This phenomenon, known as "postmortem carrying", has been observed in several cetacean species and may serve an adaptive function in maternal learning or social bonding (Bearzi *et al.* 2017). Additionally, some reports suggest that dolphins engage in what has been described as "resuscitation attempts" (Bearzi *et al.* 2017). In these cases, individuals may push, nudge, or even strike a dead dolphin, possibly as a response to distress or an attempt to revive the individual. Such behaviour parallels observations in other mammals that exhibit denial of death as part of their grieving process (Bearzi *et al.* 2017).

3.2 Cognition of bottlenose dolphins

Dolphins demonstrate many remarkable cognitive abilities, which are present in the wild but become more evident in captive individuals, who are constantly exposed to human stimuli. An example is **imitation**, both vocal and gestural, in which bottlenose dolphins are among the most proficient species. As cited in Marino *et al.* (2007: 969):

[131] Machado *et al.* (2019), while studying this dolphin community, observed that they exhibit stable bonds and spend more time with their preferred companions, even just to 'play.'

> Bottlenose dolphin abilities for both arbitrary vocal and motor imitation were demonstrated at the Kewalo Basin Marine Mammal Laboratory in Honolulu. Vocal imitation was investigated by broadcasting electronically generated "model" sounds underwater into a dolphin's habitat. In response, the dolphin vocalised into a hydrophone. [...] A variety of different waveforms were imitated accurately; the imitations of sounds G and H show spontaneous octave generalization, the imitation occurring precisely an octave above (G) or an octave below (H) the model sound. Octave generalization is a rare ability that, for example, has not been elicited from songbirds.

For gestural imitation, Marino *et al.* (2007: 969) observe:

> Social motor imitation was demonstrated first by having two dolphins side by side with a partition between them that allowed the dolphins to see each other but not their respective trainers. The "demonstrator" dolphin was instructed gesturally by its trainer to perform one of many possible behaviors, including its own self-chosen behavior. Then, the "imitator" dolphin was instructed by its trainer to either "mimic" the demonstrated behavior or to perform another behavior. Both dolphins successfully imitated familiar and novel modeled behaviors. This ability generalized easily to imitating human behaviors demonstrated either at poolside or on a television monitor placed behind an underwater window. Motor mimicry also extended to self-imitation, the imitation of one's own previous behavior. No nonhuman animal has shown the levels of diversity, flexibility, and cognitive control of imitative skill demonstrated in bottlenose dolphins.

According to Marino *et al.* (2007: 968), dolphins exhibit "an understanding of symbolic representations of things and events (declarative knowledge); an understanding of how things work or how to manipulate them (procedural knowledge); an understanding of the activities, identities, and behaviors of others (social knowledge); and an understanding of one's own image, behavior, and body parts (self knowledge)." (cf. Herman 2006). All these elements are facilitated by a well-developed memory (Marino *et al.* 2007: 968). Among social knowledge, the ability of bottlenose dolphins to comprehend the indications of other individuals is highlighted. They not only look in the direction in which a human examiner points or gazes but also observe the object being indicated. Additionally, "dolphins can use their rostrums and body alignment to point and direct a human swimmer to an object or place of interest and monitor whether the human receiver is attending to them" (Marino *et al.* 2007: 969, see also Xitco *et al.* 2001).

More specifically, experiments on captive animals showed that dolphins are capable of complex cognitive operations. For example, in two experiments conducted by Jaakkola *et al.* (2005), bottlenose dolphins demonstrated a sense of **numbers**, or at least an understanding of numerical inferiority. Initially, they were trained to select, in a comparison, the specimen with the fewer markings (for instance, when given a choice between an individual with five marks and one with two, they correctly selected the latter). Then, the dolphins were exposed to all possible pairwise combinations of numbers between one and eight, none of which had been shown during training, and they successfully generalised the rule: "the dolphins chose the exemplar with the fewer number of items at levels far above chance, showing that they could recognize and represent numerosities on an ordinal scale" (Jaakkola *et al.* 2005: 1; see also Jaakkola 2012). Furthermore, Kuczaj & Walker (2012) explore the cognitive mechanisms underlying **problem-solving** in bottlenose dolphins, emphasising the distinction between trial-and-error learning and more complex cognitive processes such as planning and insight. Their study considers various experimental approaches to evaluating dolphin intelligence, including their responses to novel challenges, the potential for syntactic processing, and the spontaneous use of pointing as a communicative strategy. One of the key aspects examined is how dolphins process gestural sequences, an ability that has led researchers to debate whether their behaviour indicates a 'syntactic' understanding or simply a sophisticated information-processing strategy. Additionally, the authors investigate whether dolphins can plan their actions in response to novel problems, a crucial indicator of higher-order cognition. Experimental evidence suggests that rather than relying solely on instinctual or conditioned responses, dolphins display a degree of foresight in their behaviours, adjusting their strategies in ways that indicate conscious decision-making. This challenges traditional views that associate such abilities exclusively with primates. The role of **play** in the ontogeny of problem-solving skills is also discussed, with findings suggesting that juvenile dolphins engage in playful behaviours that mirror more complex problem-solving strategies observed in adults. This supports the hypothesis that play serves as a mechanism for cognitive development, providing opportunities for individuals to experiment with novel interactions and refine their problem-solving abilities in a low-risk environment. Another study focused on problem-solving abilities and **coordination** was published by Yamamoto *et al.* (2019), who proposed to dolphins an experimental rope-pulling task adapted from previous research on chimpanzees. Three individuals were tested to assess their capacity to coordinate actions, i.e to pull opposite ends of a rope simultaneously, in order to obtain a reward. The dolphins first learned to manipulate the apparatus individually before progressing to cooperative trials,

where successful task completion depended on synchronization between partners. The results showed that dolphins were capable of cooperative problem-solving, with success rates improving over time as they learned to wait for their partner. The introduction of delays revealed individual differences in strategic adjustments; while some dolphins were more consistent in coordinating their actions, others displayed hesitation or misinterpretations of their partner's behaviour. These findings suggest that bottlenose dolphins possess a degree of cognitive flexibility and social awareness that allows them to adapt to cooperative challenges. This ability of coordination is expressed also through acoustic communication: King *et al.* (2022) provide evidence that dolphins can use vocalisations to solve cooperative challenges, modifying their vocal signals to facilitate coordination.

Another crucial feature is **memory**. Research suggests that dolphin memory is multifaceted, incorporating several types, including short-term, long-term, social, and spatial memory. Studies on dolphin cognition indicate that they possess a robust short-term memory system, particularly in tasks involving echolocation and auditory processing. Dolphins can retain and process echolocation clicks for up to several seconds, allowing them to construct detailed mental representations of objects in their environment (Bruck 2013). Working memory is also evident in delayed matching-to-sample tasks, where dolphins can remember and correctly identify visual and auditory stimuli after short delays (Herman *et al.* 1984). Dolphins navigate vast oceanic environments using a combination of spatial and procedural memory. They can remember foraging sites, migration routes, and feeding strategies across seasons, a skill that likely involves the hippocampus, which is well-developed in cetaceans (Manger 2006). Experimental studies have shown that dolphins can recall problem-solving strategies over extended periods, indicating the presence of procedural memory (Mercado & DeLong 2010). One of the most striking findings in dolphin cognition is their long-term social memory. Bruck (2013) demonstrated that bottlenose dolphins can remember the signature whistles of former tank mates for over 20 years, the longest recorded non-human memory. This ability suggests that dolphins possess an advanced episodic-like memory system, which is crucial for maintaining complex social relationships. Episodic memory, often associated with the conscious recall of personally experienced events, has been considered a uniquely human trait. However, Davies *et al.* (2022) provide evidence that dolphins exhibit a form of episodic-like memory, characterised by the ability to recall past events without the necessity for language-based reports. The research highlights that dolphins can remember specific past experiences and contexts, in a manner that may not be purely instinctual, but rather a cognitive process that enables them to recall and react to past events in flexible and adaptive ways.

One of the most relevant studies about dolphin cognition, connected with **self-awareness**, is that of Reiss & Marino (2001), who verified whether bottlenose dolphins were capable to recognise themselves in a mirror. The experiment, which yielded positive results, has not been demonstrated in other cetacean species. For their study, thoroughly described in Reiss & Marino (2001), the researchers used two captive bottlenose dolphins. During several videotaped sessions, the dolphins were placed in front of one or more reflective surfaces under different conditions: in some cases, they were left untouched; in others, they were marked with non-toxic ink; in a third possibility, experimenters only pretended to mark the dolphins (they were touched as if to mark them but without actually applying any ink). This was done both before and after feeding. According to Reiss & Marino (2001), when the dolphin approached the reflective surface, it paid particular attention to it only if it had been marked, engaging in self-exploratory movements to inspect itself. This idea is supported by at least three observations: first, the dolphin did not exhibit social behaviours toward the reflection, indicating that it did not perceive it as another individual; second, when multiple reflective surfaces were present, it focused on the clearest one; finally, while the dolphin initially ignored sham markings, over time, it started examining itself even in those cases, but only after multiple sessions in which it had actually been marked (Reiss & Marino 2001: 5938-5940). The dolphin, therefore, after learning that a certain action resulted in a mark on its body, inspected itself even when the action was only simulated. Compatible results were recorded in the second dolphin as well. Reiss & Marino (2001: 5942) observe that "collectively, these findings provide definitive evidence that the two dolphins in this study used the mirror (and other reflective surfaces) to investigate parts of their bodies that were marked. These findings, therefore, offer the first convincing evidence that a nonprimate species, the bottlenose dolphin, is capable of MSR." As Marino (2004: 11) summarises, this ability "is a component of the more general capacity for self-awareness", which "refers to the ability to think about one's own behaviour, physical make-up, thoughts and feelings". Furthermore, the experiment highlights the dolphins' awareness of their own body parts. Marino concludes that, although many investigators have hypothesised that the cortical circuitry of the prefrontal cortex in humans and great apes is the necessary neuroanatomical substrate for self-recognition and other dimensions of self-awareness", the 2001 experiments "show that dolphins are capable of mirror self-recognition despite possession of unelaborated frontal lobes (or homologous frontal lobe structures) and an overall cortical organizational pattern that is very different from that of primates" (Marino 2004: 12). Consequently, "this finding challenges the view that self-awareness is specific to primate frontal lobe architecture and suggests that self-awareness could be more of a holistic property

of brains of any species that achieve a certain threshold of elaboration regardless of how that elaboration occurred" (Marino 2004: 12). This also highlights the sophisticated memory abilities of cetaceans (see above), and raises the question of how these animals exhibit such mnemonic skills despite having an underdeveloped hippocampus (Marino 2004: 12). This, too, suggests that evolution can reach similar cognitive outcomes through different neural pathways. The cognitive abilities of *agency* and *ownership* – i.e., the awareness of being an agent of change and the awareness of having ownership over one's body – have been highlighted also by Herman (2012). In experiments with the bottlenose dolphin Elele (see Herman 2012: 533–539), the animal demonstrated knowledge of its own body parts and correctly associated them with signals established by human researchers. Elele obeyed various commands, such as "touch the glass with your dorsal fin" or "hit the frisbee with your tail and then jump." Beyond the taught combinations, the dolphin was also tested with novel instructions and was able to independently perform the requested actions. Once Elele had learned the repertoire of symbols associated with body parts, she could use them in tasks that had not been previously tested.[132]

3.3. Language-like skills emerged in captive dolphins

The abilities described in the previous sections raise the question of whether they are also used in complex communication, and what kind of properties might underlie them. Syntax, compositionality, and recursion are features that appear to require high cognitive potential; thus, dolphins serve as an intriguing case study for researching these traits in nonhuman animals. According to Marino (2004: 11), bottlenose dolphins "have also demonstrated extensive capacities in the ability to understand an artificial (i.e., human devised) communication system based upon symbols and rules". Before them, only some primates and African grey parrots succeeded in this task (see Part II, § 4). Furthermore, dolphins understand that word (or symbol) order can alter the meaning of a message (see below), an ability previously detected only in captive chimpanzees and bonobos (see Part II, § 4). Although Marino is cautious about comparing studies on such different animals, she believes that "the fact that studies of comprehension of artificial communication systems can only be attempted with these very few species suggests that they all have converged toward a level of cognitive complexity that allows them to understand a simple symbolic and rule-

[132] This bears similarities to Chomsky's nativist argument against behaviourism, which refers to the human ability to create novel sentences by combining known elements. Although the bottlenose dolphins demonstrated passive competence related to comprehension in this experiment, it cannot be ruled out that they use this ability actively in nature or, at the very least, that they have the potential to do so.

based system of communication" (Marino 2004: 11). It seems thus reasonable to ask whether the communicative abilities of dolphins display properties related to language, and to what extent experiments with humans can determine this. Kuczaj (2013) asserts that communication is different from conversation; while the former has been achieved in experiments with humans, the same cannot be said for the latter: "there is no compelling scientific evidence that humans and dolphins have engaged in meaningful conversations that involved mutual exchanges of information" (Kuczaj 2013: 114). The attempts of this kind of interaction have given poor results. Xitco *et al.* (1999) tried to verify the ability to understand 'syntactic-like' combinations of signals, using an underwater keyboard where both humans and captive bottlenose dolphins could activate keys representing objects, places, and activities. Without prior training, and by observing humans – who activated a symbol and then swam to the corresponding object/place or performed an action – the dolphins began using the keyboard correctly, though only with single units and opportunistically. However, the limitations of the keyboard made it difficult to create distinguishable syntactic sequences due to the slow movements in water and the time between each symbol, making it hard for dolphins to perceive sequences. Herzing *et al.* (2012) attempted to communicate with wild dolphins using a portable underwater keyboard with acoustic and visual signals, but they had no success. According to them, this failure may be due to the differences in the communication systems of animals as distant as dolphins and humans, and not to a lack of complex cognitive and communicative abilities related to language. These kinds of abilities have been confirmed in experimental results by scholars such as Herman in many of his works (1984, 1999, 2002, 2012).

The groundbreaking experiments by Herman *et al.* (1984) with the dolphins named Phoenix and Akeakamai showed that they could understand a system of gestural (Akeakamai) and vocal (Phoenix) symbols that they had been taught. The symbols referred to objects or body parts: before giving a command, the experimenters played the signature whistle related to the identity of the dolphin[133] that was to execute the task, and then transmitted (through gestures or electronically produced whistles) a combination of commands, such as *Phoenix + touch with tail + frisbee + from below. Dolphins were able to execute the commands, but not to produce the artificial language, so that only the comprehension was tested.* Figure 32 gives an example of some of the symbols utilised.

[133] See § 4 below for a detailed description and analysis of signature whistles.

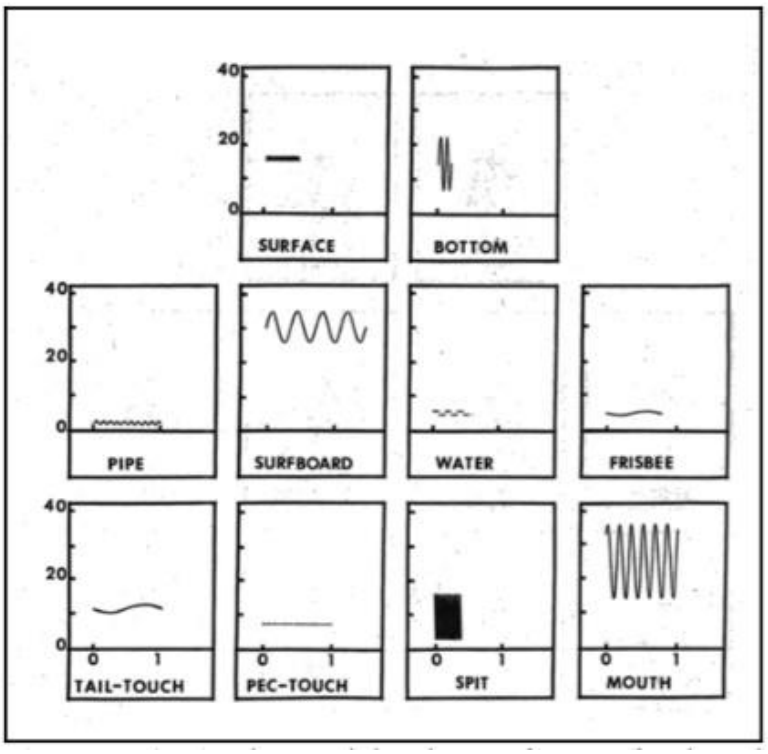

Fig. 32. Example of some of the symbols used in the experiment by Herman *et al.* (1984). Each of the represented frequencies, which were generated electronically, is associated with an element in an arbitrary manner: the first row contains modifiers (e.g., *on the surface, at the bottom*), the second row lists objects (e.g., *water, frisbee*), and the third row includes actions (e.g., *touch with tail, touch with mouth*).

The language was made by three constituents: a destination object (D), a transportable object (T) and a relational verb (R). For example, in the sequence "*basket frisbee fetch*", the dolphins had to bring the frisbee next to the basket, and in "*frisbee basket fetch*" they had to bring the basket next to the frisbee (these objects were located in the pool). The order of symbols was modified to test whether the animals understood that a different arrangement corresponded to a different type of action, thus conveying a different message. Considering that dolphins correctly executed these commands, Herman *et al.* (1984) observed that they were capable to understand the word order of this language. Table 9 highlights this contrast.

Destination object (D)	**Transport object (T)**	**Relational verb (R)**
BASKET	*FRISBEE*	*FETCH*
Bring the frisbee to the basket		
FRISBEE	*BASKET*	*FETCH*
Bring the basket to the frisbee		

Table 9. Differences in word order of the artificial language taught to dolphins by Herman *et al.* (1984).

Even more interestingly, the dolphins were sometimes deceived with false symbols or combinations that did not follow the taught "syntax". In these cases, the dolphins were able to detect and repair the sequences that did not correspond to the rules they had learned. The examiners sometimes presented syntactically incorrect or semantically impossible sequences to the dolphins. The syntactically incorrect ones had one extra constituent, e.g. "*Phoenix hoop frisbee fetch*": in this case the dolphin named Ake tried to repair the structure, by eliminating one of the extra elements. He could bring the hoop to Phoenix (the other dolphin), bring the frisbee to Phoenix, or the frisbee to the hoop. Another example of anomalies was "agrammatical" constructions, where the dolphin's signature whistle was repeated in an incorrect position within the sentence (e.g., *signature* + *touch with tail* + *signature* + *frisbee*). In these cases, the animals exhibited reactions of irritation. In the semantically impossible sequences, which were syntactically correct but impossible to execute, for example, dolphins were asked to move a non-transportable object, e.g. "*basket speaker fetch*", "bring the speaker (fixed on the pool) to the basket". In this case, the dolphins refused to perform any kind of combination, and never tried to bring the basket next to the speaker, which was physically possible but in contrast to the system they had learned. These different behaviours[134] led Herman *et al.* (1984) to claim that dolphins recognised the argument structure of that language, i.e. they knew that there were syntactic positions and relationships between words which could not be violated (see also Herman & Uyeyama 1999), hence trying other combinations would have been useless. Also in these situations, dolphins appeared to be frustrated. Table 10 summarises these observations.

Construction	**Example**	**Reaction**
Syntactically incorrect (one extra constituent)	*PHOENIX HOOP FRISBEE FETCH* *(DT_1T_2R)* *Bring the frisbee hoop to Phoenix*	**Repair** (excluding one element)
Semantically impossible (non transportable object)	*BASKET SPEAKER FETCH* *(TDR)* *Bring the speaker to the basket*	**Refusal** (no inverse combination)

Table 10. Contrast between syntactically incorrect and semantically impossible sequences in Herman *et al.* (1984) experiments.

[134] Numerous experiments on humans have shown that brain activity exhibits a certain type of alteration when exposed to syntactic errors and a different type of alteration when exposed to semantic errors. This suggests that the brain not only registers these inconsistencies in some way but also reacts differently to them (see Tettamanti 2020, for an overview). A similar study applied to bottlenose dolphins – if feasible – would be useful in identifying potential parallels in sensitivity to syntactic and semantic coherence.

Marino *et al.* (2007), summarising the results of these experiments, state that dolphins are capable of understanding word order, that they "spontaneously learn associations between sounds and temporally paired events", and "a variety of types of abstract rules" (Marino *et al.* 2007: 968), which they have demonstrated the ability to generalise. Marino *et al.* (2007: 968) further add considerations with significant implications:

> Learning of an imposed language is perhaps the most challenging cognitive task that dolphins have faced in the laboratory. Dolphins learned to understand not only the semantic features of artificial gestural and acoustic languages, but also the syntactic features (Herman *et al.* 1984). Learning of complex syntactic structures or decoding of anomalous structures was often achieved through inference, rather than through explicit instruction (Herman *et al.* 1993*)*. (Marino *et al.* 2007: 0968).

However, it is important to consider that the cited studies date back to the last century. Contrary to expectations, a crucial aspect such as learning "complex syntactic structures" has not been further explored. As a result – at least based on the available material – there are currently no new studies that either confirm or refute this ability. For this reason, I believe that claims regarding syntactic skills should be taken with caution, but they cannot be entirely dismissed as unfounded. Herman, nonetheless, maintains that his studies revealed in dolphins the ability of "acquisition of the closed-class functionality of demonstratives, prepositions, conjunctions, and locatives", as well as "sensitivity to argument structure […] also evidenced by wholly untrained and consistent interpretations of the dolphin to probes of anomalous syntactic structures." (Herman & Uyeyama 1999: 18). Regarding demonstratives, Herman & Uyeyama (1999) discuss the dolphins' understanding of gestural sequences such as "*this fetch that*", in which *this* and *that* were represented by the gesture of pointing towards a close or far object, and the verb *fetch* was produced with its corresponding symbol (which could be a gesture or a whistle, depending on the examined dolphin). Despite the ability to correctly perform the commands indicates a developed skill to understand gestural combinations (and maybe even a form of "multimodal" communication of gestures and whistles), a different approach could have led to more convincing results. For example, the scholars could have taught the dolphins an abstract symbol for *this* (a close object) and another one for *that* (a distant object) and propose the sequence without gestures. Similarly, regarding prepositions, Herman *et al.* (1984) tested the contrast between the verbs *fetch* ("put beside") and *in* ("put inside"), as in "*basket frisbee fetch*" ("put the frisbee *beside* the

basket") and "*basket frisbee in*" ("put the frisbee *inside* the basket").[135] Here, two separate prepositions and a generic verb could have been taught, i.e. a symbol for *beside*, a symbol for *inside*, and a symbol for *put*. However, Herman & Uyeyama (1999) explicitly state that they did not consider it necessary, and the contrast between the verbs *fetch* and *in* is sufficient to deduce the understanding of prepositions. More interesting is the case of conjunctions: in a sentence like "*ball and hoop over*", the dolphins had to swim over the ball *and* the hoop (objects which were present in the pool). To ensure that the signal *and* was really perceived as a symbol with a meaning, and not a mere reinforcement (in the latter case, the animals would interpret any signal between *ball* and *hoop* as a conjunction), the experimenters taught dolphins also the symbol *erase*, which implied that all the elements which came before it should be ignored. Hence, in the case of "*ball erase hoop over*", the dolphins had to swim over the hoop but not over the ball. This contrast suggest that dolphins really interpreted *and* as a conjunction, and *erase* as a sort of negation. Finally, regarding locatives, Herman *et al.* (1984) taught the symbols *left* and *right*, which were combined with the symbols corresponding to the objects, obtaining, for example, "*left frisbee*" or "*right frisbee*". Hence, a sentence as "*left frisbee basket fetch*" meant "bring the basket next to the frisbee at your left", and "*frisbee left basket fetch*" meant "bring the basket at your left next to the frisbee". Table 11 summarises these observations.

Item	Details	Example
Demonstratives	Understanding of gestural sequences	*THIS FETCH THAT*
Prepositions	*FETCH/IN* contrast	*PUT BESIDE/PUT INSIDE*
Conjunctions	*AND/ERASE* contrast	*BALL AND HOOP OVER* (O_1ANDO_2A) *BALL ERASE HOOP OVER* (O_1ERASEO_2A)
Locatives	*LEFT/RIGHT* as spatial representations	***LEFT FRISBEE*** *BASKET FETCH* ($_{LEFT}$DTR) *FRISBEE* ***LEFT BASKET*** *FETCH* (D_{LEFT}TR)

Table 11. Summary of "closed-class items" identified by Herman & Uyeyama (1999) in bottlenose dolphins' experiments.

[135] These two examples are a simplification for making the argument clearer.

The issue of locatives is particularly interesting for the topic of this dissertation, because it shows a simple and yet convincing case of combination of two meaningful units. The item *left* acts as a modifier of the object (in the example, *frisbee*), and it always precedes the object it modifies. Even in the case of "*frisbee left basket fetch*", in which *left* is adjacent to both the objects (*frisbee* and *basket*), the dolphins seemed to consider "*left basket*" as a whole unit, respecting the 'syntactic' rule according to which the locative precedes the object. From a linguistic perspective, "*left frisbee*" and "*left basket*" seem to behave as constituents. Furthermore, they are inserted into a more complex sequence (e.g. "*frisbee left basket fetch*"). It is difficult to clarify whether the whole sequence was interpreted hierarchically or in a linear way. However, sticking to the simple combination of locative + object, the question that arises is if "*left frisbee*" and "*left basket*" can be considered cases of compositionality comparable to those described in Part II (§ 3) regarding nonhuman primates and birds, although only from the perspective of comprehension. In captivity, Kanzi was able to understand combinations of symbols taught by humans (Savage-Rumbaugh 1986), which can be considered cases of 1-merge, according to the compositional interpretation (see Part II, § 3). In my opinion, there is an interesting resemblance with both the combinations observed in the wild in species as chimpanzee and Japanese tits, and in captivity with Kanzi and other apes (although only in comprehension). Hence, the cognitive ability of non-recursive 1-merge might indeed be present in bottlenose dolphins. The experiments by Herman *et al.* (1984) were not designed to test the ability to produce a complex language, so the issue whether dolphins can combine meaningful units or not remains open. The few attempts in this direction did not give the hoped results (see Xitco *et al.* 1999), mainly due to technical difficulties related to the anatomy and cognition of dolphins, which is very different from humans and nonhuman primates.

However, according to Herman, "to ignore the considerable capabilities of dolphins for complex information processing diverts attention from general cognitive processes that could in common explain dolphin and ape capabilities for syntactic processing, as well as other language-like or cognitively complex skills" (Herman & Uyeyama 1999: 21). While it should be acknowledged that dolphins have not demonstrated a syntax comparable to that of humans – and more studies would be needed to confirm or refute this – Herman highlights an aspect that I find particularly interesting: since human language is linked to cognitive aspects such as categorization, sequential processing, and referentiality, studying animal language "can help us to identify with more surety those processes in humans that may derive

from general cognitive structures rather than from language-specific structures" (Herman & Uyeyama 1999: 22).

It shall be noted that all the experiments with captive individuals described in this section do not clarify what the skills of dolphins are in their environment. However, they highlight cognitive latent abilities and a potential which is also present in wild animals, thus encouraging the research of language properties that might emerge in their ecosystems. Hence, in the next chapter I focus on the vocalisations of wild dolphins, highlighting their complexity and analysing their structure, in search for clues of compositional skills and other linguistic features.

4. Vocalisations of bottlenose dolphins

Bottlenose dolphins produce a diverse range of sounds, which have been studied for over 60 years. Their vocal repertoire spans frequencies up to 150 kHz, far exceeding human auditory perception. Jones *et al.* (2020) categorise dolphin vocalisations into clicks, burst pulses, and whistles. **Clicks**, typically used for echolocation, are broadband sounds emitted in rapid sequences. Their peak frequencies range from 20 to 130 kHz. Click trains (i.e., series of clicks) vary in inter-click interval (ICI), with shorter ICIs (i.e., clicks separated by short pauses) characterising burst pulses rather than echolocation. Some studies suggest that dolphins utilise packets of clicks during long-range echolocation. **Burst pulses** comprise rapid click sequences that exhibit graded acoustic properties. These include barks, which feature variable production rates and high frequencies (up to 16 kHz), and victory squeals, emitted post-successful task completion. Other subtypes include squawks, which typically last over 0.2 s and range in frequency from 3.94 to 21.26 kHz, and brays, a sequence of alternating sounds associated with feeding events (see below). **Single pulses**, such as pops and thunks, are isolated broadband pulses that often feature low-frequency energy (<3 kHz). Pops, sometimes produced in sequences, have been linked to aggressive or reproductive behaviours. Thunks, originally described in mother-infant separation contexts, exhibit harmonic structures extending up to 11 kHz. **Whistles**, the most studied vocalisations, are frequency-modulated tonal sounds, which range in duration (0.06–5.4 s) and fundamental frequency (0.80–41.8 kHz), with some containing harmonics extending beyond 100 kHz. Early descriptions likened them to bird calls, but later studies confirmed their production via extremely high repetition rate clicks. There are two types of whistles: signature whistles and non-signature whistles. There is no structural difference between them; however, signature whistles have been interpreted as identity markers (see § 4.1 below for a detailed description), while non-signature whistles are poorly understood, and they are believed to be context-specific shared calls (at least some of them). Furthermore, short-duration tonal emissions, including chirps and twitters, further contribute to the complexity of dolphin acoustic communication. Chirps, often sweeping in frequency, last between 0.01 and 0.25 s. The distinction between chirps, twitters, and microwhistles remains an area for further research. Finally, dolphins are capable of **biphonation**, i.e. the simultaneous production of two different types of sound at the same time. Biphonic signals arise from the independent control of the phonic lips within the dolphin's nasal complex, enabling the emission of distinct tonal components. This mechanism allows for enhanced signal complexity,

facilitating individual recognition and interspecific interactions.[136] Overall, it should be kept in mind that all the sounds produced by bottlenose dolphins, even those which appear continuous like whistles, are formed by the repetition of discrete units at differing rates and acoustic features, thus opening to the possibility that they are formed following combinatorial processes. Table 12 summarises the types of sounds classified by Jones *et al.* (2020).

Clicks	Discrete units emitted in series, also used for echolocation.
Whistles	Click series with extremely high repetition rates, modulated in frequency and amplitude (perceived as tonal whistles by humans).
Burst pulsed sounds	Click trains whose repetition rate increases and decreases, with changes of frequency (bray calls, quacks, LFN).
Single pulsed sounds	Emission of one click at a time, usually broadband, loud and intense pulses, with possible stun function (bangs, cracks, pops).
Biphonation	Production of two different sound types at the same time.

Table 12. Types of sounds categorised by Jones *et al.* (2020).

The sounds described are used by dolphins with remarkable acoustic plasticity, with vocalisations varying across populations, geographic regions, and in response to anthropogenic influences. This flexibility is evident in their ability to modify vocal repertoires to accommodate environmental conditions, social structures, and human-induced changes (see e.g. La Manna *et al.* 2020). Geographic variation in dolphin vocalisations arises due to environmental heterogeneity and population separation. For example, studies have shown that whistle structures differ among Mediterranean bottlenose dolphin populations, with Croatian dolphins displaying distinct acoustic patterns from those in Sicily and Sardinia, likely due to geographic isolation and local ecological adaptations (La Manna *et al.* 2017, 2019). Such divergence may be driven by genetic differentiation or regional acoustic environments, further corroborating the role of habitat in shaping vocal behaviour. Population-specific vocal variations have been documented extensively, with dolphins demonstrating unique whistle characteristics within distinct social groups. This variation is reinforced through vocal learning and social interactions (Janik & Sayigh 2013). Anthropogenic factors significantly impact dolphin vocal behaviour, particularly in environments with high human activity. Increased boat traffic and underwater noise

[136] Some studies identified even signature whistles composed by biphonal calls (see e.g. Papale *et al.* 2015).

pollution have been linked to modifications in whistle frequency, duration, and inflection rates, likely as a compensatory mechanism to maintain communication in noisy conditions (Rako-Gospic *et al.* 2020, La Manna *et al.* 2020). Such adaptability underscores their capacity to respond to environmental stressors, albeit with potential long-term consequences on communication efficiency and social cohesion.

Context-dependent vocalisation changes further demonstrate the dynamic nature of dolphin acoustic communication. Whistle characteristics can shift depending on behavioural context, with increased modulation and repetition observed during social interactions, foraging, and distress situations. Additionally, vocal convergence has been recorded among interacting dolphins, where individuals modify whistle contours to resemble those of social partners, reinforcing affiliative bonds and group cohesion (Quick & Janik 2012). These findings emphasise the importance of vocal learning and context-driven modifications in dolphin communication. The extensive plasticity of dolphin vocal behaviour highlights their ability to navigate complex acoustic environments and maintain effective communication despite environmental challenges. The interplay of geographic separation, social structures, and anthropogenic influences underscores the intricate and adaptive nature of their vocal repertoires.

After having clarified these general features of dolphin vocalisations, I focus on the two call types which are more interesting for the research of compositionality and other language properties: signature whistles and bray calls.

4.1 Signature whistles

Bottlenose dolphins spontaneously produce signature whistles, i.e. learned vocal labels with great inter-individual variation, which indicate the identity of the dolphin and its location. Signature whistles are narrowband frequency-modulated sounds, often with harmonic structure, that are unique to each individual and not shared with conspecifics. This makes each specimen recognizable within the group thanks to its signature (Berta *et al.* 2006: 294).[137] These signatures, first described by Melba and David Caldwell (1965) in bottlenose dolphins, have since been identified in various odontocete species (*Tursiops truncatus, Delphinus delphis, Lagenorhynchus obliquidens, Stenella plagiodon, Sousa chinensis*). It is believed that they not only convey the identity and position of the whistling individual, allowing others to locate it, but also communicate additional information related to emotions, elements of interest, or states of calm, danger, or motivation, depending on variations in the

[137] For a recording of dolphin whistles, see www.youtube.com/watch?v=yje8IYAmnl4.

sound. Kuczaj *et al.* (2015) showed that when a bottlenose dolphin produces its signature whistle in a state of distress or suffering, distant individuals react differently compared to a call produced in calm conditions. There is extensive evidence that signature whistles serve a broad social function beyond identification: dolphins whistle their signature more frequently when separated from other group members, with no visual contact, or in response to the whistles of other individuals.[138] A key aspect is that dolphins not only produce their distinctive signature whistle, but they are also capable of imitating the signature of other individuals to address them. According to King & Janik (2013), dolphins use these signatures as learned vocal labels, voluntarily employing them to call a specific group member with whom they intend to initiate a social interaction. From a linguistic perspective, they behave similarly to human names, because each dolphin has its own, and can also use those of the others in a "vocative-like" form. The "called" dolphin can also respond by imitating the signature whistle of its interlocutor, establishing contact. Each signature whistle thus refers to a specific entity of the world: in the comprehension, dolphins showed more intense reactions to the signatures of closely related or socially bonded individuals than to the signatures of more distant social contacts, though still familiar (Janik & Sayigh 2013, Janik *et al.* 2006), indicating that they can discriminate and recognise other specific dolphins through their signatures. In the production, the copying of the signatures of other individuals suggests that these animals try to reach specific dolphins, implying that they can associate the label to the identity. Furthermore, dolphins copy the frequency modulation pattern of the signatures of other specimens, but the imitation of the whistle is not identical, so that the animals are able to recognise the information about the identity and at the same time understand if it is produced by its owner or not (Janik & Sayigh 2013). Janik & Slater (1998) further showed that signature whistles are produced more frequently when a group member moves away or when individuals reunite, highlighting their function in social cohesion. According to Bruck (2013), a bottlenose dolphin can recognise a former companion lost at sea even after twenty years of separation, emphasising not only the effectiveness of signature whistles but also the exceptional long-term memory of the species (see § 3 above). It has been observed that dolphins exchange signature whistles both in the wild, when they meet at sea (Quick & Janik 2012) and in captivity, before starting more complex interactions (King *et al.* 2021).

Particularly interesting is the behaviour of whistle matching, where one dolphin mimics the frequency contour of another dolphin's whistle in real time (and may involve

[138] It has also been observed that the most sociable species of odontocetes whistle more often.

signature or non-signature whistles). This phenomenon is considered a form of social communication, potentially serving functions such as identity addressing, social bonding, and coordination during cooperative behaviours. Whistle matching is often used to directly address specific individuals: when one dolphin copies another's signature whistle, it is likely engaging in directed communication (King & Janik 2013). This behaviour has been observed in both captive and wild dolphins and appears to facilitate social recognition, and to strengthen social bonds between individuals. However, in some cases, vocal mimicry can also function in competitive or agonistic interactions, depending on the social context (Janik 2000b). Unlike aggressive signalling in some species, whistle matching in dolphins does not appear to elicit aggression but rather encourages further vocal exchanges. In the context of foraging, whistle matching may serve to enhance coordination between individuals, particularly in large groups. For example, King & Janik (2015) found that whistle matching occurred in close association with food-related bray calls, suggesting a role in cooperative foraging (see § 4.3 below). This parallels findings in other socially complex species, such as chimpanzees, where vocal interactions facilitate coordination (Fedurek *et al.* 2013). Whistle matching often follows a structured temporal pattern. Dolphins do not immediately overlap whistles but instead respond within a short delay, similar to human conversational turn-taking (King *et al.* 2014).

Signature whistles require advanced learning and vocal imitation abilities (see chapter 3 above), which manifest from an early age. The development of an individual's signature whistle is completed within the first year of life: young dolphins initially produce shaky and uncertain whistles, which gradually stabilise into their distinctive personal pattern and then remain unchanged or at least recognizable for over ten years. This confirms their role in identification, requiring the associated whistle to remain consistent (Janik *et al.* 2006). Other studies have shown that during the process of "creating" their personal signature, female calves tend to differentiate their whistle more from their mother's than male calves do. This has led to the idea that dolphins learn their signature whistles based on the various vocalisations they are exposed to, drawing from other community members rather than exclusively from their mother (Fripp *et al.* 2005).[139] It appears that females primarily adopt elements from unrelated group members, whereas male whistles, while more similar to their mother's, are not direct imitations but rather reinterpretations using the mother's signature as a model (Fripp *et al.* 2005). Considering that males derive their signature from their mother, it is understandable that females – who will later provide the model for their

[139] Miksis *et al.* (2002) claim to have discovered that dolphins in captivity are capable of incorporating artificially created human-made sounds into their signature whistles when exposed to them in experiments.

offspring's signature – develop a more distinctive signature of their own. In both cases, by copying and modifying the signals from their acoustic environment, dolphins reach such an original result that their signatures can be considered novel signals (Janik & Sayigh 2013, Janik *et al.* 2006). Furthermore, "during the first months of the calf's life, the mother apparently whistles frequently to produce a whistle model which the calf can imitate; these whistles were frequently followed by the newborn's whistle" (La Manna *et al.* 2019: 120), thus resembling a form of nonhuman "motherese". It is also confirmed that dolphins belonging to large groups develop more complex signature whistles, while dolphins from small groups (or captive ones), being exposed to less stimuli, develop acoustically poor signatures (Janik & Sayigh 2013). All these features highlight that learning and vocal imitation are fundamental processes for dolphins, who develop their vocal abilities and signatures through social interaction with conspecifics during early life. Additionally, it is interesting to note that while in many other species vocal learning is linked to aggressive competition (such as for mating or territory), in dolphins – where both males and females produce the same types of vocalisations without distinction – this process is instead associated with peaceful interaction and the formation and reinforcement of social bonds within the group (King *et al.* 2013). However, the reasons that lead an individual to develop its specific distinctive sound rather than another remain unknown.

The phenomenon of signature whistles is almost unique in the animal kingdom. Except for humans, similar features have been observed only in sperm whale codas (see § 2.3) and, according to a recent study by Pardo *et al.* (2024), in African elephants (*Loxodonta africana*). For their peculiarities, signature whistles are considered the best way to explore referentiality, concepts and mental representations in nonhuman animals (Bruck *et al.* 2022). This does not mean that dolphins are the only animals with these properties, but that they are the easiest to study from this perspective (Bruck *et al.* 2022). As we have seen in Part I (§ 4.4) referential communication, i.e. the ability to use signals to denote specific entities or concepts, is considered a foundational element of human language. Experimental studies have confirmed that dolphin signature whistles operate as arbitrary, learned signals rather than simple by-product cues. Playback experiments have shown that dolphins respond specifically to the computer-generated versions of their own signature whistles even when all voice characteristics are removed, indicating recognition of the whistle pattern itself rather than the voice of the producer (Janik *et al.* 2006). This ability closely parallels name recognition in humans, wherein a label refers to an individual regardless of the speaker's voice (Bruck 2013). Moreover, dolphins use these whistles in a manner that suggests referential intent. As we have seen, in the wild, signature whistles are exchanged in social

interactions, with individuals vocally addressing others by copying their whistles (King *et al.* 2013, Janik & Sayigh 2013). Such usage implies that dolphins are not merely producing their own identifier but are capable of employing these whistles as a means of directing social signals toward specific individuals. This contrasts with the more context-driven calls observed in many other nonhuman animals, where vocalisations primarily convey emotional states or environmental cues rather than referring to particular entities. Further evidence for referentiality in dolphin signature whistles comes from studies demonstrating their use in absentia. Dolphins have been observed producing the signature whistles of close social partners when those individuals are not present, a behaviour that suggests an ability to invoke a mental representation of the referent (King & Janik 2013). This capacity aligns with the linguistic property of displaced reference, which allows communication about entities beyond the immediate sensory environment (see Bickerton 2014 and Part I, § 3). The long-term retention of signature whistles – where dolphins recognise and respond to the calls of past associates even after decades of separation – further supports the hypothesis that these signals function as learned labels stored within long-term memory (Bruck 2013). While referentiality has been extensively debated in animal communication research, the evidence from dolphin signature whistles provides one of the strongest cases for its existence outside of human language. Unlike the contextually bound vocalisations observed in primate alarm calls (see Part II, § 3.4), dolphin signature whistles exhibit a learned, arbitrary association between signal and referent, fulfilling a key criterion for referential communication.

An experiment by Bruck *et al.* (2022) further supports these conclusions, providing empirical evidence that dolphins form cross-modal representations of conspecific identities by integrating signature whistles with gustatory information from urine samples. The results of the study indicate that dolphins possess mental representations of absent individuals: the captive dolphins involved in the research were able to recognise familiar conspecifics via taste alone, spending significantly more time sampling urine from known individuals than from unfamiliar ones. Furthermore, when urine samples were paired with acoustic playbacks of signature whistles, dolphins exhibited longer exploration times when the whistle matched the urine's owner, suggesting an expectation of correspondence between the two sensory modalities. This provides strong evidence that dolphins perceive signature whistles not merely as associated stimuli but as labels embedded within a structured mental model of social relationships. These findings contribute to a broader understanding of how nonhuman animals might internally represent social agents. The ability to form cross-modal representations and integrate identity-related information across sensory channels suggests that dolphins construct and store social concepts independently of immediate sensory input.

This capacity is critical for navigating their complex, fission-fusion social structures, where maintaining contact with dispersed individuals is essential for alliance formation and cooperative behaviours.

It should be noted again that signature whistles are not the only use of this type of sound, and they constitute only part of the whistle repertoire. For example, Cook *et al.* (2004), recording dolphins in Sarasota Bay, Florida, found that 52% of the whistles produced were signatures, revealing that nearly half of the vocal production was unrelated to them.[140] There are therefore other types of whistles, in this case shared among individuals, used in various contexts, suggesting that the identification and cohesion functions are only one aspect of the communication system conveyed through this mechanism. Moreover, comparative analyses of the bottlenose dolphin whistle repertoire have revealed that, despite sound differences, there are also similarities among individuals of the same population, who share unique traits absent in other groups (Berta *et al.* 2006: 295).

A complex and understudied aspect of dolphin whistle repertoire are multi-looped whistles. These whistles, characterised by introductory and/or final repeated elements, arranged in different orders, are thought to play a role in social communication, potentially encoding additional information beyond that of simpler tonal signals (Janik & Sayigh 2013). Their structural complexity suggests advanced vocal control and cognitive processing, further reinforcing the status of dolphins as sophisticated acoustic communicators. Research indicates that multi-looped whistles are often associated with social interactions, particularly in contexts involving group cohesion, conflict resolution, and coordination of movement (Janik & Sayigh 2013). The repetition of loops may serve to reinforce identity cues or facilitate recognition over greater distances, particularly in noisy environments where redundancy improves signal transmission (Quick & Janik 2012). Such a feature aligns with broader theories on signal evolution, where increased structural complexity may enhance communicative efficacy in dynamic social settings. Playback experiments have demonstrated that dolphins are highly responsive to the fine-scale acoustic properties of whistles, including loop structure variations, suggesting that these elements may carry individual or contextual significance (Janik *et al.* 2006). Furthermore, the ability to produce and modify multi-looped whistles highlights the flexibility of dolphin vocal learning, a trait shared only with a few other highly social and intelligent species, such as certain birds and primates (Tyack 2000, see also Part II). Understanding the functional significance of multi-

[140] In this case, signature whistles differ from human names, as vocatives certainly do not have such a strong presence in our language. However, it is understandable that in the fission-fusion societies of dolphins – where groups within the same community form and dissolve multiple times a day – individuals need to continuously remind others of their identity.

looped whistles requires further investigation, particularly regarding their potential role in referential communication. The presence of non-random repetition in vocal sequences suggests that dolphins may encode specific meanings or enhance the salience of certain messages (King & Janik 2013). Future studies should focus on the contextual analysis of multi-looped whistle usage in different social scenarios to determine whether these signals represent a transitional stage between simple calls and more elaborate referential communication.

Many other aspects of signature whistles (and whistles in general) should be further explored. Apart from the structure of multi-looped whistles, the process of formation of the signatures by the calves might contain insights about combinatorial operations. It would be interesting to verify whether the simplified models proposed by the mothers are "chunks" of their signatures, or "full" signatures deprived of some acoustic features. Similarly, it should be studied how the calves manage to obtain their complex signatures, whether from composing it by adding different segments (thus resembling a form of combinatoriality), or by refining some acoustic parameters. Another interesting analysis might be that of the signature whistles of males involved in stable alliances: it has been observed that their signatures change until they become similar, but they keep some distinctive individual features (Janik & Sayigh 2013). What kind of features are these? Do they consist in a part of their original signature which is maintained (e.g., both dolphins maintain a part of their original signature whistle, and develop a shared part)? Finally, a study by Jones *et al.* (2022) found that although the frequency contour of the signature whistle remains stable, the amplitude contour changes, possibly conveying additional information. Could this combination of identity meaning (represented by the frequency contour) and additional meaning (represented by the amplitude contour) be considered a peculiar form of compositionality?

These questions have no answer yet, but they represent promising pathways to explore the possibility of linguistic properties and 1-merge in signature whistles. As of today, it cannot be affirmed that these vocalisations contain a merge-like operation. However, some considerations can be made. First, whistles are series of clicks and not continuous sounds, and thus can only be consequence of (1) the combination or (2) the linear sequencing of discrete units; in any case the calls are stereotyped and thus the discrete units are ordered. Second, signature whistles seem to be related to other vocalisations: for example, in the experiments by Herman *et al.* (1984), the "commands" directed to the dolphins were preceded by the reproduction of the signature whistle of the individual who had to comply to them. This might be seen as a form of compositionality, i.e. ID signal + command. It

should be verified whether similar phenomena are observed in the wild, in which dolphins exchange signature whistles before starting interactions, but it is not clear whether they are used to "introduce" more complex sequences in a way similar to what emerged in Herman's experiments. Leaving these questions open, I now focus on another interesting type of vocalisation, which might help us shed light on the combinatorial potential of bottlenose dolphin communication: bray calls.

4.2 Bray calls

Bray calls are "sequences of multi-unit rhythmic signals" (Pace *et al.* 2022: 2), whose exact function is still unclear. First described by Santos *et al.* (1995), they have been identified in other populations since then (Portugal, Gulf of Mexico, Scotland, Namibia, Mediterranean Sea, see Pace *et al*. 2022). Scholars gave different descriptions of bray calls, both because of the lack of an agreement in sound classification and because these calls change among populations (e.g., Scottish and Mediterranean bray calls are different). Hence, they show a high degree of variation and can be composed by a diverse number and type of discrete units. While some of these units are shared among populations, others are variants of shared units, and others have been recorded only in certain groups (Pace *et al.* 2022). According to Pace *et al.* (2022: 15), in previous studies "this variability in the composition of the sequences was ascribed to possible different meanings of the units, suggesting the existence of diversity, combination, repetition, time, and order rules".

Although they are still poorly understood, they have been associated with food and social interaction. In particular, scholars believe that they are food calls because they are mainly produced in presence of prey and attract conspecifics (Janik 2000a, Pace *et al.* 2022). According to Janik (2000a), they might have a social function, or they could be a by-product of prey manipulation. Regarding their social function, they could serve to specifically attract closely related individuals; however, being dolphins a highly social species, it is unlikely that these calls would attract only kin members. Another possibility is thus that they attract conspecifics in general: this might represent an advantage in catching preys, because a higher number of dolphins could make the escape of fish more difficult. A third hypothesis is that they serve as cohesion and enhance the social prestige of the caller, who communicates the presence of food to the whole group; but dolphin social dynamics are not clear enough to posit a similar function. Finally, these burst pulsed signals might be used to modify the behaviour of the prey, stunning or disorienting the fish, thus working as a "sound bullet". In this case, the attraction of conspecifics would not be intentional, but other dolphins would

opportunistically reach the source of the sound to exploit the foraging opportunity. Janik (2000a), considering that bray call structure does not seem to have evolved for long-distance transmission, believes that prey manipulation is their main function. However, independently of this, there is a clear correlation between bray calls and the presence of fish. Hence, at least from the perspective of comprehension, the dolphins who hear a bray call likely associate it to fish or foraging, even if these elements are not visible. In this sense, it could be considered a form of displaced reference, because dolphins seem to understand that there is food only by hearing the call, even if the referent of the call (food) is outside of their context.

The variation of bray calls becomes evident in the comparison of different populations. The brays studied by Janik (2000a) were specifically associated with the presence of salmonids, and composed by two discrete units disposed in a rigid order, defined "a long multiband part" (called "*buzz*") and "a short downsweep" (called "*gulp*") (Janik 2000a: 925).

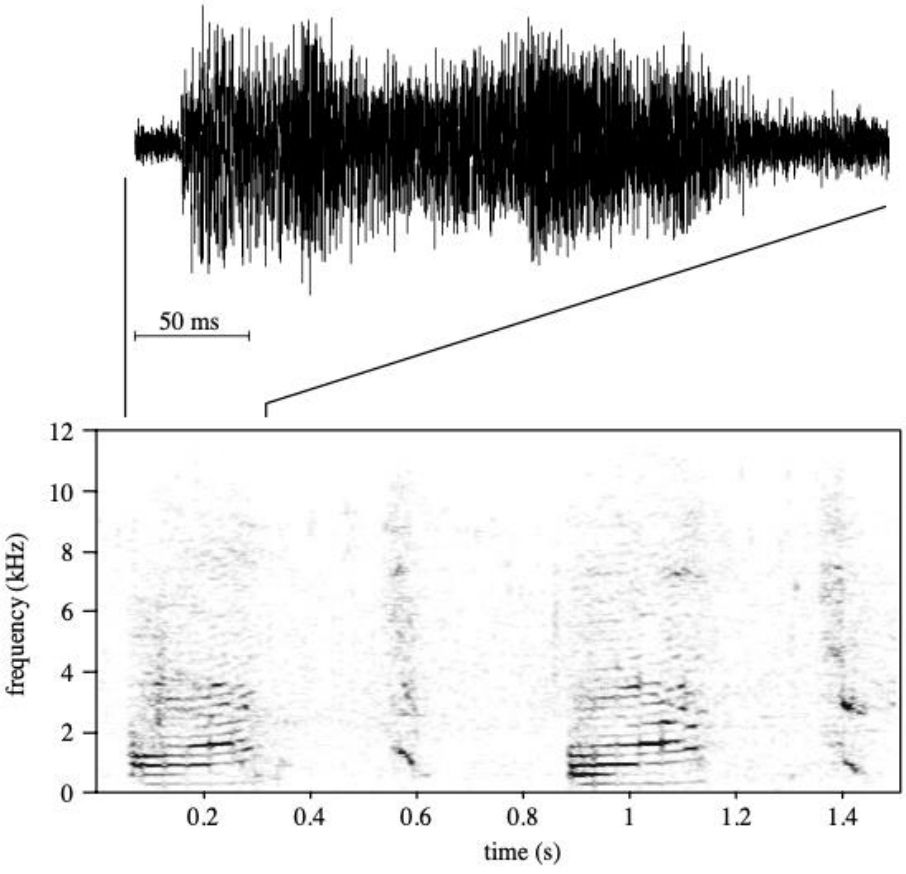

Fig. 33. Spectrogram of a bray call[141] recorded by Janik (2000a: 925) in Scottish waters. The two different discrete sounds can be seen in the lower part of the image, repeated two times.

[141] To listen to a bray call recorded in Scotland, see https://dosits.org/galleries/audio-gallery/marine-mammals/toothed-whales/bottlenose-dolphin/?vimeography_gallery=20&vimeography_video=282045101. The name "bray" is inspired by the resemblance with the sound produced by donkeys.

However, Pace *et al.* (2022) describe more complex bray calls. In their study, they provide a detailed structural analysis of bray-call sequences recorded in two Mediterranean regions: the Tyrrhenian Sea and the Strait of Sicily. They found three main types of units: "gulps", "grunts" and "squeaks", corresponding to some of the units identified in previous studies of other populations (cf. Janik 2000a, Luís *et al.* 2019). However, they also found non-bray sounds, defined "pops", "creaks", and "LFN": according to the scholars, dolphins might have added these signals to increase sequence complexity, and thus enhancing the potential for conveying more information. Furthermore, Pace *et al.* (2022) verified that the bray-call sequences identified conformed to five paradigms of information encoding, according to Kershenbaum *et al.* (2016): repetition, diversity, combination, ordering, and timing. This means that the elements can be repeated, that they are different, that they can be combined, have a non-random order and that they are arranged in precise time dispositions. The recordings revealed 13 distinct sequence typologies, as shown in figure 34 by Pace *et al.* (2022: 11).

Fig. 34. Sequences identified by Pace *et al.* (2022: 11). Each coloured box corresponds to a discrete unit. 4 variants of "gulp" and 2 variants of "grunt" can be seen. "Pop", "creak" and "LFN" are non-bray sounds, possibly added to increase sequence complexity.

Among the sequences, Pace *et al.* (2022: 14) identified "seven typologies characterized by a higher level of diversity, timing, and combination of acoustic elements (Sequences 5, 6, 7, 8, 11, 12, and 13) and four typologies characterized by a specific repetition order of pairs and triads (Sequences 2, 3, 4, and 9)." Furthermore, two sequences "are composed by the

repetition of a single acoustic element (GU1 in sequence 1 and SQ in sequence 10)". However, as Pace *et al.* (2022) observe, also the repetition of the same element is a common feature in the animal kingdom and it is used to convey information. Regarding the discrete units, the majority of sequences (80%) were composed of the "gulp" element, with its GU1 variant being present in all multi-element sequences. Further acoustic analysis revealed four distinct variants of the "gulp" element and two variants of the "grunt" element, highlighting the structural complexity of these vocalisations. Moreover, the acoustic parameters of bray-call sequences showed significant geographical variation, with lower minimum and maximum frequencies and shorter durations recorded in the Sicilian population compared to the Tyrrhenian population. A first-order Markov-chain model was employed to examine the transition probabilities between acoustic elements within sequences. The results demonstrated that transitions between elements were non-uniform, with certain combinations occurring at significantly higher probabilities than others. Specifically, "gulp" elements were most likely to be followed by another "gulp" (95% probability), while "gulp" to "grunt" transitions exhibited an 83% probability. Additionally, the model indicated that the variants GU2, GU3, and GU4 were exclusively observed in the Tyrrhenian Sea sequences, further reinforcing the regional specificity of these vocalisations. These results led Pace *et al.* (2022: 8) to observe that "overall, these results seem to suggest that Mediterranean common bottlenose dolphin sub-populations may compose sequence using several different acoustic elements and arrange them in a very flexible and complex way to produce distinct sequence types."

It should be noted that there were differences among the two Mediterranean populations: all the 13 sequences represented in Fig. 34 were present in the Tyrrhenian Sea, whereas only the sequences 1 and 10 were observed in the Strait of Sicily. These findings indicate a low degree of similarity between the sequence repertoires of the two studied populations. Notably, the Tyrrhenian population exhibited a more diversified and complex bray call repertoire compared to its Sicilian counterpart: only the two simplest sequences, composed by a single element, were detected in the latter group. Interestingly, there is a crucial difference between the two populations: the dolphins of the Tyrrhenian Sea have a "high level of site fidelity" (Pace *et al.* 2022: 15): they have been observed in the area for several years, and they can be considered resident of that region. By contrast, most dolphins of the Strait of Sicily are transient, i.e. they do not stably populate the area, and their permanence is only temporary. According to Pace et al. (2022: 15), "although common bottlenose dolphins are characterized by a fission-fusion social structure, they can establish long-term social relationships when inhabiting a specific area permanently. This

phenomenon […] may explain the development of local communicative acoustic signals used to improve the group cohesion and facilitate individuals' coordination". In other words, sequence complexity might be related to social complexity: the resident population has more contacts and stronger bonds and interactions with conspecifics, and thus a greater need of coordination and communication. Hence, the animals might have used their cerebral plasticity to compose more complex bray call sequences by combining different types of sounds, in order to convey more information. The transient population, on the other hand, would not have the need nor the time to develop such an articulated system. It is emblematic that the two bray call sequences produced by transient dolphins are composed by a single element, making it difficult to support the idea of complex combinations of meaningless, let alone meaningful, units. This perspective resembles the social complexity hypothesis by Freeberg *et al.* (2012) discussed in Part I, according to which groups with more complex social interactions develop more complex communicative systems to regulate them. If this should indeed be the case, we would have an effective example of two dolphin groups that, starting from the same cognitive and vocal skills, diverge in their acoustic production due to reasons related to external (and not genetically transmitted) dynamics. The findings of this study underscore the structural intricacy of bray call sequences and their potential functional significance within dolphin communication. The presence of unique sequence types and element variants in the Tyrrhenian Sea suggests that local environmental or social factors may influence the evolution of these vocalisations. Futhermore, "it is important to note that "sequences of sequences" have not been examined in the present study, thus more complex levels of combinations are yet to be explored." (Pace *et al.* 2022: 15). This possibility, along with other issues related to merge and compositionality in dolphins, is further discussed in the next section.

4.3 Possible analyses

The approach based on sub-units and units discussed in sections 2.1 and 2.2 can also be adopted for bottlenose dolphin vocalisations, following the same principles described in chapter 2 for humpback whales and orcas (see also Pace *et al.* 2010 and Shapiro *et al*. 2010).

Bray calls can be analysed from the perspective of the **sub-units** that compose them ("gulp", "grunt", etc., see above). In this case, they might be considered conceptually comparable to human phonemes, or to human morphemes. In the case of the comparison with phonemes, we shall assume that the sub-units constitute a finite repertoire of meaningless units that are arranged in different configurations to create meaningful calls.

From a structural perspective, this might be plausible because bray call sequences have a rigid order (e.g., always "buzz" – "gulp" in the Scottish ones), and they probably have a meaning connected with food, when considered as a whole. Furthermore, the sequences described by Pace *et al.* (2022) are more complex, but still stereotyped, and present some restrictions. Figure 35 isolates the most interesting sequences identified by Pace *et al*. (2022). Note that, for this analysis, the individual sub-units should be considered meaningless, and the sequences composed of them should be assumed to have different meanings or functions (e.g., in figure 35, sequence 2 should have a different meaning/function than sequence 3, etc.).

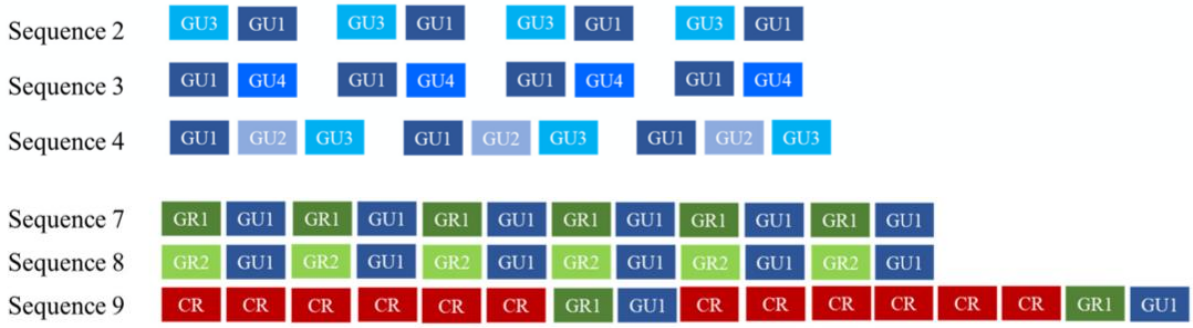

Fig. 35. Selection of the most relevant sequences for a structural analysis (adapted from Pace *et al.* 2022: 11).

As shown in figure 35, variants of the most common "gulp" sound (GU1) always occur when another "gulp" is used (i.e., in sequences 2, 3, 4,), while the latter is not always the case (in sequences 7-9 GU1 occurs without other types of "gulp"). Furthermore, sequence 7 and 8 are differentiated only by a different "grunt" variant (GR1 or GR2), resembling the linguistic concept of minimal pair, in which the difference of a single element in two words determines a different meaning (e.g., / mæd / vs / bæd /). In "human" linguistics, this contrast is used to identify phonemes (e.g., *m* vs *b* in the previous example). It should also be noted that GR elements are always followed by GU elements. This suggests that this combination may represent a "deeper" structure, with GR and GU being more strictly connected to each other than to the other units (showing a parallel with the "buzz" – "gulp" structure of the dolphins from Scottish waters). Finally, two sequences (4 and 9) are composed by three elements: in these cases, they might be represented with a cascade effect, similar to the structure of the syllable (see Rizzi 2016), in which two elements (nucleus and coda) are combined, and then the result of this operation is combined with a new item (onset), without further expanding the structure (see figure 36 below).

By contrast, to allow an analysis in which bray call sub-units are compared to morphemes, we should assume that the sub-units have a meaning *per* se, and thus these bray calls sequences would be obtained through the composition of different meanings. In this case, sequences 4 and 9, composed by three elements, would represent a higher level of complexity than the binary combinations mostly observed in nonhuman animals. The binary sequences (like sequence 7 or 8) could be represented as an instance of 1-merge combinations, and the sequences composed by three elements might be considered a cascade effect, as observed by Rizzi (2016, see also Part I, § 2.5).[142] In particular, if we analyse sequence 7 as the combination of the meaningful units GR1 and GU1, then sequence 9 could be considered a further (non-recursive) application of 1-merge, which adds to the constituent the element CR. Figure 36 gives an example of this reasoning.

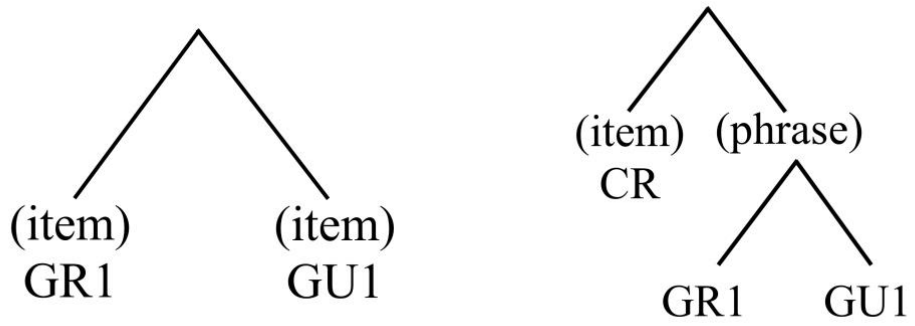

Fig. 36. Tentative graphical representation of sequences 7 and 9 following Rizzi (2016). Assuming that the sub-unit are meaningful, sequence 7 (at the left) could be considered a case of 1-merge, and sequence 9 (at the right) a non-recursive cascade effect. Assuming that the sub-units are meaningless, sequence 9 may have a structure similar to the composition of the human syllable. Note that in sequence 9 the CR unit is repeated six times, as shown in figure 35. Here it is reported only once to simplify the scheme.

However, at the moment this approach is only a speculation. Before proposing the presence of a combinatorial operation like 1-merge, it should be verified whether the single sub-units have indeed a meaning. For example, if "GR1" is used as a separate, independent call, and the same happens to "GU1", then it is likely that they have a specific meaning or function; in this case, their combination might produce a more complex meaning, and thus be a case of compositionality comparable to those observed in other nonhuman animals (see Part II). Nonetheless, if these sub-units do not appear in isolation, but are present in other vocalisations (e.g., in context specific calls), then it would be more probable that they have

[142] In this case, although having the same structure of the syllable, it would not be totally comparable to that because the combined units would be meaningful instead of meaningless.

no semantic meaning *per se*: the only meaningful signal would be the complete "GR1-GU1" call, and they should be considered a case of combinatoriality. According to the data from Pace *et al.* (2022), it is evident that single sub-units occur in more sequences, and thus it seems more likely that they are a case of combinatoriality: the sub-units would be meaningless, but the stereotyped sequences obtained by combining them might have different meanings (related to food or social cohesion). Furthermore, Pace *et al.* (2022) compared the Mediterranean sub-units with those of other populations, finding some similarities which suggest that at least some elements are shared and used in different dispositions to create different calls, that vary depending on the group. This corroborates the combinatorial interpretation and weakens the compositional one.

New studies might shed light on the issue, focusing on different repertoires, like those of the Scottish population, in which bray calls are always composed by the sub-units "buzz" and "gulp", in rigid order. In this case, it should be verified whether the sub-units "buzz" and "gulp" 1) also appear in isolation, thus suggesting that they have a meaning, and that the "buzz-gulp" sequence is obtained by a compositional operation like 1-merge, and 2) are present in other vocalisations, leading to the combinatorial interpretation described above. To this end, an effective approach would be to calculate the transition probabilities between the different sub-units produced, i.e. to verify how probable it is that two elements co-occur, identifying the co-occurrences which happen with a frequency above chance level. This can be obtained by performing a collocation analysis to a dataset with raw recordings of bottlenose dolphin vocalisation (including bray calls, whistles, and any other sound type recorded in the session). This method, generally used in computational linguistics to measure the relative attraction between two units and how exclusively they associate with each other (e.g. in English, how *going* associates with *to*), has been adapted to the study of call combinations in chimpanzees (Bosshard *et al.* 2022, Leroux *et al.* 2021, Leroux *et al.* 2022), and it could be adapted to bottlenose dolphin sub-units as well. This would help to identify the frequency and ordering of non-random combinations between sub-units, which might be governed by rules. Until a similar analysis is made, it would be premature to claim or refute the presence of compositionality in bray calls.

A similar approach could be reserved to the level of **units**, i.e. the whole calls which compose the dolphin repertoire (whistles, brays, etc.). From this perspective, the calls should be considered meaningful units conceptually comparable to morphemes and words, and their combination would be a case of compositionality. The same type of collocation analysis could be thus applied to the call level, verifying whether there are non-random co-occurrences of whole calls (e.g., whistle + bray call), which could be compared with the

cases discussed in Part II in nonhuman primates and birds. As of today, only King & Janik (2015) explored this phenomenon in bottlenose dolphins. In their study, they analysed in bottlenose dolphins of Scottish waters the co-occurrence of whistles and food-related bray calls in foraging contexts, and found a significant correlation between them. The authors aimed to determine whether specific social signals, particularly whistle matching, were systematically associated with food calls, and whether these combinations followed a structured sequential order. Whistle matching is a phenomenon in which two dolphins produced identical frequency-modulated whistles in overlapping succession (e.g., one dolphin produces a whistle and the other one replies with the same whistle, until they overlap producing the whistle together, see also chapter 3 above). Using hydrophone arrays, the researchers identified 81 braying events and examined their temporal association with overlapping whistle matching events, defined as instances where two dolphins produced identical frequency-modulated whistles in overlapping succession. Overlapping but non-matching whistles were also recorded for comparison. Statistical analyses revealed a significant correlation between bray calls and whistle matching, with 73% of overlap matching events occurring within one minute of a bray call. This was significantly higher than expected by chance and contrasted with non-matching overlaps, which occurred within one minute of a bray call only 49% of the time. Permutation tests further confirmed that the proximity between bray calls and overlap matching events was significantly smaller than expected under a random distribution, indicating a non-random association between these vocalisations. Regarding the order of correlations, the study found no evidence of a fixed sequential structure in which one call type consistently preceded the other. While bray calls and whistle matches co-occurred with high frequency, whistle matches did not consistently occur before or after bray calls in a predictable pattern. This suggests that their association is functional rather than strictly hierarchical. One possibility is that dolphins use whistle matching to establish social coordination before or during the emission of food calls, potentially facilitating cooperative foraging strategies. Interestingly, the majority (60%) of whistle matching events involved variations of a single whistle pattern, with another type accounting for 20% of cases. This suggests that specific whistle types may play a role in foraging coordination, though the study was unable to determine whether these were individually distinctive signature whistles or shared context-specific calls used by multiple individuals. However, the researchers are keen to believe in the second option, because they recorded fewer whistles than dolphins (and each dolphin has a unique signature whistle). In any case, these sequences could be interpreted as an unordered combination of meaningful units: context-specific call + food call, or ID + food call. This would be similar to the pant-

hoot + food call combinations studied in chimpanzees (Leroux *et al*. 2021), and might be considered a case of 1-merge. However, the correlation between whistles and bray calls does not necessarily imply a compositional interpretation. The results only suggest that bottlenose dolphins integrate social and food-related communication during foraging. The co-occurrence of bray calls and whistle matching thus highlights a social dimension to food calling that extends beyond prey manipulation or simple recruitment, but further studies are necessary to better understand the issue.

The main problems to solve to shed light on the matter are those of caller identification, and observation of behaviour to assign a function to each call. As we have seen, in marine mammals this is a complex task to solve. Apart from the studies on captive animals, a solution is to use data of tagged animals. Other "creative" solutions have been adopted: for example, Stephanie King, when studying the dolphins of the shallow Australian waters, performs playback experiments from a boat, and then videotapes the animals from above with a drone. Being the water transparent, the reactions of the dolphins can be observed. However, many limitations remain, so that statistical analyses are probably the best way to identify patterns and rules in cetacean vocalisations.

All the phenomena described in this chapter do not clarify whether an operation like non-recursive merge is present in bottlenose dolphins. Nonetheless, they highlight several properties related to language, that are summarised in the next chapter. Before delving into a more general analysis of linguistic properties, I focus on the compositional and non-compositional interpretations proposed by Schlenker *et al.* (2016, 2023) and discussed in Part II (§ 5) for nonhuman primates and birds, to verify how they apply to bottlenose dolphins. Recall that the compositional interpretation involves the combination of two meaningful units, which create a more complex structure with a more complex meaning, derivable from its parts. The non-compositional interpretations propose that the "combination" of meaningful units is not a combination at all, and it can be just *one-word,* only acoustically complex, or the linear sequence of two *separate utterances*. Regarding the vocalisations described in this chapter, and observed in the wild, **signature whistles** seem to be perceived as single calls, with the main function of conveying the identity of their owners (see § 4.1). Although they are composed by smaller units and are stereotyped – so that the elements that compose them have a precise order – it seems more likely that they are *one-word*. They are hardly *separate utterances*, because they cannot be split into two parts that are used independently, in different contexts. Furthermore, each dolphin has its

own unique signature, so they are not shared signals.[143] It should be kept in mind that there are properties that might suggest a compositional operation, as discussed in section 4.1: the multi-looped whistles are composed by segments that can be arranged in different orders, and researchers believe that they might contain additional information (Janik & Sayigh 2013). A similar suggestion was proposed by Jones *et al.* (2022) regarding the amplitude contour of the whistles, which is variable and might convey different meanings, while the stable frequency contour conveys information about identity, so that we would have a combination of identity (frequency contour) and other information (amplitude contour). This would not be a composition of meaningful units in the traditional sense, because there are no separate elements, but only two acoustic parameters whose modifications change the overall meaning. Although these phenomena might be considered compositional, and the perspective of 1-merge cannot be excluded (at least in the comprehension), it seems more likely that the whistles, although complex, are the consequence of combinatorial operations (and not compositional ones).

Regarding **bray calls**, the issue appears more difficult to clarify: as discussed in section 4.2, the order of the discrete units plays a crucial role in the stereotyped sequences, but they could be considered the consequence of 1-merge operations only if the elements that compose them are meaningful, and this has to be proven yet. In my opinion, the most cautious position is to consider them *one-word*, i.e. complex combinations of meaningless discrete units that create a meaningful food-related call. However, the *separate utterances* and the compositional hypotheses cannot be discarded. Furthermore, even if we consider them one-word, and thus a case of combinatoriality, there might be combinations of more bray calls, as pointed out by Pace *et al.* (2022), which could suggest compositionality. The analyses proposed at the beginning of this section would help shed light on this matter.

Another possible, more complex combination is that of whole calls. In this case I can only refer to the study by King & Janik (2015) described above. According to these scholars, there is a correlation between **bray calls and whistles**, but no order, because whistle could precede or follow the bray calls. In this case, the two elements cannot be considered *one-word*, because they are clearly independent calls. However, it is not clear if they should be analysed as *separate utterances* or with a compositional (1-merge like) interpretation. The fact that there is no word order suggests that this feature does not affect the overall meaning; however, the researchers did not directly observe the behaviour of the dolphins, so that it is

[143] However, further analyses should be directed towards non-signature whistles, which appear to be context-specific shared calls. That said, it looks improbable that these whistles are composed by meaningful units, because no "half-whistles" have been recorded. A more promising path would be to analyse the possible combinations of "complete" whistles.

unknown whether a different order determined a different reaction. In any case, despite the potential similarity with call combinations described in chimpanzees (see Leroux *et al.* 2022), the presence of 1-merge cannot be postulated nor refuted, due to the well-known limitations of the studies on marine animals.

Turning to captive individuals, in this case the experiments by Herman *et al.* (1984, see § 3.3) support the idea that bottlenose dolphins are at least capable of comprehending combinations of meaningful units. In the case of the already discussed **locative + object** combination, e.g. *left frisbee*, neither of the non-compositional hypotheses seems convincing: *left frisbee* cannot be *one-word* because both *left* and *frisbee* were taught by researchers as different meaningful signals. Moreover, it seems improbable that they were interpreted by dolphins as *separate utterances* because the locative (in this case, *left*) was an element which was not used in isolation, and needed an object to modify (in this case, *frisbee*, resembling the *-oo* element of Campbell's monkeys, which behaves as an affix). Here, the compositional interpretation appears the most consistent with the fact that dolphins seemed to interpret *left frisbee* as a whole block, similarly to the constituents of human language, recognising it even when inserted in more complex sequences (e.g., *left frisbee basket fetch*). Moreover, *left* is combined also with other elements (e.g., *basket*): this further shows that it is understood as a separate word.[144] Regarding the **entire sequences** reproduced by Herman *et al.* (1984), like *left frisbee basket fetch* above ("bring the basket to the frisbee at your left"), they were composed by three "constituents": destination object (here *left frisbee*), transport object (here *basket*) and relational verb (here *fetch*). In these sequences word order has been proven to be fundamental to the understanding, to the extent that a different disposition of the same elements conveyed a different message (e.g. *frisbee left basket fetch*, "bring the basket at your left to the frisbee"). The *one-word* hypothesis cannot be applied, because all the signals are separate meaningful units, but the interpretation of *separate utterances* cannot be excluded, if we consider that the dolphins might have interpreted the sequence linearly. [145] However, this does not seem plausible because the artificial language was structured in a way that compromised the linear execution of the commands: the first item was the destination object, the second was the object to be transported, and the verb was the last one. For example, in *frisbee basket fetch*, a linear interpretation would imply that the dolphin first focused on the frisbee, then on the basket, and finally on the verb *fetch*,

[144] It might even be argued that *left frisbee* resembles a predicative structure, somehow similar to a reduced relative clause (left frisbee → the frisbee that stays at the left).
[145] This reasoning applies to sequences composed by "single" signals, such as *basket frisbee fetch*. Nonetheless, in cases like *left* + X (e.g. *left frisbee*, *left basket*), the *separate utterances* interpretation does not seem convincing.

but the linear execution in this case would have been impossible. A linear interpretation would make more sense if the structure was *basket fetch frisbee*, reflecting the temporal steps of the task (first select an object, then perform the action which connects it with the other object). As proposed by Herman *et al.* (1984), the idea that dolphins recognised the "argument structure" and the "syntactic" rules of the language seems the most convincing interpretation. This would imply a hierarchical organisation which is based on the operation merge, so in my opinion the compositional interpretation is compatible with the results of this experiment.

After having discussed these possible interpretations focused on the topic of compositionality, I can now put together all the data about cetaceans presented in the previous chapters (2-4), highlighting the features related to language on a wider perspective. By using the same classification presented in Part I (chapters 4-5) and Part II (chapters 3-4), I summarise both the linguistic properties and the general traits important for language observed in cetaceans. This will help me clarify which features are present in these animals, and will allow me to compare them with the skills of the species discussed in Part II, and with the abilities of humans highlighted in Part I.

5. The properties of language in cetaceans

Following the systematic classification presented in Part I (§ 4), the properties of human language can be divided in *signal*, *signal learning*, *structure*, *meaning*, and *contextual usage*. I now summarise which features exist in humpback whales, killer whales, sperm whales and bottlenose dolphins.

5.1 Signal

As for humans and other nonhuman species, cetaceans exchange signals. These animals mostly rely on vocal communication, so that the features of ***vocal auditory channel***, ***broadcast transmission and directional reception***, ***rapid fading***, and ***total feedback*** proposed by Hockett (1960a, see also Part I, § 4.1) are present in all cetacean species. Visual communication (i.e., gestures) is not a common trait of cetaceans, because in the marine environment sound waves travel better than light. Nonetheless, captive dolphins have been capable to understand an artificial gestural language (Herman *et al.* 1984), and physical contact is a way to maintain social cohesion in most cetacean species (Marino *et al.* 2007). ***Linearity*** (Berruto & Cerruti 1997) is also present, since vocal signals are inevitably ordered in time (one must follow the other). However, as pointed out in Part II (§ 3.1) and in section 4 above, bottlenose dolphins exhibit the phenomenon of biphonation, in which two sounds are produced at the same time (Jones *et al*. 2020), thus representing a potential case of non-linear vocal communication. This applies also to humpback whales (Doh *et al.* 2025) and killer whales (Filatova *et al.* 2010), but not to sperm whales.

5.2 Signal learning

Cetaceans are among the most proficient animals in signal learning, due to their developed imitative skills (see § 2-3). Their ability relates both to learning discrete meaningless units (conceptually comparable to human phonemes) and meaningful units (conceptually comparable to human morphemes and words). They are capable of both ***vocal contextual learning*** (to use an existing signal with a new function) and ***vocal production learning*** (to learn a new acoustic signal absent in their repertoire) (Janik & Slater 2000). In the previous chapters I gave several examples of these abilities, both in the wild and in captivity. In the wild, for example, male humpback whales change their songs by imitating conspecifics, adopting new units (Cholewiak *et al*. 2013); orcas have different sets of stereotyped calls which change according to the group and can learn new ones (Ford 1991),

and bottlenose dolphins' vocalisations vary in different regions (e.g., Scottish vs Mediterranean bray calls, see § 4.2). Vocal production learning is particularly evident in dolphin signature whistles, which are novel signals spontaneously developed by modifying elements from their acoustic environment, which are copied by conspecifics (Janik & Sayigh 2013, see also § 4.1). Sperm whales emit only clicks, so they are not capable to produce novel sounds; however, they can develop and learn new codas (series of clicks with a stereotyped temporal arrangement), which are novel signals absent in their repertoire (see § 2.3). In captivity, the best examples come from bottlenose dolphins, which spontaneously imitate artificial sounds produced by humans (Marino *et al.* 2007). These features are rooted in ***cultural transmission*** (Hockett 1958), here intended as the ability to teach or culturally learn novel signals (not dictated by instinct): humpback whales, orcas, sperm whales and bottlenose dolphins all display this property, because social learning plays a role in most of the vocalisations described above (§ 2-4). This interpretation is also supported by the fact that many other behaviours, such as orcas' hunting techniques or dolphin tool use, seem to be voluntarily transmitted to the offspring or to selected members of the community (see § 2-4 above). Finally, a phenomenon comparable to ***learnability*** (Hockett 1966), i.e. the ability to learn a new communication system, has only been documented in wild orcas by Filatova *et al.* (2010), which describe some individuals who learned a new "dialect" (a whole repertoire of discrete units) to be accepted in a new pod (see also Part II, § 3.2). In captivity, the bottlenose dolphins studied by Herman *et al.* (1984) learned an artificial language made by whistles and gestures, although only in comprehension (see § 3.3).

5.3 Structure

Cetacean vocalisations are highly structured, as we have seen in the previous chapters. All the species described have stereotyped calls composed by discrete units arranged in specific orders, and in certain cases the same units are rearranged to compose different calls. Scholars believe that these calls are meaningful (see e.g. Marino *et al.* 2007), although more studies are needed to assign precise meanings to them. I now describe the different aspects of the structure of these vocalisations.

5.3.1 Duality of patterning

No nonhuman animal has ever shown the property of duality of patterning (Hockett 1958, Martinet 1955), and cetaceans here are not an exception. Although several species present examples of combinatoriality and (possibly) compositionality (see below), the data

are not sufficient to claim the presence of both properties. However, this might be due to limitations of the studies in marine animals. The hypothesis of duality of patterning, for example, would be convincing if dolphin bray calls (see § 4.2) were composed by meaningless discrete units (1st level of articulation), and then if they were further combined with other meaningful calls obtained by the same process (2nd level of articulation), e.g. bray call + whistle or other bray call sequences. Although this cannot be excluded, there is no proof yet of a similar behaviour.

5.3.2 Phonology

Some cetacean vocalisations have a structure that can be conceptually compared with phonology, to different degrees. The songs of humpback whales are considered a form of phonocoding, or phonological syntax (Engesser & Townsend 2019), and they have been compared to birdsong: they have a holistic meaning related to sexual selection and group cohesion, but the message is not obtained by the sum of smaller meaningful parts. Their complexity serves to enhance the same simple message and to make it more effective (Cholewiak *et al.* 2013). Humpback whales have also other call types, but they are poorly explored and it is not clear whether they are formed by smaller units. Mechanisms more similar to 'proper' phonology are those observed in killer whales and bottlenose dolphins. As discussed in section 2.2, the orca populations analysed by Shapiro *et al*. (2010) seemed to have a repertoire of discrete units which is recombined to produce several calls, and the bottlenose dolphins studied by Pace *et al.* (2022) in the Mediterranean Sea display 13 stereotyped bray call sequences composed by different arrangements of the same units (see § 4.2). Regarding dolphins, also the whistles are composed by smaller discrete units; notably, multi-looped whistles have elements which are repeated or can appear in different configurations (see § 4.1). Sperm whales, despite producing only one type of sound, arrange it in temporal sequences which create different meaningful calls: although it cannot be considered a form of phonology in the traditional sense, the process might have a conceptual resemblance with it. In any case, all the vocalisations of the species described are formed by discrete units: although the linguistic property of ***discreteness*** is not presented with this term in the cetacean literature, researchers are clear in stating that whale and dolphin vocalisations are not continuous, and that the units are distinct and distinguishable from each other (see e.g. Shapiro *et al.* 2010, Pace *et al.* 2010, Marino *et al.* 2007). Following the same reasoning, these vocalisations are obtained through a ***combinatorial ability***, i.e. the animals are capable to combine the discrete units into more complex, meaningful structures.

5.3.3 Morphosyntax

Syntax is the most debated aspect of 'language' in nonhuman animals (see Part II, § 4.3), and there is no definitive proof of its presence in cetaceans. The strongest claim has been made by Suzuki *et al.* (2006) and Kennedy (2018), which proposed that humpback whale songs have a syntactic, hierarchical structure. However, although structurally complex, these songs have no complex meaning, so that the research of syntactic properties might be more promising in odontocetes. The possibility of ***compositionality*** in cetaceans has been discussed in the previous chapters (2-4). The strongest case comes again from bottlenose dolphins. In the wild, the co-occurrences of whistles and bray calls analysed by King & Janik (2015) might be considered compositional, but more data which associate calls with behaviours should be collected before claiming this (§ 4.3). In captivity, the bottlenose dolphins Ake and Phoenix trained by Herman *et al*. (1984, see also Herman & Uyeyama 1999) were able to understand an artificial language, and to obey to the commands given to them through this system. Among other abilities, there was the capacity to understand combinations like *left frisbee* or *left basket,* inserted in a more complex sequence like *frisbee left basket fetch*, which was interpreted as *left basket* and not as *frisbee left*. The experiment has been thoroughly described in section 3.3, and it seems to highlight a cognitive compositional skill, also related to structural restrictions, because the modifier (left or right) always preceded the item it modified. These abilities appear similar to those showed by Kanzi and other trained nonhuman primates (see Part II, § 3.2), and the fact that dolphins were not able to produce this language might be due to the anatomical and environmental differences between them and humans (and other primates), and not to cognitive limitations. Unfortunately, an experiment with a language suitable to dolphin physical skills has never been designed, and attempts like the keyboard of Xitco *et al.* (1999) failed because of technical difficulties. Hence, the issue whether bottlenose dolphins and other cetaceans are capable to combine two meaningful units, thus displaying non-recursive 1-merge, remains open. ***Productivity*** (Hockett 1958) is another property not observed in cetaceans, because it is not clear whether they produce novel utterances to enlarge their repertoire. It is true that dolphin signature whistles are novel signals spontaneously created by their owners, but they are limited to one for each individual, thus showing at best a very limited productive skill. Cetaceans have a developed ability to understand and imitate novel signals, but their repertoires are confined to a modest number of calls. More studies are needed to further explore the degree of "creativity" of cetacean vocalisations, to verify whether some species

use this ability to expand their set of signals. The cerebral and behavioural plasticity of these animals, and the variation of vocalisations among populations (observed in humpback whales, killer whales, sperm whales and bottlenose dolphins) might suggest a positive answer, but cetacean calls are understudied and poorly understood, so that it is too early to propose productivity in the species described. The same reasoning applies to ***open-endedness***: if cetaceans lack productivity (or it cannot be demonstrated), then their ability to create new signals is certainly not open-ended, nor potentially infinite. Similarly, ***recursion*** is not observed in their vocalisations, although it might be present in other domains, such as visual processing and orientation, as proposed by Hauser *et al.* (2002). This could suggest that this property is not uniquely human, but only in humans it is applied to communication. Furthermore, in sequences composed of three units, like some dolphin bray calls described by Pace *et al.* (2022), there is no trace of recursive operations: even if one wants to propose a compositional interpretation, the combination can be analysed with a non-recursive 1-merge cascade effect (see § 4.3). It is thus difficult to hypothesise the presence of ***hierarchical structures***: apart from the case of humpback whale songs, which have been represented hierarchically, although with no complex meaning (see § 2.1), any other hierarchical analysis of cetacean calls (e.g. dolphin bray calls) would be arbitrary. However, the social dynamics of cetaceans reflect a hierarchical organisation (as those of other animals), suggesting that this feature, like recursion, might be present outside communication (e.g., dolphin alliances organised in three levels, where each level is embedded in the subsequent one, see § 3). The matter appears more complex in the experiments by Herman *et al.* (1984): the sequences might be interpreted in a hierarchical way, as in [frisbee [[left basket] fetch]]. Furthermore, the facts that a different disposition of the signals conveyed a different message, and that dolphins detected the anomalies (i.e., semantically impossible and syntactically incorrect sequences), might suggest that the artificial language had a ***structure dependency*** which was correctly identified by the trained specimens. In the wild, it is not clear whether the structure of the sequences alters their meaning. However, it is evident that ***word order*** (or signal order) has a crucial role in most of the stereotyped calls, both in Herman's experiments with trained dolphins and in the wild, regarding all the vocalisations analysed (see e.g. Marino *et al.* 2007). This applies to humpback whale songs, orca "dialects", sperm whale codas and bottlenose dolphin repertoires. However, the only explored combination of meaningful units, i.e. whistle + bray call observed by King & Janik (2015), had no order restrictions, thus likely shrinking the importance of order to the disposition of the meaningless units which create the meaningful calls (resembling more phonology than morphosyntax). However, more studies are needed

to further explore this topic. Similarly, there is no literature about ***non-adjacent dependencies*** in cetaceans: the experiments by Herman *et al*. (1984), which are the most consistent case of learning of an artificial language, did not test this aspect. However, from the study it emerged that dolphins were capable of ***rule generalisation***, because they tried to repair the incorrect sequences, and they performed novel combinations of signals which had never been tried before, but followed the same rules that were taught to them. A similar result was obtained also with the dolphin Elele (Herman 2012). Overall, it can be observed that – although there is no definitive proof of merge and compositionality in cetaceans – bottlenose dolphins are the species with the most interesting syntactic-like properties, which mainly emerged in captivity. This raises the question whether other cetacean species have similar abilities, which might not have been found yet due to lack of experiments (and difficulties in testing less sociable or bigger animals).

5.4 Meaning

Like other vocalisations discussed in Part II (§ 3.4), the structured calls of the cetaceans seem to have a meaning, i.e. the signals are associated with an object or concept, which elicits a reaction or triggers a mental representation (see below). The fundamental building blocks of communication are represented by the ***lexicon***, which constitutes the repertoires of the calls used by a species. As we have seen, cetaceans have different sets of calls depending on the populations, so that these repertoires diverge even in the same species (see chapters 2-4). Notably, despite being so diverse among populations, each group seems to have a restricted set of calls, limited to a small number: for example, orcas live in different pods with different vocalisations, but each pod has only 10-20 calls. Similarly, dolphin and sperm whale populations only have few calls (which rise in quantity if we also consider the signature whistles and identity codas). Apart from shared songs, humpback whales appear to have a limited set of calls as well. However, in all these animals most repertoires have never been classified, so it is not possible to give a precise estimate. The common denominator in all the cetaceans analysed is that lexicon is restricted, a trend which is observed also in the other nonhuman species classified in Part II (§ 3.4).

Although it has not been explicitly claimed in the literature, the cetaceans described in the previous chapters show the property of ***referentiality***, at least to the extent proposed by researchers for nonhuman primates and birds discussed in Part II. In captivity, bottlenose dolphins demonstrated to understand the meaningful signals taught by Herman *et al.* (1984), so this property is present at least as a latent ability. In the wild – excluding humpback whale

songs – the calls of orcas, sperm whales and dolphins seem to be associated to a meaning, although the precise function is not understood (social signals, coordination, group cohesion, etc.).[146] Bottlenose dolphins have the most convincing examples of referentiality: bray calls are associated with food (see § 4.2), and signature whistles convey the identity of the individual (see § 4.1), similarly to sperm whale identity codas (see § 2.3). These identity signals, in particular, refer to a specific entity of the world, i.e. the signature owner. Furthermore, not being related to any precise feature of the individual, they show a high degree of ***arbitrariness*** (Hockett 1958). The concept of arbitrariness can be extended to all signals which have no natural relationship with the element they refer to; hence, the variation of the repertoire among different populations of orcas, sperm whales and dolphins aligns with this definition. Among the different levels of arbitrariness, none of the cetaceans analysed shows ***iconicity***, while following Fitch (2010) all the context-specific calls – which need the context to be understood, and are not "onomatopoeic" – can be considered to have the property of ***indexicality***. Furthermore, dolphins are capable to spontaneously point and understand pointing, which is an indexical ability strictly related to language (see § 3 above). Finally, ***symbolism*** has been proposed only for sperm whale identity codas, which according to Hersh *et al.* (2022) contain a symbolic cultural reference to the clan the animals belong to. As observed in Part II (§ 3.3), animals seem to have a degree of arbitrariness which is not related to the context (thus not indexical), but less complex than human symbolism: for example, bottlenose dolphin bray calls trigger an approach to a food source even if the animals do not see the prey. This leads to the concept of ***mental representations***. According to King & Janik (2013) and Bruck *et al.* (2022) when dolphins hear a signature whistle, it triggers the mental representation of the individual who 'owns' that signature, even when the individual is out of sight (see chapter 3). Similarly, it is possible that when dolphins hear a bray call, the concept of food is triggered in their mind, or that the captive specimens studied by Herman *et al*. (1984) had mental images of the objects and actions that were represented by the artificial signals which were taught to them. I already pointed out that Bruck *et al.* (2022) believe that concepts and mental representations are spread in the animal kingdom, but dolphin signature whistles allow to study them in an unprecedented way. Hence, considering their cognitive complexity, this trait might be present also in orcas and sperm whales, although to my knowledge no study has addressed this issue. Another unexplored property is ***equivocity*** (Berruto & Cerruti 1997), which refers to the fact that a word can be associated to more than one meaning, and vice versa a meaning can be

[146] The same reasoning also applies to the other calls of humpback whales and of other cetaceans not discussed here, although they have never been studied from this perspective.

associated to more than one word. Although not described in cetaceans, they are proficient vocal learners and this ability might emerge in captive animals, if trained to associate a pre-existing vocalisation to a new function. ***Displacement*** is not openly described either; however, following the argument of arbitrariness and mental representations, there are reasons to believe that at least dolphins exhibit this property: if it is true that dolphin signature whistles indicate the presence of labelled mental representations of the individuals (Bruck *et al*. 2022), and if they use these labels to address those individuals when they are absent (King & Janik 2013), this could be considered as communicating about elements outside of the context, not displaced in time (like humans that can talk about past or future events) but in space ("talking" about an absent individual). Furthermore, Premack (1984) observes that mental representations lead to displacement, so the latter could be considered an externalization of the former. If displacement is difficult to demonstrate, even harder is to analyse ***detachment from stimuli*** (Berruto & Cerruti 1997): it is not possible to determine whether the animals produce a signal without having an internal or external need that triggers it, so I leave the question open. ***Specialisation***, however, seems to be present at least in bottlenose dolphins, because their calls are not strictly related to the context: the reaction that an individual has when it hears a food related bray call is comparable with the reaction that it would have when seeing a prey. According to Hockett (1958), this is an example of specialisation. It is not clear whether other cetaceans have this property: although it is possible, the poor understanding of their calls is not sufficient to postulate its presence. Similarly, ***prevarication*** (Hockett 1966) has never been described in cetaceans. Other properties which seem absent are ***reflexiveness*** (Hockett 1966) and ***semantic omnipotence*** (Berruto & Cerruti 1997, see also Part I, § 4.4 for a detailed description of these features).

5.5 Contextual usage

Language and communication always take place in an external context, which involves the interaction of different individuals. As cetaceans are very social species, pragmatics plays a pivotal role in their survival.

5.5.1 Intentionality

Intentionality has only been proposed for a few nonhuman primate and bird species, and only in its most simple level, first order intentionality, defined as the presence of desires, beliefs, intentions and the will to modify others' behaviour (Dennet 1983). Cetaceans might seem to fit in this definition, but no evidence of first-order intentionality is described in these

animals, although they show a high degree of social, communicative and cognitive complexity, and live in complex social groups. Furthermore, cetaceans seem to have complex behaviours that they actively transmit to the offspring (e.g. orcas' intentional stranding) and in captivity showed signals of intentionality (e.g. intentional use of imitation to solve a task in bottlenose dolphins, see Whitehead & Rendell 2015). Moreover, captive bottlenose dolphins have been found to be sensible to attentional states of human examiners, performing a task differently when they were looking or not (Davies & Garcia-Pelegrin 2023). Hence, the possibility of first-order intentionality in cetaceans should be further explored. Without claims about first-order intentionality, the more complex ***Theory of Mind*** cannot be postulated, because it requires that animals have an understanding of other minds and other individualities. This skill, described in Part I (§ 4.5), has never been proposed for cetaceans; however, animals with such a complex cognition and social relationships, which – in the case of bottlenose dolphins – are even capable to recognise themselves in a mirror, are among the best candidates to have it. ***Shared intentionality***, i.e. the will to cooperate with other "minds" to pursue a common goal, is considered to be uniquely human (Tomasello 2008). Nonetheless, like for Theory of Mind, it seems plausible that cetaceans have at least some aspects of it, having a pronounced attitude towards cooperation. Therefore, in my opinion these abilities related to intentionality cannot be ruled out.

5.5.2 Signal exchanges

The social inclination of cetaceans implies that they have frequent interactions with other individuals, which involve signal exchanges. These signals are transmitted in a way that avoids overlapping: ***turn-taking*** is observed at least in bottlenose dolphins and sperm whales (King *et al.* 2014), and presents similarities with the human turns in conversation, indicating complex social cognition, enhancing communication efficiency, and supporting social bonding. Other properties present in cetaceans are ***interchangeability*** **(**Hockett 1958), because any member of the species can produce and perceive any type of signal of their repertoire, with the exception of humpback whales, in which only males sing; and ***parity*** (Arbib 2005), which requires that every message has the same meaning for sender and receiver. This property seems present in any cetacean context-specific call, and it seems evident in the case of signature whistles, which refer to the identity of a dolphin both in the perception of its owner and the other individuals who hear it. Finally, the ***cooperative principle*** (Grice 1975) is not demonstrated in cetaceans (and neither in other species), because according to Grice it requires the will to cooperate in the communication, generating

an effort by both sides, while in animals it is generally only the receiver who adapts to obtain the information. However, dolphin turn-taking behaviour (and other complex cognitive abilities) might represent a first step towards this ability.

5.5.3 Pragmatic functions

In this subsection, I describe the pragmatic functions that can be covered by language (and communication). ***Multimodality*** (Kendon 2004) does not have a crucial role in cetaceans, because they mostly rely on vocal communication (see § 1.4). Hence, despite gestures and touch, which are used to strengthen the bonds inside the group, cetacean communication cannot be considered multimodal, being almost only acoustic. ***Modality-independence*** (Berruto & Cerruti 1997), i.e. the ability to send the same message in two different channels (e.g. vocal and gestural) is not present in cetaceans (nor in other animals), unless we want to consider the perspective of perception. In this case, in the experiment by Bruck *et al.* (2022), the cross-modal recognition of the identity of others bottlenose dolphins with both the auditory stimulus of their signature whistle and the taste stimulus of their urine might be a case of modality-independence, because both signature (hearing) and urine (taste) conveyed the same message. Finally, ***multifunctionality*** is a widespread property referred to the fact that a communication system is used for different functions, and cetaceans clearly use their repertoire to communicate a diverse range of information. Also, some of the six functions of human language by Jakobson (1960) have an equivalent in cetacean communication: the *referential* function seems to be present in all the four species analysed, whose signals have meanings; the *poetic* function is absent; the *emotive* function is present to the extent that vocalisations can communicate also about internal states as hunger, fear, etc.; the *conative* function might be considered present only in bottlenose dolphins signature whistles, whose function to address the hearer can be considered "vocative" from a linguistic perspective, and perhaps in sperm whale identity codas; the *phatic* function emerges in all the vocalisations used to maintain social contact and coordination, widespread in cetaceans, and the *metalingual* function seems absent, like in all the species examined in Part II.
Table 13 summarises the observation presented in this chapter.

<table>
<tr><td colspan="2">Cetacean linguistic traits</td></tr>
<tr><td colspan="2">Signal
Physical aspects of the sign</td></tr>
<tr><td>Vocal-auditory channel
All cetacean species described</td><td>Rapid fading
All cetacean species described</td></tr>
<tr><td>Total feedback
All cetacean species described</td><td>Linearity
All cetacean species described</td></tr>
<tr><td colspan="2">Broadcast transmission and directional reception
All cetacean species described</td></tr>
<tr><td colspan="2">Signal learning
Ability to learn new signals and use old signals in new contexts</td></tr>
<tr><td>Vocal contextual learning
All cetacean species described</td><td>Vocal production learning
All cetacean species described</td></tr>
<tr><td>Cultural transmission
(depends on the definition of culture)

All cetacean species described</td><td>Learnability
(learning of a new “language”)
• Wild orcas
• Captive bottlenose dolphins</td></tr>
<tr><td colspan="2">Structure
Complex structures created by combining smaller units</td></tr>
<tr><td colspan="2">Duality of patterning (double articulation)
No proof (only partial clues)</td></tr>
<tr><td colspan="2">Phonology
Combination of meaningless units</td></tr>
<tr><td>Discreteness
All the cetacean species described have discrete signals (although discreteness is not defined as a linguistic property in cetacean literature)</td><td>Combinatorial ability
• Phonocoding: humpback whales
• Combinations of discrete meaningless units to obtain meaningful signals: dolphins, orcas, sperm whales (?)</td></tr>
<tr><td colspan="2">Morphosyntax
Combination of meaningful units</td></tr>
<tr><td>Compositionality
• Captive dolphins (?)
• Wild dolphins (?)
• More studies are needed for other species</td><td>Non-compositional syntax

Not explored in cetaceans</td></tr>
<tr><td>Productivity
Apparently, no cetacean species (at best a very limited productivity in dolphin signature whistles)</td><td>Open-endedness
No cetacean species
(maybe in humpback whales, but without meaning)</td></tr>
<tr><td>Recursion
No cetaceans species (possibly in non-linguistic domains, e.g. vision, orientation)</td><td>Merge
Non-recursive 1-merge: possible in bottlenose dolphins, but not confirmed
(comparable with compositionality)</td></tr>
<tr><td>Recursive merge
No cetacean species</td><td>Hierarchical structuring
• Humpback whales (no complex meaning)
• Dolphins, orcas (in social organisation, but not demonstrable in communication)</td></tr>
</table>

<table>
<tr><td>Structure dependency

• Humpback whales (with no meaning)
• Captive dolphins (?)</td><td>Importance of word order
• Bottlenose dolphins (wild and captive)
• Orcas
• Sperm whales (?)</td></tr>
<tr><td>Non-adjacent dependencies
Never studied in cetaceans</td><td>Rule generalisation
• Captive bottlenose dolphins</td></tr>
<tr><td colspan="2">Meaning
Presence of meaning in language structures (semantics)</td></tr>
<tr><td>Lexicon (words)

All cetacean species described</td><td>Referentiality
Not explicitly claimed, but likely present in:
• Bottlenose dolphins
• Orcas
• Sperm whales
(more evident in captive animals and in ID signals)</td></tr>
<tr><td>Arbitrariness
• Low: all cetaceans with referentiality
• High: not observed</td><td>Equivocity
Never explored in cetaceans, but possibly present in proficient vocal learners</td></tr>
<tr><td>Iconicity
No cetacean species</td><td>Indexicality
All cetacean species described (they have context-specific calls, dolphins can also point)</td></tr>
<tr><td>Symbolism
Not to the extent of humans
• Sperm whales (?)
• However: cetaceans might have a level of arbitrariness higher than indexicality but lower than symbolism</td><td>Mental representations
Controversial: likely but not demonstrated
• Bottlenose dolphins
• Sperm whales
• Orcas (?)</td></tr>
<tr><td>Displacement
Controversial. Possible at least in:
• Bottlenose dolphins</td><td>Detachment from stimuli
Difficult to demonstrate.
Possible in species with displacement</td></tr>
<tr><td>Specialization
All cetaceans with referential calls</td><td>Prevarication
No cetacean species</td></tr>
<tr><td>Reflexiveness
No cetacean species</td><td>Semantic omnipotence
No cetacean species</td></tr>
<tr><td colspan="2">Context
Use of language in external communication (pragmatics)</td></tr>
<tr><td colspan="2">Intentionality
Drive to share meanings
• First-order intentionality: not claimed in the literature, but possible in cetaceans</td></tr>
<tr><td>Theory of Mind
Controversial: difficult to demonstrate</td><td>Shared intentionality
No cetacean species</td></tr>
<tr><td colspan="2">Signal exchanges
Active exchange of linguistic signals</td></tr>
<tr><td>Turn-taking
• Bottlenose dolphins
• Sperm whales</td><td>Interchangeability
• Bottlenose dolphins
• Orcas
• Sperm whales</td></tr>
<tr><td>Parity
All cetacean species described</td><td>Cooperative principle
No proof in any cetacean species</td></tr>
</table>

Pragmatic functions *Functions covered by language*	
Multimodality *Scarce: cetaceans rely on vocal communication*	Modality-independence • *Production: no cetaceans species* • *Understanding: possibly dolphins*
Multifunctionality *All cetacean species described*	Jakobson's functions • *Referential: all cetacean species* • *Poetic: no cetacean species* • *Emotive: all cetacean species* • *Conative: bottlenose dolphins* • *Phatic: all cetacean species* • *Metalingual: no cetacean species*

Table 11. Summary of linguistic properties of cetaceans.

After having discussed the linguistic properties of cetaceans, I now focus on the general traits described also in Part II, to obtain a complete picture of their skills, which will be useful for my comparative analysis.

6. The general traits related to language in cetaceans

As I have clarified in Part I and II, linguistic properties are not the only features involved in language. To investigate the possible presence of language complexity, it is necessary to consider also those general traits which, despite not being linguistic *per se*, are fundamental for the production and comprehension of language, and thus have to be researched in nonhuman animals. Here I summarise the most relevant traits related to biology, brain and cognition, and sociality observed in cetaceans, and discussed in the previous chapters.

6.1 Biological traits

All cetacean species are characterised by a remarkable ***longevity***, a feature that usually implies plasticity and adaptability, as discussed in Part II (§ 4.1). Humpback whales live up to 90 years, orcas 80 years, sperm whales 70 years, and bottlenose dolphins 60 years (see Berta *et al.* 2006 and § 1.2). These animals also have a ***long gestation*** period (11.5 months for humpback whales, 15-18 months for orcas, 14-16 months for sperm whales, and 12 months for bottlenose dolphins (see Berta *et al.* 2006 and § 1.2). Furthermore, they give birth to a ***few offspring***, generally one, placing cetaceans among the K selected animals (Pianka 1970), i.e. the species which invest energy and time in the care of a few calves, which are trained to survive in the environment, instead of generating numerous descendants among which only a few have success (see Part II, § 4.1, and § 1.2 above).

6.2 Brain and cognition

As I have thoroughly discussed in the previous chapters, cetaceans have a complex brain (see § 1.3) and a complex cognition (see § 2-4). Their ***encephalisation quotient*** is particularly high in the case of delphinids, with bottlenose dolphins (4.14 or 5.3, depending on the research) and orcas (2.57) well above 1. Notably, dolphins have the highest EQ after humans, while orcas' values are comparable to those of nonhuman anthropoid primates (1.02-3.21, with a mean of 2, see Marino 2004). In terms of absolute dimensions, orcas have the second-largest brain among mammals, second only to that of sperm whales, which, on the other hand, have the lowest EQ for this group (0.58). According to Marino (2004), this is not a marker of scarce intelligence, because sperm whales had a considerable body growth during their evolution (see also § 1.3). Finally, humpback whales, like the other mysticetes, have an exceptionally low EQ (0.18), but similarly to sperm whales, their big dimensions

are not correlated with an increase of brain mass, so this value does not imply a lack of intelligence[147] (see § 1.3). ***Globularity***, i.e. the globular shape of the brain, which puts different areas close, is described only in humans, but Marino (2004) generically observes that the cetacean brain also has this characteristic. This feature should be further explored to verify whether there are similarities with the human brain. A complex trait shared with humans is the gyrification of the brain, which is highly convoluted, so that cetaceans have a ***large neocortex***, with a larger surface area than humans, although it is thinner (§ 1.3). According to Marino (2004), this represents a case of "convergent complexity", with a structure unique in the animal kingdom, and yet comparable to humans to the extent that it also allows to perform complex cognitive operations. Another interesting feature is the ***FOXP2 gene***, which has been found in delphinids and whales (see § 1.3). Some neurons related to higher brain functions are also present: ***spindle neurons*** have been identified in bottlenose dolphins, orcas, sperm whales and humpback whales, in an amount even superior to that of humans (Butti *et al*. 2019), while ***mirror neurons*** have never been found. However, researchers believe that – considering cetacean complex imitative abilities – mirror neurons or an analogous system play a role their cognition (see Jeon & Lee 2018, Kilner & Lemon 2013, Herman 2012). A feature connected with these neurons is ***self-awareness***, which is difficult to identify; nonetheless, bottlenose dolphins' and orcas' ability to pass the mirror self-recognition experiment supports the hypothesis of its presence in these animals. Considering the cognitive similarities between dolphins and other cetaceans (and the phylogenetic relations between them), it is possible that other species like sperm whales and even humpback whales have this ability, although a test like that of Reiss & Marino (2001) would be impossible to replicate in animals as large as whales. What is certain is that all these animals have a developed ***memory***, as discussed in chapters 2 and 3 above (see also Marino *et al.* 2007, King *et al.* 2022). This emerges in observations in the wild, with individuals capable to remember other entities or events even after decades, and it is further confirmed by experiments with captive animals, especially dolphins (§ 3). All cetaceans also have a high degree of ***cerebral and behavioural plasticity***, which are widely documented and have been already discussed (see § 1-3). ***Tool use***, however, has been observed only in a community of bottlenose dolphins (Smolker *et al*. 1997). Although other dolphin groups and other cetacean species display the same level of flexibility and plasticity, as of today "sponging" appears to be a *unicum* in the Shark Bay population. However, the presence of this skill tells us that – at least – dolphins are capable of using tools, in presence of the right

[147] In these animals, body growth might be connected to physiological necessities related to marine life.

conditions. Finally, ***pointing*** is documented only in dolphins (see chapter 4 above, cf. Xitco *et al.* 2001), and ***pantomime***, here intended as imitation, is a fundamental feature of all the cetacean species described (see § 2-3).

6.3 Sociality

Sociality is one of the most evident features of cetaceans, and it is a common trait in the species with complex communication. Orcas, bottlenose dolphins and sperm whales have some of the most ***complex social structures*** in the animal kingdom (e.g. orca pods, dolphin alliances, sperm whale clans), and also humpback whales, despite being more "independent" animals, have complex social dynamics. Therefore, all these species have a strong ***reliance on cooperation and interaction***, which emerges in their complex coordinated hunting and foraging techniques, vocalisations, and social environment (see § 2-3 above, cf. Whitehead & Rendell 2015, Marino *et al.* 2007). An aspect of this cooperation observed at least in orcas, sperm whales and bottlenose dolphins is alloparental care, a form of ***cooperative breeding*** in which the animals help in the care of the calves of other individuals: for example, while the mother dives in search for food, her calf remains in surface with other members of the group (see § 2-3, cf. Whitehead & Rendell 2015). As I have discussed in the previous chapters (2-3), many behaviours, vocalisations and techniques seem to be manifestations of ***cultural transmission***, i.e. active teaching and sharing of knowledge (see for example techniques as orca intentional stranding in § 2.2 or dolphin sponging in § 3). Similar features are described in all the cetacean species analysed in this dissertation. The same applies to ***social learning***, closely related to cultural transmission: cetaceans are considered capable to learn not only from their mothers (i.e. vertically), but also from other adults (i.e. horizontally), through imitation of conspecifics, plasticity, and memory. All these skills make cetaceans one of the most articulated groups of animals who exist today, and represent an important layer for more complex communicative and cognitive capacities, which display even linguistic properties, as shown in the previous chapters. Table 14 summarises the general traits presented in this chapter.

<table>
<tr><th colspan="2">Cetacean general traits</th></tr>
<tr><td colspan="2">Biological traits
General features related to main aspects of biology</td></tr>
<tr><td>Complex vocal auditory channel

All cetacean species described</td><td>Longevity
• Humpback whales: up to 90 years
• Orcas: up to 80 years
• Sperm whales: up to 70 years
• Bottlenose dolphins: up to 60 years</td></tr>
<tr><td>Long gestation
• Humpback whales: 11.5 months
• Orcas: 15-18 months
• Sperm whales: 14-16 months
• Bottlenose dolphins: 12 months</td><td>Few offspring

All cetacean species described (usually one)</td></tr>
<tr><td colspan="2">Brain and cognition
Structural aspects of the brain and cognitive features of the mind</td></tr>
<tr><td>Encephalization quotient (high)
• Bottlenose dolphins: 4.14-5.3
• Orcas: 2.57</td><td>Globularity
According to Marino cetacean brain is globular, but this feature should be further explored</td></tr>
<tr><td>Large neocortex
All the cetacean species described</td><td>FOXP2 gene
All the cetacean species described (although humans have different substitutions)</td></tr>
<tr><td>Spindle neurons
All the cetacean species described</td><td>Mirror neurons
Not found in cetaceans. However, analogous systems might be present.</td></tr>
<tr><td>Self-awareness
Animals who passed the mirror test:
• Bottlenose dolphins
• Orcas</td><td>Memory
Developed in all the cetacean species described.

Captive dolphins also showed episodic-like memory</td></tr>
<tr><td>Cerebral plasticity
All cetacean species</td><td>Behavioural plasticity
All cetacean species</td></tr>
<tr><td>Tool use
• Bottlenose dolphins</td><td>Pointing
• Bottlenose dolphins</td></tr>
<tr><td>Pantomime/imitation
All the cetacean species described</td><td>Theory of Mind
Controversial: difficult to demonstrate</td></tr>
<tr><td>Shared intentionality

Not demonstrated in any cetacean species</td><td>Mental representations
Controversial: likely but not demonstrated.
• Bottlenose dolphins (?)</td></tr>
<tr><td colspan="2">Sociality
Aspects related to social groups and interactions</td></tr>
<tr><td>Complex social structure
All the cetacean species described</td><td>Reliance on cooperation and interaction
All the cetacean species described</td></tr>
<tr><td>Cooperative breeding
In weak sense (alloparental care):
• Sperm whales
• Orcas
• Bottlenose dolphins</td><td>Cultural transmission

All the cetacean species described</td></tr>
<tr><td>Social learning
All the cetacean species described</td><td>Self-domestication
Not explored in cetaceans</td></tr>
</table>

Table 14. Summary of the general traits related to language in cetaceans.

7. Cetaceans, other animals, and humans: a comparison

In the Part III of this dissertation, I showed that cetaceans are very complex animals that display several features related to language, both from the perspective of general traits (chapter 6) and of linguistic properties (chapter 5). Among the species analysed, bottlenose dolphins emerge as the most promising candidates for linguistic research for many reasons: first, they are the most studied, and this provides a huge amount of data that are not available for other animals. Second, they are sociable animals, widespread in numerous seas of the world (including the Mediterranean), so that comparative studies have been made, and it is possible to witness the great diversity in vocalisations and behaviours. Furthermore, they are the most examined animals in captivity, thanks to their high tolerance: the results which come from the experiments, although not indicating the skills that are used in the wild, show a range of cognitive abilities which are present in their organisms. It should be noted that this does not necessarily mean that dolphins are more intelligent than the other cetaceans described: orcas, sperm whales and humpback whales display several traits of complexity in all the aspects analysed; however, many experiments have never been performed on these species. For example, the issue of mental representations has been explored only in bottlenose dolphins, but there are reasons to believe that other species have this feature as well. Furthermore, also other cetacean species show complex communicative, cognitive and social skills (e.g., belugas), but they were not treated in this dissertation. This is why bottlenose dolphins have received a "special" attention in this thesis, although they share most of their traits with other cetaceans.

Interestingly, most of the general and linguistic features discussed in the previous chapters are shared also with the animals described in Part II, in particular with some nonhuman primate and bird species. Notably, all the species with the most complex communication systems, which present call combinations that can be considered a form of compositionality or 1-merge (see Part II, § 4.3), also have complex biological, cognitive and social traits. These traits are present in *Homo sapiens* as well, which displays them in the most complex level observed in nature. The challenge, here, is to understand if they should be considered qualitatively similar (and thus homologies or analogies), or completely different, and whether they were present also in hominins. In figure 37 I compare the most important features related to language (both linguistic and general traits) of the most relevant species described throughout this thesis: bottlenose dolphins, Campbell's monkeys, chimpanzees, Japanese tits and Southern pied babblers, hominins (in general), and humans.

Trait/Species	Dolphin	Camp. monkey	Chimp	Jap. tit/ babbler	Hominin	Human
Longevity	v	v	v	v	v	v
Long gestation	v	v	v	x	v	v
Complex vocal production/Imitation	v	x	x	v	?	v
Cerebral and behavioural plasticity	v	v	v	v	v	v
Complex neuroanatomical traits/Memory	v	v	v	x	?	v
Learning and cultural transmission	v	v	v	v	v	v
Complex social structure	v	v	v	v	v	v
Mental representations	v	?	?	?	?	v
Referentiality	v	v	v	v	?	v
Word order	v	v	v	v	?	v
Compatibility with 1-merge	?	v	v	v	?	v
Recursion (linguistic)	x	x	x	x	?	v
Productivity	x	x	x	x	?	v
Lexicon (large)	x	x	x	x	?	v

Fig. 37. Comparison of the most important features of language in humans and other species.

As we can see, all the species in figure 37 are long-lived and, with the obvious exception of birds, they have a long gestation period of a few individuals at a time (generally one), so that they are K-selected, i.e. they invest energy in the care of the offspring. Dolphins (as other cetaceans), birds and humans also have complex vocal production and developed imitative skills. Notably, nonhuman primates lack this ability; however, they are still capable to produce vocalisations and apparently to combine them: hence, complex vocal production and imitation are not a requirement for compositionality and 1-merge. Regarding hominins, there is a long-standing debate whether they were vocal learners or not (see Part II, § 1): according to some scholars (e.g. Berwick & Chomsky 2016) they were not, according to others (e.g. Fitch 2010), they were. As of today, it is not possible to clarify this issue, but it should be considered that if a trait is present in humans and in other primates, then there are good chances that it developed in a common ancestor of both groups (i.e. homology), and then it would be plausible that also hominins had it. Conversely, if a trait is present in humans and in other distantly related species as birds, but not in primates, then it could be a case of convergent evolution (i.e. analogy), independently developed by different animals subject to the same pressures. In this case, the issue remains open. What is shared by all the species in the figure is cerebral and behavioural plasticity: all of them have flexible brains and are capable to adapt to different situations. They also have complex neuroanatomical traits and a developed memory; however, Japanese tits and Southern pied babblers have very different brain circuits and a lower memory capacity, so that – despite being complex – I did not

consider them at the same level of the other animals in the figure. It is not possible to claim the presence of these traits in hominins because, like for vocal production, the fossils do not allow to confirm it (although there are reasonable chances that they had complex brains and memory). All these animals are capable of learning and cultural transmission, as described in the previous chapters (§ 3 for dolphins, Part II, § 3-4 for the other species). In this case, fossils and archaeological findings confirm it also for hominins (see Part II, § 1). Similarly, all these species have a complex social structure, which involves coordination, sharing of knowledge, frequent interactions, and a success that depends on the group, more than on the single individual. It is evident that in humans these skills are present to a higher extent than in the other animals described.

Moving on more "linguistic" traits, the strongest case for mental representations and concepts is that of bottlenose dolphins raised by Bruck *et al.* (2022, see also § 3); however, although it is difficult to demonstrate, it is possible that this property is present also in other animals, especially in the species who have referential calls, in which a signal is associated to a meaning. Following this definition, referentiality is a feature shared by all the species in the figure, because all of them produce meaningful signals (see § 5.4 above, and Part II, § 4.4). Regarding hominins, however, none of the linguistic properties (from mental representations to the others that complete this list) can be confirmed. Another trait shared by all the species is word order: the calls (and the units that compose them) follow specific sequential rules, according to which only one disposition of the calls is understood as a combination (e.g. Campbell's monkey's *krak-oo*, Japanese tit's *alert-recruitment*). Regarding dolphins, this has been demonstrated in captivity with an artificial language by Herman *et al.* (1984); in the wild, although no word order was found by King & Janik (2015), other sequences like bray calls respect order restrictions (although they might be meaningless units).

Finally, all the call combinations described in this Part III (§ 2-4) and in Part II (§ 3.3) are compatible with the compositional interpretation of 1-merge (cf. Rizzi 2016), according to which the two calls merge to create a more complex, meaningful structure: in the wild, Campbell's monkey's *krak-oo* and *hok-oo*, chimpanzee's, Japanese tit's and Southern pied babbler's *alert-recruitment* represent a convincing case of compositionality, although in certain cases the deflationary theories proposed by Schlenker *et al.* 2023 cannot be excluded (see Part II, § 3.3 and chapter 5). Regarding bottlenose dolphins and other cetaceans, as observed in chapter 5 above, there are not enough data to claim the presence of 1-merge with certainty. However, the experiments by Herman *et al.* (1984) showed that captive dolphins were capable to understand combinations of meaningful units as *left frisbee*

and *left basket*, inserted in more complex sequences (§ 3.3 above). It is important to observe that, even admitting the compositional interpretation for all the species described here (including other cetaceans), all the combinations are limited to only two meaningful units. In the cases with three units, like some bottlenose dolphin bray calls analysed by Pace *et al.* (2022), they are probably meaningless, but even if they were not, they could be described with a cascade effect, i.e. a combination of more 1-merge systems. Hence, none of the call combinations observed or proposed in nonhuman animals is recursive, i.e. the combinatorial operation cannot be applied to its own output to generate more complex sequences, while in humans this is possible – and it is the foundation of our linguistic complexity. Therefore, despite possibly having merge (more precisely, 1-merge, following Rizzi 2016), all nonhuman animal calls lack recursion, so that their combinations are limited to two elements. This further restricts the possibility to expand their call repertoire, and thus to convey more information by combining pre-existing elements to form new words, associated with new meanings. It does not seem a coincidence that, while having referentiality, these animals do not display the property of productivity: they have it at best with a very limited extent (e.g., to create signature whistles), but they do not use it in a "creative" way to forge wide sets of words. Even the different repertoires of groups of the same species (e.g. orcas' "dialects") do not seem to increase with time. This is connected with lexicon size: none of these animals has a large number of calls: furthermore, not all the calls of the repertoire are combined; generally this is restricted only to a few (for example, Campbell's monkeys also have other calls, but they only combine *krak* and *hok* with *-oo*). In humans, the same recursive combinatorial processes allow to combine a finite set of phonemes to create words, and a potentially infinite set of words (among which there are lexical and functional ones) to create phrases and sentences. Hence, the lack of recursion in the communication of nonhuman animals might limit them to only a few calls, thus reducing their possibility of expression, planning, etc., even in the presence of a non-recursive merge operation. As observed in Part I (§ 2.5 and 3.5) the difference between humans and nonhuman animals might not be the presence of merge *per se*, but of recursive merge, which is the reiterate application of the same, simple combinatorial operation. If that is the case, what happened to humans (or human ancestors) which led them to this apparently unique and unprecedented feature?

Although there is no clear answer yet – and a definitive proof could never be found –, what emerges from the comparison between humans and nonhuman animals is that cooperation and interaction are essential elements for all the species which display compositionality and linguistic properties. Sociality seems strictly related to language: there can be complex cognition without communication (e.g. in corvids, who are very intelligent

but not social animals), but there are no animals which have complex communication but lack sociality. This seems logical, because for communication (and language) to exist there must be the need to use it, and thus to exchange information with conspecifics. However, it cannot be excluded that an operation like 1-merge exists even in intelligent non-social animals, on a cognitive, internal level, although it is not used in communication. What seems clear is that complex cognition does not automatically generate complex communication (or language), but sociality is required. Sociality cannot be the sole vehicle for complex cognition – considering that there are cognitively complex, non-social species – but it surely is vehicle of complex communication. Hence, linguistic properties should emerge only in social animals (and humans are among the most social species in the animal kingdom). Notably, there are highly social species with a limited cognition, as ants and bees, which display complex linguistic properties like displaced reference, according to Bickerton (2014). One could then ask whether the social dynamics are not responsible for the existence of complex linguistic operations like 1-merge, but only for their externalization. The internal cognitive ability to combine concepts, even without labels, could be present in the animal kingdom also in non-social animals like corvids, which are considered capable of reasoning and planning, among other abilities (see e.g. Fitch 2010). In this scenario, 1-merge might be more common than it is believed, but it would be evident only in the few species which adopt it in communication, pushed by social pressures. Then it is obvious that these dynamics are intertwined, and cognition, sociality and communication likely interact and reinforce each other on evolutionary timescales, following the principles of natural selection. However, a cause-effect relationship can be found, in which sociality leads to complex communication, which leads to the development of linguistic properties – not necessarily dependent on a developed cognition, although in complex animals like mammals this connection seems always present.

This reasoning leads us to a possible explanation to linguistic complexity in humans, and communicative complexity in other species: the *social complexity hypothesis* (Freeberg *et al*. 2012), according to which groups with complex social systems tend to develop more complex communicative systems to regulate the complex interaction inside the groups (see also Part I, § 5.3). Considering what has been discussed in this dissertation, it seems reasonable to believe that the selective pressures that led to human language should be searched in sociality and interactions, intended not only as cooperation, but also as competition among groups (as proposed by Pinker & Bloom 1990, see § 3.4). By starting from a potentially shared, non-recursive 1-merge operation, social pressures might have led hominin ad human evolution towards its recursive application, allowed by an increase in

memory and brain power, through quantitative or qualitative physical mutations, gradually reaching the full generative potential of 3-merge. This will be further discussed in the conclusion below. However, evolution does not work as a linear process in which all the structures and behaviours of a species are determined by the sole genes. As discussed in Part I (§ 2.4), the perspective of *evolutionary developmental biology*, or evo-devo, proposes that genes and environmental factors interact, leading to a development of the organisms which is dictated by both phylogenesis and ontogenesis (see also Pleyer & Hartmann 2019). In the case of language evolution, this approach suggests a feedback loop: in evolutionary timescales, the biologically inherited linguistic capacities could have determined the communicative and cognitive abilities at the species level (in phylogenesis), while exposure to a linguistic environment could have shaped neural connectivity and cognitive plasticity at the individual level (in ontogenesis). This reciprocal interaction could have enabled humans to progressively refine their language abilities over generations, as the brain adapted through experience, ultimately leading to the complexity of modern language. In this scenario, the evolution of language would have resulted from both inherited predispositions and experiential modifications, dissolving rigid nature/nurture distinctions. Furthermore, cerebral plasticity would have facilitated the progressive development of linguistic structures, with non-language-specific cognitive mechanisms contributing to the emergence of complex communication.

After having discussed these relevant hypotheses, I can move forward and focus on the research questions presented at the beginning of this dissertation.

7.1 Answers to the research questions

In this thesis I analysed in detail many aspects of the nature of language and its evolution, and several features of animal communication, to verify whether nonhuman species have some properties that can be considered linguistic. This work had the goal to explore the topic of the origin of language, and find out if its core traits are uniquely human or shared with other animals. After having identified merge and recursion as the most relevant properties of human language, I thus asked whether the merge operation is present in nonhuman animals in a more simple, non-recursive form, from which human recursive merge might have originated. Then, more specifically, I asked whether this operation is present in bottlenose dolphins and in other cetaceans, animals particularly promising for their several complex skills. Furthermore, I wanted to verify whether recursion is indeed uniquely human, and if its presence is related to lexicon, considering that one of the greatest

differences between humans and other animals is the size of call repertoire. Now I answer to these questions to the best of my knowledge, considering everything that has been written in this dissertation.

- Is non-recursive merge present in nonhuman animals?

Yes. Considering the experiments with captive nonhuman primates like Kanzi (Savage-Rumbaugh 1986) and Chantek (Miles 1990), and the studies with wild animals such as Campbell's monkeys (Ouattara *et al.* 2009), chimpanzees (Leroux *et al.* 2023), Japanese tits (Suzuki *et al.* 2016, Suzuki & Matsumoto 2022) and Southern pied babblers (Engesser *et al.* 2024), it can be claimed that some nonhuman species combine meaningful calls into more complex structures, whose meaning is derivable from the sum of their parts (e.g. alert + recruitment = mobbing sequence). This aligns with the interpretation of Rizzi (2016), which observes that according to Chomsky's definition of merge, it is difficult to claim that these combinations are not a simple form of it. Although other non-compositional interpretations have been proposed, like the deflationary theories of *one-word* and *separate utterances* by Schlenker *et al.* (2023), in my opinion the combinations cited above should be considered compositional. As discussed in several chapters of this thesis (e.g. Part II, § 4.3 and § 5), there are no reasons to exclude the compositional interpretation, while the deflationary theories appear to be more unlikely, and in certain cases they do not apply (see Part II, chapter 5, for a detailed discussion of the hypotheses compared with bird and primate call combinations). This idea is reinforced by the data from captive animals, which were trained to use and comprehend arbitrary meaningful signals: the fact that nonhuman primates were capable to understand combinations of signals further supports the idea that this ability is present in their cognition, and thus it is potentially accessible also in the wild. Furthermore, as discussed above, there is the possibility that 1-merge is present in other species as a cognitive internal operation utilised to combine concepts, which does not manifest in communication. However, more studies are needed to obtain definitive proof of 1-merge in nonhuman animals, and to exclude without doubt other interpretations.

- Is non-recursive merge present in bottlenose dolphins?

Unclear. From the perspective of cognition, as for other species, there are reasons to believe that it is present; however, it cannot be claimed with certainty. Regarding cetaceans as a whole (thus including humpback whales, orcas, and sperm whales), their cerebral, cognitive

and social complexity supports the hypothesis that these animals combine concepts. Furthermore, combinatorial abilities have been observed in all the cetacean species described in this thesis: humpback whale songs, orca vocalisations and sperm whale codas are all formed by the combination of smaller discrete units (see § 2 and § 5.3 above). Although it is not clear whether the units are meaningful or not, their formation might involve a combinatorial process conceptually comparable to phonology. Regarding bottlenose dolphins, there is more convincing proof, especially coming from captivity (§ 3.3 above): the trained animals were capable to understand combinations of meaningful units arranged in rigid order, to recognise that changes in word order conveyed a different message, and to discriminate between semantically impossible and syntactically incorrect sequences, trying to repair them to save the structure (see also Herman *et al.* 1984, Herman & Uyeyama 1999). These dolphin skills have been discussed also at the beginning of this chapter, so I will not delve into details again. What should be clarified is that the presence of this ability in captive animals supports the possibility that similar processes occur in the wild as well, or at least that they could develop following the right circumstances, considering that they are already part of dolphins' latent capacities. In the wild, as discussed in section 4.3, bray calls indicate at least a combinatorial process, while more complex call combinations are possible (see King & Janik 2015) but should be further explored. Moreover, the study of cetaceans faces limitations that are not present in land animals, so that many unclear issues might be due to lack of data, instead of lack of complexity. It should also be noted that statistical analyses and automatic systems of pattern recognition could help to shed light on these matter in the future.

- Is recursion uniquely human?

Yes. As of today, evaluating all the data at disposal and the literature on nonhuman animals, recursion in communication seems to be a unique trait of humans. As discussed above, in the comment of figure 37, all call combinations of nonhuman animals seem to be limited to only two elements, thus excluding even the possibility of cascade 1-merge systems. This does not mean that recursion is a uniquely human properties: as suggested by Hauser *et al.* (2002), it is likely present in other domains, such as visual and musical processing, and orientation, both in humans and other species. This is particularly plausible for animals like birds and cetaceans who face long-distance migrations, and have to keep in mind pathways composed by several "recursive" steps (first one stage, then another, and so on). In this scenario, recursion might have been exapted by humans – and only by humans – in

communication and in the combination of abstract concepts, starting from a shared substratum of 1) non-recursive merge at cognition level (and in some species also in communication), and 2) non-linguistic recursion utilised by several animals.

- Is recursion involved in the formation of a large lexicon?

Yes. As discussed at the beginning of this chapter, there recursion, productivity, and lexicon seem to be related. There is no nonhuman species with a large lexicon, and we know that learning a vast number of different words, associated with different meanings, requires a high amount of energy and memory. Hence, the human combinatorial mechanism (which combines a few discrete meaningless units like phonemes in complex morphemes and words, and then words in more complex structures as sentences) is very effective in terms of optimisation. The human property of duality of patterning, not found in other species, allows to generate a potentially infinite set of words, through a relatively simple process. Merge is the fundamental operation which performs the combination, recursion is the property that allows to repeat it, and lexicon is the result. Furthermore, this combinatorial process is present not only at the syntactic, but also at the semantic level: this ability also produces complex meanings, thus expanding the expressive potential of language, leading to more concepts, more ability to plan, organise, etc. We have seen that combinatoriality and compositionality are present in nature, but as of today there is no evidence of both layers in any nonhuman species. Compositionality is limited to two units. Combinatoriality seems to be more extended, and songbirds and cetacean create sequences formed by more discrete (likely) meaningless units (e.g. humpback whale songs). These animals might be the best candidates to further explore recursive abilities: their ability to combine meaningless units, even without creating semantic complexity, might indicate that they use their (recursive?) merge-like operation in this process. However, human language is characterised by the ability to merge not only phonemes, but also morphemes and words: without this, there would be no linguistic complexity. Hence, this feature (recursion applied to merge, which generates a large lexicon) appears the most relevant difference between humans and other species. Surely this depends – at least in part – by some general traits, such as memory, brain size, computational power. But has this "superior" complexity been achieved through a gradual process, or as the consequence of something exceptional?

In the conclusion below, I discuss the implications of this reasoning and the possible answers to this final question, after a summary of the topics treated in this dissertation.

Conclusion

This thesis has been driven by three fundamental questions since the beginning. What is language? Where has it come from? Is it uniquely human? This ancient topic has no simple answer, and my work did not expect to definitively solve an issue which has been puzzling humans for millennia. However, I aimed to shed light on some aspects of language related to its presumed uniqueness, trying to put together the knowledge collected by researchers so far, and to evaluate the explanations proposed by theoretical linguistics considering the most recent studies regarding animal cognition and communication. To reach this goal, I divided the dissertation in three parts, each one dedicated to a different field: Part I was focused on human language, Part II on the linguistic features of the most studied nonhuman animals, and Part III on cetaceans, which have been less explored from a linguistic perspective.

More specifically, in Part I, I introduced the issue of the uniqueness of language and its fundamental properties (chapter 1), moving from the topic of the origin of spoken languages (qualitatively modern, which use the same properties despite being superficially different) to the topic of the origin of the faculty of language, i.e. the biological endowment which allows *Homo sapiens* to produce all its possible languages (§ 1.1). Then, I identified the main features of language according to Hockett (1958, 1960, 1966), the first scholar to discuss the presence of language properties in nonhuman animals (§ 1.2). I also discussed some limitations of Hockett's classification, and I presented a different one, by Berruto & Cerruti (1997), which highlights the generative power of language, despite not being sufficient to describe all its characteristics. Subsequently, I focused on the key role of syntax (§ 1.3), i.e. the aspect of language traditionally believed to be the stronghold of human uniqueness, and I described its two layers of combinatoriality (combination of meaningless units like phonemes) and compositionality (combination of meaningful units, as morphemes and words). In chapter 2, I presented the main theories about the nature of human language proposed in the literature, showing that different accounts have different views of this phenomenon. I discussed what I defined innatist frameworks, which tend to see language as an internal phenomenon, biologically determined, focusing on Chomskyan generative theory (§ 2.1); then I discussed the emergentist frameworks, which highlight the external and communicative aspects of language related to experience (§ 2.2). After that, I compared the two approaches (§ 2.3) and I underlined their points of contact, presenting the evo-devo theories that try to solve the dichotomy of nature vs nurture through a more flexible view of the genes, which are seen as interacting with the environment (§ 2.4). Finally, I tried to reduce the distance between the opposite views of innatism and emergentism, by bridging the concept of generative merge with the idea of gradual development (§ 2.5). To do so, I presented the account of Rizzi (2016), which proposes a typological classification of merge

systems of increasing complexity (from 1-merge to 3-merge), in which the simplest operation is non recursive. After having sketched the state of the art of the theories about language, in chapter 3 I focused on the theories about its origin and evolution: first, I presented the division between faculty of language in the broad sense (FLB) and in the narrow sense (FLN) proposed by Hauser *et al.* (2002), which represents a starting point for most of the analyses (§ 3.1). Then, I discussed how innatism (§ 3.2) and emergentism (§ 3.3) apply to language evolution: while in their strong versions the two approaches are incompatible, with innatism advocating for an evolutionary leap (§ 3.2.1) and emergentism considering language only as a form of social interaction (§ 3.3.1), I showed that some intermediate views might present a common ground, which recognises both the role of biology and experience (§ 3.4). I thus discussed "moderate" emergentism (§ 3.4.1, 3.4.2) and "moderate" innatism (§ 3.4.3), and I compared the different approaches (§ 3.5). To conclude, I tried to bridge the concept of generative merge with gradual evolution (§ 3.6), following the insights from Rizzi (2016) presented in the previous chapter (§ 2.5), observing that different levels of merge might reflect different stages of the evolution of human language. In chapter 4 I extended my analysis beyond merge and recursion, observing that there are many other properties involved in language, which cannot be reduced to a single core operation, no matter how important it might be. Hence, I proposed a novel classification of the properties of human language, identifying the elements which characterise it: signal (§ 4.1), signal learning (§ 4.2), structure – including phonology and morphosyntax – (§ 4.3), meaning – related to semantics – (§ 4.4), and contextual usage – related to pragmatics – (§ 4.5). Each "layer" included several properties that have been discussed in detail. After that, I dedicated chapter 5 to a similar classification, this time focused on the general traits related to language, i.e. those features that, despite not being purely linguistics, are fundamental for language production and perception: biological traits (§ 5.1), aspects of brain and cognition (§ 5.2), and sociality (§ 5.3). The classifications presented in chapters 4-5 are important not only to better define language, but also to identify those traits and properties that should be searched in nonhuman species to verify the proposal of human uniqueness of language, which is the topic of the following parts. Before focusing on animals, however, in chapter 6 I discussed again the "core" traits of language in light of the many properties highlighted in Part I, and I presented the research questions: is non-recursive merge present in nonhuman animals? Is it present in dolphins? Is recursion uniquely human? Is this property related to the size of lexicon?

In Part II, I addressed these questions focusing on nonhuman animals in general. First, in chapter 1, I briefly discussed some alternative approaches to verify the issue of

human uniqueness of linguistic features: I thus focused on the research of language properties in human non-linguistic domains, showing that recursion seems to be present also in visual and musical processing, and in orientation, reasonably also in other species (§ 1.1); then, I sketched the main issues and limitations of the research of language properties in extinct hominins, who have left at best indirect proof of their abilities, which cannot be considered definitive (§ 1.2). Subsequently, I dedicated chapter 2 to nonhuman animals, indicating the advantages of the comparative method: considering that hominins cannot be studied, other living species are the best candidates to verify whether some linguistic traits exist outside humans, in more or less rudimental forms. In this case, they could be a case of homology or analogy. I explained why it is important to consider animal skills, which can help insert human specificities into the wider picture of general evolution, although some scholars disagree with this position (e.g. Berwick & Chomsky 2016). Hence, in chapter 3 I presented the same classification of language properties proposed in Part I (§ 4): following the same layers of signal (§ 3.1), signal learning (§ 3.2), structure (§ 3.3), meaning (§ 3.4) and contextual usage (§ 3.5), I reviewed the literature to identify the features present in nonhuman animals (and in which species), thus highlighting the elements that are not unique, and those that apparently are. In chapter 4, I presented the classification of the general traits related to language (the same as Part I, chapter 5), which should be researched in animals and considered pre-requisites for more complex linguistic abilities: I thus reviewed biological traits (§ 4.1), brain and cognition (§ 4.2) and sociality (§ 4.3) of several species. After these two classifications, it emerged that – despite many similarities – most of the linguistic properties are present in nonhuman animals only in more rudimental forms, and that some "core" traits are indeed not found outside humans: the most relevant are duality of patterning, productivity, recursion, and a large lexicon. However, simple call combinations of two meaningful units have been identified in some nonhuman primate and bird species, which might reflect the 1-merge operation described by Rizzi (2016). In chapter 5 I summarised the results of the classifications, and I discussed in detail the most promising animals: Campbell's monkeys, chimpanzees, Japanese tits and Southern pied babblers. I verified whether the compositional interpretation (1-merge) could explain their call combinations, and I also evaluated two deflationary theories proposed by Schlenker *et al.* (2023): *one-word* and *separate utterances*. What emerged is that the compositional interpretation seems convincing and cannot be excluded, while the deflationary theories in some cases are incompatible, and in others they appear to be more unlikely than compositionality. Elaborating from the results of the classifications of linguistic and general properties, I then focused on a group of animals that has not been thoroughly studied from

the perspective of language, but presents many interesting traits of complexity: the cetaceans.

Part III is entirely dedicated to cetaceans. First, I explained why these animals are so important for language research, clarifying that they combine complex vocal production, complex cognition and social dynamics, i.e. all the general features related to language that coexist only in humans (but not in other nonhuman primates or bird species who have traditionally been studied in language research). I also clarified the limitations to the studies, due to the marine nature of these species, which makes the observations more challenging. Then, in chapter 1, I outlined the biology of cetaceans, describing their anatomy (§ 1.1), physiology (§ 1.2), brain structure (§ 1.3), and systems of vocal production and perception (§ 1.4). After this overview, in which I considered cetaceans as a whole, in chapter 2 I focused on some of the most promising animals: humpback whales (§ 2.1), killer whales (§ 2.2) and sperm whales (§ 2.3). For each species I described their complex cognition, behaviours and social structures, and I paid special attention to the complex vocalisations that might contain linguistic properties: humpback whale calls (§ 2.1.1) and songs, which appear to be hierarchically structured (§ 2.1.2), orca "dialects", which vary according to the group (§ 2.2.1), and sperm whale codas, which are temporally arranged in a morse-like way to convey complex information (§ 2.3.1). For each animal, I also discussed the possibility of 1-merge and recursion, based on the data available. In chapters 3 and 4 I focused on bottlenose dolphins, because they are the species that can provide the most insights into language properties in cetaceans: they are widespread and well-studied, particularly intelligent and sociable, and there is a large amount of data available from both captivity and the wild. First, I described their fission-fusion society and their behaviours (§ 3.1), then I outlined their cognition, discussing some studies with captive individuals that showed the presence of complex abilities (§ 3.2), and finally I presented the experiments by Herman *et al.* (1984), which demonstrated very interesting and relevant language-like skills, including the capacity to understand combinations of meaningful signals and to generalise the rules of an artificial language (§ 3.3). After having assessed these latent abilities, I analysed dolphin vocalisations in the wild, focusing on signature whistles (§ 4.1), i.e. learned vocal labels referred to the identity of the individual, and food-related bray calls (§ 4.2). For both types of signals, I identified the linguistic properties of combinatoriality and referentiality. I concluded the chapters dedicated to dolphins by discussing the possibility of 1-merge, recursion and other features, and proposing some statistical approaches which could be made in the future to shed light on the unclear issues. From these analyses it emerged that there is no sufficient evidence to claim the presence of 1-merge in wild dolphins; however, this might

be due to lack of data. Furthermore, the experiments in captivity indicate that dolphins have the cognitive skill to understand call combinations, performing an operation very similar to non-recursive merge. After having discussed cetacean species separately, in chapters 5 and 6 I applied the classifications presented in Part I (chapters 4-5) and Part II (chapters 3-4) to these animals, summarising their linguistic properties regarding signal (§ 5.1), signal learning (§ 5.2), structure (§ 5.3), meaning (§ 5.4), contextual usage (§ 5.5), and their general features of biological traits (§ 6.1), brain and cognition (§ 6.2), and sociality (§ 6.3). This facilitated the comparison among cetaceans, which confirmed that, like other nonhuman animals, even admitting the possibility of 1-merge (and thus of meaningful call combinations), they do not display recursion, and their repertoires are characterised by a low number of calls. In the final chapter (chapter 7), I discussed these findings, and I compared the results of cetaceans with those of the nonhuman animals described in Part II, also considering humans and hominins. It emerged that cetaceans, in particular dolphins, have more general traits related to language than other species (e.g. complex vocal production, cognition, social dynamics, developed memory), and some linguistic properties which are difficult to identify in other animals (e.g. mental representations, calls related to specific individuals, complex combinatorial systems). I also observed that all the animals confirm the same trend: they seem to have only binary call combinations, thus reflecting at best 1-merge, they do not display recursion, and they have a limited lexicon. The discussion led me to answer to the research questions of this thesis (§ 7.1). Is non-recursive merge present in nonhuman animals? Considering what I have exposed in this dissertation, the most convincing answer is yes (although there is no definitive proof). Is non-recursive merge present in dolphins (and other cetaceans)? As of today, it is not possible to claim it with certainty, although studies on captive animals lean towards an affirmative answer (at least in comprehension). Is recursion uniquely human? In language (and communication), it appear that it is; however, it is likely that this property is present in other species in different domains (e.g. vision, orientation). Is recursion related to lexicon size? As discussed in Part III (chapter 7), there seems to be a clear relationship between merge, recursion and lexicon, because the recursive application of the 1-merge operation is responsible to the expansion of lexicon, a feature observed only in humans.

All the topics discussed in this dissertation allow me to return to the main questions related to the nature, the origin, and the uniqueness of language. The study of nonhuman animals is a useful tool to better understand these issues, because it permits to evaluate the accounts of theoretical linguistics in light of the data coming from ethology: for example, if a theory is based on the lack of certain abilities in nonhuman animals, and research identifies

these abilities, then the theory should be reconsidered. Keeping this in mind, I am now ready to evaluate the main reconstructions about the origin of language, discussed in Part I (chapters 3-4). Considering the compositional abilities of nonhuman animals (at least the most convincing cases of nonhuman primates and birds), in my opinion it is hard to believe that merge suddenly "emerged" in its recursive, non-divisible form, only in anatomically modern humans, as Berwick & Chomsky (2016) suggest: the similarities in the operation indicate that a non-recursive form of merge was probably already present in human ancestors, as it seems to be in other extant primates. The proposal that animal call combinations are totally unrelated with human combinations, and use different properties, does not seem convincing. Even more unlikely is the idea that hominins did not have a phase of protolanguage (be it gestural, lexical or musical). In Berwick & Chomsky's (2016) theory – which is part of the "strong" innatism and advocates for an evolutionary leap –, hominins and ancient humans did not produce complex vocalisations, hence the origin of lexicon is not explained, and it is considered a "mistery" (see Part I, § 3.2.1). This lack of explanation is a weakness too big to be ignored, and cannot be relegated as a secondary question: without explaining the origin of the words, the operation merge that combines them would remain useless. The only realistic way to explain lexicon is that hominins already produced vocalisations, whether they combined them in a simple way or not. Furthermore, as of today a growing body of studies confirms that nonhuman primates produce simple meaningful vocalisations, and that certain species combine them in more complex structures (see Part II, § 4.3). In the past, this ability had not been studied, and nonhuman primates were considered devoid of vocal calls. However, now this perspective has been reversed: despite being more limited with respect to humans, nonhuman primates produce calls, and the fact that they do makes it probable that this ability was present in a common ancestor with humans (and hominins). In this case, the ability to produce vocal calls would be a homology, and the same might apply to the ability to combine meaningful units, if it should be proven that all nonhuman primates have some form of merge (at least on an internal, cognitive level): in humans it could have become more complex, but it would not have appeared from scratch. Therefore, if we want to hypothesise an evolutionary leap, in my opinion the only possible saltation would be a case of exaptation, i.e. a pre-existing trait used for a new function. It could be the application of the cognitive operation 1-merge to combine calls, or, if this skill was already present, another possibility would be the application of recursion to this existing system. Anyway, this can happen only through the necessity of utilising a certain trait, in this case in communication (while in Berwick & Chomsky's 2016 theory, in a first phase recursive merge remained as an internal phenomenon, not externalised). On the other hand,

the theories of strong emergentism (e.g. Tomasello 2008, see Part I, § 3.3.1), despite being in my opinion more convincing (at least in principle), they relegate language to one of the many complex human skills achieved by using domain-general traits (the same utilised, for example, to learn cultural behaviours or complex techniques). However, it seems to me that a specific biological endowment should be present, although it could have evolved gradually following selective pressures, and not through a saltation. Regarding the "intermediate" theories, Bickerton (2014) proposes that the first step towards language was a linear protolanguage dictated by the necessity of the hominins, 2 million years ago, to communicate the location of megafauna carcasses to conspecifics, because they needed help to scavenge them and protect them from other predators (see Part I, § 3.4.1). Using their latent abilities, hominins would have gradually started to communicate about elements outside of the context, developing the linguistic property of displaced reference, in a way that shows peculiar similarities with ants and bees. Such a reconstruction appears more credible than that of Berwick & Chomsky (2016), even if Bickerton (2014) proposes a second phase in which a mutation occurred, endowing humans with merge: it is easier to accept the idea of a saltation in animals that already had a series of shared vocalisations (or gestures), than in animals totally incapable of signal production. Also, Pinker & Bloom (1990) raise important questions with their argument according to which a language faculty, even biologically predetermined, could be the consequence of a gradual evolution perfectly fitting in the boundaries of natural selection (see Part I, § 3.4.1). They propose that our modern ability derived from the clash of different hominin species and subspecies, in a linguistic arms race characterised by cooperation and competition among groups, which would have slowly selected the individuals more capable of communication, of planning and of defeating the rivals. In this scenario, modern humans would be the outcome of this selection, the sole survivors of this fight which would have killed or incorporated the other species – and today there is evidence that Neanderthals and Denisovans did not completely disappeared, because Sapiens bred with them so that now part of their DNA is included in some human populations. These reconstructions appear credible and compatible with the wider picture of evolution, because they see language as a complex adaptive system, thus classifying it at the same level of other complex specialisations observed in nonhuman animals, in the process of niche construction. In my opinion some aspects of Tomasello's (2008) account are also convincing, in particular where he considers language a form of social interaction: the dynamics related to the need of cooperation and to the development of shared intentionality, with the pivotal role of the theory of mind, are in fact compatible with many observations from Bickerton (2014) and Pinker & Bloom (1990). The three theories put selective

pressures towards sociality and cooperation at the centre. Furthermore, I agree with Tomasello's (2008) consideration that any other species subject to the same pressures faced by ancient humans (or hominins) would develop a form of language: in nature there are countless cases of convergent evolution which demonstrate that, with time, similar traits can be obtained by very different animals. It is thus important to always keep in mind the evolutionary dynamics that apply to all species, from which humans cannot escape, and for which we do not represent an exception, no matter how exceptional the results that we obtained with language and with our civilisation might appear. Another perspective is that of the scholars which advocate for the "moderate" innatism, and propose the existence of a "language-ready brain" (see Part I, § 3.4.3). In this view, a series of gradual structural modifications of the brain would have led to an "open" language faculty, i.e. an attitude to learn, manage and organise linguistic structures (and not a fixed set of rules). This account, like the ones previously discussed, is compatible with the idea of gradual evolution following selective pressures toward higher efficiency. Overall, the theories do not appear to be completely incompatible: the most irreconcilable aspects are the idea of saltation, which excludes gradual development, and the opposite one which considers language as a form of culture, excluding the species-specific biological endowment. As I have previously discussed (see Part I, § 2.4, but also Part III, chapter 7), the evo-devo approaches solve part of the contradictions, with the gene-culture interaction that can gradually lead to a complex biological endowment without excluding the role of experience. In this scenario, the external cultural context would lead the organism, in ontogenesis, to develop certain brain connections which facilitate language, and in phylogenesis might have selected the individuals more capable to perform these cerebral processes, or the most plastic and flexible ones. On the other hand, the biological endowment would allow humans – and only humans – to manage this complex communicative and cognitive system, so that evolution really equipped humans with a species-specific ability which can be considered a faculty of language. From a philosophical point of view, trying to disentangle this relationship would be like asking which came first, the chicken or the egg. So we can only go back to the main question: if the dynamics that led to language are not exceptional, what is the core aspect that differentiates us from the other species, if there is one? Considering all that has been written in this dissertation, it does not seem merge, but recursion. According to Rizzi (2016, 2021) it could be the consequence of a qualitative change, i.e. a mutation which endowed humans with two new brain areas called temporary memory storages, which allow to store information of previous combinations and "unlock" recursion – leading from 1-merge to 3-merge (see Part I, § 2.5 and § 3.6). As I already discussed, a similar mutation seems to me

less drastic than the appearance of merge, and it is consistent with the abilities observed in other animals. Regarding the origin of recursion, it could have been exapted from other cognitive systems. However, it should be noted that the recursion observed in human language is different from that present in other domains: in visual processing (see Part II, § 1.1), it seems to be a "matrioska" structure, in which one element includes another (and so on); in orientation, it adds more and more elements in an apparently "linear" way (first one step, then the other, e.g. from the nest to the river, then from the river to the tree, etc.). By contrast, in human language, recursion allows to apply merge to its own output in a hierarchical way in which some elements control others (e.g. through c-command): this hierarchical organisation might thus represent a peculiar type of recursion and could be considered a qualitative difference between humans and other species. A second possibility would be that there are no specific linguistic memory storages in the brain, and that the information of the combinations is instead stored in general memory: if that were the case – and if compositional skills derive from general processing mechanisms rather than specifically linguistic ones –, then there would be no need to postulate a mutation that endowed humans with novel dedicated memory slots. However, before proposing a purely quantitative difference between human 3-merge and animal 1-merge, it should be first noted that recursion in language shows different properties than in other domains (as discussed above). Hence, it should be assessed whether these different properties can be obtained through a quantitative modification (e.g., more memory or computational power), or if they represent a case of discontinuity. As of today, there is no way to demonstrate or contradict these two hypotheses, and the same applies to most of the theories presented here (and, more in general, in the literature). Therefore, I conclude this dissertation by proposing my speculative reconstruction of language evolution, which stems from all the data and the observations that I reported in these pages.

In light of everything discussed in this thesis, I propose that merge, intended as the non-recursive cognitive ability of combining two concepts, is a property widespread in the animal kingdom, which reflects binary neural operations. However, very few species use it in communication, so that it rarely emerges in the external environment. Similarly, recursion is a natural property present in other domains such as vision and orientation (although in these domains it appears less complex than in language), and it is likely ancient and widespread in different species. Hominins were complex animals with complex brains, which struggled to survive and were facing strong selective pressures toward cooperation, in a situation similar to that proposed by Bickerton (2014). Their success depended on their ability to coordinate among conspecifics, and they already had a rudimental protolanguage,

in which they combined meaningful units in a way comparable to that of modern nonhuman primates. During the course of evolution, they have always been in competition with other groups and subspecies, in a scenario comparable to that of Pinker & Bloom (1990). Hence, social pressures gradually selected the individuals and the species more capable to communicate and cooperate, leading to bigger brains capable of performing more complex operations, endowed with more memory and computational power, following the social complexity hypothesis (Freeberg *et al.* 2012). With these complex brains already equipped with "linguistic" 1-merge operation, actively used in a protolanguage, the increase of memory and computational power would have led to more complex combinations, allowing to apply 1-merge in a wider way, first with a cascade effect (combining three single items) and then with recursion. This would have allowed to 1) form more words combining morphemes, and 2) combine more words in complex structures, thus expanding the lexicon. This ability would not be confined only to syntax, but would also involve semantics, so that the combinations would have led to a more complex conceptual world, and thus to a higher ability to plan, communicate, defeat enemies, in a chain reaction which would have finally led to modern humans. The step towards recursive merge could be a qualitative saltation, as Rizzi (2016) proposes, or a gradual, quantitative change (if we consider the other "types" of recursion as precursors of that used in language). In both cases, during the course of the evolution of language, the interaction between biology and external environment would have been a feedback loop following the evo-devo theory, so that both aspects interacted and led to 1) the selection of the individuals more capable to produce language, ultimately conducing to a biological language faculty in phylogenesis, and 2) the development of a linguistic ecosystem made by more and more complex interactions, which also shaped the individuals in ontogenesis. This view could explain why, despite having a faculty of language, exposure to language is still necessary for its acquisition (while animals with hardwired calls produced them even in captivity), highlighting the role of experience and ontogenesis. Moreover, it could also explain why only humans can produce a language with these properties, while other species are not capable to do it even after training, highlighting the role of the biological endowment and phylogenesis. Finally, it should also be considered that our faculty of language does not imply a superiority with respect of other species, but only a specificity, because humans are not capable to learn the communication systems of other animals exactly like they cannot learn ours.

This reconstruction takes several aspects from different theories, and tries to put together the most convincing observations coming from innatist and emergentist frameworks. It is an attempt to put these frameworks closer, advocating for a common

starting point of analysis, from which diverse approaches can then focus on different aspects of language. Most of them can be considered valid and need not be in complete opposition. In the complex and uncertain field of language evolution, what is most needed is cooperation between different views, ideas, and perspectives. This does not refer only to theoretical linguistics, but also to other disciplines which should interact to shed light on this fascinating and unsolved topic. Only an interdisciplinary approach that conjugates linguistics, biology, ethology, statistics, philosophy, palaeoanthropology (and many other fields) can help in reconstructing the origin of language. With this dissertation, I hope that I have demonstrated that the study of nonhuman animals is not only interesting *per se*, but it is an important tool to explore human cognition and communication, and to better understand our place in the wider picture of nature. This work is just a small step: most of the path of this research has yet to be written.

References

Abe, K., Watanabe, D. (2011). 'Songbirds possess the spontaneous ability to discriminate syntactic rules'. *Nature Neuroscience 14 (8)*. 1067-74.

Abramson, J. Z., Hernández-Lloreda, M. V., García, L. *et al*. (2018). 'Imitation of novel conspecific and human speech sounds in the killer whale (*Orcinus orca*)'. *Proc. R. Soc. B,* 285: 20172171.

Akmajian, A., Demers, R.A., Farmer, A.K., Harnish, R.M. (2010). *Linguistics (6th ed.)*. Cambridge, MA: The MIT Press.

Aitchison, J. (1987, 2012). *Words in the Mind: An Introduction to the Mental Lexicon* (1st ed. 1987, 4th ed. 2012). Hoboken, NJ: Wiley.

Allen, T., Fortin, N. (2013). 'The evolution of episodic memory'. *Proceedings of the National Academy of Sciences of the United States of America, 110.* 10.1073/pnas.1301199110.

Allen, J. A., Garland, E. C., Dunlop, R. A., Noad, M. J. (2019). 'Network analysis reveals underlying syntactic features in a vocally learnt mammalian display, humpback whale song'. *Proc. R. Soc. B, 286*: 20192014.

Allman, J., Hakeem, A., Watson, K. (2002). Book Review: Two Phylogenetic Specializations in the Human Brain. *The Neuroscientist: a review journal bringing neurobiology, neurology and psychiatry, 8*. 335-46. 10.1177/107385840200800409.

Ambridge, B., Lieven, E.V.M. (2011). *Child language acquisition: Contrasting theoretical approaches*. Cambridge/New York: Cambridge University Press.

Anderson, J.R., Gallup, G.G. (2015). 'Mirror self-recognition: a review and critique of attempts to promote and engineer self-recognition in primates'. *Primates, 56*. 317–326. https://doi.org/10.1007/s10329-015-0488-9

Andreas, J., Begus, G., Bronstein, M. M. *et al*. (2021). 'Cetacean Translation Initiative: a roadmap to deciphering the communication of sperm whales'. *ArXiv* abs/2104.0861

Arbib, M. (2005). 'From Monkey-Like Action Recognition to Human Language: An Evolutionary Framework for Neurolinguistics'. *The Behavioral and brain sciences, 28.* 105-124.

Arbib, M.A. (2011). 'From mirror neurons to complex imitation in the evolution of language and tool use'. *Ann. Rev. Anthropol., 40*. 257–273. doi: 10.1146/annurev- anthro-081309-145722

Arbib, M.A. (2012). *How the brain got language: The mirror system hypothesis*. Oxford: University Press.

Arbib, M. (2018). 'Computational challenges of evolving the language-ready brain: 2. Building towards neurolinguistics'. *Interaction Studies, 19.* 22-37. Baker, M.C. (2008). *The syntax of agreement and concord*. Cambridge: Cambridge University Press.

Ari, C., D'Agostino, D.P. (2016). 'Contingency checking and self-directed behaviors in giant manta rays: Do elasmobranchs have self-awareness?'. *J Ethol*, *34*. 167–174. https://doi.org/10.1007/s10164-016-0462-z

Arnold, K., Zuberbühler, K. (2006a). 'The alarm-calling system of adult male puttynosed monkeys, *Cercopithecus nictitans martini*.' *Animal Behaviour, 72*, 643e653.

Arnold, K., Zuberbühler, K. (2006b). 'Language evolution: Semantic combinations in primate calls'. *Nature, 441(7091)*. 303. https://doi.org/10.1038/441303a

Arnold, K., Zuberbühler, K. (2008). 'Meaningful call combinations in a non-human primate'. *Current Biology, 18(5)*. R202–R203. https://doi.org/10.1016/j.cub.2008. 01.040

Arnold, K., Zuberbühler, K. (2012). 'Call combinations in monkeys: Compositional or idiomatic expressions?' *Brain and Language, 120(3)*. 303–309. https://doi.org/10.1016/j.bandl.2011.10.001

Babb, S.J., Crystal, J.D. (2006). 'Episodic-like memory in the rat'. *Curr. Biol. 16*. 1317–1321.

Balari, S., Lorenzo, G. (2013). *Computational Phenotypes. Towards an Evolutionary Developmental Biolinguistics*. Oxford: Oxford University Press.

Banovac, I., Sedmak, D., Judas, M., Petanjek, Z. (2021). 'Von Economo Neurons – Primate-Specific or Commonplace in the Mammalian Brain?'. *Frontiers in Neural Circuits, 15*. 10.3389/fncir.2021.714611.

Baron-Cohen, S. (1991). 'Precursors to a theory of mind: Understanding attention in others'. In Whiten, A. (Ed.). *Natural theories of mind: evolution, development, and simulation of everyday mindreading*. Oxford, UK Cambridge, Massachusetts: B. Blackwell. pp. 233–251.

Baron-Cohen, S. (2020). *The Pattern Seekers. How Autism Drives Human Invention*. New York: Basic Books.

Bearzi, G., Eddy, L., Piwetz, S., Reggente, M., Cozzi, B. (2017). 'Cetacean Behavior Toward the Dead and Dying'. 10.1007/978-3-319-47829-6_2023-1.

Beckers, G.J.L., Berwick, R.C., Okanoya, K., Bolhuis, J.J. (2017). 'What do animals learn in artificial grammar studies?' *Neurosci. Biobehav. Rev., 81*. 238–246. doi:10.1016/j.neubiorev.2016.12.021

Benítez-Burraco, A., Longa, V.M. (2010). 'Evo-devo – of course, but which one? Some comments on Chomsky's analogies between the biolinguistic approach and evo-devo'. *Biol. Ther. Dent., 4.* 308–323.

Benítez-Burraco, A., Boeckx, C. (2014). 'Universal Grammar and biological variation: an evo-devo agenda for comparative biolinguistics'. *Biological Theory, 9*. 122-134.

Benítez-Burraco, A., Boeckx, C. (2015). 'Possible functional links among brain- and skull- related genes selected in modern humans'. *Frontiers in Psycology, 6.* 794.

Benítez-Burraco, A. (2017). 'Grammaticalization and language evolution: focusing the debate'. *Lang. Sci., 63.* 60–68. doi: 10.1016/j.langsci.2017.03.003

Benítez-Burraco, A. (2020). 'Prehistoric languages and human self-domestication'. *Language Dynamics and Change, 10*. 27-58.

Bentley-Condit, V., Smith, E.O. (2010). 'Animal tool use: Current definitions and an updated comprehensive catalog'. *Behaviour, 147*. 185

Berger, L.R., Hawks, J., de Ruiter, D.J., *et al*. (2015). '*Homo naledi*, a new species of the genus *Homo* from the Dinaledi Chamber, South Africa'. *eLife, 4*. e09560

Berger, L.R., Hawks, J., Dirks, P.H., Elliott, M., Roberts, E.M. (2017). '*Homo naledi* and Pleistocene hominin evolution in subequatorial Africa'. *eLife*, *6*. e24234.

https://doi.org/10.7554/eLife.24234

Berruto, G. (1976). *Nozioni di linguistica generale*. Napoli: Liguori.

Berruto, G., Cerruti, M. (1997). *La linguistica. Un corso introduttivo*. Torino: Utet.

Berta, A., Sumich, J., Kovacs, K. (2006). *Marine Mammals: Evolutionary Biology (2nd ed)*. Cambridge, MA: Academic Press.

Berruto, G. (2012). *Sociolinguistica dell'italiano contemporaneo*. Roma: Carocci.

Berthet, M., Mesbahi, G., Pajot, A., Cäsar, C., Neumann, C., Zuberbühler, K. (2019). 'Titi monkeys combine alarm calls to create probabilistic meaning'. *Science advances*, *5*(5). eaav3991. https://doi.org/10.1126/sciadv.aav3991

Berwick, R.C*., Chomsky,* N. (*2016*)*. Why Only Us. Language and Evolution*. Cambridge, MA: MIT Press.

Berwick, R., Chomsky, N. (2019). 'All or nothing: No half-Merge and the evolution of syntax'. *PLOS Biology, 17*. e3000539. 10.1371/journal.pbio.3000539.

Biberauer, T., Roberts, I., Sheehan, M. (2014). 'No-Choice Parameters and the Limits of Syntactic Variation'. In Santana-LaBarge, R.E., *Proceedings of the 31st West Coast Conference on Formal Linguistics*. Somerville, MA: Cascadilla Proceedings Project. 46-55.

Biberauer, T. (2018). 'Pro-drop and emergent parameter hierarchies'. In Cognola, F., Casalicchio, J. (Eds.), *Null subjects in Generative Grammar. A Synchronic and Diachronic Perspective*. Oxford: Oxford University Press. 94-135.

Bickerton, D. (2014). *More Than Nature Needs: Language, Mind and Evolution.* Cambridge: Harvard University Press.

Boeckx, C. (2013). 'Biolinguistics: forays into human cognitive biology'. *J. Anthropol. Sci., 91.* 63–89.

Boeckx, C. (2021). *Reflections on language evolution: From minimalism to pluralism.* (Conceptual Foundations of Language Science 6). Berlin: Language Science Press. DOI: 10.5281/zenodo.5524633.

Bohn, M., Call, J. Tomasello, M. (2015). 'Communication about absent entities in great apes and human infants'. *Cognition, 145*. 63–72.

Bohn, M., Call, J., Tomasello, M. (2016). 'Comprehension of iconic gestures by chimpanzees and human children'. *J Exp Child Psychol, 142*. 1-17. doi: 10.1016/j.jecp.2015.09.001.

Bolhuis, J., Tattersall, I., Chomsky, N., Berwick, R. (2014). 'How Could Language Have Evolved?' *PLoS biology, 12 (8)*. 1- 6.

Bolhuis, J.J., Beckers, G.J.L., Huybregts, M.A.C., Berwick, R.C., Everaert, M.B.H. (2018). 'Meaningful syntactic structure in songbird vocalizations?'. *PLoS biology*, *16*(6). e2005157. https://doi.org/10.1371/journal.pbio.2005157

Bonfanti, L., Charvet, C. (2021). 'Brain Plasticity in Humans and Model Systems: Advances, Challenges, and Future Directions'. *International Journal of Molecular Sciences, 22*. 9358. 10.3390/ijms22179358.

Boros, M., Magyari, L., Morval, B., Hernández-Pérez, R., Dror, S., Andics, A. (2024). 'Neural evidence for referential understanding of object words in dogs'. *Current Biology, 34*. 1–5. doi.org/10.1016/j.cub.2024.02.029

Branchini, C., Donati, C. (2016). 'Assessing lexicalism through bimodal eyes'. *Glossa: a journal of general linguistics, 1(1):* 48. 1–30.

Brenowitz, E., Beecher, M. (2023). 'An ecological and neurobiological perspective on the evolution of vocal learning'. *Frontiers in Ecology and Evolution, 11*. 10.3389/fevo.2023.1193903.

Brighton, H. (2002). 'Compositional syntax from cultural transmission'. *Artificial Life, 8*. 25-54.

Brooks, L., Franklin, T., Franklin, W., Harrison, P., Corkeron, P., Pollock, K.H. (2024) 'Humpback whales (Megaptera novaeangliae) in Hervey Bay, Australia: a stopover for females early in their southern migration'. *Front. Mar. Sci. 11*:1426248. doi: 10.3389/fmars.2024.1426248

Brown, R. (1973). *A First Language: The Early Stages.* Harvard: University Press.

Brown, S. (2017). 'A joint prosodic origin of language and music'. *Frontiers in Psychology, 8: 1894.*

Brownell, R. L. Jr. *et al*. (1987). 'External morphology and pigmentation of the vaquita, *Phocoena sinus* (*Cetacea: Mammalia*)'. *Marine Mammal Science, 3 (1)*. 22-30.

Bruck, J. N. (2013). 'Decades-long social memory in bottlenose dolphins'. *Proc R Soc B, 280*: 20131726.

Bruck, J., Walmsley, S., Janik, V. M. (2022). 'Cross-modal perception of identity by sound and taste in bottlenose dolphins'. *Science Advances, 8*. 7684.

Bugnyar, T., Reber, S.A., Buckner, C. (2016). 'Ravens attribute visual access to unseen competitors. *Nature Communications*, *7*. 10506.

Burkart, J.M., van Schaik, C.P. (2010). 'Cognitive consequences of cooperative breeding in primates?' *Animal Cognition, 13(1)*. 1–19.

Burkart, J.M., van Schaik, C.P. (2016). 'Revisiting the consequences of cooperative breeding'. *J. Zool. (Lond.), 299.* 77–83.

Burkart, J.M., Hrdy, S.B., van Schaik, C.P. (2009). 'Cooperative breeding and human cognitive evolution'. *Evol. Anthropol., 18.* 175–186.

Butti, C., Sherwood, C.C., Hakeem, A.Y., (2009). 'Total number and volume of Von Economo neurons in the cerebral cortex of cetaceans'. *J Comp Neurol 515.* 243-259.

Cairó, O. (2011). 'External Measures of Cognition'. *Frontiers in human neuroscience, 5*. 108. 10.3389/fnhum.2011.00108.

Caldwell, M. C., Caldwell, D. K. (1965). 'Individualized whistle contours in bottlenosed dolphins (*Tursiops truncatus*)'. *Nature, 207.* 434-435.

Call, J., Tomasello, M. (1998). 'Distinguishing intentional from accidental actions in orangutans (*Pongo pygmaeus*), chimpanzees (*Pan troglodytes*), and human children (*Homo sapiens*)'. *Journal of Comparative Psychology, 112(2).* 192–206.

Candiotti, A., Zuberbühler, K., Lemasson, A. (2012). 'Context-related call combinations in female Diana monkeys'. *Animal cognition*, *15*(3). 327–339. https://doi.org/10.1007/s10071-011-0456-8

Carls-Diamante, S. (2022). ‘Where Is It Like to Be an Octopus?’ *Front Syst Neurosci.,14*. 16:840022. doi: 10.3389/fnsys.2022.840022. PMID: 35401127, PMCID: PMC8988249.

Caro, T. M., Hauser, M. D. (1992). ‘Is There Teaching in Nonhuman Animals?’. *The Quarterly Review of Biology*, *67(2*). 151–174.

Carrillo, M., Han, Y., Migliorati, F., Liu, M., Gazzola, V., Keysers, C. (2019). ‘Emotional Mirror Neurons in the Rat's Anterior Cingulate Cortex’. *Curr Biol., 29(8*). 1301–1312.

Cavalli-Sforza, L. L. (1997). 'Genes, peoples, and languages'. *Proceedings of the National Academy of Sciences, 94.* 7719-7724.

Cazau, D., Adam, O., Laitman, J., Reidenberg, J. (2013). 'Understanding the intentional acoustic behavior of humpback whales: A production-based approach'. *The Journal of the Acoustical Society of America, 134*. 2268-73. 10.1121/1.4816403.

Chambers, J. K., & Trudgill, P. (1998). *Dialectology*. Cambridge: Cambridge University Press.

Chapman, C.A., Chapman, L.J., Wrangham, R.W. (1995). ‘Ecological constraints on group size: an analysis of spider monkey and chimpanzee subgroups’. *Behavioral Ecology and Sociobiology, 36(1).* 59–70.

Chen, P. (2009). ‘Aspects of referentiality’. *Journal of Pragmatics, 41(8)*. 1657–1674.

Chen, J., ten Cate, C. (2017). ‘Bridging the gap: Learning of acoustic nonadjacent dependencies by a songbird’. *Journal of Experimental Psychology: Animal Learning and Cognition, 43*(3). 295–302. https://doi.org/10.1037/xan0000145

Cherniak, C. (2009). 'Brain wiring optimization and nongenomic nativism'. In Piattelli- Palmarini, M., Uriagereka, J., Salaburu, P. (Eds.), *Of Minds and Language. A Dialogue with Noam Chomsky in the Basque Country.* New York: Oxford University Press. 108-119.

Cholewiak, D., Sousa-Lima, R., Cerchio, S. (2013). ‘Humpback whale song hierarchical structure: Historical context and discussion of current classification issues’. *Marine Mammal Science*, 29. E312-E332. 10.1111/mms.12005.

Chomsky, N. (1956). 'Three models for the description of language'. *IRE Transcriptions of Information Theory, 2*. 113-124.

Chomsky, N. (1957). *Syntactic structures*. Paris: Mouton.

Chomsky, N. (1959). Review of B. F. Skinner's Verbal Behavior. *Language, 35*(1), 26-58.

Chomsky, N. (1965). *Aspects of the Theory of Syntax*. Cambridge, MA: MIT Press.

Chomsky, N. (1968). *Language and Mind.* New York: Harper and Row.

Chomsky, N. (1980). *Rules and representations*. New York: Columbia University Press.

Chomsky, N. (1981). *Lectures on Government and Binding.* Dordrecht: Foris Publications.

Chomsky, N. (1986). *Knowledge of Language: Its Nature, Origin, and Use*. New York: Praeger Publishers.

Chomsky, N. (1995). *The Minimalist Program.* Cambridge, MA: MIT Press.

Chomsky, N. (2001). ‘Beyond explanatory adequacy’. *MIT Occasional Papers in Linguistics, 20*.

Chomsky, N. (2005). ‘Three factors in language design’. *Linguistic Inquiry, 36*(1), 1-22.

Chomsky, N. (2007). ‘Approaching UG from below’. In Gartner, H.M., Sauerland, U., *Interfaces + recursion = language?: Chomsky’s minimalism and the view from syntax-semantics.* Berlin/New York: De Gruyter (pp. 1–29).

Chomsky N. (2010). ‘Some simple evo devo theses: how true might they be for language?’. In Larson, R.K., Déprez ,V., Yamakido, H., (Eds.). *The Evolution of Human Language: Biolinguistic Perspectives*. Approaches to the Evolution of Language. Cambridge: University Press (pp. 45–62).

Chomsky, N., Halle, M. (1968). *The Sound Pattern of English*. New York: Harper & Row.

Christiansen, M.H., Chater, N. (2015). ‘The language faculty that wasn’t: a usage-based account of natural language recursion’. *Front. Psychol., 6*. 1182.

Chung, D. (2016). ‘The Basic Principles of Kin Sociality and Eusociality: Human Evolution’. *Natural Science, 8.* 8–19. doi: 10.4236/ns.2016.81002.

Chusyd, D.E., Ackermans, N.L., Austad, S.N., Hof, P.R., Mielke, M.M., Sherwood, C.C., Allison, D.B. (2021). ‘Aging: What We Can Learn From Elephants’. *Front Aging, 2*. 726714. doi: 10.3389/fragi.2021.726714. PMID: 35822016, PMCID: PMC9261397.

Clark, C. W. (1983) 'Communication and behavior of the Southern Right Whale (*Eubalaena australis*). In Payne, R. (Eds.), *Communication and behavior of whales, AAAS selected symposium 76*. Boulder: Westview Press. 163-198.

Clark, C. W., Ellison, W. T. (2004). 'Potential Use of Low-Frequency Sound by Baleen Whales for Probing the Environment: Evidence from Models and Empirical Measurements.' In Thomas, J., Moss, C., Water, M. (Eds.), *Echolocation in Bats and Dolphins.* Chicago: University Press. 564-581.

Cockburn, A. (2006) ‘Prevalence of different modes of parental care in birds’. *Proc. R. Soc. B., 273.* 1375–1383. http://doi.org/10.1098/rspb.2005.3458

Collier, K., Townsend, S. W. & Manser, M. B. (2017). ‘Call concatenation in wild meerkats’. *Animal Behaviour, 134*. 257–269.

Collins, C. (2002). ‘Eliminating Labels’. In D. Epstein & D. Seely (eds.), *Derivation and Explanation in the Minimalist Program*, 42-64. Oxford: Blackwell.

Conde-Valverde, M., Martínez, I., Quam, R.M., Rosa, M., Velez, A.D., Lorenzo, C., Jarabo, P., Bermúdez de Castro, J.M., Carbonell, E., Arsuaga, J.L. (2021). 'Neanderthals and *Homo sapiens* had similar auditory and speech capacities.' *Nature Ecology Evolution. 1-7.*

Connor, R.C., Mann, J., Tyack, P.L., Whitehead, H. (1998). 'Social evolution in toothed whales'. *Trends Ecol Evol. 13(6)*. 228–232.

Connor, R.C., Heithaus, M.R., Barre, L.M. (2001). ‘Complex social structure, alliance stability and mating access in a bottlenose dolphin ‘super-alliance’’. *Proc. R. Soc. Lond. B, 268*. 263-267

Cook, M. L. H., Sayigh, L. S., Blum, J. E., Wells, R. S. (2004). 'Signature-whistle production in undisturbed free-ranging bottlenose dolphins (*Tursiops truncatus*)'. *Proc. R. Soc. B, 271.* 1043-1049.

Corballis, M.C. (2011). *The Recursive Mind: The Origins of Human Language, Thought, and Civilization*. Princeton, NJ: Princeton University Press.

Corballis, M.C. (2019). ‘Language, Memory, and Mental Time Travel: An Evolutionary Perspective’. *Frontiers in Human Neuroscience, 13*. 217. 10.3389/fnhum.2019.00217.

Corkeron, P. J., Connor, R. C. (1999). 'Why do baleen whales migrate?'. *Marine Mammal Science, 15 (4).* 1228-1245.

Coseriu, E. (1980). “Historische Sprache” und “Dialekt”. In Göschel, J, Ivić, P., Kehr, K. (Eds.): *Dialekt und Dialektologie*. Steiner, Wiesbaden, 106–22.

Coudé, G., Ferrari, P.F., Rodà, F., Maranesi, M., Borelli, E., Veroni, V., Monti, F., Rozzi, S., Fogassi, L. (2011). ‘Neurons controlling voluntary vocalization in the macaque ventral premotor cortex’. *PLoS One, 6.* e26822.

Coye, C., Townsend, S.W., Lemasson, A. (2018). ‘From animal communication to linguistics and back: insight from combinatorial abilities in monkeys and birds’. In Boë, L., Fagot, J., Perrier, P., Schwarz, J.L. (Eds.), *Origins of Human Language: Continuities and Discontinuities with Nonhuman Primates*, pp. 187–232. Bern: Peter Lang.

Crockford, C., Wittig, R.M., Zuberbühler, K. (2017). ‘Vocalizing in chimpanzees is influenced by social-cognitive processes’. *Science advances*, *3*(11). e1701742.

Croft, W. (2015). ‘Functional approaches to grammar’. In Wright, J. (Ed.). *International Encyclopedia of the Social & Behavioral Sciences*. Elsevier.

Croft, W., Cruse, D.A. (2004). *Cognitive Linguistics*. Cambridge: Cambridge University Press.

Cum, M., Santiago Pérez, J.A., Wangia, E. (2024). ‘A systematic review and meta-analysis of how social memory is studied’. *Sci Rep*, 14. 2221.

Dąbrowska, E. (2015). ‘What exactly is universal grammar, and has anyone seen it?’ *Front. Psychol., 6*. 852.

Dancygier, B. (Ed.) (2017). *The Cambridge handbook of cognitive linguistics.* Cambridge: University Press.

Darwin , C. (1871). *The descent of man and selection in relation to sex*. London: John Murray.

Davies, J.R., Garcia-Pelegrin, E. (2023). ‘Bottlenose dolphins are sensitive to human attentional features, including eye functionality’. *Sci Rep, 13*. 12565. https://doi.org/10.1038/s41598-023-39031-7

Davies, J.R., Garcia-Pelegrin, E., Baciadonna, L., Pilenga, C., Favaro, L., Clayton, N.S. (2022). ‘Episodic-like memory in common bottlenose dolphins’. *Curr Biol., 32(15)*. 3436–3442.

Dawkins, R., Pyle, L. (1986). *The blind watchmaker*. London: Longman.

Dawkins, R. (2006). *The Selfish Gene* (30th anniversary ed.). London: Oxford University Press.

de la Mora, D.M., Toro, J.M. (2013). ‘Rule learning over consonants and vowels in a non-human animal’. *Cognition, 126 (2)*. 307-312.

Deacon, T. W. (1997). *The Symbolic Species: The Co-Evolution of Language and the Brain*. New York, NY: W. W. Norton & Company.

Deecke, V. B., Ford, J. K. B., Spong, P. (2000). 'Dialect change in resident killer whales: implications for vocal learning and cultural transmission'. *Animal Behaviour, 40.* 629-638.

Deecke, V. B., Slater, P. J. B., Ford, J. K. B. (2002). 'Selective habituation shapes acoustic predator recognition in harbour seals'. *Nature, 420*. 171-173.

Delfour, F., Marten, K. (2001). 'Mirror image processing in three marine mammal species: killer whales (*Orcinus orca*), false killer whales (*Pseudorca crassidens*) and California sea lions (*Zalophus californianus*)'. *Behavioural processes*, *53*(3). 181–190.

Dennett, D.C. (1983). 'Intentional systems in cognitive ethology: the 'Panglossian paradigm' defended'. *Behav. Brain Sci, 6*. 343–390. (doi:10.1017/ S0140525X00016393)

Dere, E., Kart-Teke, E., Huston, J.P., De Souza Silva, M.A. (2006). 'The case for episodic memory in animals'. *Neuroscience & Biobehavioral Reviews, 30(8).* 1206–1224.

Derex, M. (2022). 'Human cumulative culture and the exploitation of natural phenomena'. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, *377(1843)*. 20200311. https://doi.org/10.1098/rstb.2020.0311

Di Sparti, A. (2006). 'I cocci di Babele. Tra metafore e neurolinguistica.' In Di Sparti, A. (Ed.), *Tradurre. I cocci di Babele*. Palermo: Compostampa. (pp. 7-16).

Dias Martins, M. (2012). 'Distinctive signatures of recursion'. *Phil. Trans. R. Soc., 367.* 2055-2064.

Dias Martins, M. (2014). 'Recursion is not language domain-specific: interim results of a research program'. In Cartmill, E., Roberts, S., Lyn, H., Cornish, H. (Eds.), *The Evolution of language. Proceedings of the 10th International conference*. 177-184.

Diaz, A.L., Gleeson, J.G. (2009). 'The molecular and genetic mechanisms of neocortex development'. *Clinics in Perinatology, 36(3)*. 503–12.

Dik, S.C. (1981). *Functional Grammar*. Dordrecht/Cinnaminson NJ: Foris.

Dik, S.C. (1997). *The Theory of Functional Grammar, Part 1: The Structure of the Clause* (2 ed.). Berlin: Mouton de Gruyter.

Dirks, P.H., Berger, L.R., Roberts, E.M., Kramers, J.D., *et al*. (2015). 'Geological and taphonomic context for the new hominin species *Homo naledi* from the Dinaledi Chamber, South Africa'. *eLife*, *4*, e09561. https://doi.org/10.7554/eLife.09561

Dirks, P.H., Roberts, E.M., Hilbert-Wolf, H., Kramers, J.D., *et al*. (2017). 'The age of *Homo naledi* and associated sediments in the Rising Star Cave, South Africa'. *eLife*, *6*. e24231. https://doi.org/10.7554/eLife.24231

Dixon, M.M., Jones, P.L., Ryan, M.J., Carter, G.G., Page, R.A. (2022). 'Long-term memory in frog-eating bats'. *Curr Biol., 32(12*). R557–R558.

Doh, Y., Cazau, D., Lamaj, G., *et al*. (2025). 'Study of Non-Linearities in Humpback Whale Song Units'. *Journal of Marine Science and Engineering, 13*. 215. 10.3390/jmse13020215.

dos Santos, M.E.; Ferreira, A.J.; Harzen, S. (1995). 'Rhythmic sound sequences emitted by aroused bottlenose dolphins in the Sado Estuary, Portugal'. In Evans, P.G.H., Nice, H. (Eds.), *Sensory Systems of Aquatic Mammals.* De Spil Publishers: Woerden, The Netherlands. 325–334.

Drew, P., Heritage, J. (2006). *Conversation Analysis*. Vol. I. London: SAGE Publications Ltd. (pp. xxxiv).

Dudzinski, Kathleen & Sakai, Mai & Masaki, K. & Kogi, K. & Hishii, T. & Kurimoto, M. (2003). 'Behavioural observations of bottlenose dolphins towards two dead conspecifics'. *Aquatic Mammals, 29*. 108-116. 10.1578/016754203101023951.

Emery, N. J. (2004). 'Are corvids 'feathered apes'? Cognitive evolution in crows, jays, rooks and jackdaws'. In S. Watanabe (Ed.) *Comparative analysis of minds*, pp. 181–213. Tokyo, Japan: Keio University Press.

Egeland, C.P., Domínguez-Rodrigo, M., Pickering, T.R., Menter, C.G., Heaton, J.L. (2018). 'Hominin skeletal part abundances and claims of deliberate disposal of corpses in the Middle Pleistocene'. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(18). 4601–4606. https://doi.org/10.1073/pnas.1718678115

El-Amary, N., Balbaa, A., Swief, R., Abdel-Salam, T. (2018). 'A Reconfigured Whale Optimization Technique (RWOT) for Renewable Electrical Energy Optimal Scheduling Impact on Sustainable Development Applied to Damietta Seaport, Egypt'. *Energies, 11*. 535.

Ellis, N.C. (1998). 'Emergentism, Connectionism, and Language Learning'. *Language Learning, 48*. 631–664.

Ellis, Nick & Wulff, Stefanie. (2015). Usage-based approaches in second language acquisition. In VanPatten, B., Williams, J. (Eds.) *Theories in Second Language Acquisition: An Introduction*. New York: Routledge (pp.75-93).

Embick, D., & Noyer, R. (2001). 'Movement Operations after Syntax'. *Linguistic Inquiry, 32*. 555-595.

Emery, N., Clayton, N. (2005). 'The Mentality of Crows: Convergent Evolution of Intelligence in Corvids and Apes'. *Science, 306*. 1903-7. 10.1126/science.1098410.

Emery, N.J. (2006). 'Cognitive ornithology: the evolution of avian intelligence'. *Philos Trans R Soc Lond B Biol Sci, 361(1465)*. 23-43.

Enard, W., Przeworski, M., Fisher, S., Lai, C.S.L., Wiebe, V., Kitano, T., Monaco, A., Pääbo, S. (2002). 'Molecular Evolution of FOXP2, a Gene Involved in Speech and Language'. *Nature, 418*. 869–72. 10.1038/nature01025.

Engesser, S., Crane, J.M.S., Savage, J.L., Russell, A.F., Townsend, S.W. (2015). 'Experimental evidence for phonemic contrasts in a nonhuman vocal system'. *PLoS Biology, 13*. e1002171.

Engesser, S., Ridley, A., Townsend, S.W. (2016). 'Meaningful call combinations and compositional processing in the southern pied babbler'. *Proceedings of the National Academy of Sciences of the United States of America, 113*.

Engesser, S., & Townsend, S. W. (2019). 'Combinatoriality in the vocal systems of nonhuman animals'. *Wiley Interdisciplinary Reviews: Cognitive Science, 10*(4), e1493.

Evans, N., Levinson, S. (2009). 'The myth of language universals: Language diversity and its importance for cognitive science'. *Behavioral and Brain Sciences, 32*. 429-448.

Everett, D. (2005). 'Cultural Constraints on Grammar and Cognition in Pirahã: Another Look at the Design Features of Human Language'. *Current Anthropology, 46.* 621-646.

Evrard, H.C., Forro, T., Logothetis, N.K. (2012). 'Von Economo Neurons in the Anterior Insula of the Macaque Monkey'. *Neuron, 74(3)*. 482–489.

Fedurek, P., Machanda, Z. P., Schel, A. M., & Slocombe, K. E. (2013). 'Pant hoot chorusing and social bonds in male chimpanzees'. *Animal Behaviour, 86*(1). 189-196. https://doi.org/10.1016/j.anbehav.2013.05.010

Fedutin, I. D., Filatova, O. A., Mamaev, E. G. *et al*. (2015). 'Occurrence and social structure of Baird's beaked whales, *Berardius bairdii*, in the Commander Islands, Russia'. *Mar Mam Sci, 31.* 853-865.

Feeney, M.C., Roberts, W.A., Sherry, D.F. (2009). 'Memory for what, where, and when in the black-capped chickadee (*Poecile atricapillus*)'. *Anim. Cognit., 12*. 767–777.

Ferretti, F., Adornetti, I., Chiera, A., Cosentino, E., Nicchiarelli, S. (2018). 'Introduction: origin and evolution of language: an interdisciplinary perspective'. *Topoi, 37*. 219–234.

Ferrigno, S., Cheyette, S.J., Piantadosi, S.T., Cantlon, J.F. (2020). 'Recursive sequence generation in monkeys, children, U.S. adults, and native Amazonians'. *Sci. Adv., 6*. eaaz1002.

Filatova, O., Hoyt, E., Burdin, A. (2010). 'Horizontal transmission of call features in killer whale (*Orcinus orca*) dialects'. *Biology Bulletin, 37.* 965-971.

Fischer, J., Hammerschmidt, K. (2019). 'Towards a new taxonomy of primate vocal production learning'. *Phil. Trans. R. Soc. B, 375*. 20190045. http://dx.doi.org/10.1098/rstb.2019.0045

Fishlock, V., Lee, P.C. (2013). 'Forest elephants: fission–fusion and social arenas'. *Animal Behaviour*, *85(2).* 357–363.

Fishman, J. (1972). *The Sociology Of Language: An Interdisciplinary Social Science Approach to Language in Society*. New York: Newbury House Publishers.

Fitch, W. T. (2000). 'The evolution of speech: A comparative review'. *Trends in Cognitive Science, 4*. 258-267.

Fitch, W. T. (2010). *The Evolution of Language*. Cambridge: Cambridge University Press.

Fitch, W. T. (2013). 'Musical Protolanguage: Darwin's Theory of Language Evolution Revisited'. In Bolhuis J., Everaert, M. (Eds.), *Birdsong, Speech, and Language: Exploring the Evolution of Mind and Brain*. Cambridge, MA: MIT Press. 489-503.

Fitch, W. T. (2017). 'Empirical approaches to the study of language evolution'. *Psychon. Bull. Rev, 24.* 3–33.

Fitch, W. T., Hauser, M. D., Chomsky, N. (2005). 'The evolution of language faculty: clarifications and implications'. *Cognition, 97*. 179-210.

Flower, T.P., Gribble, M., Ridley, A.R. (2014). 'Deception by flexible alarm mimicry in an African bird'. *Science, 344(6183).* 513-6. doi: 10.1126/science.1249723.

Fodor, J.A. (1983). *The Modularity of Mind: An Essay on Faculty Psychology*. Cambridge, MA: MIT Press.

Fongaro, E., Rose, J. (2020). 'Crows control working memory before and after stimulus encoding'. *Sci Rep*, *10*. 3253.

Ford, J. K. B. (1991). 'Vocal traditions among resident killer whales (*Orcinus orca*) in coastal waters of British Columbia'. *Can J Zool, 69.* 1454-1483.

Fournet, M., Gabriele, C., Culp, D. *et al*. (2018). 'Some things never change: multi-decadal stability in humpback whale calling repertoire on Southeast Alaskan foraging grounds'. *Scientific Reports, 8*.

Freeberg, T.M., Dunbar, R.I.M., Ord, T.J. (2012). 'Social complexity as a proximate and ultimate factor in communicative complexity'. *Philos Trans R Soc Lond B Biol Sci. 367.* 1785–1801.

Friedmann, N., Belletti, A. & Rizzi, L., (2021) "Growing trees: The acquisition of the left periphery", *Glossa: a journal of general linguistics* 6(1). 131.
doi: https://doi.org/10.16995/glossa.5877

Fripp, D., Owen, C., Quintana-Rizzo, E. *et al*. (2005). 'Bottlenose dolphin (*Tursiops truncatus*) calves appear to model their signature whistles on the signature whistles of community members'. *Animal Cognition*, *8* (1). 17-26.

Fröhlich, M., van Schaik, C.P., van Noordwijk, M.A., Knief, U. (2022). 'Individual variation and plasticity in the infant-directed communication of orang-utan mothers'. *Proceedings. Biological sciences*, *289(1975)*. 20220200. https://doi.org/10.1098/rspb.2022.0200

Frohoff, T., Oriel, E. (2017). 'Conversing with Dolphins: The Holy Grail of Interspecies Communication?' In Gordon, R., Seckbach, J. (Eds.), *Biocommunication: Sign- Mediated Interactions between Cells and Organisms*. Singapore: World Scientific Publishing Co.

Frumhoff, P. (1983). 'Aberrant songs of humpback whales (*Megaptera novaeangliae*): Clues to the structure of humpback songs'. In Payne, R. (Eds.), C*ommunication and behavior of whales. AAAS Selected Symposium, 76*. Boulder: Westview Press. 81-127.

Fujita, K., Morisaki, A., Takaoka, A., Maeda, T., Hori, Y. (2012). 'Incidental memory in dogs (*Canis familiaris*): adaptive behavioral solution at an unexpected memory test'. Anim. Cogn., 15. 1055–1063.

Gallese, V., Fadiga, L., Fogassi, L., Rizzolatti, G. (1996). 'Action recognition in the premotor cortex'. *Brain, 119(2).* 593-609.

Gallup, G.G. Jr. (1970). 'Chimpanzees: Self recognition'. *Science, 167(3914).* 86–87.

Gardner, R. A., Gardner, B.T., Van Cantfort, T.E. (1989). *Teaching Sign Language to Chimpanzees*. State University of New York Press.

Gariépy, J., Watson, K., Du, E., Xie, D., Erb, J., Amasino, D., Platt, M. (2014). 'Social learning in humans and other animals'. *Frontiers in neuroscience, 8*. 58. 10.3389/fnins.2014.00058.

Garland, E. C., Rendell, L., Lamoni, L., Poole, M. M., Noad, M. J. (2017). 'Song hybridization events during revolutionary song change provide insights into cultural transmission in humpback whales'. *Proceedings of the National Academy of Sciences of the United States of America*, *114*(30), 7822–7829. https://doi.org/10.1073/pnas.1621072114

Garland, E., Mcgregor, P. (2020). 'Cultural Transmission, Evolution, and Revolution in Vocal Displays: Insights From Bird and Whale Song'. *Frontiers in Psychology, 11*. 10.3389/fpsyg.2020.544929.

George, J. C., Bada, J., Zeh, J. *et al*. (1999). 'Age and growth estimates of bowhead whales (*Balaena mysticetus*) via aspartic acid racemization'. *Canadian Journal of Zoology*, 77 (4). 571-580.

Gero, S., Bøttcher, A., Whitehead, H., Madsen, P.T. (2016). 'Socially segregated, sympatric sperm whale clans in the Atlantic Ocean'. *R. Soc. open sci., 3*: 160061.

Ghalambor, C.K., Angeloni, L.M., Carroll, S.P. (2010). 'Behavior as phenotypic plasticity'. In Westneat, D.F., Fox, C.W. (Eds.), *Evolutionary behavioral ecology* (pp. 90e107). New York: Oxford University Press.

Gil, D. (2022). 'Bare and Constructional Compositionality'. *International Journal of Primatology.* https://doi. org/ 10. 1007/ s10764-022-00343-6.

Gilbert, S.F. (2003). 'The morphogenesis of evolutionary developmental biology'. *International Journal of Developmental Biology, 47(7-8).* 467–477.

Gill, S.A., Bierema, A.M.K. (2013). 'On the meaning of alarm calls: a review of functional reference in avian alarm calling'. *Ethology*, *119*(6). 449-461.

Girard-Buttoz, C., Zaccarella, E., Bortolato, T. *et al*. (2022). 'Chimpanzees produce diverse vocal sequences with ordered and recombinatorial properties'. *Commun Biol* 5, 410. https://doi.org/10.1038/s42003-022-03350-8

Givón, T. (1995). *Functionalism and Grammar.* Amsterdam: John Benjamins.

Goldberg, A.E. (1995). *Constructions: a construction grammar approach to argument structure*. Chicago, London: The University of Chicago Press.

Goldberg, A.E. (2006). *Constructions at Work: The Nature of Generalization in Language*. Oxford: Oxford University Press.

Gómez-Moreno, J. (2019). 'The 'Mimic' or 'Mimetic' Octopus? A Cognitive-Semiotic Study of Mimicry and Deception in *Thaumoctopus Mimicus'*. *Biosemiotics.* 10.1007/s12304-019-09362-y.

Gontier, N., Hartmann, S., Pleyer, M., Rodrigues, E. (2024). 'Combinatoriality and Compositionality in Communication, Skills, Tool Use, and Language'. *International Journal of Primatology, 10*.1007/s10764-024-00416-8.

Graffi, G. (2019). 'Origin of language and origin of languages'. *Evolutionary Linguistics Theory, 1:1,* 2019: 6-23.

Graffi, G., Scalise, S. (2002). *Le lingue e il linguaggio. Introduzione alla linguistica*. Bologna: Il Mulino.

Graham, K.E., Wilke, C., Lahiff, N.J., Slocombe, K.E. (2019). 'Scratching beneath the surface: intentionality in great ape signal production'. *Phil. Trans. R. Soc. B, 375*. 20180403. dx.doi.org/10.1098/rstb.2018.040
Greco, B., Brown, T., Andrews, J., Swaisgood, R., Caine, N. (2012). 'Social learning in captive African elephants (*Loxodonta africana africana*)'. *Animal cognition, 16.* 10.1007/s10071-012-0586-7.

Green, P., Brandley, N., Nowicki, S. (2020). 'Categorical perception in animal communication and decision-making'. *Behavioral Ecology, 31*(4) 859–867. 10.1093/beheco/araa004.

Green, R. E., Krause, J., Briggs, A.W., Maricic, T., Stenzel, U., Kircher, M., Patterson, N., Li, H., Zhai, W., Fritz, M.H., Hansen, N.F., Durand, E.Y., Malaspinas, A.S., Jensen, J.D., Marques-Bonet, T., Alkan, C., Prüfer, K., Meyer, M., Burbano, H.A., Good, J.M., … Pääbo, S. (2010). 'A draft sequence of the Neanderthal genome.' *Science,* 328. 710–722.

Greenfield, P.M., Savage-Rumbaugh, E.S. (1990). 'Grammatical combination in Pan paniscus: Processed of learning and invention in the evolution and development of language'. In Parker,

S.T., Gibson, K.R. (Eds.) *"Language" and Intelligence in Monkeys and Apes: Comparative developmental perspectives*, pp. 540–578. New York: Cambridge University Press.

Grice, P. (1975). 'Logic and Conversation'. In Cole, P., Morgan, J.J., *Syntax and Semantics 3: Speech Acts*. New York, NY: Academic Press (pp. 41–58).

Grove, M., Pearce, E., Dunbar, R. (2012). 'Fission-fusion and the evolution of hominin social systems'. *Journal of Human Evolution, 62(2)*. 191–200.

Groves, C.P., Napier, J.R. (2024, March 14). 'primate'. *Encyclopedia Britannica*. https://www.britannica.com/animal/primate-mammal

Guinet, C. (1991). 'Intentional stranding apprenticeship and social play in killer whales (*Orcinus orca*)'. *Canadian Journal of Zoology. 69*. 2712-2716.

Guinet, C., Bouvier, J. (1995). 'Development of intentional stranding hunting techniques in killer whale (*Orcinus orca*) calves at Crozet Archipelago'. *Canadian Journal of Zoology. 73.* 27-33.

Hakeem, A.Y., Sherwood, C.C., Bonar, C.J., Butti, C., Hof, P.R., Allman, J.M. (2009). 'Von Economo neurons in the elephant brain'. *The Anatomical Record, 292(2).* 242–248.

Halle, M., Marantz, A. (1993). 'Distributed Morphology and the Pieces of Inflection'. In Hale, K., Jay Keyser, S. (Eds.), *The view from building 20*. Cambridge, MA: MIT Press. 111-176.

Halliday, M.A.K. (1984). *A Short Introduction to Functional Grammar*. London: Arnold.

Halliday, M.A.K., Matthiessen, C. (2014) *An introduction to functional grammar* (4th edition). Oxon: Routledge.

Hanks, W.F. (1999). 'Indexicality'. *Journal of Linguistic Anthropology, 9 (1/2).* 124–126.

Hardus, M.E., Lameira, A.R., Van Schaik, C.P., Wich, S.A. (2009). 'Tool use in wild orang-utans modifies sound production: A functionally deceptive innovation?' *Proc. Biol. Sci., 276.* 3689–3694.

Harley, H. (2014). 'On the identity of roots'. *Theoretical Linguistics*, *40*(3-4). 225-276. https://doi.org/10.1515/tl-2014-0010

Haspelmath, M. (1999). 'Why is grammaticalization irreversible?'. *Linguistics*, *37*(6). 1043-1068.

Haug, H. (1987). 'Brain sizes, surfaces, and neuronal sizes of the cortex cerebri: a stereological investigation of man and his variability and a comparison with some mammals (primates, whales, marsupials, insectivores, and one elephant)'. *Am. J. Anat, 180*. 126–142

Hauser, M.D. (1989). 'Ontogenetic changes in the comprehension and production of vervet monkey (*Cercopithecus aethiops*) vocalizations'. J. Comp. Psychol., 103. 149–158.

Hauser, M.D., Chomsky, N., Fitch, W.T. (2002). 'The faculty of language: What is it, who has it, and how did it evolve?' *Science, 298.* 1569–1579.

Hayes, C. (1951). *The Ape in Our House*. New York: Harper.

Heberlein, M.T.E., Manser, M.B., Turner, D.C. (2017). 'Deceptive-like behaviour in dogs (*Canis familiaris*)'. *Anim Cogn., 20(3).* 511-520. doi: 10.1007/s10071-017-1078-6.

Hedwig, D., Mundry, R., Robbins, M.M., Boesch, C. (2015). ‘Contextual correlates of syntactic variation in mountain and western gorilla close-distance vocalizations: Indications for lexical or phonological syntax?’ *Anim. Cogn., 18*. 423–435.

Heinrich, B., Marzluff, J.M. (1991). ‘Do common ravens yell because they want to attract others?’ *Behavioral Ecology and Sociobiology, 28*. 13-21.

Helfman, G.S, Schultz, E.T. (1984). ‘Social transmission of behavioural traditions in a coral reef fish’. *Animal Behaviour, 32(2).* 379–384.

Hemila, S., Nummela, S., Reuter, T. (2001). 'Modeling Whale Audiograms: Effects of Bone Mass on High- Frequency Hearing.' *Hear. Res., 151.* 221-226.

Hengeveld, K., Mackenzie, J.L. (2008). *Functional Discourse Grammar: A Typologically-Based Theory of Language Structure.* Oxford: University Press.

Herculano-Houzel, S. (2017). *The Human Advantage: How our brains became remarkable*. Cambridge, MA: MIT Press.

Herman, L.M., Richards, D.G., Wolz, J.P. (1984). 'Comprehension of sentences by bottlenosed dolphins'. *Cognition, 16*. 129-219.

Herman, L.M., Forestell, P.H. (1985). ‘Reporting presence or absence of named objects by a language-trained dolphin’. Rutgers University Symposium: The question of animal cognition (1985, New Brunswick, New Jersey). *Neuroscience and Biobehavioral Reviews, 9*. 667–681.

Herman, L. M., Kuczaj II, S. A., Holder, M. D. (1993). 'Responses to anomalous gestural sequences by a language-trained dolphin: Evidence for processing of semantic relations and syntactic information'. *J Exp Psychol, 122*. 184-194.

Herman, L.M., Uyeyama, R.K. (1999). 'The dolphin's grammatical competency: Comments on Kako (1999)'. *Animal Learning & Behavior, 27*(1).

Herman, L.M., Pack, A.A. (2001). ‘Laboratory evidence for cultural transmission mechanisms.’ *Behavioral and Brain Sciences, 24*. 335–36.

Herman, L. M. (2006). 'Intelligence and rational behaviour in the bottlenosed dolphin'. In Hurley, S., Nudds, M. (Eds.), *Rational animals.* Oxford: University Press. 439-467.

Herman, L. M. (2012). 'Body and self in dolphins'. *Conscious Cognition, 21(1)*. 526-545.

Hersh, T.A., Gero, S., Rendell, L., Cantor, M., Weilgart, L., Amano, M., *et al*. (2022). ‘Evidence from sperm whale clans of symbolic marking in non-human cultures’. *Proc. Natl. Acad. Sci., 119*. e2201692119. doi: 10.1073/pnas.2201692119

Herzing, D., Johnson, C. M. (1997). 'Interspecific interactions between Atlantic Spotted dolphins (*Stenella frontalis*) and Bottlenose dolphins (*Tursiops truncatus*) in the Bahamas, 1985–1995'. *Aquatic Mammals, 23*. 85-99.

Herzing, D., Delfour, F., Pack, A. (2012). ‘Responses of Human-Habituated Wild Atlantic Spotted Dolphins to Play Behaviors Using a Two-Way Human/Dolphin Interface’. *International Journal of Comparative Psychology, 25*. 137-165.

Herzing, D. L. (2015). 'Making sense of it all: Multimodal dolphin communication'. In Herzing, D. L., Johnson, C. M. (Eds.), *Dolphin communication and cognition: Past, present, and future*. Boston: MIT Press. 139-171.

Hildebrandt, T., Drews, B., Gaeth, A.P., Goeritz, F., Hermes, R., Schmitt, D., Gray, C., Rich, P., Streich, W. J., Short, R.V., Renfree, M.B. (2007). 'Foetal age determination and development in elephants'. *Proceedings. Biological sciences*, *274(1608)*. 323–331. https://doi.org/10.1098/rspb.2006.3738

Hillix, W.A., Rumbaugh, D.M. (2004). *Animal Bodies, Human Minds: Ape, Dolphin, and Parrot Language Skills*. (Developments in primatology). New York: Kluwer Academic/Plenum Publishers.

Hjelmslev, L. (1953) [1943]. *Prolegomena to a Theory of Language*. Baltimore: Indiana University Publications in Anthropology and Linguistics.

Hockett, C.F. (1958). *A course in modern linguistics*. New York: Macmillan.

Hockett, C.F. (1959). 'Animal 'languages' and human language'. *Human Biology, 31*. 32-39.

Hockett, C.F. (1960a). 'The origin of speech'. *Scientific American, 203*. 88-111.

Hockett, C.F. [1960b] (1977). Logical considerations in the study of animal communication. In the view from language: selected essays 1948–1974. Athens, GA: The University of Georgia Press. 124–162. Reprinted from: 'Animal "languages" and human language'. In W. E. Lanyon & W. N. Tavolga (eds.), *Animal sounds and communication* (pp. 392–430). Washington: American Institute of Biological Sciences.

Hockett, C.F. (1966). 'The problem of universals in language'. In J. Greenberg (ed.), *Universals of language* (pp. 1–29). Cambridge: The MIT Press.

Hockett, C.F., Altmann, S.A. (1968). 'A note on design features'. In T. Sebeok (ed.), *Animal communication: techniques of study and results of research* (pp. 61–72). Bloomington: Indiana University Press.

Hoeksema, N. *et al*. (2021). 'Neuroanatomy of the grey seal brain: bringing pinnipeds into the neurobiological study of vocal learning'. *Phil. Trans. R. Soc. B, 376*. 20200252. https://doi.org/10.1098/rstb.2020.0252

Hof, P. R., Chanis, R., Marino, L. (2005). 'Cortical complexity in cetacean brains'. *The anatomical record. Part A, Discoveries in molecular, cellular, and evolutionary biology. 287*. 1142-1152.

Hof, P. R., Van Der Gucht, E. (2007). 'Structure of the cerebral cortex of the humpback whale, *Megaptera novaeangliae* (Cetacea, Mysticeti, Balaenopteridae)'. *The Anatomical Record, 290.* 1-31.

Hoff, E. (2009). Language Development (4th ed.). Belmont, CA: Wadsworth Cengage Learning.

Hoffmann, D.L., Angelucci, D.E., Villaverde, V., Zapata, J., Zilhão, J. (2018). 'Symbolic use of marine shells and mineral pigments by Iberian Neanderthals 115.000 years ago'. *Science advances, 4(5255)*.

Hoffmann, D.L., Standish, C. D., García-Diez, M., Pettitt, P.B., Milton, J. A., Zilhão, J., Alcolea-González, J.J., Cantalejo-Duarte, P., Collado, H., de Balbín, R., Lorblanchet, M., Ramos-Muñoz, J., Weniger, G.C., Pike, A.W.G. (2018). 'U-Th dating of carbonate crusts reveal Neanderthal origin of Iberian cave art'. *Science, 359(6378)*. 912-915.

Hopper, P.J. (1987). 'Emergent grammar'. *Berkeley Linguistics Society, 13.* 139–157.

Hoppitt, W., Brown, G., Kendal, R., Rendell, L., Thornton, A., Webster, M., Lala, K. (2008). 'Lessons from animal teaching'. *Trends in ecology & evolution, 23*. 486-93. 10.1016/j.tree.2008.05.008.

Humboldt, W. (1998) [1836]. *Über die Verschiedenheit des menschlichen Sprachbaues und ihren Einfluss auf die geistige Entwicklung des Menschengeschlechts*. Paderborn: Schöningh.

Hunt, G., Gray, R., Taylor, A. (2013). 'Why is tool use rare in animals?'. In Sanz, C.M., Call, J., Boesch, C. *Tool Use in Animals Cognition and Ecology*. Cambridge: University Press.

Hurford, J.R. (1990). 'Nativist and functional explanations in language acquisition'. In Roca, I.M (Ed.), *Logical Issues in Language Acquisition*. Foris: Dordrecht: (pp. 85–136).

Hurford, J.R. (2003). 'The language mosaic and its evolution'. In M. Christiansen & K. Simon (eds.), *Language evolution* (pp. 38–57). New York: Oxford University Press.

Hurford, J.R. (2007). *The Origins of Meaning: Language in the Light of Evolution.* New York: Oxford University Press.

Hurford, J.R. (2011). *The origins of grammar: language in the light of evolution II*. New York: Oxford University Press.

Irish, J.D., Bailey, S.E., Guatelli-Steinberg, D., Delezene, L.K., Berger, L.R. (2018). 'Ancient teeth, phenetic affinities, and African hominins: Another look at where *Homo naledi* fits in'. *Journal of human evolution*, *122*. 108–123. https://doi.org/10.1016/j.jhevol.2018.05.007

Jaakkola, K., Fellner, W., Erb, L. *et al*. (2005). 'Understanding of the concept of numerically "less" by bottlenose dolphins (*Tursiops truncatus*)'. *Journal of comparative psychology,119*. 296-303.

Jaakkola, K., Guarino, E., Donegan, K., King, S. L. (2018) 'Bottlenose dolphins can understand their partner's role in a cooperative task'. *Proc. R. Soc. B, 285*: 20180948.

Jabr, F. (2012). 'Self-Awareness with a Simple Brain'. *Scientific American Mind, 23(5).* 28–29.

Jackendoff, R. (2002). *Foundations of language: Brain, Meaning, Grammar, Evolution*. Oxford: Oxford University Press.

Jackendoff, R. (2010). 'Your theory of language evolution depends on your theory of language.' In R. Larson, V. Déprez, H. Yamakido (eds.), *The evolution of Human Language: Biolinguistic Perspectives (Approaches to the Evolution of Language)*. 63-72. Cambridge: Cambridge University Press.

Jackendoff, R., Pinker, S. (2005). 'The nature of the language faculty and its implications for evolution of language (Reply to Fitch, Hauser e Chomsky)'. *Cognition, 97*. 211-225.

Jackendoff, R., Wittenberg, E. (2014). 'What can you say without syntax? A hierarchy of grammatical complexity.' In Newmeyer, F.J. e Preston, L.B. (Eds.), *Measuring Grammatical Complexity*. Oxford: University Press.

Jacobs, A. M. (2011). 'Pointing and Mind Reading: On *Origins of Human Communication* by Michael Tomasello'. *Journal of Literary Theory Online*.

Jakobson, R. (1960). 'Linguistics and Poetics'. In T. Sebeok (Ed.), *Style in Language* (pp. 350-377). Cambridge: MIT Press.

Janik, V. M., Slater, P. J. B. (1998). 'Context-specific use suggests that bottlenose dolphin signature whistles are cohesion calls'. *Animal Behaviour, 56.* 829-838.

Janik, V. M. (2000a). ‘Food-related bray calls in wild bottlenose dolphins (*Tursiops truncatus*)’. *Proc Biol Sci., 267*. 923-927.

Janik, V. M. (2000b). ‘Whistle matching in wild bottlenose dolphins (*Tursiops truncatus*)’. *Science, 289*(5483). 1355–1357.

Janik, V. M., Slater, P. (2000). ‘The different roles of social learning in vocal communication’. *Animal behaviour, 60*. 1–11. 10.1006/anbe.2000.1410.

Janik, V. M., Sayigh, L., Wells, R. (2006). 'Signature whistle shape conveys identity information to bottlenose dolphins'. *Proc Natl Acad Sci USA, 103*. 8293-8297

Janik, V. M., Sayigh, L.S. (2013). ‘Communication in bottlenose dolphins: 50 years of signature whistle research’. *J. Comp. Physiol. A, 199.* 479–489. doi: 10.1007/ s00359-013-0817-7

Janik, V. M., Knörnschild, M. (2021). ‘Vocal production learning in mammals revisited’. *Phil. Trans. R. Soc. B, 376.* 20200244. https://doi.org/10.1098/rstb.2020.0244

Jeon, H., Lee, S. H. (2018). 'From Neurons to Social Beings: Short Review of the Mirror Neuron System Research and Its Socio-Psychological and Psychiatric Implications'. *Clinical psychopharmacology and neuroscience: the official scientific journal of the Korean College of Neuropsychopharmacology, 16.* 18-31.

Jerison, H.J. (1973). *Evolution of the Brain and Intelligence*. Academic Press, New York.

Jespersen , O. (1922). *Language: Its nature, development and origin.* New York: Norton.

Jo, H., McCune, K.B., Jablonski, P.G. (2023). ‘Long-term memory of experienced jays facilitates problem-solving by naïve group members in the wild’. *Sci Rep*, *13*. 21593.

Johnson-Frey, S.H. (2003). ‘What’s so special about human tool use?’ *Neuron, 39*. 201–204.

Jones, B., Zapetis, M., Samuelson, M., Ridgway, S. (2020). ‘Sounds produced by bottlenose dolphins (*Tursiops*): a review of the defining characteristics and acoustic criteria of the dolphin vocal repertoire’. *Bioacoustics, 29 (4)*. 399-440.

Jones, B., Tufano, S., Daniels, R., (2021). ‘Non-stereotyped amplitude modulation across signature whistle contours’. *Behavioural Processes, 194.*

Jorgewich-Cohen, G., Townsend, S., Padovese, L., Klein, N., Praschag, P., Ferrara, C., Ettmar, S., Menezes, S., Varani, A., Serano, J., Sánchez-Villagra, M. (2022). ‘Common evolutionary origin of acoustic communication in choanate vertebrates’. *Nature Communications, 13*. 10.1038/s41467-022-33741-8.

Jozet-Alves, C., Bertin, M., Clayton, N.S. (2013). ‘Evidence of episodic- like memory in cuttlefish’. *Curr. Biol., 23*. R1033–1035.

Kappeler, P.M. (2019). ‘A framework for studying social complexity’. *Behav Ecol Sociobiol*, *73(13).* https://doi.org/10.1007/s00265-018-2601-8

Karg, K., Schmelz, M., Call, J., Tomasello, M. (2015). ‘Chimpanzees strategically manipulate what others can see’. *Anim Cogn., 18(5).* 1069-76. doi: 10.1007/s10071-015-0875-z.

Kashima, K., Ohtsuki, H., Satake, A. (2013). ‘Fission-fusion bat behavior as a strategy for balancing the conflicting needs of maximizing information accuracy and minimizing infection risk’. *Journal of Theoretical Biology*, *318.* 101–109.

Kaufmann, L., Becker, R., Ochs, A., Brecht, M. (2023). 'Elephant banana peeling'. *Current Biology, 33.* R257-R258. 10.1016/j.cub.2023.02.076.

Kendon, A. (2004). *Gesture: visible action as utterance*. Cambridge: Cambridge University Press.

Kendon, A. (2011). 'Some modern considerations for thinking about language evolution: A discussion of The Evolution of Language by Tecumseh Fitch'. *Public J. Semiotics, 3*. 79- 108.

Kenney, R. D., Mayo, C. A., Winn, H. E. (2001). 'Migration and foraging strategies at varying spatial scales in western North Atlantic right whales: a review of hypotheses'. *J. Cetacean res. Manage (special issue), 2.* 251-260.

Kershenbaum, A., Ilany, A., Blaustein, L., Geffen, E. (2012). 'Syntactic structure and geographical dialects in the songs of male rock hyraxes'. *Proceedings. Biological sciences*, *279*(1740). 2974–2981. https://doi.org/10.1098/rspb.2012.0322

Kershenbaum, A., Blumstein, D.T., Roch, M.A. (2016). 'Acoustic sequences in non-human animals: A tutorial review and prospectus'. *Biol. Rev.*, *91.* 13–52.

Kesebir, S. (2012). 'The Superorganism Account of Human Sociality: How and When Human Groups Are Like Beehives'. *Personality and Social Psychology Review, 16(3)*. 233–261. https://doi.org/10.1177/1088868311430834

Ketten, D. R. (1992). 'The Marine Mammal Ear: Specializations for Aquatic Audition and Echolocation'. In Webster, D. B., Popper, A. N., Fay, R. R. (Eds.), *The Evolutionary Biology of Hearing.* New York: Springer. 717-750.

Kilner, J. M., Lemon, R. N. (2013). 'What we know currently about mirror neurons'. *Current biology*, *23.* 1057–1062.

King, S, Janik, V. M. (2013). 'Bottlenose dolphins can use learned vocal labels to address each other'. *PNAS, 110*.

King, S, Janik, V. M. (2015). 'Come dine with me: food-associated social signalling in wild bottlenose dolphins (*Tursiops truncatus*). *Anim Cogn.,*18*(4).* 969-74.

King, S. L., Sayigh, L. S., Wells, R. S. *et al.* (2013) 'Vocal copying of individually distinctive signature whistles in bottlenose dolphins'. *Proc Biol Sci, 22*.

King, S., Harley, H. E., Janik, V. M. (2014). 'The role of signature whistle matching in bottlenose dolphins (*Tursiops truncatus*)'. *Animal Behaviour*, *96*. 79-86. https://doi.org/10.1016/j.anbehav.2014.07.019

King, S., Connor, R.C., Krützen, M., Allen, S.J. (2021). 'Cooperation-based concept formation in male bottlenose dolphins'. *Nature communications*, *12*(1). 2373. https://doi.org/10.1038/s41467-021-22668-1

King, S., Connor, R., Montgomery, S.H. (2022). 'Social and vocal complexity in bottlenose dolphins'. *Trends in Neurosciences, 45 (12).*

Kirby, S. (2000). 'Syntax without natural selection: How compositionality emerges from vocabulary in a population of learners'. In Knight, C., Studdert-Kennedy, M., Hurford, J.R. (Eds.), *The evolutionary emergence of language: Social function and the origins of linguistic form* (pp. 303-323). Cambridge: Cambridge University Press.

Kirby, S., Griffiths, T. L., & Smith, K. (2015). 'Iterated learning and the evolution of language'. *Current Opinion in Neurobiology, 28*, 108-114.

Kyriazis, M. (2020). 'Ageing Throughout History: The Evolution of Human Lifespan'. *J Mol Evol, 88*. 57–65. https://doi.org/10.1007/s00239-019-09896-2

Kohda, M., Bshary, R., Kubo, N., Awata, S., Sowersby, W. Kawasaka, K., Kobayashi, T., Sogawa, S. (2023). 'Cleaner fish recognize self in a mirror via self-face recognition like humans'. *Proceedings of the National Academy of Sciences of the United States of America, 120*. e2208420120. 10.1073/pnas.2208420120.

Konrad, C.M., Gero, S., Frasier, T., Whitehead, H. (2018). 'Kinship influences sperm whale social organization within, but generally not among, social units'. *R. Soc. open sci. 5*: 180914.

Krause, M., Udell, M., Leavens, D., Skopos, L. (2018). 'Animal pointing: changing trends and findings from 30 years of research'. *Journal of Comparative Psychology, 132*. 10.1037/com0000125.

Krupenye, C., Call, J. (2019). 'Theory of mind in animals: Current and future directions'. *WIREs Cogn Sci*., 10:e1503. https://doi.org/10.1002/wcs.1503

Krupenye, C., Kano, F., Hirata, S., Call, J., Tomasello, M. (2016). 'Great apes anticipate that other individuals will act according to false beliefs'. *Science*, *354(6308)*. 110–114.

Krützen, M., Mann, J., Heithaus, M. (2005). 'Cultural Transmission of Tool Use in Bottlenose Dolphins'. *Proceedings of the National Academy of Sciences of the United States of America, 102.* 8939-8943.

Kuczaj, S. A. II, Walker, R. T. (2012). 'Dolphin problem solving'. In T. R. Zentall & E. A. Wasserman (Eds.), *The Oxford handbook of comparative cognition.* (736–756). Oxford: University Press.

Kuczaj, S. (2013). 'Are Conversations Between Dolphins and Humans Possible? Why the Distinction Between Communication and Conversation is Important'. *International Journal of Comparative Psychology, 26.* 114-123.

Kuczaj, S. A., Eskelinen, H. C. (2014). 'Why do dolphins play?'. *Animal Behavior and Cognition, 1(2).* 113-127.

Kuczaj, S., & Frick, E., Jones, B. *et al*. (2015). 'Underwater observations of dolphin reactions to a distressed conspecific'. *Learning & behavior, 43*.

La Manna, G., Rako Gospić, N., Manghi, M., Sarà, G. (2017). 'Assessing geographical variation on whistle acoustic structure of three Mediterranean populations of common bottlenose dolphin (*Tursiops truncatus*)'. *Behaviour, 154*.

La Manna, G., Rako Gospić, N., Manghi, M., Ceccherelli, G. (2019). 'Influence of environmental, social and behavioural variables on the whistling of the common bottlenose dolphin (*Tursiops truncatus*)'. *Behavioral Ecology and Sociobiology*. 73:121.

La Manna, G., Rako-Gospić, N., Sarà, G., (2020) 'Whistle variation in Mediterranean common bottlenose dolphin: The role of geographical, anthropogenic, social, and behavioral factors'. *Ecology and Evolution, 10.* 1971-1987.

Lakoff, G., Johnson, M. (1980). *Metaphors We Live By*. Chicago: University Press.

Lakoff, G., Johnson, M. (1999). *Philosophy in the flesh : the embodied mind and its challenge to Western thought*. New York: Basic Books.

Laland, K.N., Uller, T., Feldman, M.W., Sterelny, K., Müller, G.B., Moczek, A., (2015). 'The extended evolutionary synthesis: its structure, assumptions and predictions'. *Proc. R. Soc. B Biol. Sci., 282.* 20151019.

Lambert, K.G., Bardi, M., Landis, T., Hyer, M.M., Rzucidlo, A., Gehrt, S., Anchor, C., Jardim Messeder, D., Herculano-Houzel, S. (2014). ' Behind the Mask: Neurobiological indicants of emotional resilience and cognitive function in wild raccoon (Procyon lotor). *Society for Neuroscience*

Lambrechts, M.M., Perret, P., Maistre, M., Blondel, J. (1999). 'Do experiments with captive non-domesticated animals make sense without population field studies? A case study with blue tits' breeding time'. *Proceedings of the Royal Society B: Biological Sciences*, *266*(1426). 1311. https://doi.org/10.1098/rspb.1999.0780

Lameira, A.R., Hardus, M.E., Kowalsky, B., de Vries, H., Spruijt, B.M., Sterck, E.H.M., Shumaker, R.W., Wich, S.A. (2013). 'Orangutan (*Pongo* spp.) whistling and implications for the emergence of an open-ended call repertoire: A replication and extension'. *J. Acoust. Soc. Am., 134*. 2326–2335.

Lameira, A.R. (2017). 'Bidding evidence for primate vocal learning and the cultural substrates for speech evolution'. *Neurosci. Biobehav. Rev., 83*. 429–439.

Lameira, A.R., Call, J. (2018). 'Time-space–displaced responses in the orangutan vocal system'. *Science Advances, 4*. eaau3401. 10.1126/sciadv.aau3401.

Langacker, R.W. (1987) *Foundations of Cognitive Grammar*, Volume 1, *Theoretical Prerequisites*. Stanford: University Press.

Leavens, D.A., Hostetter, A.B., Wesley, M.J., Hopkins, W.D. (2004). 'Tactical use of unimodal and bimodal communication by chimpanzees, *Pan troglodytes*'. *Anim. Behav., 67*. 467–476. doi: 10.1016/j.anbehav.2003.04.007

Lee, P.C., Fishlock, V., Webber, C.E., Moss, C.J. (2016). 'The reproductive advantages of a long life: longevity and senescence in wild female African elephants'. *Behavioral ecology and sociobiology*, *70*. 337–345. https://doi.org/10.1007/s00265-015-2051-5

Lenneberg, E.H. (1967). *Biological Foundations of Language*. New York: John Wiley & Sons.

Leroux, M., Townsend, S.W. (2020). 'Call combinations in great apes and the evolution of syntax'. *Animal Behavior and Cognition, 7(2).* 131–139. doi: https://doi.org/10.26451/abc.07.02.07.2020

Leroux, M., Bosshard, A., Chandia, B., Manser, A., Zuberbühler, K., Townsend, S. (2021). 'Chimpanzees combine pant hoots with food calls into larger structures'. *Animal Behaviour, 179*. 41-50. 10.1016/j.anbehav.2021.06.026.

Leroux, M., Schel, A.M., Wilke, C. (2023). 'Call combinations and compositional processing in wild chimpanzees'. *Nat Commun, 14*. 2225. https://doi.org/10.1038/s41467-023-37816-y

Levine, A., Jewett, D., Kotz, C., Olszewski, P. (2022). 'Behavioral plasticity: Role of neuropeptides in shaping feeding responses'. *Appetite, 174.* 106031. 10.1016/j.appet.2022.106031.

Levinson, S.C. (2006). 'On the human "interaction engine"'. In Enfield, N. J., Levinson, S. C. (Eds.), *Roots of Human Sociality*. Oxford: Berg. 39-69.

Lewis, J. (2003). 'Mud plume feeding, a unique foraging behavior of the bottlenose dolphin in the Florida Keys'. *Gulf of Mexico Science, 21*. 92-97.

Liao, J.B., Buhimschi, C.S., Norwitz, E.R. (2005). 'Normal labor: mechanism and duration'. *Obstetrics and Gynecology Clinics of North America, 32(2).* 145–64, vii. doi:10.1016/j.ogc.2005.01.001

Liao, D.A., Brecht, K.F., Johnston, M., Nieder, A. (2022). 'Recursive sequence generation in crows'. *Sci Adv, 8(44)*. eabq3356. doi: 10.1126/sciadv.abq3356.

Lightfoot, D. (1999). *The Development of Language: Acquisition, Change, and Evolution*. Hoboken: Blackwell.

Lomac-MacNair, K., Smultea, M. A. (2016). 'Blue whale (*Balaenoptera musculus*) behavior and group dynamics as observed from an aircraft off Southern California'. *Animal Behavior and Cognition, 3*(1). 1-21.

Longa, V. M. (2013). 'The evolution of the Faculty of Language from a Chomskyan perspective: bridging linguistics and biology.' *Journal of Anthropological Sciences, 91*. 15- 62.

Lopez, J. C., Lopez, D. (1985). 'Killer Whales (*Orcinus orca*) of Patagonia, and Their Behavior of Intentional Stranding While Hunting Nearshore' *Journal of Mammalogy, 66* (1). 181-183.

Lopez, J., Scott, J. (2000). *Social Structure*. Buckingham: Open University Press.

Lovette, I.J., Fitzpatrick, J.W. (Eds.) (2016). *The Handbook of Bird Biology (3rd Ed).* Hoboken: Wiley-Blackwell.

Lukas, D., Clutton-Brock, T. (2012). 'Cooperative breeding and monogamy in mammalian societies'. *Proceedings. Biological sciences*, *279(1736)*. 2151–2156. https://doi.org/10.1098/rspb.2011.2468

Lukas, D., Clutton-Brock, T. (2018). 'Social complexity and kinship in animal societies'. *Ecol Lett, 21.* 1129–1134. https://doi.org/10.1111/ele.13079

Luís, A.R., Alves, I.S.; Sobreira, F.V. (2019). 'Brays and bits: Information theory applied to acoustic communication sequences of bottlenose dolphins'. *Bioacoustics, 28*. 286–296.

Lust, B.C. (2006). *Child Language: Acquisition and Growth.* Cambridge, MA: University Press.

Lyn, H., Russel, J.L., Leaves, D.A., Bard, K.A., Boysen, S.T., Schaeffer, J.A., Hopkins, W.D. (2014). 'Apes communicate about absent and displaced objects: Methodology matters'. *Anim. Cogn., 17.* 85–94.

Macedonia, J.M., Evans, C.S. (1993). 'Variation among mammalian alarm call systems and the problem of meaning in animal signals,' *Ethology, 93*. 177–197.

Machado, A. M. S., Cantor, M., Costa, A. P. B. *et al*. (2019). 'Homophily around specialized foraging underlies dolphin social preferences'. *Biology Letters, 15*.

MacWhinney, B. (2015). 'Introduction: language emergence'. In MacWhinney, B., O'Grady, W. (Eds.), *The handbook of language emergence.* Malden, MA: Wiley-Blackwell (pp. 1–31).

MacWhinney, B., O'Grady, W. (Eds.) (2015). *The handbook of language emergence*. Malden, MA: Wiley-Blackwell.

Maes, E., De Filippo, G., Inkster, A.B., Lea, S.E., De Houwer, J., D'Hooge, R., Beckers, T., Wills, A.J. (2015). 'Feature- versus rule-based generalization in rats, pigeons and humans'. *Anim Cogn., 18(6)*. 1267-84. doi: 10.1007/s10071-015-0895-8.

Manger, P. R. (2006). 'An examination of cetacean brain structure with a novel hypothesis correlating thermogenesis to the evolution of a big brain'. *Biological Reviews, 81.* 293-338.

Marino, L. (2002). 'Convergence of Complex Cognitive Abilities in Cetaceans and Primates'. *Brain, behavior and evolution. 59*. 21-32.

Marino, L. (2004). 'Cetacean Brain Evolution: Multiplication Generates Complexity'. *International Journal of Comparative Psychology, 17.* 1-16.

Marino, L., Connor, R., Fordyce, R. *et al*. (2007). 'Cetaceans Have Complex Brains for Complex Cognition'. *PLoS biology*, 5: e139.

Marshall J. (2014). 'Mirror neurons'. *Proceedings of the National Academy of Sciences*, *111* (18), 6531.

Martin-Ordas, G., Haun, D., Colmenares, F., Call, J. (2010). 'Keeping track of time: evidence for episodic-like memory in great apes'. *Anim. Cognit., 13*. 331–340.

Martinet, A. (1949). 'La double articulation linguistique'. *TCLC, V*, 30-37.

Martinet, A. (1960). *Éléments de linguistique générale* (1st ed.). Paris: Colin.

Martins, P., Leivada, E., Benítez-Burraco, A., Boeckx, C. (2016). 'Biological pluralisms in service of biolinguistics'. In Fujita, K., Boeckx, C. (Eds.), *Advances in biolinguistics: The human language faculty and its biological basis*. New York: Routledge (pp. 153–169).

Martins, P., Boeckx, C. (2019). 'Language evolution and complexity considerations: The no half-Merge fallacy'. *PLOS Biology, 17*. e3000389. 10.1371/journal.pbio.3000389.

Mayford, M., Siegelbaum, S., Kandel, E. (2012). 'Synapses and Memory Storage'. *Cold Spring Harbor perspectives in biology, 4.* 10.1101/cshperspect.a005751.

McClain, C. *et al*. (2015). 'Sizing ocean giants: patterns of intraspecific size variation in marine megafauna'. *PeerJ,* 3:e715

McDonald, M. A., Hildebrand, J. A., Mesnick, S. L. (2009). 'Worldwide decline in tonal frequencies of blue whale songs'. *Endangered Species Research, 9*. 13-21.

McNeill, D. (2012). *How language began: Gesture and speech in human evolution*. Cambridge: Cambridge University Press.

Mercado, E. III., Murray, S. O., Uyeyama, R. K. *et al*. (1998). 'Memory for recent actions in the bottlenosed dolphin (*Tursiops truncatus*): repletion of arbitrary behaviors using an abstract rule'. *Animal Learning and Behavior*, *26.* 210-218.

Mercado, E. III, Uyeyama, R. K., Pack, A. A., Herman, L. M. (1999). 'Memory for action events in the bottlenosed dolphin'. *Animal Cognition, 2.* 17-25.

Mercado, E. III, DeLong, C.M. (2010). 'Dolphin cognition: Representations and processes in memory and perception'. *International Journal of Comparative Psychology, 23(3)*. 344–378.

Mercado, E. III, Handel, S. (2012). 'Understanding the structure of humpback whale songs (L)'. *J Acoust Soc Am, 132*. 2947-2950.

Mercado, E. III, Perazio, C. E. (2021). 'Similarities in composition and transformations of songs by humpback whales (*Megaptera novaeangliae*) over time and space'. *Journal of Comparative Psychology, 135*(1), 28-50.

Mesoudi, A., Whiten, A. (2008). 'The multiple roles of cultural transmission experiments in understanding human cultural evolution'. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences, 363*. 3489–501. 10.1098/rstb.2008.0129.

Mey, J.L. (2006). 'Pragmatics: Overview'. In Brown, K. (Ed.), *Encyclopedia of Language & Linguistics (Second Edition).* Amstedam: Elsevier (pp. 51–62).

Miksis, J. L., Tyack, P. L., Buck, J. R. (2002). 'Captive dolphins, *Tursiops truncatus*, develop signature whistles that match acoustic features of human-made model sounds'. *J. Acoust. Soc. Am., 112.* 728-739.

Mitchell, R.W. (1986), 'A framework for discussing deception'. In Mitchell, R.W., Thompson, N.S. (Eds.), *Deception, Perspectives on Human and Nonhuman Deceit*. Albany: SUNY Press (pp. 21–29).

Miyagawa, S., Clarke, E. (2019). 'Systems underlying human and old world monkey communication: One, two, or infinite'. *Frontiers in Psychology, 10.* Article 1911. https://doi.org/10.3389/fpsyg.2019.01911

Miyagawa, S., DeSalle, R., Nóbrega, V., Nitschke, R., Okumura, M., Tattersall, I. (2025). 'Linguistic capacity was present in the *Homo sapiens* population 135 thousand years ago'. *Frontiers in Psychology,* 16. 10.3389/fpsyg.2025.1503900.

Mobley, J. R., Herman, L. M., Frankel, A. S. (1988). 'Responses of wintering humpback whales (*Megaptera novaeangliae*) to playbacks of winter and summer vocalizations and synthetic sound'. *Behav Ecol Sociobiol, 23*. 211-223.

Mooney, A. T., Yamato, M., Branstetter, B. K. (2012). 'Hearing in Cetaceans: From Natural History to Experimental Biology'. In Lesser, M. (Eds.), *Advances in Marine Biology, 63.* Cambridge, MA: Academic Press. 197-246.

Neubauer, S., Hublin, J., Gunz, P. (2018). 'The evolution of modern human brain shape'. *Science advances, 4(1)*. eaao5961.

Newmeyer, F.J. (1999). 'Some remarks on the functionalist–formalist controversy in linguistics'. In Darnell, M., Moravcsik, E.A., Noonan, M., Newmeyer, F.J., Wheatley, K. (Eds.). *Functionalism and Formalism in Linguistics, Vol. 1*. Amsterdam: John Benjamins (pp. 469–486).

Newport, E.L., Hauser, M.D., Spaepen, G., Aslin, R.N. (2004). 'Learning at a distance II. Statistical learning of non-adjacent dependencies in a non-human primate'. *Cognitive Psychology, 49(2)*. 85–117.

Nieukirk, S. L., Fregosi, S., Mellinger, D. K., Klinck, H. (2016). 'A complex baleen whale call recorded in the Mariana Trench Marine National Monument'. *J Acoust Soc Am, 140 (3).* EL274.

Nimchinsky, E., Gilissen, E., Allman, J.M., Perl, D.P., Erwin, J.M., Hof, P.R. (1999). 'A neuronal morphologic type unique to humans and great apes'. *Proc Nat Acad Sci, 96*. 5268–73.

Noad, M. J., Cato, D. H., Bryden, M. M. *et al*. (2000) 'Cultural revolution in whale songs'. *Nature, 408*. 537.

Nowak, M.A., Tarnita, C.E., Wilson, E.O. (2010). 'The evolution of eusociality'. *Nature*, *466(7310)*. 1057–1062. https://doi.org/10.1038/nature09205

Ntalampiras, S. (2017). 'Deep Learning Framework for Classifying Sounds of Mysticete Whales'. In Samui, S. P., Sekhar, S., Balas, V. E. (Eds.), *Handbook of Neural Computation*. London/San Diego: Academic Press. 403-415.

Nudel, R., Newbury, D. (2013). 'FOXP2'. *WIREs Cognitive science*, 4. 547–560. 10.1002/wcs.1247.

O'Corry-Crowe, G., Suydam, R., Quakenbush, L. *et al*. (2020). 'Group structure and kinship in beluga whale societies'. *Scientific Reports, 10*. 11462.

O'Grady, W. (2005). *Syntactic Carpentry: An Emergentist Approach to Syntax*. Mahwah, NJ: Lawrence Erlbaum Associates.

O'Leary, M. A., Allard, M., Novacek, J. *et al*. (2004). 'Building the mammalian sector of the Tree of Life'. In Cracraft, J. E Donoghue, M. J. (Eds.), *Assembling the Tree of Life.* New York: Oxford University Press.

Oleson, E. M., Calambokidis, J., Burgess, W. C. *et al*. (2007). 'Behavioral context of call production by eastern North Pacific blue whales'. *Mar Ecol Prog Ser, 330*. 269-284.

Oliveira, C., Wahlberg, M., Silva, M.A., Johnson, M., Antunes, R., Wisniewska, D.M., Fais, A., Gonçalves, J., Madsen, P.T. (2016). 'Sperm whale codas may encode individuality as well as clan identity'. *The Journal of the Acoustical Society of America*, *139*(5). 2860. https://doi.org/10.1121/1.4949478

Ortega, J. (2016). *Sociality in bats*. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-38953-0

Oseki, Y. (2015). 'Eliminating Pair-Merge'. In U. Steindl *et al*. (eds.), *Proceedings of the 32nd West Coast Conference on Formal Linguistics* (pp. 303-312). Somerville, MA: Cascadilla Proceedings.

Ouattara, K., Lemasson, A., & Zuberbühler, K. (2009). 'Campbell's monkeys concatenate vocalizations into context-specific call sequences'. *Proceedings of the National Academy of Sciences, 106(51)*, 22026e22031.

Oxford English Dictionary, "language (*n.*)", March 2024. Oxford University Press. https://doi.org/10.1093/OED/1079050695

Paabo, S. (2014). *Neanderthal Man: In Search of Lost Genomes*. New York: Basic Books.

Pace, F, Benard, F., Glotin, H., Adam, O., and White, P. (2010). Subunit definition and analysis for humpback whale call classification. *Applied acoustics, 71*. 1107-1112.

Pace, D.S., Tumino, C., Silvestri, M. *et al*. (2022). 'Bray-Call Sequences in the Mediterranean Common Bottlenose Dolphin *(Tursiops truncatus)* Acoustic Repertoire'. *Biology 2022, 11, 367.*

Pahl, M., Zhu, H., Pix, W., Tautz, J., Zhang, S. (2007). 'Circadian timed episodic-like memory – a bee knows what to do when, and also where'. *The Journal of Experimental Biology*, 210 (20). 3559–3567.

Panoz-Brown, D., Iyer, V., Carey, L., *et al*. (2018). 'Replay of Episodic Memories in the Rat'. *Current Biology, 28*. 10.1016/j.cub.2018.04.006.

Papale, E., Buffa, G., Filiciotto, F., *et al*. (2015). 'Biphonic calls as signature whistles in a free-ranging bottlenose dolphin'. *Bioacoustics: The International Journal of Animal Sound and its Recording*.
DOI: 10.1080/09524622.2015.1041158

Paradis, M. (2004). *A Neurolinguistic Theory of Bilingualism*. Philadelphia: John Benjamins.

Pardo, M.A., Fristrup, K., Lolchuragi, D.S. (2024). 'African elephants address one another with individually specific name-like calls'. *Nat Ecol Evol* 8- 1353–1364. https://doi.org/10.1038/s41559-024-02420-w

Park, T., Evans, A. R., Gallagher, S. J., Fitzgerald E. M. G. (2017). 'Low-frequency hearing preceded the evolution of giant body size and filter feeding in baleen whales'. *Proc. R. Soc. B, 284*.

Parks, S. E., Searby, A., Celerier, A. *et al*. (2011). 'Sound production behavior of individual North Atlantic right whales: implications for passive acoustic monitoring'. *Endangered Species Research, 15*. 63-76.

Payne, R. S., McVay, S. (1971) 'Songs of humpback whales'. *Science, 173*. 585-597.

Payne, K., Tyack, P., Payne, R. S. (1983). 'Progressive changes in the song of humpback whales (*Megaptera novaengliae*): a detailed analysis of two seasons in Hawaii'. In Payne, R. (Eds.), *Communication and behavior of whales, AAAS selected symposium 76*. Boulder: Westview Press. 9-57.

Payne, K., Payne, R. (1985). 'Large scale changes over 19 years in songs of humpback whales in Bermuda'. Zeitschrift für Tierpsychologie, 68. 89-114.

Pedrazzi, G., Giacomini, G., Pace, D. S. (2022). 'First Report of Epimeletic and Acoustic Behavior in Mediterranean Common Bottlenose Dolphins (*Tursiops truncatus*) Carrying Dead Calves'. *Biology*, *11*(2). 337. https://doi.org/10.3390/biology11020337

Peirce, C.S. (1932 [1897]). 'Division of signs', in *Collected Papers.*

Pepperberg, I. (1999). *The Alex Studies: Cognitive and communicative abilities of grey parrots*. Cambridge: Harvard University Press.

Pepperberg, I. (2009). *Alex & Me: How a scientists and a parrot discovered a hidden world of animal intelligence – and formed a deep bond in the process.* New York, NY: Harper Perennial

Petkov, C.I., ten Cate, C. (2020). 'Structured sequence learning: Animal abilities, cognitive operations, and language evolution'. *Topics in Cognitive Science, 12*(3). 828–842. https://doi.org/10.1111/tops.12444

Pfenning, A.R., Hara, E., Whitney, O., *et al*. (2014). 'Convergent transcriptional specializations in the brains of humans and song-learning birds'. *Science*, *346(6215).* 1256846. https://doi.org/10.1126/science.1256846

Pianka, E.R. (1970). 'On r and K selection'. *American Naturalist*, *104(940)*. 592–597.

Pigliucci, M., Müller, G.B. (Eds.) (2010). *Evolution: The extended synthesis*. Cambridge, MA: MIT Press.

Pika, S., Wilkinson, R., Kendrick, K., Vernes, S. (2018). 'Taking turns: Bridging the gap between human and animal communication'. *Proceedings of the Royal Society B: Biological Sciences, 285*. 20180598. 10.1098/rspb.2018.0598.

Pinker, S., Bloom, P. (1990). 'Natural language and natural selection.' *Behavioral and Brain Sciences, 13*. 707-727.

Pinker, S. (1994). *The language instinct: How the mind creates language.* New York: HarperCollins.

Planer, R.J. (2019). 'What is Symbolic Cognition?'. *Topoi, 40*. 1–12. 10.1007/s11245-019-09670-5.

Pleyer, M., Hartmann, S. (2019). 'Constructing a Consensus on Language Evolution? Convergences and Differences Between Biolinguistic and Usage-Based Approaches'. *Frontiers in Psychology, 10.* 496334.

Plotnik, J.M., de Waal, F.B., Reiss, D. (2006). 'Self-recognition in an Asian elephant'. *Proceedings of the National Academy of Sciences of the United States of America*, *103(45)*. 17053–17057. https://doi.org/10.1073/pnas.0608062103

Poole, J. (1996). *Coming of Age with Elephants.* Chicago: Trafalgar Square.

Pougnault, L., Levréro, F., Leroux, M., Paulet, J., Bombani, P., Dentressangle, F., Deruti, L., Mulot, B., Lemasson, A. (2022). 'Social pressure drives "conversational rules" in great apes'. *Biological reviews of the Cambridge Philosophical Society*, *97*(2). 749–765. https://doi.org/10.1111/brv.12821

Prather, J., Peters, S., Nowicki, S. (2008). 'Precise auditory–vocal mirroring in neurons for learned vocal communication'. *Nature, 451***.** 305–310. https://doi.org/10.1038/nature06492

Premack, D., Woodruff, G. (1978). 'Does a chimpanzee have a theory of mind?'. *Behavioral and Brain Sciences, 1*. 515–526. 10.1017/S0140525X00076512.

Prior, H., Schwarz, A., Güntürkün, O. (2008). 'Mirror-Induced Behavior in the Magpie (*Pica pica*): Evidence of Self-Recognition'. *PLOS Biology,6(8)*. e202. https://doi.org/10.1371/journal.pbio.0060202

Quick, N., Janik, V. M. (2012). 'Bottlenose dolphins exchange signature whistles when meeting at sea'. *Proceedings Biological sciences / The Royal Society*, 279. 2539-2545. 10.1098/rspb.2011.2537.

Raghanti, M.A., Spurlock, L.B., Robert Treichler, F. (2015). 'An analysis of von Economo neurons in the cerebral cortex of cetaceans, artiodactyls, and perissodactyls'. *Brain Struct Funct*, *220*. 2303–2314. https://doi.org/10.1007/s00429-014-0792-y

Raihani, N.J., Ridley, A.R. (2008). 'Experimental evidence for teaching in wild pied babblers'. *Animal Behaviour, 75(1).* 3–11.

Raimondi, V. (2014). 'Social interaction, languaging and the operational conditions for the emergence of observing'. *Frontiers in Psychology, 5(899)*. 1-12.

Rako Gospić, N., La Manna, G., Ceccherelli, G. (2020). 'Influence of Foraging Context on the Whistle Structure of the Common Bottlenose Dolphin.' *Behavioural Processes, 182.*

Ravignani, A., Sonnweber, R.S., Stobbe, N. Fitch, W.T. (2013). 'Action at a distance: dependency sensitivity in a New World primate'. *Biology Letters, 9*. 20130852.

Ravignani, A., Verga, L., Greenfield, M.D. (2019). 'Interactive rhythms across species: the evolutionary biology of animal chorusing and turn-taking'. *Annals of the New York Academy of Sciences*, *1453*(1). 12–21. https://doi.org/10.1111/nyas.14230

Reader, S.M. (2016). ‘Animal social learning: associations and adaptations’. *F1000Research, 5*. 2120. doi: 10.12688/f1000research.7922.1. PMID: 27635227; PMCID: PMC5007745.

Reddy, R.B., Mitani, J.C. (2019). ‘Social relationships and caregiving behavior between recently orphaned chimpanzee siblings’. *Primates; journal of primatology*, *60(5)*. 389–400. https://doi.org/10.1007/s10329-019-00732-1

Reeves, R., Smith, B., Crespo, E., Notarbartolo di Sciara, G. (2003). *Dolphins, Whales and Porpoises Dolphins, Whales and Porpoises: 2002-2010 Conservation Action Plan for the World's Cetaceans*.

Reidenberg, J. S., Laitman, J. T. (2007). ‘Discovery of a low frequency sound source in Mysticeti (baleen whales): Anatomical establishment of a vocal fold homolog’. *Anat. Rec.* 290. 745–760.

Reidenberg, J. S., Laitman, J. T. (2008). ‘Sisters of the sinuses: Cetacean air sacs’. *Anat. Rec.* 291. 1389-1396.

Reinhart, T. (1981). ‘Definite NP anaphora and C-command domains’. *Linguistic Inquiry*, 12(4), 605–635.

Reiss, D., Marino, L. (2001). 'Mirror self-recognition in the bottlenose dolphin: A case of cognitive convergence'. *PNAS, 98 (10).* 5937-5942.

Rendell, L., Whitehead, H. (2001). 'Culture in Whales and Dolphins'. *The Behavioral and brain sciences, 24.* 309-324.

Rendell, L. E., Whitehead, H. (2003). 'V ocal clans in sperm whales *(Physeter macrocephalus)'. Proc Biol Sci, 270*. 225–231.

Risch, D., Gales, N., Gedamke, J. *et al*. (2014). 'Mysterious bio-duck sound attributed to the Antarctic minke whale (*Balaenoptera bonaerensis*)'. *Biol. Lett.* 10: 20140175.

Risch, D., Norris, T., Curnock, M., Friedlaender, A. (2019) 'Common and Antarctic Minke Whales: Conservation Status and Future Research Directions'. *Front. Mar. Sci., 6.* 247.

Rizzi, L. (2016). ‘Monkey morpho-syntax and merge-based systems’. *Theoretical Linguistics, 42*.

Rizzi, L. (2021). *'*Some thoughts on merge: Typology, labeling, freezing effects*'. Presentation held at Goethe-Universität of Frankfurt, on 05/04/2021.*

Rizzolatti, G., Craighero, L. (2004). ‘The Mirror-Neuron System’. *Annual review of neuroscience, 27*. 169–92. 10.1146/annurev.neuro.27.070203.144230.

Roberts, I. (2005). *Principles and Parameters in a VSO Language: A Case Study in Welsh*. New York: Oxford University Press

Roberts, I. (2012). ‘On the nature of syntactic parameters: a programme for research’. In Galves, C., Cyrino, S., Lopes, R., Sandalo, S., Avelar, J. (Eds.), *Parameter Theory and Linguistic Change.* Oxford: OUP. 319-334.

Robins, R.H, Crystal, D. (2024, March 25). *language*. *Encyclopedia Britannica*. https://www.britannica.com/topic/language

Rodriguez-Vidal, J., d'Errico, F., Giles Pacheco, F., Blasco, R., Rosell, J., Jennings, R.P., Queffelec, A., Finlayson, G., Fa, D.A., Gutiérrez López, J.M., Carrión, J.S., Negro, J.J., Finlayson, S., Cáceres, L.M., Bernal, M.A., Fernández Jiménez, S., Finlayson, C. (2014). 'A rock engraving

made by Neanderthals in Gibraltar'. *Proceedings of the National Academy of Sciences, 111*. 13301-13306.

Roth, G., Dicke, U. (2005). 'Evolution of the brain and intelligence'. Trends in Cognitive Sciences, 9(5). 250–257.

Ruhlen, M. (1994). *The origin of language. Tracing the evolution of the mother tongue.* New York: John Wiley & Sons

Sakai, M., Hishii, T., Takeda, S., Kohshima, S. (2006). 'Flipper rubbing behaviors in wild bottlenose dolphins (*Tursiops aduncus*)'. *Marine Mammal Science, 22*. 966-978.

Saidel, E. (2009). 'Attributing mental representations to animals'. In Lurz, R.W. (Ed.), *The philosophy of animal minds.* Cambridge: University Press. https://doi.org/10.1017/CBO9780511819001

Samo, G. (2021) 'N-merge systems in adult and child grammars: a quantitative study on external arguments'. *Qulso, 7*. 103–130.

Saussure, F. (1916) *Course de Linguistique Générale*. Paris: Payot & Cie.

Savage-Rumbaugh, E. S. (1986). *Ape Language: From conditioned response to symbol*. New York: Columbia University Press.

Sayigh, L. S. (2013). 'Cetacean Acoustical Communication'. In Witzany, G. (Eds.), *Biocommunication of Animals*. Dordrecht: Springer. 275-297.

Sayigh, L.S., El Haddad, N., Tyack, P.L., Janik, V.M., Wells, R.S., Jensen, F.H. (2023). 'Bottlenose dolphin mothers modify signature whistles in the presence of their own calves'. *Proceedings of the National Academy of Sciences of the United States of America*, *120*(27). e2300262120. https://doi.org/10.1073/pnas.2300262120

Schamberg, I., Cheney, D. L., Clay, Z., Hohmann, G., Seyfarth, R.M. (2017). 'Bonobos use call combinations to facilitate inter-party travel recruitment'. *Behav. Ecol. Sociobiol., 71*. 75.

Schel, A.M., Townsend, S.W., Machanda, Z., Zuberbühler, K., Slocombe, K.E. (2013). 'Chimpanzee alarm call production meets key criteria for intentionality'. *PloS one*, *8*(10). e76674. https://doi.org/10.1371/journal.pone.0076674

Schleicher, A. (1861-1862). *Compendium der vergleichenden Grammatik der indogermanischen Sprachen*, 2 voll. Weimar: H. Böhlau.

Schlenker, P., Chemla, E., Schel, A. (2016). 'Formal monkey linguistics'. *Theoretical Linguistics, 42 (1)*.

Schlenker, P., Coye, C., Steinert-Threlkeld, S., Klinedinst, N., Chemla, E. (2022). 'Beyond Anthropocentrism in Comparative Cognition: Recentering Animal Linguistics'. *Cognitive Science, 46*. e13220. https://doi.org/10.1111/cogs.13220

Schlenker, P., Coye, C., Leroux, M., Chemla, E. (2023). 'The ABC-D of animal linguistics: are syntax and compositionality for real?'. *Biological reviews of the Cambridge Philosophical Society.* 10.1111/brv.12944.

Schmidt, J. (1872). *Die Verwandtschaftsverhältnisse der indogermanischen Sprachen.* Weimar: H. Böhlau.

Schwartz, B. D., Sprouse, R.. (1996). ‘L2 cognitive states and the full transfer/full access model’. *Second Language Research* 12: 40–72.

Seed, A., Byrne, R. (2010). ‘Animal tool-use’. *Curr Biol, 20(23).* R1032-9. doi: 10.1016/j.cub.2010.09.042. PMID: 21145022.

Seyfarth, R.M., Cheney, D.L. (1980). ‘The ontogeny of vervet monkey alarm calling behavior: a preliminary report’. *Zeitschrift für Tierpsychologie, 54*. 37–56.

Seyfarth, R.M., Cheney, D.L., Marler, P. (1980a). “Monkey responses to three different alarm calls: Evidence of predator classification and semantic communication,” *Science, 210*. 801–803.

Seyfarth, R.M., Cheney, D.L., Marler, P (1980b). “Vervet monkey alarm calls: Semantic communication in a free-ranging primate,” *Animal Behavior, 28*. 1070–1094.

Seyfarth, R.M., Cheney, D.L. (2016). ‘The origin of meaning in animal signals’. *Animal Behaviour, 124*. 339–346.

Shapiro, A., Tyack, P., Seneff, S. (2010). ‘Comparing call-based versus subunit-based methods for categorizing Norwegian killer whale, *Orcinus orca*, vocalizations’. *Animal Behaviour, 81*. 377-386. 10.1016/j.anbehav.2010.09.020.

Shaw, R.C., Harvey, A. (2020). ‘Long-term memory for a learned behaviour in a wild bird’. *Biol. Lett., 16***.** 2019091220190912

Simon, M., Stafford, K. M., Beedholm, K. *et al*. (2010). 'Singing behavior of fin whales in the Davis Strait with implications for mating, migration and foraging'. *J Acoust Soc Am, 128*. 3200-3210.

Sirovic, A., Hildebrand, J. A., Wiggins, S. M. (2007). 'Blue and fin whale call source levels and propagation range in the southern ocean'. *J Acoust Soc Am, 122.* 1208-1215.

Sirovic, A., Williams, L., Kerosky, S. *et al*. (2013). 'Temporal separation of two fin whale call types across the eastern North Pacific'. *Mar Biol, 160.* 47-57.

Skinner, B.F. (1976). *About Behaviorism*. New York: Random House.

Slagsvold, T., Wiebe, K.L. (2011). ‘Social learning in birds and its role in shaping a foraging niche’. *Philos Trans R Soc Lond B Biol Sci., 366(1567*). 969-77. doi: 10.1098/rstb.2010.0343.

Slater, G.J., Goldbogen, J. A., Pyenson, N. D. (2017). 'Independent evolution of baleen whale gigantism linked to Plio-Pleistocene ocean dynamics'. *Proc. R. Soc. B, 284.*

Slocombe, K.E., Zuberbühler, K. (2005). ‘Functionally referential communication in a chimpanzee’. *Curr. Biol., 15*. 1779–1784.

Slocombe, K.E., Zuberbühler, K. (2007). ‘Chimpanzees modify recruitment screams as a function of audience composition’. *Proc. Natl. Acad. Sci. U.S.A., 104*. 17228–17233.

Smet, A.F., Byrne, R.W. (2014). ‘Interpretation of human pointing by African elephants: generalisation and rationality’. Anim Cogn., 17(6). 1365–74.

Smith, T. G., Siniff, D. B., Reichle, R., Stone, S. (1981). 'Coordinated behavior of killer whales, *Orcinus orca,* hunting a crabeater seal, *Lobodon carcinohagus'. Canadian Journal of Zoology,* 59. 1185–1189.

Smolker, R., Richards, A., Connor, R. *et al*. (1997). 'Sponge carrying by dolphins (Delphinidae, *Tursiops sp.*): A foraging specialization involving tool use?' *Ethology, 103*(6). 454-465.

Song, Jae Jung (2012). *Word Order*. Cambridge: University Press

Spillmann, B., Dunkel, L.P., van Noordwijk, M.A., Amda, R.N.A., Lameira, A.R., Wich, S.A., van Schaik, C.P. (2010). 'Acoustic properties of long calls given by flanged male orang-utans (*Pongo pygmaeus wurmbii*) reflect both individual identity and context'. *Ethology, 116*. 385–395.

Stewart, J.D., Stevens, G.M.W., Marshall, G.J. (2017). 'Are mantas self aware or simply social? A response to Ari and D'Agostino 2016'. *J Ethol, 35*. 145–147. https://doi.org/10.1007/s10164-016-0491-7

Stoeger, A.S., Baotic, A. (2021). 'Operant control and call usage learning in African elephants'. *Phil. Trans. R. Soc. B, 376*. 20200254

Suzuki, R., Buck, J. R., Tyack, P. L. (2006). 'Information entropy of humpback whale songs'. *The Journal of the Acoustical Society of America, 119* (3). 1849-1866.

Suzuki, T.N., Matsumoto, Y.K. (2022). 'Experimental evidence for core-Merge in the vocal communication system of a wild passerine'. *Nat Commun, 13: 5605.*

Suzuki, T.N., Wheatcroft, D., Griesser, M. (2016). 'Experimental evidence for compositional syntax in bird calls'. *Nat Commun., 7*. 10986.

Suzuki, T.N., Wheatcroft, D., Griesser, M. (2019). 'The syntax–semantics interface in animal vocal communication'. *Phil. Trans. R. Soc. B, 375*. 20180405. http://dx.doi.org/10.1098/rstb.2018.0405

Szabó, Z.G. (2020). 'Compositionality'. In E.N. Zalta (ed.), *The Stanford encyclopedia of philosophy (Fall 2000 edition)*. Stanford: Metaphysics Research Lab.

Taglialatela, J.P., Reamer, L., Schapiro, S.J., Hopkins, W.D. (2012). 'Social learning of a communicative signal in captive chimpanzees'. *Biol. Lett., 8.* 498–501.

Takagi, S., Tsuzuki, M., Chijiiwa, H., Arahori, M., Watanabe, A., Saito, A., Fujita, K. (2017). 'Use of incidentally encoded memory from a single experience in cats'. *Behavioural Processes, 141(3).* 267–272.

Tattersall, I. (2016). 'A tentative framework for the acquisition of language and modern human cognition'. *Journal of Anthropological Sciences*. 94. 157-166.

Tauzin, T., Bohn, M., Gergely, G. (2020). 'Context-sensitive adjustment of pointing in great apes'. *Sci Rep*, *10*. 1048.

Tervo, O. M., Christoffersen, M. F., Parks, S. E. *et al*. (2012). 'Evidence for simultaneous sound production in the bowhead whale (*Balaena mysticetus*)'. *J Acoust Soc Am, 130*. 2257-2262.

Tettamanti, M. (2020). *Il cervello sintattico*. Roma: Carocci.

Théoret, H., Pascual-Leone, A. (2002). 'Language acquisition: do as you hear'. *Current Biology*, *12(21).* R736–R737

Thornton, A., McAuliffe, K., Dall, S.R., Fernandez-Duque, E., Garber, P.A., Young, A.J. (2016). 'Fundamental problems with the cooperative breeding hypothesis. A reply to Burkart & van Schaik'. *J Zool (1987), 299(2).* 84–88. doi: 10.1111/jzo.12351.

Tokuyama, N., Toda, K., Poiret, M.L., Iyokango, B., Bakaa, B., Ishizuka, S. (2021). 'Two wild female bonobos adopted infants from a different social group at Wamba'. *Scientific reports*, *11(1)*. 4967. https://doi.org/10.1038/s41598-021-83667-2

Tomasello, M. (2003). *Constructing a Language: A Usage-Based Theory of Language Acquisition*. Cambridge, MA: Harvard University Press.

Tomasello, M. (2008). *Origins of human communication.* Cambridge, MA: The MIT Press.

Tomasello, M., Call, J. (1997). *Primate cognition.* Oxford: University Press.

Tomasello, M., Call, J. (2019). 'Thirty years of great ape gestures'. *Anim Cogn*, *22*. 461–469. https://doi.org/10.1007/s10071-018-1167-1

Torres Ortiz, S., Smeele, S.Q., Champenois, J. (2022). 'Memory for own actions in parrots'. *Sci Rep*, *12*. 20561.

Townsend, S.W., Manser, M.B. (2013). 'Functionally referential calls, the past, present and the future'. *Ethology*, *119*(1). 1-11.

Townsend, S. W., Engesser, S., Stoll, S., Zuberbühler, K., & Bickel, B. (2018). Compositionality in animals and humans. PLoS Biology, 16(8), e2006425.

Townsend, S.W., Koski, S.E., Byrne, R.W., Slocombe, K.E., Bickel, B., Boeckle, M., Braga Goncalves, I., Burkart, J.M., Flower, T., Gaunet, F., Glock, H.J., Gruber, T., Jansen, D.A.W.A.M., Liebal, K., Linke, A., Miklósi, Á., Moore, R., van Schaik, C.P., Stoll, S., Vail, A., Waller, B.M., Wild, M., Zuberbühler, K., Manser, M.B. (2017). 'Exorcising Grice's ghost: an empirical approach to studying intentional communication in animals'. *Biol Rev, 92*. 1427-1433. https://doi.org/10.1111/brv.12289

Tricot, M., Cammaerts, R. (2015). 'Are ants (Hymenoptera, Formicidae) capable of self recognition?'. *Journal of science, 5*. 521–532.

Tyack, P. L. (1983) 'Differential response of humpback whales, *Megaptera novaeagliae*, to playback of song or social sounds'. *Behav Ecol Sociobiol, 13*. 49-55.

Tyack, P. (2000). 'Functional aspects of cetacean communication'. In Mann, J. *et al*. (Eds.), *Cetacean societies: field studies of dolphins and whales.* Chicago: University Press. 270-307.

Ullman, M. T. (2001). The declarative/procedural model of lexicon and grammar. Journal of psycholinguistic research, 30, 37-69.

Vakili, P. (2023). 'Criticism of the Poverty of the Stimulus Argument (POSA)'. *European Journal of Language and Culture Studies, 2*(2), 77–81. https://doi.org/10.24018/ejlang.2023.2.2.84

Valsecchi, E., Hale, P., Corkeron, P., Amos, W. (2002). 'Social structure in migrating humpback whales (*Megaptera novaeangliae*)'. *Molecular ecology, 11*. 507-518.

van Leeuwen, E.J.C., Cronin, K.A., Haun, D.B.M. (2014). 'A group-specific arbitrary tradition in chimpanzees (*Pan troglodytes*)'. *Anim. Cogn., 17.* 1421–1425.

van Schaik, C.P. (1999). 'The socioecology of fission-fusion sociality in Orangutans'. *Primates, 40(1)*. 69–86.

Vanduffel, W. (2018). 'Long-term value memory in primates'. Proc Natl Acad Sci U S A., 115(9). 1956–1958.

Vernes, S., Kriengwatana, B., Beeck, V., Fischer, J., Tyack, P., ten Cate, C., Janik, V. M. (2021). 'The multi-dimensional nature of vocal learning'. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences, 376.* 20200236. 10.1098/rstb.2020.0236.

Veselinović, D., Candiotti, A., & Lemasson, A. (2014). *Female Diana monkeys (Cercopithecus Diana) have complex calls*. Manuscript, New York University.

Villar, F. (1997). *Gli indoeuropei e le origini dell'Europa. Lingua e storia*. Bologna: Il Mulino [original edition: *Los indoeuropeos y los origenes de Europa. Lenguaje e historia.* 1992. Barcelona: Gredos].

Visser, I., Smith, T., Bullock, I. *et al*. (2007). 'Antarctic Peninsula killer whales (*Orcinus orca*) hunt seals and a penguin on floating ice'. *Marine Mammal Science, 24*. 225-234.

Wacewicz, S., Żywiczyński, P. (2015). 'Language evolution: Why Hockett's design features are a non-starter'. *Biosemiotics, 8*. 29-46.

Wacewicz, S., Żywiczyński, P., Hartmann, S., Pleyer, M., Benítez-Burraco, A. (2020). 'Language in Language Evolution Research: In Defense of a Pluralistic View.' *Biolinguistics, 14*. 59-101.

Watson, K. K., Jones, T. K., Allman, J. M. (2006). 'Dendritic architecture of the Von Economo neurons'. *Neuroscience, 141 (3).* 1107–1112.

Watson, S.K., Townsend, S., Schel, A., Wilke, C., Wallace, E., Cheng, L., West, V., Slocombe, K. (2015). 'Vocal Learning in the Functionally Referential Food Grunts of Chimpanzees'. *Current Biology, 25.* 1-5. 10.1016/j.cub.2014.12.032.

Watson, S.K., Burkart, J.M., Schapiro, S.J., Lambeth, S.P., Mueller, J.L., Townsend, S.W. (2020). 'Nonadjacent dependency processing in monkeys, apes, and humans'. *Sci. Adv., 6.* eabb0725.

Watson, S.K., Burkart, J.M., Schapiro, S.J., Lambeth, S.P., Mueller, J.L., Townsend, S.W. (2021). Reply to comment on 'Nonadjacent dependency processing in monkeys, apes, and humans'. *Science Advances, 7(30).* eabj1517. DOI: 10.1126/sciadv.abj1517.

Watson, S.K., Filippi, P., Gasparri, L., Falk, N., Tamer, N., Widmer, P., Manser, M., Glock, H. (2022). 'Optionality in animal communication: a novel framework for examining the evolution of arbitrariness'. *Biological reviews of the Cambridge Philosophical Society, 97*. 10.1111/brv.12882.

Webb, D. M., Zhang, J. (2005). '*FoxP2* in Song-Learning Birds and Vocal-Learning Mammals'. *Journal of Heredity*, 96(3). 212–216.

Weilgart, L., Whitehead, H. (1997). 'Group-specific dialects and geographical variation in coda repertoire in South Pacific sperm whales'. *Behav Ecol Sociobiol*, *40*. 277-285.

Westermann, G., Ruh, N., Plunkett, K. (2009). 'Connectionist approaches to language learning'. *Linguistics, 47*. 413–452.

White, L. (2003). *Second Language Acquisition and Universal Grammar.* Cambridge: University Press.

Whitehead, H., Dillon, M., Dufault, S. *et al*. (1998). 'Non-geographically based population structure of South Pacific sperm whales: dialects, fluke-markings and genetics'. *Journal of Animal Ecology, 67.* 253-262.

Whitehead, H., Rendell, L. (2015). *The Cultural Lives of Whales and Dolphins.* Chicago: University Press.

Whiten, A., Byrne, R. (1988). ‘Tactical deception in primates’. *Behav. Brain Sci.*, *11.* 233–244.

Whiten, A., van de Waal, E. (2018). ‘The pervasive role of social learning in primate lifetime development’. *Behav Ecol Sociobiol., 72(5).* 80. doi: 10.1007/s00265-018-2489-3.

Whiten, Andrew (2021). ‘The burgeoning reach of animal culture’. *Science, 372(6537).*

Wich, S.A., Krützen, M., Lameira, A.R., Nater, A., Arora, N., Bastian, M.L., Meulman, E., Morrogh-Bernard, H.C., Atmoko, S.S.U., Pamungkas, J., Perwitasari-Farajallah, D., Hardus, M.E., van Noordwijk, M., van Schaik, C.P. (2012). ‘Call cultures in orang-utans?’ *PLoS One, 7.* e36180.

Wild, S., Chimento, M., Mc Mahon, K., Farine, D., Sheldon, B. Aplin, L. (2021). ‘Complex foraging behaviours in wild birds emerge from social learning and recombination of components’. *Philosophical Transactions of the Royal Society B: Biological Sciences, 377*. 10.1098/rstb.2020.0307.

Wiley, D., Ware, C., Bocconcelli, A. *et al*. (2011). 'Underwater components of humpback whale bubble-net feeding'. *Behaviour, 148*. 575-602.

Wilke, C., Lahiff, N.J., Sabbi, K.H., Watts, D.P., Townsend, S.W., Slocombe, K.E. (2022). ‘Declarative referential gesturing in a wild chimpanzee (*Pan troglodytes*)’. *Proceedings of the National Academy of Sciences of the United States of America*, *119*(47). e2206486119. https://doi.org/10.1073/pnas.2206486119

Williams, T. M., Worthy, G. A. (2002). 'Anatomy and physiology: the challenge of aquatic living'. In Hoelzel, A. R. (Eds.), *Marine Mammal Biology: An Evolutionary Approach.* Hoboken: Blackwell. 73-97.

Williams, H., Levin, I.I., Norris, D.M., Newman, A.E.M., Wheelwright, N.T. (2013). ‘Three decades of cultural evolution in Savannah sparrow songs’. *Anim. Behav., 85*. 213–223.

Wilson, E.O. (1971). *The Insect Societies.* Cambridge, MA: Belknap Press of Harvard University Press.

Wilson, B., Spierings, M., Ravignani, A., Mueller, J.L., Mintz, T.H., Wijnen, F., van der Kant, A., Smith, K., Rey, A. (2020). ‘Non-adjacent Dependency Learning in Humans and Other Animals’. *Top Cogn Sci., 12(3).* 843–858. doi: 10.1111/tops.12381.

Winn, H. E., Winn, L. K. (1978). 'The song of the humpback whale (*Megaptera novaengliae*) in the West Indies'. *Marine Biology, 47*. 97-114.

Wittemyer, G., Douglas-Hamilton, I., Getz, W. M. (2005). ‘The Socioecology of Elephants: Analysis of the Processes Creating Multitiered Social Structures’. *Anim. Behav, 69*. 1357–1371.

Wulff, S., Ellis, N. (2018). ‘Usage-based approaches to second language acquisition’. In Miller, D.T., Bayram, F., Rothman, J., Serratrice, L. (Eds.), *Bilingual cognition and language: the state of the science across its subfields*. Amsterdam: John Benjamins (pp.37-56)

Wursig, B., Clark, C. (1990) 'Behavior'. In Burns, J. J., Montague, J. J., Cowles, C. J. (Eds.), *The bowhead whale, Society for Marine Mammalogy, special publication 2*. Lawrence: Allen Press. 157-199.

Xitco, M., Gory, J., Kuczaj, S. (1999). ‘An introduction to the Living Sea’s dolphin keyboard communication system’. In Ramirez, K. (Eds.), *Animal training.* Chicago: Shedd Aquarium Press. 493-498.

Xitco, M., Gory, J., Kuczaj, S. (2001). 'Spontaneous pointing by bottlenose dolphins (*Tursiops truncatus*)'. *Animal Cognition, 4.* 115-123.

Yamamoto C, Kashiwagi N, Otsuka M, Sakai M, Tomonaga M. (2019). 'Cooperation in bottlenose dolphins: bidirectional coordination in a rope-pulling task'. *PeerJ* 7:e7826 https://doi.org/10.7717/peerj.7826

Yamato, M., Ketten, D. R., Arruda, J., Cramer, S., Moore, K. M. E. (2012). 'The auditory anatomy of the minke whale (*Balenapera acutorstrata*): Insights into the potential sound reception pathways in a baleen whale'. *Anatomical Record, 295.* 991-998.

Yurk, H., Barrett-Lennard, L., Ford, J. K. B., Matkin, C. O. (2002). 'Cultural transmission within maternal lineages: vocal clans in resident killer whales in southern Alaska'. *Animal Behaviour*, *63*(6). 1103–1119.

Zamorano-Abramson, J., Michon, M., Hernández-Lloreda, M.V., Aboitiz, F. (2023). 'Multimodal imitative learning and synchrony in cetaceans: A model for speech and singing evolution.' *Front. Psychol.* 14:1061381. doi: 10.3389/fpsyg.2023.1061381

Zentall, T.R. (2006). 'Imitation: definitions, evidence, and mechanisms'. *Anim Cogn*, 9. 335–353.

Zentall, T.R., Wasserman, E.A., Urcuioli, P.J. (2014). 'Associative concept learning in animals'. *J. Exp. Anal. Behav., 101.* 130–151.

Zinkivskay, A., Nazir, F., Smulders, T.V. (2009). 'What-where-when memory in magpies (*Pica pica*)'. *Anim. Cognit., 12*. 119–125.

Zlatev, J. (2014). 'Human uniqueness, bodily mimesis and the evolution of language'. *Humana.Mente Journal of Philosophical Studies 7*, 197–219.

Zuberbühler K. (2000). 'Referential labelling in Diana monkeys'. *Animal behaviour*, *59*(5). 917–927. https://doi.org/10.1006/anbe.1999.1317

Zuberbühler, K. (2008). 'Audience effects'. *Curr. Biol., 18.* R189–R190.

Zuberbühler, K. (2018). 'Intentional communication in primates'. *Revue Tranel (Travaux neuchâtelois de linguistique) 68.* 69-75.

Zuberbühler, K. (2019). 'Syntax and compositionality in animal communication'. *Phil. Trans. R. Soc. B, 375*: 20190062. http://dx.doi.org/10.1098/rstb.2019.0062

Zuberbühler, K., Cheney, D.L., Seyfarth, R.M. (1999). 'Conceptual semantics in a nonhuman primate'. *Journal of Comparative Psychology, 113*(1). 33–42. https://doi.org/10.1037/0735-7036.113.1.33

Zuberbühler, K., Léon, J., Deshpande, A., Quintero, F. (2022). 'Socially scripted vocal learning in primates'. *Current Opinion in Behavioral Sciences, 46*. 101153. 10.1016/j.cobeha.2022.101153.