Parallels between Machine and Brain Decoding

Lorenzo Dell'Arciprete[1], Brian Murphy[2], and Fabio Massimo Zanzotto[1]

[1] Artificial Intelligence Research, University of Rome Tor Vergata, Rome, Italy
fabio.massimo.zanzotto@uniroma2.it
[2] Machine Learning Department, Carnegie Mellon University, Pittsburgh, USA
brianmurphy@cmu.edu

Abstract. We report some existing work, inspired by analogies between human thought and machine computation, showing that the informational state of a digital computer can be decoded in a similar way to brain decoding. We then discuss some proposed work that would leverage this analogy to shed light on the amount of information that may be missed by the technical limitations of current neuroimaging technologies.

1 Introduction

Analogies have often been drawn between machine computation and human thought. In computer science biological principles are sometimes appealed to when designing machines. For instance in cybernetics [1], artificial adaptive machines – that is machines that can control their states – have been studied with respect to the adaptivity of natural living organisms. And both the Von Neumann architecture [2], and the neural-based computing architecture originally introduced by Turing [3] show the influence of concepts coming from the study of the mind and of the brain. In contemporary cognitive sciences the analogy is also widespread, a computational theory of mind [4] being central to the "cognitive revolution" in the second half of the last century. For example in the work of Chomsky [5, 6] we see cognitive operations (in this case linguistic) being rooted in formal mathematical models, and in various guises this idea continues to be an influential model for understanding cognition more generally (see e.g. [7]).

Here we try to explore whether the analogy may provide *concrete* insights into the nature of the task of decoding brain states from neuroimaging data, describing in this paper some preliminary work that might inform the fields of cognitive neuroscience, and machine learning from neuroimaging data.

In neural decoding analyses, we take recordings of brain activity, which are noisy and limited in resolution, and use machine learning methods to determine which patterns of activity consistently co-occur with the cognitive states and processes that are active in the minds of participants. One could take a similar "black box", data-driven approach to reading the informational state of a computer, by learning the relationship between particular computational tasks and snapshots of its contemporary RAM contents. But of course in the case of a computer, unlike in the human brain, the ground-truth is well-defined: the precise informational state of the machine is known at all times, and the distinction between process (code) and representation (data) is unambiguous. Previous papers have demonstrated that such decoding of RAM-states is possible for classical computational tasks [8, 9] and that it is even possible to produce real chip scanners that could capture RAM activation images [10].

In this paper, we propose to use the more cognitively realistic tasks of linguistic processing for classification of semantic and syntactic categories (analogous to cognitive-neuroscience studies [11–15]). In our view, observing this computational task can give concrete insights on the task of decoding neuroimaging data. In section 2 we discuss how such an analysis could give us a better handle on the limitations of resolution imposed by physiological properties of the brain and technical limitations of imaging devices: since in computer decoding we have access to the "true" informational state, we can investigate the extent to which we may be missing information in the relatively impoverished brain data to which we have access. In section 3, we describe the computational task of linguistic processing from the distributional perspective. In section 4, we estimate the size of the brain areas involved in particular linguistic processing tasks. And, finally, in section 5, we discuss the experiments on decoding computational processing while exploring different blurring levels. With this experiments, we aim to investigate how decoding performances decrease with blurring. This is an useful insight for the understanding of brain state decoding.

2 Modelling Resolution Limitations of Imaging Technologies

Current neuroimaging methods have impressive spatial and temporal resolution, but are still working at a level far from the actual physical phenomenon of interest: millions or billions of single neuron firing events. For example, MEG has a temporal resolution that can capture the full temporal dynamics of neuronal firing, but spatially a single channel may aggregate the activity of the order of 10^9 individual neurons.[3] EEG is further limited in that its sensors are sensitive to larger overlapping areas of cortex, and that the low-pass filtering properties of the skull make higher frequencies hard to record. Turning to fMRI, its effective sampling rate of under 0.5Hz (due to the sluggish blood-oxygen level response) is well below the firing rates found in neurons (ca. 10-1000Hz), and even at higher limits of its spatial resolution (voxels of size 1mm^3) it still samples the order of 10^4 neurons at a time.[4] Capabilities for recording individual neuron activity is limited to very small numbers of cells, typically under 100 and is usually random in the particular neurons it samples in a brain locality of interest. As a result we do not have a detailed understanding of the functional activity of large neuronal populations, nor consequently how the patterns of activity seen in neuroimaging experiments relate to it. Furthermore, from the machine learning perspective,

[3] Assuming 100-300 channels, and 100 billion neurons in total
[4] Assuming cortex covered by ca. 700,000 voxels of size 1mm^3

we do not have a handle on how impoverished our data is, relative to the full neuronal population activity that we would ideally have access to.

Here we propose to make use of the computer-brain analogy in the following way. We will choose cognitive tasks that are studied in neuroscience, and can be emulated successfully by a computer – in this case detecting the semantic and syntactic categories involved in noun-phrase composition. While such a task is being performed by the computer we will take snapshots of RAM state. Given this very rich data, we expect that high-accuracy decoding should be achievable. The question we will then ask of the data, is how much accuracy is degraded as the input data is downsampled in ways that reflect neuroimaging technologies, such as MEG, EEG or fMRI.

3 Computer-based Distributional Semantic Processing

Processing natural language is one of the key activities of artificial intelligence. Many morphological, syntactic, and semantic formal models are available. Even if many of these approaches are not *cognitively inspired*, we can rely on a good basis of computational models to experiment with our idea of a comparative study between brain activities and machine activities.

Among the others, distributional semantics is an attractive model for our comparative study. Unlike symbolic formal semantics for natural language [16], word meaning is represented in the memory as vectors of real numbers. These vectors can be easily seen as activation images like the activation images of the brain. These idea of observing vectors as images has been also used in the slightly different context of distributed knowledge representation (see [17]). The recently revitalized trend of compositional models for distributional semantics [18–22] produces interesting computational processing models for semantic interpretation of natural language utterances. This is a semantic process and the model along with the distributional semantic vectors can be easily seen as activation images.

The rest of the section is organized as follows. First, section 3.1 introduces to distributional semantic principles and to linear compositional distributional semantic models. Then, section 3.2 describes how these computational models can be easily transformed in activation matrices and, consequently, activation images.

3.1 A Linear Compositional Distributional Semantic Model

Lexical distributional semantics has been largely used to model word meaning in many fields as computational linguistics [23, 24], linguistics [25], corpus linguistics [26], and cognitive research [27]. The fundamental hypothesis is the distributional hypothesis (DH): "similar words share similar contexts" [25]. Recently, this hypothesis has been operationally defined in many ways in the fields of physiology, computational linguistics, and information retrieval [28–30].

Given the successful application to words, distributional semantics has been extended to word sequences. This has happened in two ways: (1) via the reformulation of DH for specific word sequences [31]; and (2) via the definition of compositional distributional semantics (CDS) models [18, 19]. These are two different ways of addressing the problem.

Lin and Pantel [31] propose the *pattern distributional hypothesis* that extends the distributional hypothesis for specific patterns, i.e. word sequences representing partial verb phrases. Distributional meaning for these patterns is derived directly by looking to their occurrences in a corpus. Due to data sparsity, patterns of different length appear with very different frequencies in the corpus, affecting their statistics detrimentally. On the other hand, compositional distributional semantics (CDS) propose to obtain distributional meaning for sequences by composing the vectors of the words in the sequences [18, 19]. This approach is fairly interesting as the distributional meaning of sequences of different length is obtained by composing distributional vectors of single words.

A compositional distributional semantic model aims to compute the distributional meaning of word sequences by composing distributional vectors of individual words. Focussing on 2-word sequences, e.g., $z =$*close contact*, the CDS model has to compute the distributional vector $\boldsymbol{z}$ for the entire sequence using the distributional vectors $\boldsymbol{u}$ and $\boldsymbol{v}$ of, respectively, *close* and *contact*.

Among all the models, we focus here on the generic *additive* model that sums the vectors $\boldsymbol{u}$ and $\boldsymbol{v}$ in a new vector $\boldsymbol{z}$:

$$A\boldsymbol{u} + B\boldsymbol{v} = \boldsymbol{z} \tag{1}$$

where A and B are two square matrices capturing a particular syntactic relation R between the two words, e.g., adjective-noun (JN) for *close contact*. This linear model for semantic processing is extremely interesting as A and B activated by $\boldsymbol{u}$ and $\boldsymbol{v}$ can be easily seen as activation images.

For a good CDS model, we estimate matrices A and B using the methodology described in [20]. We can then have different CDS models for different syntactic relations. In the experiments, we use three different pairs of matrices for three different syntactic relations: adjective-noun (JN), noun-noun (NN), and verb-noun (VN).

3.2 Image-based Interpretation

Linear CDS are then interesting computational semantic processing models as we can easily interpret them as activation images. This idea is sketched in Figure 1. The original matrices A and B represent a Composition Matrix that is in active process. The Stimulus Vector, that represents the two distributional vectors for the two words, activate the process by producing the Activated Computation Matrix. This latter is then transformed in the final composition distributional vector. Looking the process in this way, we can easily derive an image representing the active state of the semantic composition computational model.

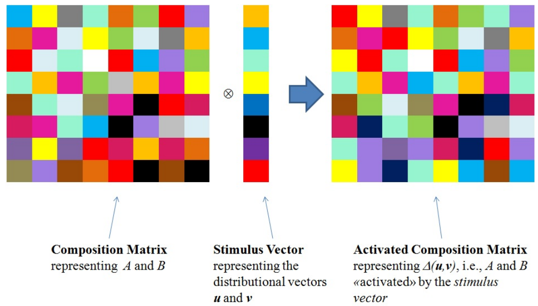

Fig. 1. Compositional Distributional Semantics: Activation of the Compositional Matrix

We can derive the above view by looking at the linear equations of the model:

$$A\boldsymbol{u} + B\boldsymbol{v} = \begin{bmatrix} a_{0,0} & a_{0,1} & a_{0,2} & \dots & a_{0,n} \\ a_{1,0} & a_{1,1} & a_{1,2} & \dots & a_{1,n} \\ a_{2,0} & a_{2,1} & a_{2,2} & \dots & a_{2,n} \\ & & & \ddots & \\ a_{n,0} & a_{n,1} & a_{n,2} & \dots & a_{n,n} \end{bmatrix} \begin{pmatrix} u_0 \\ u_1 \\ u_2 \\ \vdots \\ u_n \end{pmatrix} + \begin{bmatrix} b_{0,0} & b_{0,1} & b_{0,2} & \dots & b_{0,n} \\ b_{1,0} & b_{1,1} & b_{1,2} & \dots & b_{1,n} \\ b_{2,0} & b_{2,1} & b_{2,2} & \dots & b_{2,n} \\ & & & \ddots & \\ b_{n,0} & b_{n,1} & b_{n,2} & \dots & b_{n,n} \end{bmatrix} \begin{pmatrix} v_0 \\ v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix} =$$

$$= \begin{bmatrix} a_{0,0}u_0 + b_{0,0}v_0 & a_{0,1}u_1 + b_{0,1}v_1 & a_{0,2}u_2 + b_{0,2}v_2 & \dots & a_{0,n}u_n + b_{0,n}v_n \\ a_{1,0}u_0 + b_{1,0}v_0 & a_{1,1}u_1 + b_{1,1}v_1 & a_{1,2}u_2 + b_{1,2}v_2 & \dots & a_{1,n}u_n + b_{1,n}v_n \\ a_{2,0}u_0 + b_{2,0}v_0 & a_{2,1}u_1 + b_{2,1}v_1 & a_{2,2}u_2 + b_{2,2}v_2 & \dots & a_{2,n}u_n + b_{2,n}v_n \\ & & & \ddots & \\ a_{n,0}u_0 + b_{n,0}v_0 & a_{n,1}u_1 + b_{n,1}v_1 & a_{n,2}u_2 + b_{n,2}v_2 & \dots & a_{n,n}u_n + b_{n,n}v_n \end{bmatrix} \begin{pmatrix} 1 \\ 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} =$$

$$= \Delta(\boldsymbol{u}, \boldsymbol{v}) \begin{pmatrix} 1 \\ 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} = \boldsymbol{z}$$

The above derivation describes how it is possible to model the activation of the A and B matrices as a separate step that is done before the final sum. The matrix $\Delta(\boldsymbol{u}, \boldsymbol{v})$ is then a matrix representing the activation of A and B with respect to the input vectors $\boldsymbol{u}$ and $\boldsymbol{v}$. Matrices $\Delta(\boldsymbol{u}, \boldsymbol{v})$ are then easily transformed in activation images as done in [8, 9].

3.3 Decoding Tasks for Computational Semantic Processing

The previous compositional and distributional semantic processing model opens two possible decoding tasks:

T_1) decoding the kind of composition, e.g., deciding whether activation images $\Delta(\boldsymbol{u}, \boldsymbol{v})$ are related to JN, NN, or VN semantic composition activity;

T_2) decoding the semantic types of the involved vectors $\boldsymbol{v}$ and $\boldsymbol{u}$ in the composition looking at the activation matrix $\Delta(\boldsymbol{u}, \boldsymbol{v})$, e.g., decide *animals* vs. *tools*.

4 Estimating the Size of Correlated Brain Areas

We have the opportunity to use the two computational semantic processing tasks to investigate our primary problem. We need now to understand the size of the brain areas that are related to these two tasks. The literature suggests that the two most relevant areas are:

- Brocas area for the syntactic composition task (T_1)
- the fusiform gyri for the animal-vs-tool task (T_2)

We need now to estimate the size of each area in order to tune the size of the computational semantic composition matrices A and B.

4.1 Deciding on Scale of Neural Areas

To model the scale of the neural decoding problem, we should first consider several candidates for basic units of neural hardware: neurons, minicolumns, and macrocolumns.

Neurons instantaneously can be either on or off (like a bit). The cerebrum (that is the cerebral hemispheres, or neocortex), is where the great majority of higher cognition takes place, and it contains about 20 billion of neurons. The precise number varies with age and gender, and estimates are affected by the methodologies adopted by each study. Pakkenberg and colleagues [32] found a wide range across individuals (also of similar ages), from 15 billion to 32 billion, with the mean for females at 18 billion, and for males at 23 billion. Stark and colleagues [33] have very similar average figures for females and males (20 billion and 23 billion). Here we choose a particularly careful whole-brain estimate which found an average of 16 billion over four individuals [34].

Minicolumns are collections of about 100 neurons (so there are 200 million in the cerebrum). One byte could be used to represent the firing rate of each of these neural assemblies [35].

Macrocolumns (also termed hypercolumns, or simply cortical columns) are larger collection of neurons that contain approximately 5000-10000 neurons (so 3 million in cerebrum) [36]. Optimistically, macrocolumns are the smallest structure (3mm deep, spaced at 0.5mm) that can be detected individually with fMRI or MEG. We estimate that the state of a single macrocolumn would require one or more bytes to be represented.

4.2 Dimensions of relevant brain areas

Based on these cortical densities, and volume estimates for functionally relevant parts of the brain, we can estimate the dimensionality needed to model them (see table 1). These numbers are estimated as such. The young healthy brain

Cerebrum	neurons	minicolumns	macrocolumns
size	$2 \cdot 10^{10}$	$2 \cdot 10^{8}$	$3 \cdot 10^{6}$
Ram size (instantaneous)	2.5Gb	200MB	3MB
Fusiform gyri (2.8% of cerebrum)			
N	$5.6 \cdot 10^{8}$	$5.6 \cdot 10^{6}$	$8.4 \cdot 10^{4}$
Ram size (instantaneous)	452Mb	4.5 MB	68 kB
Brocas areas (0.5% of cerebrum)			
size	$1 \cdot 10^{8}$	$1 \cdot 10^{6}$	$1.5 \cdot 10^{4}$
Ram size (instantaneous)	80Mb	0.8MB	12kB

Table 1. Brain areas in bytes

has a volume of about 1250 cm^3 (see [37]). The cortical activity detected by fMRI, MEG and EEG is in the cortex, or grey matter surface of the brain tissues. Estimates of grey/white matter proportions vary, and here we use the figure of 1.35 derived from [38] (estimated over 80 individuals of both genders). Taking this ratio, and the estimated volumes from [34], we estimate the amount of grey matter in a typical participant to be approximately 710 cm^3, containing 16 billion neurons. This leads to an estimate of cerebral neural density of 23 million neurons per cm^3, which we assume to be uniform. This density can be used to estimate the number of neurons (and hence micro- and macrocolumns) in a functionally relevant part of the brain.

For the semantic category decoding task we take the fusiform gyri (left and right, part of Brodmann area 37), which have a grey matter volume of approximately $20cm^3$ [37]. These areas are involved in high level processing of visual categories (including distinguishing between pictures of living vs non-living things), and which have been shown to have a similar discriminative activity in linguistic tasks (with word stimuli), even in congenitally blind participants who have no visual experience [39].

For the syntactic task, we consider Broca's area, which is generally agreed to be central to processing linguistic structures [40]. Here we take this to cover the pars opercularis and pars triangularis (Brodmann areas 44 and 45) of the left inferior frontal gyrus, with grey-matter volume of $5cm^3$ [41].

5 Experimental Investigation

5.1 Experimental Set-up

Classification Tasks Two classification tasks have been designed in the context of compositional distributional semantics processing.

The first one aims at distinguishing between the different syntactic relations among the concepts being composed. We considered three syntactic relations: noun-noun (NN), adjective-noun (JN) and verb-object (VN). As specified in Section 3.1, the three classes of compositions are performed by means of different pairs of matrices A and B. The data set included 100 term pairs for each class, selected at random from the set of terms whose distributional vectors were produced. The set was split into a 70% training set and a 30% testing set, uniformly distributed among the three classes. For this experiment, the classifiers were used in the context of a multi-class classification.

In the second experiment, only the JN syntactic relation was considered, but the adjective-noun pairs were chosen among two different semantic domains. The domains are those of *animals* versus *objects*. *Aggressive anteater* and *shy reindeer* are examples of the *animals* class, while *elegant dress* and *modern pot* are examples of the *objects* class. In this case, the experiment tried to relate the composition process to a semantic domain rather than to a syntactic relation. While this task could have been performed on nouns only, we still used noun-adjective pairs to work in the same setting of the first experiment and of the general model presented in the paper. The data set included 123 pairs for each class, and was again split into uniformly distributed 70% training set and a 30% testing set.

Feature extraction from activation images For learning and applying a classifier, we need to extract specific features from images generated using the activation matrices $\Delta(\boldsymbol{u}, \boldsymbol{v})$. We then used two major classes of features: chromatic and energetic. Chromaticity features express the color properties of the image. They determine, in particular, a n-dimensional vector representation of the 2D chromaticity histograms. Since chromaticity is invariant against changes in the illuminant color, the intensity information can also be evaluated using simple color histograms, one for each color component R, G and B and the luminance L. Energy features emphasize the background properties and their composition. They are extracted by generating a texture-energy image, and calculating the energy images for the three color channels R, G and B. A more detailed discussion of the theoretical and methodological aspects behind each feature set are presented in [42].

Classifier learners For finally building the classifiers of the "*cognitive task*" that the machine is performing, we used three alternative machine learning models. This is useful to see whether or not results are confirmed for any kind of classification method. We then used: a decision tree based learner [43], a simple Naive Bayes classifier (for more information see [44]), and, finally, an instance based learner (IBk) [45]. These machine learning methods have been used in the context of Weka [46].

The three models are different. Decision tree learners capture and select the best features for doing the classification. Naive Bayes learners instead use a simple probabilistic model that considers all the features to be independent. The instance based learner defines a distance in the feature space, does not make any abstraction of the samples, and classifies new instances according to the distance of these new elements with respect to training samples. While the first model makes a sort of feature selection, the second and the third use all the features for taking the final decision.

Distributional Vector Extraction Finally, we describe how the distributional vectors were obtained for all experiments of this paper. Raw frequency distributional vectors were obtained from the UKWaC British English web corpus[5]. We considered as contextual window the sentence in which each target word or linguistic unit occurs. Features are contextual words and the weighting scheme is term frequency times inverse document frequency ($tf \times idf$). We applied a crude feature selection using $tf \times idf$ and keeping the first 10.000 dimensions. The resulting distributional vectors constitute a high dimensional vector space model. An SVD reduction with $k = 250$ was then applied to the vector space to build our final distributional vector set.

Learning of the Compositional Matrices For learning the compositional matrices A and B, we used the methodology described in [9]. As we operated in English, we used the definitions in WordNet [47] to extract training instances for the dictionary-based method. We extracted bigram training instances that follow three syntactic structures, i.e. noun-noun (NN-WN), adjective-noun (JN-WN) and verb-object (VN-WN). Respectively, we used 1220, 6131, and 1317 triples to learn different A and B matrices for NN, JN, and VN. For the details of the methodology, refer to [9].

Dimension of the CDS matrices The CDS process is represented by an activation matrix of 500x250 decimal values (in double format, i.e. 8 bytes each). This leads to a byte matrix of size 4000x250, for a total of 1000000 bytes ($\sim$1 MB). Looking at Table 1, the modeled process has an order of magnitude similar to what we would get by observing the Broca's area or the fusiform gyri at the level of the minicolumns. Applying a blurring factor $\sim$10 to the resulting images can simulate the observation at the level of the macrocolumns.

5.2 Results and Analysis

Tables 2 and 3 report the results of the experiments. Both are organized in the same way. The first column reports the different blurring levels we experimented on. The second and third columns give an estimate of the corresponding level of blurring that would arise by observing the brain at the levels of minicolumns and macrocolumns respectively. The estimates are based on the orders of magnitude reported in table 1 for the sizes of the brain areas. The last three columns report the results of the classification task obtained by the three considered classifiers.

[5] http://trac.sketchengine.co.uk/wiki/Corpora/UKWaC

RAM blurring factor	Minicolumns blurring factor	Macrocolumns blurring factor	Decision Tree	Naive Bayes	IBk
1	1	-	96.67%	98.89%	98.89%
2	2	-	95.56%	98.89%	96.67%
4	4	-	96.67%	98.89%	96.67%
10	10	1	93.33%	98.89%	92.22%
100	100	10	88.89%	92.22%	94.44%

Table 2. Test 1: classify JN vs NN vs VN

The results of the first experiment are reported in Table 2. The accuracies scored by all the classifiers are very high, even though the task requires a multi-class classification. This comes from the fact that in this experiment both the inactive process (matrices A and B) and the input stimuli (vectors $\boldsymbol{u}$ and $\boldsymbol{v}$) are different. The introduction of blurring factors has an impact on the results, but still allows for high accuracies.

RAM blurring factor	Minicolumns blurring factor	Macrocolumns blurring factor	Decision Tree	Naive Bayes	IBk
1	1	-	59,46%	75,68%	54,05%
2	2	-	63,51%	70,27%	64,86%
4	4	-	62,16%	70,27%	63,51%
10	10	1	60,81%	66,22%	52,70%
100	100	10	51,35%	52,70%	54,05%

Table 3. Test 2: classify animals vs objects

The results of the second experiment are reported in Table 3. In this case the accuracies are much lower, especially considering that this task requires a binary classification. The introduction of blurring factors has different impacts, depending on the considered classifier. The Naive Bayes classifier is the one that scores higher accuracies, but its performances degrade rapidly when introducing higher blurring levels. The other two classifiers, instead, score lower accuracies, but are less affected by the introduction of blurring factors. In fact, they seem to benefit from a slight blurring.

Notice that the activation matrices in themselves appear to contain enough information to perform a correct classification. In fact, the same experiment run using the explicit activation matrices as features yields an accuracy of 100% for Naive Bayes and IBk, and 91.89% for Decision Trees.

These results are very encouraging, and we hope to build upon them in future work. A priority will be to directly compare our simulations of decoding performance with classification accuracies achieved for neuroimaging data recorded during similar tasks. We also hope to examine the dimension of time (in which EEG and MEG have an advantage over fMRI) and compare the effect of different trade-offs of temporal and spatial resolution that neuroimaging technologies provide.

References

1. Wiener, N.: Cybernetics: Or the Control and Communication in the Animal and the Machine. MIT Press, Cambridge, MA (1948)
2. von Neumann, J.: First draft of a report on the EDVAC. IEEE Ann. Hist. Comput. **15**(4) (1993) 27–75
3. Turing, A.: Intelligent machinery. In Meltzer, B., Michie, D., eds.: Machine Intelligence. Volume 5. Edinburgh University Press, Edinburgh (1969) 3–23
4. Putnam, H.: Minds and Machines. In Hook, S., ed.: Dimensions of Mind. Collier Books, New York and London (1960) 130–164
5. Chomsky, N.: Syntactic Structures. Mouton, The Hague (1957)
6. Chomsky, N.: A Review of B. F. Skinner's Verbal Behavior. Language **35**(1) (1959) 26–58
7. Pinker, S.: How the Mind Works. Norton and Company (2009)
8. Zanzotto, F.M., Croce, D.: Reading what machines "think". In: BI 2009: Proceedings of the Brain Informatics Conference - Bejing, China, October 2009. Volume LNAI 5819. Springer-Verlag, Berlin, Heidelberg (2009) 159–170
9. Zanzotto, F.M., Croce, D.: Comparing EEG/ERP-like and fMRI-like techniques for reading machine thoughts. In: BI 2010: Proceedings of the Brain Informatics Conference - Toronto. (2010) 133–144
10. Prezioso, S., Croce, D., Zanzotto, F.M.: Reading what machines think: a challenge for nanotechnology. Journal of Computational and Theoretical Nanoscience **8**(10) (2011) 1–6
11. Haxby, J.V., Gobbini, M.I., Furey, M.L., Ishai, A., Schouten, J.L., Pietrini, P.: Distributed and overlapping representations of faces and objects in ventral temporal cortex. Science **293**(5539) (2001) 2425–2430
12. Murphy, B., Baroni, M., Poesio, M.: EEG responds to conceptual stimuli and corpus semantics. In: Proceedings of EMNLP, ACL (2009) 619–627
13. Murphy, B., Poesio, M., Bovolo, F., Bruzzone, L., Dalponte, M., Lakany, H.: EEG decoding of semantic category reveals distributed representations for single concepts. Brain and Language **117**(1) (April 2011) 12–22
14. Chan, A.M., Halgren, E., Marinkovic, K., Cash, S.S.: Decoding word and category-specific spatiotemporal representations from MEG and EEG. NeuroImage **54**(4) (February 2011) 3028–39
15. Sudre, G., Pomerleau, D., Palatucci, M., Wehbe, L., Fyshe, A., Salmelin, R., Mitchell, T.: Tracking Neural Coding of Perceptual and Semantic Features of Concrete Nouns. NeuroImage **62**(1) (May 2012) 451–463
16. Montague, R.: English as a formal language. In Thomason, R., ed.: Formal Philosophy: Selected Papers of Richard Montague. Yale University Press, New Haven (1974) 188–221
17. Plate, T.A.: Distributed Representations and Nested Compositional Structure. PhD thesis (1994)
18. Mitchell, J., Lapata, M.: Vector-based models of semantic composition. In: Proceedings of ACL-08: HLT, Columbus, Ohio, Association for Computational Linguistics (June 2008) 236–244
19. Jones, M.N., Mewhort, D.J.K.: Representing word meaning and order information in a composite holographic lexicon. Psychological Review **114** (2007) 1–37
20. Zanzotto, F.M., Korkontzelos, I., Fallucchi, F., Manandhar, S.: Estimating linear models for compositional distributional semantics. In: Proceedings of the 23rd International Conference on Computational Linguistics (COLING). (August, 2010)

21. Baroni, M., Zamparelli, R.: Nouns are vectors, adjectives are matrices: Representing adjective-noun constructions in semantic space. In: Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing, Cambridge, MA, Association for Computational Linguistics (October 2010) 1183–1193
22. Guevara, E.: A regression model of adjective-noun compositionality in distributional semantics. In: Proceedings of the 2010 Workshop on GEometrical Models of Natural Language Semantics, Uppsala, Sweden, Association for Computational Linguistics (July 2010) 33–37
23. McCarthy, D., Carroll, J.: Disambiguating nouns, verbs, and adjectives using automatically acquired selectional preferences. Comput. Linguist. **29**(4) (2003) 639–654
24. Manning, C.D., Raghavan, P., Schütze, H.: Introduction to Information Retrieval. Cambridge University Press, Cambridge, UK (2008)
25. Harris, Z.: Distributional structure. In Katz, J.J., Fodor, J.A., eds.: The Philosophy of Linguistics, New York, Oxford University Press (1964)
26. Firth, J.R.: Papers in Linguistics. Oxford University Press., London (1957)
27. Miller, G.A., Charles, W.G.: Contextual correlates of semantic similarity. Language and Cognitive Processes **VI** (1991) 1–28
28. Li, P., Burgess, C., Lund, K.: The acquisition of word meaning through global lexical co-occurrences. In: Proceedings of the 31st Child Language Research Forum. (2000)
29. Pado, S., Lapata, M.: Dependency-based construction of semantic space models. Computational Linguistics **33**(2) (2007) 161–199
30. Deerwester, S.C., Dumais, S.T., Landauer, T.K., Furnas, G.W., Harshman, R.A.: Indexing by latent semantic analysis. Journal of the American Society of Information Science **41**(6) (1990) 391–407
31. Lin, D., Pantel, P.: DIRT-discovery of inference rules from text. In: Proceedings of the ACM Conference on Knowledge Discovery and Data Mining (KDD-01), San Francisco, CA (2001)
32. Pakkenberg, B., Gundersen, H.J.: Neocortical neuron number in humans: effect of sex and age. Journal of Comparative Neurology **384**(2) (1997) 312–320
33. Stark, A., Toft, M., Pakkenberg, H., Fabricius, K., Eriksen, N., Pelvig, D., Mller, M., Pakkenberg, B.: The effect of age and gender on the volume and size distribution of neocortical neurons. Neuroscience **150**(1) (2007) 121 – 130
34. Azevedo, F.A.C., Carvalho, L.R.B., Grinberg, L.T., Farfel, J.M., Ferretti, R.E.L., Leite, R.E.P., Jacob Filho, W., Lent, R., Herculano-Houzel, S.: Equal numbers of neuronal and nonneuronal cells make the human brain an isometrically scaled-up primate brain. Journal of Comparative Neurology **513**(5) (2009) 532–541
35. Buxhoeveden, D.P., Casanova, M.F.: The minicolumn hypothesis in neuroscience. Brain : a journal of neurology **125**(Pt 5) (May 2002) 935–51
36. Mountcastle, V.B.: The columnar organization of the neocortex. Brain: A journal of neurology **120**(4) (1997) 701–722
37. Raz, N., Gunning-Dixon, F., Head, D., Rodrigue, K.M., Williamson, A., Acker, J.D.: Aging, sexual dimorphism, and hemispheric asymmetry of the cerebral cortex: replicability of regional differences in volume. Neurobiol Aging **25**(3) (Mar 2004) 377–396
38. Gur, R.C., Turetsky, B.I., Matsui, M., Yan, M., Bilker, W., Hughett, P., Gur, R.E.: Sex differences in brain gray and white matter in healthy young adults: correlations with cognitive performance. J Neurosci **19**(10) (May 1999) 4065–4072
39. Mahon, B.Z., Anzellotti, S., Schwarzbach, J., Zampini, M., Caramazza, A.: Category-specific organization in the human brain does not require visual experience. Neuron **63**(3) (2009) 397–405

40. Hagoort, P.: On Broca, brain, and binding: a new framework. Trends in cognitive sciences **9**(9) (September 2005) 416–23
41. Yamasaki, S., Yamasue, H., Abe, O., Suga, M., Yamada, H., Inoue, H., Kuwabara, H., Kawakubo, Y., Yahata, N., Aoki, S., Kano, Y., Kato, N., Kasai, K.: Reduced gray matter volume of pars opercularis is associated with impaired social communication in high-functioning autism spectrum disorders. Biol Psychiatry **68**(12) (Dec 2010) 1141–1147
42. Alvarado, P., Doerfler, P., Wickel, J.: Axon^2 - a visual object recognition system for non-rigid objects. In: IASTED International Conference-Signal Processing, Pattern Recognition and Applications (SPPRA), Rhodes, IASTED (July 2001) 235–240
43. Quinlan, J.: C4:5:programs for Machine Learning. Morgan Kaufmann, San Mateo (1993)
44. John, G.H., Langley, P.: Estimating continuous distributions in bayesian classifiers. (1995) 338–345
45. Aha, D.W., Kibler, D., Albert, M.K.: Instance-based learning algorithms. Mach. Learn. **6**(1) (1991) 37–66
46. Witten, I.H., Frank, E.: Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations. Morgan Kaufmann, Chicago, IL (1999)
47. Miller, G.A.: WordNet: A lexical database for English. Communications of the ACM **38**(11) (November 1995) 39–41