Harm Brouwer, Matthew W. Crocker and Noortje Venhuizen. Neural Semantics. In: Wieling, M., Kroon, M., Van Noord, G., and Bouma, G. (eds.), From Semantics to Dialectometry: Festschrift for John Nerbonne, pp. 75–83. College Publications, 2017.

Chapter 8

Neural semantics

Harm Brouwer
Saarland University

Matthew W. Crocker
Saarland University

Noortje J. Venhuizen
Saarland University

The study of language is ultimately about meaning: how can meaning be constructed from linguistic signal, and how can it be represented? The human language comprehension system is highly efficient and accurate at attributing meaning to linguistic input. Hence, in trying to identify computational principles and representations for meaning construction, we should consider how these could be implemented at the neural level in the brain. Here, we introduce a framework for such a *neural semantics*. This framework offers meaning representations that are neurally plausible (can be implemented in neural hardware), expressive (capture negation, quantification, and modality), compositional (capture complex propositional meaning as the sum of its parts), graded (are probabilistic in nature), and inferential (allow for inferences beyond literal propositional content). Moreover, it is shown how these meaning representations can be constructed incrementally, on a word-by-word basis in a neurocomputational model of language processing.

1 Introduction

Language is about meaning. The aim of the study of language, therefore, is to capture and represent meaning, as well as to understand how it is constructed from linguistic input. Hence, albeit with different approaches and for different proximate goals, the fields of theoretical linguistics, computational linguistics, and psycholinguistics, all pursue systems that comprehend language. A successful comprehension system requires the representation of grammar, meaning, and knowledge about the world, be it in either a heterogeneous or homogeneous way (Nerbonne 1992). One system that is particularly effective and accurate at attributing meaning to linguistic input is the human language comprehension system. Crucially, this system is implemented in the neural hardware of the brain. This suggests that, in trying to identify optimal computational principles and representations for meaning derivation, we may want to turn to how those principles and representations are implemented in neural hardware; that is, we may want to understand meaning construction and representation in terms of 'brain-style computation' by identifying a *neural semantics*.

A framework for neural semantics should minimally meet the following requirements:

- **neural plausibility**: the assumed computational principles and representations should be implementable at the neural level (cf. Rumelhart 1989);
- **expressivity**: the representations should capture necessary dimensions of meaning, such as negation, quantification, and modality (cf. Frege 1892);
- **compositionality**: the meaning of complex propositions should be derivable from the meaning of its parts (cf. Partee 1984);
- **gradedness**: meaning representations are probabilistic, rather than discrete in nature (cf. Spivey 2008);
- **inferential**: the derivation of utterance meaning entails (direct) inferences that go beyond literal propositional content (cf. Johnson-Laird 1983);
- **incrementality**: as natural language unfolds over time, representations should allow for incremental construction (cf. Tanenhaus et al. 1995).

In the present paper, we will introduce a framework for neural semantics that offers meaning representations that meet these requirements. Moreover, we show how these representations can be used within a neurocomputational model of language processing, to derive them incrementally, on a word-by-word basis for unfolding linguistic input.

2 A framework for neural semantics

In order to model story comprehension, Golden & Rumelhart (1993) developed a framework for modeling mental representations as points in a high-dimensional space called "situation-state space" (see also Golden et al. 1994). In their model, there is a one-to-one mapping between dimensions of the situation-state space and propositional meaning. Frank et al. (2003) extended this localist model for story comprehension by incorporating a distributed notion of propositional meaning. In what follows, we will introduce this Distributed Situation-state Space (DSS) model and show that it captures the aforementioned requirements for a neural semantics.

Table 1: Distributed Situation-state Space.

	proposition_1	proposition_2	proposition_3	...	proposition_n
observation_1	1	0	0	...	1
observation_2	0	1	1	...	1
observation_3	1	1	0	...	0
...	.	.	.	...	.
observation_m	0	1	0	...	0

2.1 Distributed situation-state space

A DSS is an $m \times n$ matrix that is constituted of a large set of m observations of states-of-affairs in the world, defined in terms of n atomic propositions (e.g., $enter(john, restaurant)$ and $order(ellen, wine)$)—the smallest discerning units of propositional meaning. Each of the m observations in this matrix is encoded by setting atomic propositions that are the case in a given observation to 1/True and those that are not to 0/False (see Table 1). The resulting situation-state space matrix is then effectively one big truth table, in which each column represents the *situation vector* for its corresponding atomic proposition—a point in situation-state space on a Euclidean perspective.

Situation vectors encode the meaning of propositions in terms of the observations in which they are the case. As a result, propositions that are the case in a similar set of observations obtain a similar meaning, whereas propositions that mostly occur in different observations obtain a dissimilar meaning. Crucially, the co-occurrence of propositions across the entire set of m observations in a DSS naturally captures world knowledge; that is, some propositions may never co-occur (hard constraints; e.g., a person can only be a single place at any given time), and some propositions may co-occur more often than others (probabilistic constraints; e.g., a person may prefer certain activities over other).

2.2 The DSS model as a neural semantics

The DSS-derived situation vectors inherently satisfy, the aforementioned requirements. Firstly, situation vectors are **neurally plausible**. They can be represented at the neural level as firing patterns over neural ensembles, where vector components correspond to either the firing of single neurons or to the collective firing of neural populations.

Secondly, situation vectors are also **expressive** and **compositional**. The meaning of the negation of an atomic proposition a, for instance, is given by the situation vector $\vec{v}(\neg a)$ that assigns a 0 to all observations in which a is the case, and a 1 otherwise (thus resulting in a maximally different situation vector relative to $\vec{v}(a)$); this vector can be directly derived from $\vec{v}(a)$, the situation vector of a, as follows: $\vec{v}(\neg a) = 1 - \vec{v}(a)$. In a similar manner, the meaning of the conjunction between propositions a and b will be described by the situation vector that assigns 1 to all observations in which both a and b are the case, and 0 otherwise; this vector can be calculated by the pointwise multiplication of the situation vectors of a and b: $\vec{v}(a \wedge b) = \vec{v}(a)\vec{v}(b)$ for $a \neq b$, and $\vec{v}(a \wedge a) = \vec{v}(a)$.[1] Since the negation and conjunction operators together define a functionally complete system, the meaning of any other logical combination between propositions in situation-state space can be described using these two operations (in particular, the situation vector representing the disjunction between p and q, $\vec{v}(p \vee q)$, is defined as $\vec{v}(\neg(\neg p \wedge \neg q))$, which assigns a 1 to all observations in which either p or q is the case, and a 0 otherwise). Hence, we can combine atomic propositions into complex propositions, which can in turn be combined with other atomic and complex propositions, thus allowing for situation vectors of arbitrary complexity (i.e., both a and b can be either atomic or complex propositions in the aforementioned equations). This means that we can *minimally* capture all meanings expressible in propositional logic.

Thirdly, situation vectors constitute **graded** representations; that is, they inherently encode the (co-)occurrence probability of propositions. On the basis of the m observations in the situation-state space matrix, we can estimate the *prior probability* of the occurrence of each (atomic or complex) proposition a in the microworld from its situation vector $\vec{v}(a)$: $P(a) = \frac{1}{m}\sum_i \vec{v}_i(a)$. Indeed, this probability is simply the number of observations in which proposition a is the case, divided by the total number of observations constituting the situation-state space. Similarly the co-occurrence probability of two propositions a and b can also be estimated from their corresponding vectors $\vec{v}(a)$ and $\vec{v}(b)$: $P(a \wedge b) = \frac{1}{m}\sum_i \vec{v}_i(a)\vec{v}_i(b)$ for $a \neq b$, and $P(a \wedge a) = P(a)$. Crucially, this means that we can also compute the conditional probability of proposition a given b: $P(a|b) = \frac{P(a \wedge b)}{P(b)}$. Hence, given a proposition b, we can **infer** any proposition a that depends on b. Taking this one step further, this allows us to define a comprehension score $cs(a, b)$ that quantifies how much a proposition a is 'understood' from b: if $P(a|b) > P(a)$, then $cs(a, b) = \frac{P(a|b) - P(a)}{1 - P(a)}$, otherwise $cs(a, b) = \frac{P(a|b) - P(a)}{P(a)}$; this score yields a value ranging from $+1$ to -1, where $+1$ indicates that event a is perfectly 'understood' to be the case form b, whereas a value of -1 indicates that a is perfectly 'understood' *not* to be the case from b (Frank, Haselager & van Rooij 2009).

In sum, DSS-derived situation vectors offer meaning representations that are neurally plausible, expressive and compositional, as well as graded and inferential. Hence, the DSS model meets five of the six requirements for a neural semantics. But what about the requirement of **incrementality**?

[1] This is to account for real-valued situation vectors, which may result from applying dimension reduction to a DSS.

3 Neural semantics in a neurocomputational model

Natural language unfolds over time, and the human language comprehension system incrementally attributes meaning to this unfolding input (see e.g., Tanenhaus et al. 1995). In what follows, we will show that the DSS-derived meaning representations allow for such incremental meaning construction; that is, we will show how situation vectors for linguistic input can be derived on a word-by-word basis in a neurocomputational model of language processing.

3.1 A neurocomputational model

Our comprehension model is a Simple Recurrent Network (SRN; Elman 1990), consisting of three groups of artificial logistic dot-product neurons: an INPUT layer (22 units), HIDDEN layer (100), and OUTPUT layer (150). Time in the model is discrete, and at each processing time-step t, activation flows from the INPUT through the HIDDEN layer to the OUTPUT layer. In addition to the activation pattern at the INPUT layer, the HIDDEN layer also receives its own activation pattern at time-step $t-1$ as input (effectuated through an additional CONTEXT layer, which receives a copy of the activation pattern at the hidden layer prior to feedforward propagation). The HIDDEN and the OUTPUT layers both receive input from a bias unit. We trained the model using bounded gradient descent (Rohde 2002) to map sequences of localist word representations constituting the words of a sentence, onto a DSS-derived situation vector representing the meaning of that sentence (initial weight range: $(-.5, +.5)$; zero error radius: 0.05; learning rate: 0.1, momentum: 0.9; epochs: 5000). After training, the overall performance of the model was perfect (each output vector has a higher cosine similarity to its target vector than to any other target vector in the training data).

3.2 A microworld approach to DSS construction

The sentences on which the model is trained describe situations in a confined microworld (cf. Frank, Haselager & van Rooij 2009). This microworld is defined in terms of two persons $P = \{john, ellen\}$, two places $X = \{restaurant, bar\}$, and two types of food $F = \{pizza, fries\}$ and drinks $D = \{wine, beer\}$, which can be combined into 26 atomic propositions using the following 7 predicates: *enter(P,X)*, *ask_menu(P)*, *order(P,F/D)*, *eat(P,F)*, *drink(P,D)*, *pay(P)* and *leave(P)*. A DSS was constructed from these atomic propositions by sampling 10K observations (using a non-deterministic inference-based sampling algorithm), while taking into account hard and probabilistic constraints on proposition co-occurrence; for instance, a person can only enter a single place (hard), and *john* prefers to drink *beer* over *wine* (probabilistic). In order to employ situation vectors derived from this DSS in the SRN, we algorithmically selected 150 observations from these 10K that adequately reflected the structure of the world. Situations in the microworld were described using sentences from a microlanguage consisting of 22 words. The grammar of this microlanguage generates a total of 176 sentences, including simple (NP VP) and coordinated (NP VP *And* VP) sentences. The sentence-initial NPs may be *john*, *ellen*, *someone*, or *everyone*, and the VPs map onto the aforementioned propositions. The corresponding situation vectors for these sentences were derived using the machinery discussed above. In particular, existentially quantified sentences such as *Someone entered the restaurant and left* map onto a vector corresponding to a disjunctive semantics: $(enter(john, restaurant) \wedge leave(john)) \vee (enter(ellen, restaurant) \wedge leave(ellen))$. Universally quantified sentences, in turn, obtain a conjunctive semantics, e.g., *Everyone left* maps onto $leave(john) \wedge leave(ellen)$.

3.3 Incremental Neural Semantics

On the basis of its linguistic input, the model incrementally constructs a situation vector capturing its meaning; that is, the model effectively navigates DSS on a word-by-word basis. This means that we can study what it 'understands' at each word of a sentence by computing comprehension scores for relevant propositions (i.e., $cs(a, b)$, where a is the vector of a proposition of interest, and b the output vector of the SRN). Figure 1 shows the word-by-word comprehension scores for the sentence *John entered the restaurant and ordered wine* with respect to 6 propositions. First of all, this figure shows that by the end of the sentence, the model has understood its meaning: $enter(john, restaurant) \wedge order(john, wine)$. What is more, it does so on an incremental basis: at the word *restaurant*, the model commits to the inference $enter(john, restaurant)$, which rules out $enter(john, bar)$ (since these do not co-occur in the world; $P = 0$). At the word *ordered*, the model finds itself in state that is closer to the inference that $order(john, beer)$ than $order(john, wine)$ (as John prefers beer over wine; ($P = 0.63$) > ($P = 0.26$)). However, at the word *wine* this inference is reversed, and the model understands that $order(john, wine)$ is the case, and that $order(john, beer)$ is (probably) not the case. In addition, the word *wine* also leads the model to understand $drink(john, wine)$, even though this proposition is not explicitly part of the semantics of the sentence (John ordering wine is something that co-occurs relatively often with John drinking wine; $P = 0.25$). Finally, no significant inferences are drawn about the unrelated proposition $leave(ellen)$.

4 Discussion

We have shown how the DSS model of story comprehension developed by Frank et al. (2003) can serve as a framework for neural semantics. This framework offers neurally plausible, expressive and compositional, as well as graded and inferential meaning representations. Moreover, we have shown how these meaning representations can be derived on a word-by-word basis in a neurocomputational model of language processing (see also Frank, Haselager & van Rooij 2009). Building in this direction, we are currently employing the framework to increase the coverage of a neurocomputational model of the electrophysiology of language comprehension (Brouwer 2014; Brouwer, Hoeks & Crocker 2015), to model script-based surprisal

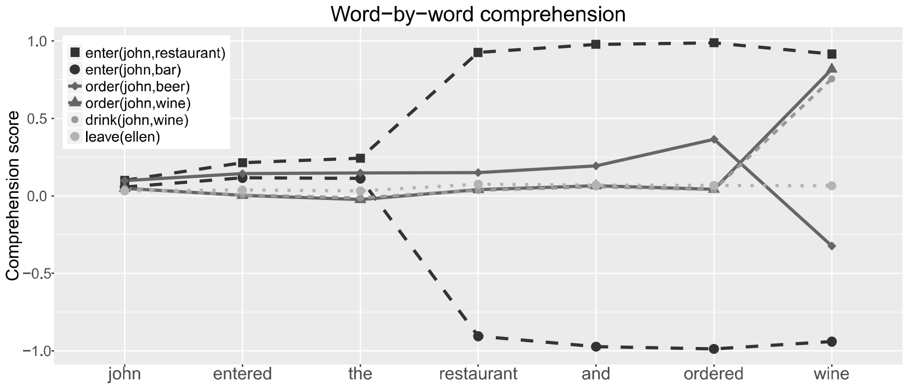

Figure 1: Word-by-word comprehension scores of selected propositions for the sentence *John entered the restaurant and ordered wine* with the semantics: $enter(john, restaurant) \land order(john, wine)$ (see text for details).

(Venhuizen, Brouwer & Crocker 2016), and to model language production (Calvillo, Brouwer & Crocker 2016).

Scalability. The meaning representations that we employed in our neurocomputational model were derived from a DSS constituted of observations sampled from a microworld. For cognitive modeling, this microworld-strategy has the advantage that it renders it feasible to make all knowledge about the world available to a cognitive model, which is preferred over omitting or selecting relevant world knowledge. The outlined framework for neural semantics does, however, not hinge upon this microworld-strategy; that is, all machinery naturally scales up to larger DSSs. Crucially, this also true if situation vectors obtain real-valued components when dimension reduction techniques are used to render very large DSSs computationally manageable (i.e., all machinery extends to the domain of fuzzy logic). Hence, it is interesting to see how larger DSSs can be automatically constructed from large corpora of annotated data, such that the framework can be employed in wide-coverage natural language processing (NLP).

Comparison to Distributional Semantics. The use of the framework in large scale NLP raises the question how its distributed representations relate to those commonly used in the field of distributional semantics. In representations derived using techniques like Latent Semantic Analysis (LSA; Landauer & Dumais 1997), the representational currency are *words*. In DSSs, by contrast, the representational currency are *propositions*. Thus, instead of defining the meaning of a *word* in terms of the *words* that is co-occurs with, in the DSS model the meaning of a *proposition* is defined in terms of the *propositions* it co-occurs with. As result, the meaning representations naturally capture inferences driven by world knowledge, and are in addition expressive and compositional in nature.

On the nature of atomic propositions. In the DSS model, as presented in this paper, the smallest meaning-discerning units are atomic propositions (e.g.,

$order(john, beer)$). However, the DSS model does not enforce these units to be propositional in nature; that is, one may think of these units as the smallest meaning-discerning atoms in any relevant domain. For instance, if one were to model an embodied cognition perspective on language, certain atoms may reflect action-related meaning, others sensory-related meaning, and again others conceptual meaning, the co-occurrence of which encodes embodied meaning. Again, all machinery extends beyond propositional atoms.

5 Conclusion

We have described a framework for neural semantics that offers neurally plausible meaning representations. These representations directly reflect experience with the world, in terms of observations over meaning-discerning atoms. Complex meaning can be directly derived from these atoms (offering expressivity and compositionality). Moreover, the resulting meaning representations inherently carry probabilistic information about themselves and their relation to each other (gradedness and inferentiality). Finally, it was shown how these representations can be constructed on a word-by-word basis in a neurocomputational model of language processing (incrementality). This framework—which unifies ideas and techniques from theoretical linguistics, computational linguistics, and psycholinguistics—paves the way for a more comprehensive neural semantics. In future work, we will investigate how the approach can be extended with regard to formal semantic properties, linguistic coverage, and as part of larger neurocomputational models.

References

Brouwer, Harm. 2014. *The electrophysiology of language comprehension: a neurocomputational model*. University of Groningen PhD thesis.

Brouwer, Harm, John C. J. Hoeks & Matthew W. Crocker. 2015. The electrophysiology of language comprehension: a neurocomputational model. In *Society for the neurobiology of language (SNL2015)*.

Calvillo, Jesús, Harm Brouwer & Matthew W. Crocker. 2016. Connectionist semantic systematicity in language production. In Anna Papafragou, Daniel Grodner, Daniel Mirman & John C. Trueswell (eds.), *Proceedings of the 38th Annual Conference of the Cognitive Science Society*, 2555–3560. Austin, TX.

Elman, Jeffrey L. 1990. Finding structure in time. *Cognitive Science* 14(2). 179–211.

Frank, Stefan L., Willem F. G. Haselager & Iris van Rooij. 2009. Connectionist semantic systematicity. *Cognition* 110(3). 358–379.

Frank, Stefan L., Mathieu Koppen, Leo G. M. Noordman & Wietske Vonk. 2003. Modeling knowledge-based inferences in story comprehension. *Cognitive Science* 27(6). 875–910.

Frege, Gottlob. 1892. Über Sinn und Bedeutung. *Zeitschrift für Philosophie und philosophische Kritik* 100. 25–50.

Golden, Richard M. & David E. Rumelhart. 1993. A parallel distributed processing model of story comprehension and recall. *Discourse Processes* 16(3). 203–237.

Golden, Richard M., David E. Rumelhart, Jospeh Strickland & Alice Ting. 1994. Markov random fields for text comprehension. *Neural networks for knowledge representation and inference*. 283–309.

Johnson-Laird, Philip N. 1983. *Mental models*. Cambridge University Press.

Landauer, Thomas K. & Susan T. Dumais. 1997. A solution to Plato's problem: the latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review* 104(2). 211–240.

Nerbonne, John A. 1992. Representing grammar, meaning and knowledge. In Susanne Preuß & Birte Schmitz (eds.), *Proceedings of the Berlin Workshop on Natural Language Processing and Knowledge Representation*. Berlin.

Partee, Barbara. 1984. Compositionality. *Varieties of formal semantics* 3. 281–311.

Rohde, Douglas L. T. 2002. *A connectionist model of sentence comprehension and production*. Carnegie Mellon University PhD thesis.

Rumelhart, David E. 1989. The architecture of mind: a connectionist approach. In Michael I. Posner (ed.), *Foundations of cognitive science*, 133–159. Cambridge, MA: The MIT Press.

Spivey, Michael J. 2008. *The continuity of mind*. Oxford University Press.

Tanenhaus, Michael K., Michael J. Spivey-Knowlton, Kathleen M. Eberhard & Julie C. Sedivy. 1995. Integration of visual and linguistic information in spoken language comprehension. *Science* 268(5217). 1632.

Venhuizen, Noortje J., Harm Brouwer & Matthew W. Crocker. 2016. When the food arrives before the menu: modeling event-driven surprisal in language comprehension. In *Pre-CUNY Workshop on Events in Language and Cognition (ELC 2016)*.