Neural Semantic Pointers in Context

Alessio Plebe and Arianna Pavone
Department of Cognitive Science, University of Messina, v. Concezione 8, 98121, Messina, Italy
{*aplebe, apavone*}*@unime.it*

Keywords: Context, Neural Model, Semantic Disambiguation, Brain Model.

Abstract: Resolving linguistic ambiguities is a task frequently called for in human communication. In many cases, such task cannot be solved without additional information about an associated context, which can be often captured from the visual scene referred by the sentence. This type of inference is crucial in several aspects of language, communication in the first place, and in the grounding of language in perception. This paper focuses on the contextual effects of visual scenes on semantics, investigated using neural computational simulation. Specifically, here we address the problem of selecting the interpretation of sentences with an ambiguous prepositional phrase, matching the context provided by visual perception. More formally, provided with a sentence, admitting two or more candidate resolutions for a prepositional phrase attachment, and an image that depicts the content of the sentence, it is required to choose the correct resolution depending on the image's content. From the neuro-computational point of view, our model is based on Nengo, the implementation of Neural Engineering Framework (NEF), whose basic semantic component is the so-called Semantic Pointer Architecture (SPA), a biologically plausible way of representing concepts by dynamic neural assemblies. We evaluated the ability of our model in resolving linguistic ambiguities on the LAVA (Language and Vision Ambiguities) dataset, a corpus of sentences with a wide range of ambiguities, associated with visual scenes.

1 INTRODUCTION

In recent years, a number of different disciplines have begun to investigate the fundamental role that context plays in different cognitive phenomena. The problem of context spans from the abstract level of semantics down to the level of neural representations. It has increasingly been studied also for its role in influencing mental concepts and, more specifically, linguistic communication has been the area of study that has traditionally explored these issues. The term context is not easy to define: it is something that cannot be specified independently of a specific frame and it may play quite different roles within alternative research paradigms. In wider terms, as stated by Goodwin and Duranti (1992), we can define the context as a "frame that surrounds the event and provides resources for its appropriate interpretation". However, in order to obtain a more complete understanding of what context stands for, it is necessary to investigate how it interacts with cognitive phenomena at three different levels: the linguistic, the cognitive and the neural level (Plebe and De La Cruz, 2020).

There is a long tradition in linguistics and pragmatics which invokes context to help account for aspects of meaning in language that go beyond the scope of semantics. The main elements of context has roots that dates back to the past, and regards the degree to which truth-functional semantics depends on context. Gottlob Frege raised the point in his uncompleted 1897 volume Logik, and though he was not explicitly using the term context, he underlined how for many expressions, fixing their truth value requires supplemental information, coming from the circumstances, the "frames", in which such expressions are pronounced. The first clear elucidation of the dependence of language on context was proposed by Searle (1978). He takes up Frege's idea that a word has meaning only if related to the meaning of the whole sentence and if its meaning is perceived by both interlocutors, speaker and listener. At a cognitive level the issue regards concepts and the degree to which they are dependent on context. Barsalou (1983) has been one of the first to underline how difficult it is to conceive them as stable subjective entities, while it appears more appropriate to think of categories as dynamically constructed and tailored to specific contexts, or as *ad hoc* categories. A recent review of studies of the cognitive perspective on the linguistic issue of context can be found in (Airenti and Plebe, 2017). Just as in the strictly linguistic domain, they find in the wider cognitive view a variety of positions, some that minimize the destabilizing effect context has on concepts, such as that of Machery (2015), or others that assume a more intermediate position such as. that of Mazzone and Lalumera (2009), that while acknowledging the fundamental. role context might play in concepts, sustain that a characterizing stable nucleus of mental concepts is also a part.

On the other hand, cognitive neuroscience is now starting to consider in a systematic way how context interacts with neural responses (Stark et al., 2018). The way context drives language comprehension depends on the effects of context on the conceptual scaffolding of the listener, which in turn, is the result of his neural responses in combination to context.

The kind of ambiguity addressed in this paper is the canonical case of structural ambiguity, technically known as Prepositional Phrase Attachment, where a sentence includes a prepositional phrase that can be attached to more than one higher level phrases (Hindle and Rooth, 1993). The attachment resolution is context dependent, we deal specifically with the case when depends on the visual context.

Specifically, provided with a sentence, admitting two or more candidate interpretations, and an image that depicts its content, it is required to choose the correct interpretation of the sentence depending on the image's content. Thus we address the problem of selecting the interpretation of an ambiguous sentence matching the content of a given image.

This type of inference is frequently called for in human communication occurring in a visual environment, and is crucial for language acquisition, when much of the linguistic content refers to the visual surroundings of the child (Bates et al., 1995; Bornstein and R.Cote, 2004).

This kind of task is also fundamental to the problem of grounding vision in language, by focusing on phenomena of linguistic ambiguity, which are prevalent in language, but typically overlooked when using language as a medium for expressing understanding of visual content. Due to such ambiguities, a superficially appropriate description of a visual scene may in fact not be sufficient for demonstrating a correct understanding of the relevant visual content.

From the neurocomputational point of view, our model is based on Nengo (`https://www.nengo.ai`), the implementation of Eliasmith's Neural Engineering Framework (NEF) (Eliasmith, 2013). The basic semantic component within NEF is the so-called Semantic Pointer Architecture (SPA) (Thagard, 2011), which determines how the concepts are represented as dymanic neural assemblies. The model works by extracting the three involved entities from the input sentence and identifying the categories involved.

Early experimental results show that the presented computational model achieves a reliable ability to disambiguate sentences.

1.1 A Framework for Neural Semantics

The two main requirements we seek in the identification of a suitable neural framework to be adopted all along this work is the biological plausibility and the possibility of modeling at a level enough abstract to deal with full images and with words in sentences. The two requirements are clearly in stark contrast.

Today the legacy of connectionism has been taken up by the family of algorithms collected under the name *deep learning*. Unlike the former artificial neural networks, deep learning models succeeds in highly complex cognitive tasks, reaching even human-like performances in some visual tasks (VanRullen, 2017). However, the level of biological plausibility of deep learning algorithms is in general even lower than in connectionism, these models were developed with engineering goals in mind, and exploring cognition is not in the agenda of this research community (Plebe and Grasso, 2019). In our model we will also include a very simple deep learning component, but only for the low-level analysis of the images. This choice makes the model simpler, by exploiting the ease of deep learning model in processing visual stimuli. It would have been easy to solve also the crucial part of our problem, the semantic disambiguation, through deep learning, but this would have been of little value as a cognitive model.

Currently, the neural framework that can simulate the widest range of cognitive tasks, by adopting a unified methodology with a reasonable degree of biological plausibility, is Nengo (Neural ENGineering Objects) (Eliasmith, 2013). The idea behind Nengo dates back to 2003, thanks to the former NEF (Neural Engineering Framework) (Eliasmith and Anderson, 2003), which defines a general methodology for the construction of large cognitive models, informed by a number of key neuroscientific concepts. In brief, the three main such concepts are the following:

- The *Representation Principle*: neural representations are defined by the combination of nonlinear encoding of spikes over a population of neurons, and weighted decoding over the same populations of neurons and over time;
- The *Transformation Principle*: transformations of neural representations are functions of the variables represented by neural populations. Transformations are determined using an alternately weighted decoding;

- The *Dynamic Principle*: neural dynamics are characterized by considering neural representations as state variables of dynamic systems. Thus, the dynamics of neuro-biological systems can be analyzed using control theory.

According to the listed principles, the basic computational object in Nengo is a population of neurons that collectively can represent a multidimensional entity. The meaningful entity is retrieved from the neural activation by the equation (Eliasmith, 2013):

$$\vec{x} = \sum_{i=1}^{N} \sum_{j=1}^{M} e^{-\frac{t-t_{i,j}}{\tau}} \vec{d_i} \tag{1}$$

where N is the number of neurons in the population, and M is the number of spikes that happen in the time windows of the computation; $t_{i,j}$ is the time when the i-th neuron in the population has fired for the j-th time; $\vec{d_i}$ is the i-th row of the $N \times D$ *decoding* matrix $\vec{D}$ with D the dimension of the entity to be represented; τ is the time constant of decay of the postsynaptic activation. The activity of the neurons in a population depends from the *encoding* of their input that can be multidimensional with a dimension different from D.

A fundamental extension of the general neural population, ruled by equation (1), is the the Semantic Pointer Architecture (SPA), used when representing entities at higher cognitive level, i.e. conceptual and linguistic. In addition to the encoding and the decoding features, SPA structures allow a number of high level operations, that may correspond to conceptual manipulation, with some degree of biological plausibility. The foundation of these conceptual operations is in the mathematics of *holographic* representations, as theorized by Plate (2003). One of the basic operations is the *binding* of two SPA, computed by circular convolution (Eliasmith, 2013):

$$\vec{x} \circledast \vec{y} = \mathscr{F}^{-1} \left(\mathscr{F}(\vec{x}) \cdot \mathscr{F}(\vec{y}) \right) \tag{2}$$

where $\mathscr{F}$ is the discrete Fourier transform and $\mathscr{F}^{-1}$ is its inverse, and $\cdot$ is the element-wise multiplication. Informally, SPA together with its associated operations, can be thought as a neural process that compresses information in other neural processes to which it points and into which it can be expanded when needed, providing shallow meanings through symbol-like relations to the world and other representations, and expanding to provide deeper meanings with relations to perceptual, motor, and emotional information, support complex syntactic operations. They also help to control the flow of information through a cognitive system to accomplish its goals. Thus semantic pointers have semantic, syntactic, and pragmatic functions, just like the symbols in a rule-based system, but with a highly distributed, probabilistic operation.

2 THE AMBIGUITY TESTBED

The testbed of our study is the LAVA (Language and Vision Ambiguities) corpus, recently introduced by Berzak et al. (2015).

Such corpus contains sentences with linguistic ambiguities that can only be resolved using external information. The sentences are paired with short videos that visualize different interpretations of each sentence. Such sentences encompass a wide range of syntactic, semantic and discourse ambiguities, including ambiguous prepositional and verb phrase attachments, conjunctions, logical forms, anaphora and ellipsis.

Overall, the corpus contains 237 sentences, with 2 to 3 interpretations per sentence, and an average of 3.37 videos that depict visual variations of each sentence interpretation, corresponding to a total of 1679 videos. Each sentence involves two or more entities in one among four categories (person, bag, telescope and chair).

In their paper Berzak et al. (2015) also addressed the problem of selecting the interpretation of an ambiguous sentence that matches the content of a given video. In our case also, the road to solve the ambiguities is in pairing sentences with images that visualize the corresponding scene. In order to simplify the task, we limited on sentences where exactly three distinct entities belonging to three distinct categories are involved (among person, bag, telescope and chair).

For instance, given the sentence ”Sam approached the chair with a bag”, three different categories involved: person (Sam), chair and bag. In addition two different linguistic interpretations are plausible: the first interpretation assumes that Sam has the bag while approaching the chair, while the second one assumes that the bag is on the chair while Sam is approaching.

In addition, in this preliminary work, the ambiguous phrases examined by the system are limited to the preposition “with”: the system is able to solve the ambiguity thanks to the given image and therefore to understand who the proposition refers to.

For example in the following sentence

“Dany approached the chair with a yellow bag”

the system is able to recognize to whom it refers “with” and specifically if Dany brings a yellow bag while approaching the chair or if the bag is on the chair while Dany is approaching them.

Our corpus contains 81 sentences, with 2 to 3 interpretations per sentence.

3 METHOD

Without restricting the general case we can assume that the world is populated by objects that are grouped into categories $\mathscr{C} = \{C_1, C_2, \cdots\}$. The small world of LAVA is populated with a limited number of categories, whose instances can appear in images, corresponding to scenes viewed by the agent. Images are matrices $\mathbf{I}$ of pixels in two dimensions. Inside an image $\mathbf{I}$ it is possible to cut submatrices $\mathbf{B}^{(C)}_{x,y}$ with centers at coordinates x, y, and size suitable to contain objects of category C. The submatrices $\mathbf{B}^{(C)}_{x,y}$ bear a resemblance with foveal images captured during saccadic movements gazing over different portion of the visual scene, with the purpose of recognizing each single object. Unlike natural saccades, in our model the centers of the detailed views are not driven by top-down mechanisms, are instead sampled at fixed regular intervals, scanning the whole image:

$$\begin{aligned} X &= \langle x_1, x_1 + \Delta_x, \cdots, x_1 + N\Delta_x \rangle, & (3)\\ Y &= \left\langle y_1, y_1 + \Delta_y, \cdots, y_1 + M\Delta_y \right\rangle. & (4) \end{aligned}$$

The visual component of the model is made by a set of deep convolutional neural networks, tuned to recognize one of the categories $C \in \mathscr{C}$. Each network is a function $f_{(C)}(\cdot)$ estimating the probability of an object of category C to be inside a submatrix $\mathbf{B}_{x,y}$:

$$f_{(C)}() : \mathbb{R} \times \mathbb{R} \to [0..1] \qquad (5)$$

By applying the deep convolutional neural network $f_C(\cdot)$ to all the submatrices of an image, a vector $\vec{p}^{(C)}$ of probabilities to find an object of category C in the discrete horizontal positions X is constructed. An element $p^{(C)}_i$ of $\vec{p}^{(C)}$ is computed as:

$$p^{(C)}_i = \max_{y \in Y} \left\{ f_{(C)} \left(\mathbf{B}^{(C)}_{x_i,y} \right) \right\} \qquad (6)$$

The rationale behind equation (6) is that in an interior environment the displacement of objects – therefore their spatial relationship – appears mainly in the horizontal dimension of the retinal projection of the scene. It is therefore possibly to capture the probabilistic locations of objects as vectorial representations corresponding to scanning the scene horizontally along X. We can now compose equation (6) into a family of functions $\phi^{(C)}(\cdot)$ that, given and image $\mathbf{I}$, return probability vectors $\vec{p}^{(C)}$:

$$\phi^{(C)}() : \mathbb{R} \times \mathbb{R} \to [0..1]^N \qquad (7)$$

Let us move on the linguistic part of LAVA and of the stimuli to the model. The full set of sentences in LAVA use words from a closed lexicon $\mathscr{L}$, and within this lexicon there are two subsets relevant for our model. One is the lexicon of words $\mathscr{W} \subset \mathscr{L} = \{W_1, W_2, \cdots\}$ used to name objects of the categories in $\mathscr{C}$. In the case of LAVA we can assume a deterministic reference function:

$$c(W) : \mathscr{W} \to \mathscr{C} \qquad (8)$$

associating every word W to a category C. There is then a smaller lexicon of prepositions, the grammatical category responsible for the contextual ambiguities: $\mathscr{P} \subset \mathscr{L} = \{\texttt{with}, \cdots\}$.

A sentence in the LAVA is an ordered set $\mathscr{S}$, with elements $S_i \in \mathscr{L}$, from which a simple preprocessing extracts three key words:

$$\mathscr{S} \to \begin{cases} W_\mathrm{P} & \text{noun under the head of } \mathscr{S} \\ W' & \text{first noun possible head of } \mathscr{S} \\ W'' & \text{second noun possible head of } \mathscr{S} \end{cases} \qquad (9)$$

The noun W_P is easily found by searching in $\mathscr{S}$ the first element $S_i \in \mathscr{P}$, and then searching the first element S_j with $j > i$ such that $S_j \in \mathscr{W}$. The other two nouns W' and W'' are the only two possible elements $S_{l,k} \in \mathscr{W}$ with $l \neq j, k \neq j$. Let us call $W_\mathrm{H} \in \{W', W''\}$ the correct head of the prepositional phrase.

The three key words W_P, W', W'' find a correspondence in the model in terms of three Nengo SPA items: $\vec{V}_\mathrm{P}$, $\vec{V}'$ and $\vec{V}''$. The processed sentence $\mathscr{S}$ is linked with an image $\mathbf{I}$ in which the objects of categories referred by W_P, W' and W'' are searched:

$$\begin{aligned} \vec{p}_\mathrm{P} &= \phi^{(c(W_\mathrm{P}))}(\mathbf{I}), & (10)\\ \vec{p}' &= \phi^{\left(c\left(W'\right)\right)}(\mathbf{I}), & (11)\\ \vec{p}'' &= \phi^{\left(c\left(W''\right)\right)}(\mathbf{I}). & (12) \end{aligned}$$

These vectors, expressing probabilities of locations of the three categories along the horizontal view of the agent, are bind to the corresponding Nengo SPA items, using NEF $\circledast$ operator, introduced in equation (2). We can express the binding in our case as a function $b(\cdot) : \mathbb{R}^N \to \mathbb{R}^N$:

$$\begin{aligned} b\left(\vec{V}_\mathrm{P}\right) &= \vec{V}_\mathrm{P} \circledast \vec{p}_\mathrm{P}, & (13)\\ b\left(\vec{V}'\right) &= \vec{V}' \circledast \vec{p}', & (14)\\ b\left(\vec{V}''\right) &= \vec{V}' \circledast \vec{p}'', & (15) \end{aligned}$$

and the disambiguate item $\vec{V}^*$ is selected as following:

$$\vec{V}^* = \arg \min_{\vec{V} \in \{\vec{V}', \vec{V}''\}} \left\{ \zeta \left(b\left(\vec{V}_\mathrm{P}\right), b\left(\vec{V}\right) \right) \right\} \qquad (16)$$

where $\zeta\left(\vec{V}_1, \vec{V}_2\right)$ is a measure of similarity between the two Nengo SPA items $\vec{V}_1$ and $\vec{V}_2$. Therefore, the predicted head of the prepositional phrase W^*_H is the lexical item associated with $\vec{V}^*$.

The combined deep convolutional and Nengo SPA neural processes are sketched in Fig. 1.

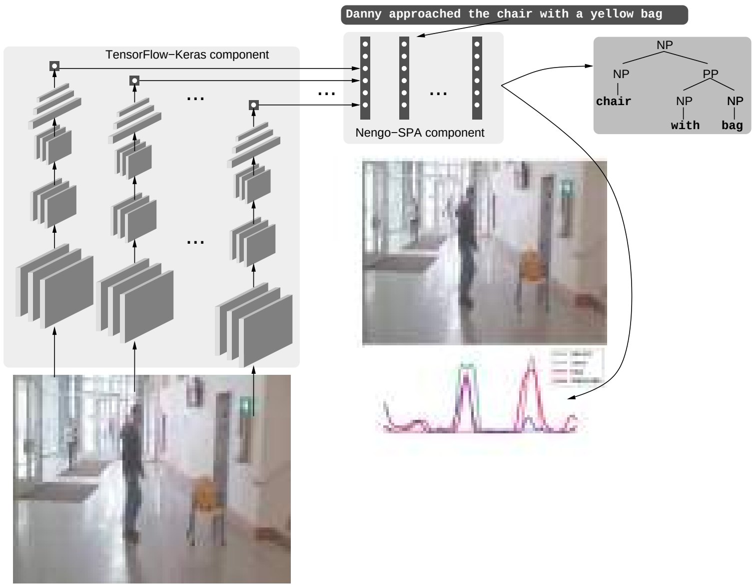

Figure 1: Sketch of the neural model.

4 RESULTS

We evaluated the performance of our neural model, described in the previous section, on the LAVA dataset. In this section we report our main preliminary results. Each sentence and its associated picture in the dataset was processed, predicting the lexical item that most likely is the head of the ambiguous propositional phrase. We recall that in this preliminary work, the ambiguous phrases examined by the system are limited to the preposition "with". The code we used for constructing the neural model and computing our experimental results is available for download at `http:www.github.com/alex-me/nencog`.

Here we will present first a set of qualitative results useful for illustrating the neural processes performed by the model, and then the quantitative results of the disambiguation task.

Fig. 2 illustrates few qualitative results of the intermediate stage of the process, when the vectors $\vec{p}^{(C)}$ of probabilities to find an object of category C in the discrete horizontal positions X are computed. Vectors are generated by applying the deep convolutional neural network $f_C(\cdot)$ to all the submatrices of a given image according to equation (6). Fig. 2 include the following 6 examples:

LAVA ID code	sentence
`22-9570-9660`	Danny approached the chair with a yellow bag
`22-18590-18700`	Danny left the chair with a yellow bag
`22-22420-22510`	Danny left the chair with a green bag
`22-54050-54160`	Danny approached the chair with a blue telescope
`22-55780-55850`	Danny approached the chair with a blue telescope
`29-24110-24210`	Danny looked at the chair with a blue telescope

The vectors of the kind shown in Fig. 2 that are relevant in the sentence, are then associated with the three Nengo SPA units $\vec{V}_{\mathrm{P}}$, $\vec{V}'$, $\vec{V}''$ by means of equations (13), (14), (15). All SPA units are populations of spiking neurons, which vectors evolve in time following equation (1). This evolution is shown in Fig. 3 for a small number of examples. All plots show the evolution in time of the three vectors related with Nengo SPA units $\vec{V}_{\mathrm{P}}$, $\vec{V}'$, $\vec{V}''$. The crucial aspect for the purpose of the disambiguation is that the final shape of the vectors is such that between $\vec{V}'$ and $\vec{V}''$ the most similar to $\vec{V}_{\mathrm{P}}$ will be the SPA unit associated with the correct word W_{H}. This final similarity can be appreciated in the four examples of Fig. 3.

Table 1 presents the quantitative results of the model over all the processed LAVA sentences. The total set of sentences has been divided into two categories, those with W_{P}=`bag` and those with W_{P}=`telescope`. In the first set the possible correct heads of the prepositional phrase W_{H} can be `person` or `chair`, while in the second set the possibilities are `bag`, `person`, and `chair`. For each of the sets the matrix of errors is reported, showing the fractions of lexical element that the model has predicted as head of the prepositional phrase, given the correct head word. The overall accuracy of the model is good, over 80%, and slightly lower when W_{P}=`telescope`.

As in the case of the experiments by Berzak et al. (2015), the most significant source of failures are poor object detection. Objects in the LAVA corpus are often rotated and presented at angles which turns out to be difficult to recognize.

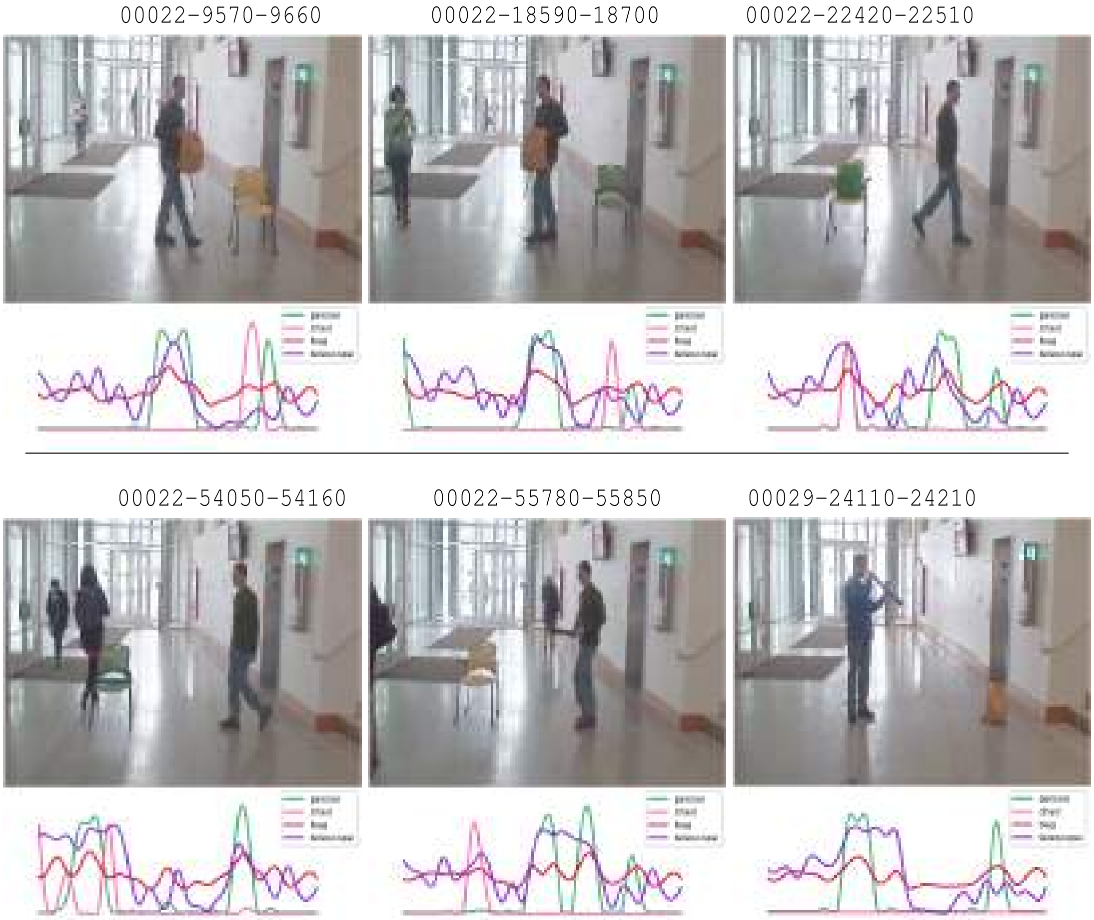

Figure 2: The vectors $\vec{p}^{(C)}$ of probabilities to find an object of category $C \in \{$ `person`, `bag`, `telescope`, `chair` $\}$ in the discrete horizontal positions X computed for 6 different images of the LAVA corpus. Vectors are generated by applying the deep convolutional neural network $f_C(\cdot)$ to all the submatrices of a given image according to equation (6). Images are labeled with their corresponding codes in the LAVA corpus.

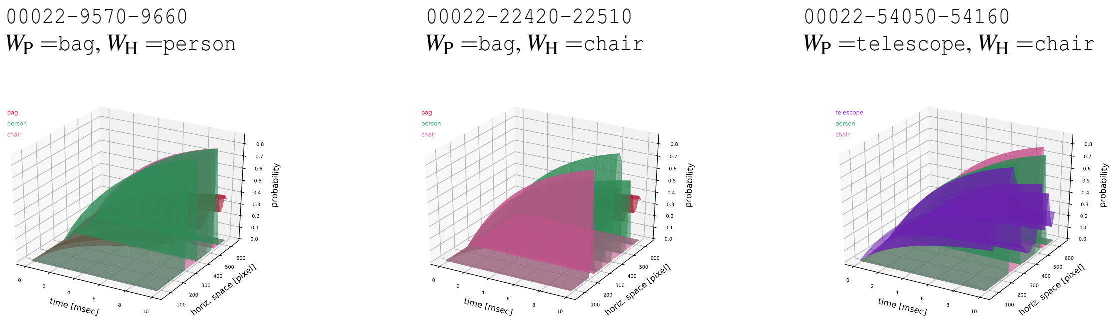

Figure 3: Evolution in time of the Nengo SPA neural populations associated with the three key words in the disambiguation tasks. In the example in the left the noun under the head of the preposition is `bag`, and the head of the prepositional phrase is `person`, it is visible how at the end of the evolution the SPA vector associated with `bag` become more similar to the vector associated with `person`, with respect to the vector associated with `chair`. Exactly the opposite happen in the example in the center, where this time the noun under the head of the preposition, `bag` has `chair` as head of the prepositional phrase. The scenes corresponding to the three examples can be seen in Fig. 2.

Table 1: Overall and detailed accuracy obtained by the model when tested on the LAVA dataset. The results are grouped for the two possible W_P, and for each one the matrix of errors is shown, with the true W_H as rows, and the predicted W_H^* as columns. The overall accuracy is of 0.81.

W_P =`bag`

$W_H \setminus W_H$	`person`	`chair`
`person`	.73	.27
`chair`	.00	1.00
accuracy	0.87	

W_P =`telescope`

$W_H \setminus W_H$	`bag`	`person`	`chair`
`bag`	.25	.75	.00
`person`	.00	.96	.04
`chair`	.07	.20	.73
accuracy	0.77		

It turns out moreover that some object classes, like the telescope and the bag, are much more difficult to be recognized. It can be observed in Fig. 2 that objects of the classes `bag` and `telescope` are the most difficult to be recognized due to their small size and to the fact that hands tend, in most cases, to largely cover them. Conversely objects of the classes `person` and `chair` are generally well detected and generate a much more accurate probability vector. We have assessed this source of error by evaluating the pure visual recognition accuracy, which is of 80% for person objects, of 79% for chair object, of 67% for the bag object, and as low as 60% for telescope.

Note that we deliberately avoided to include in the model a state-of-the-art deep learning model that would have easily achieved better recognition rates, but loosing biological plausibility.

Moreover, with our model we have been able to evaluate the disambiguation performances that takes into account uncertainty in the visual process. As seen in the performances shown in Table 1, disambiguation is more reliable that the pure visual object recognition.

5 CONCLUSIONS

We described a biologically plausible neural cognitive model able to resolve linguistic ambiguities in a sentences by selecting the interpretation of an ambiguous sentence matching the content of a given image. The model has been based on Nengo, using SPA for representing concepts.

The component of our model dealing with visual object recognition is based on deep convolutional networks, which are less biological plausible than Nengo. This solution is motivated by the marginal significance of object recognition in our objectives, and by the well known performances of deep convolutional networks.

Our neural model has been evaluated on the subset of sentences in the LAVA dataset, in which the preposition `with` is responsible for the contextual ambiguities.

Our model achieve an ability to resolve linguist ambiguities of the kind described, in the LAVA dataset, over 80%. It turns out that the most significant source of failures is poor object detection, as in the case of Berzak et al. (2015), however, the obtained disambiguation accuracy is greater that the visual recognition error.

REFERENCES

Airenti, G. and Plebe, A. (2017). Editorial: Context in communication: A cognitive view. *Frontiers in Psychology*, 8:115.

Barsalou, L. W. (1983). Ad hoc concepts. *Memory and Cognition*, 11:211–217.

Bates, E., Dal, P. S., and Thal, D. (1995). Individual differences and their implications for theories of language development. In Fletcher, P. and Whinney, B. M., editors, *Handbook of child language*, pages 96–151. Basil Blackwell, Oxford (UK).

Berzak, Y., Barbu, A., Harari, D., Katz, B., and Ullman, S. (2015). Do you see what i mean? visual resolution of linguistic ambiguities. In *Conference on Empirical Methods in Natural Language Processing*, pages 1477–1487.

Bornstein, M. H. and R.Cote, L. (2004). Cross-linguistic analysis of vocabulary in young children: Spanish, dutch, french, hebrew, italian, korean, and american english. *Child Development*, 75:1115–1139.

Eliasmith, C. (2013). *How to build a brain: a neural architecture for biological cognition*. Oxford University Press, Oxford (UK).

Eliasmith, C. and Anderson, C. H. (2003). *Neural Engineering Computation, Representation, and Dynamics in Neurobiological Systems*. MIT Press, Cambridge (MA).

Goodwin, C. and Duranti, A. (1992). Rethinking context: an introduction. In Duranti, A. and Goodwin, C., editors, *Rethinking context: Language as an interactive phenomenon*, pages 1–42, Cambridge (UK). Cambridge University Press.

Hindle, D. and Rooth, M. (1993). Structural ambiguity and lexical relations. *Cognitive Linguistics*, 19:103–120.

Machery, E. (2015). By default: Concepts are accessed in a context-independent manner. In Laurence, S. and Margolis, E., editors, *The Conceptual Mind: New Directions in the Study of Concepts*. MIT Press, Cambridge (MA).

Mazzone, M. and Lalumera, E. (2009). Concepts: Stored or created? *Minds and Machines*, 20:47–68.

Plate, T. (2003). *Holographic reduced representations*. CSLI Publication, Stanford, (CA).

Plebe, A. and De La Cruz, V. M. (2020). Neural representations in context. In Pennisi, A. and Falzone, A., editors, *The Extended Theory of Cognitive Creativity – Interdisciplinary Approaches to Performativity*, pages 285–300. Springer, Berlin.

Plebe, A. and Grasso, G. (2019). The unbearable shallow understanding of deep learning. *Minds and Machines*, 29:515–553.

Searle, J. R. (1978). Literal meaning. *Erkenntnis*, 13:207–224.

Stark, S. M., Reagh, Z. M., Yassa, M. A., and Stark, C. E. (2018). What's in a context? cautions, limitations, and potential paths forward. *Neuroscience Letters*.

Thagard, P. (2011). Cognitive architectures. In Frankish, K. and Ramsey, W., editors, *The Cambridge Handbook of Cognitive Science*, pages 50–70. Cambridge University Press, Cambridge (UK).

VanRullen, R. (2017). Perception science in the age of deep neural networks. *Frontiers in Psychology*, 8:142.