

Distributional Semantics Meets Construction Grammar. Towards a Unified Usage-Based Model of Grammar and Meaning

Giulia Rambelli, Emmanuele Chersoni, Philippe Blache, Chu-Ren Huang,
Alessandro Lenci

► To cite this version:

Giulia Rambelli, Emmanuele Chersoni, Philippe Blache, Chu-Ren Huang, Alessandro Lenci. Distributional Semantics Meets Construction Grammar. Towards a Unified Usage-Based Model of Grammar and Meaning. First International Workshop on Designing Meaning Representations (DMR 2019), Aug 2019, Florence, Italy. hal-02146537

HAL Id: hal-02146537

<https://hal.archives-ouvertes.fr/hal-02146537>

Submitted on 4 Jun 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Distributional Semantics Meets Construction Grammar. Towards a Unified Usage-Based Model of Grammar and Meaning

Giulia Rambelli

University of Pisa

giulia.rambelli@phd.unipi.it emmanuelechersoni@gmail.com

Emmanuele Chersoni

The Hong Kong Polytechnic University

Philippe Blache

Aix-Marseille University

blache@lpl-aix.fr

Chu-Ren Huang

The Hong Kong Polytechnic University

churen.huang@polyu.edu.hk

Alessandro Lenci

University of Pisa

alessandro.lenci@unipi.it

Abstract

In this paper, we propose a new type of semantic representation of Construction Grammar that combines constructions with the vector representations used in Distributional Semantics. We introduce a new framework, Distributional Construction Grammar, where grammar and meaning are systematically modeled from language use, and finally, we discuss the kind of contributions that distributional models can provide to CxG representation from a linguistic and cognitive perspective.

1 Introduction

In the last decades, usage-based models of language have captured the attention of linguistics and cognitive science (Tommasello, 2003; Bybee, 2010). The different approaches covered by this label are based on the assumptions that linguistic knowledge is embodied in mental processing and representations that are sensitive to context and statistical probabilities (Boyland, 2009), and that language structures at all levels, from morphology to syntax, emerge out of facts of actual language usage (Bybee, 2010).

A usage-based framework that turned out to be extremely influential is **Construction Grammar** (CxG) (Hoffman and Trousdale, 2013), a family of theories sharing the fundamental idea that language is a collection of form-meaning pairings called *constructions* (henceforth *Cxs*) (Fillmore, 1988; Goldberg, 2006). Cxs differ for their degree of schematicity, ranging from morphemes (e.g., *pre-*, *-ing*), to complex words (e.g., *daredevil*) to filled or partially-filled idioms (e.g., *give the devil his dues* or *Jog (someones) memory*) to more abstract patterns like the ditransitive Cxs [Subj V

Obj1 Obj2]). It is worth stressing that, even if the concept of construction is based on the idea that linguistic properties actually emerge from language *use*, CxG theories have typically preferred to model the semantic content of constructions in terms of hand-made, formal representations like those of Frame Semantics (Baker et al., 1998). This leaves open the issue of how semantic representations can be learned from empirical evidence, and how do they relate to the usage-based nature of Cxs. In fact, for a usage-based model of grammar based on a strong syntax-semantics parallelism, it would be desirable to be grounded on a framework allowing to learn the semantic content of Cxs from language use.

In this perspective, a promising solution for representing constructional semantics is given by an approach to meaning representations that has gained a rising interest in both computational linguistics and cognitive science, namely **Distributional Semantics** (henceforth DS). DS is a usage-based model of word meaning, based on the well-established assumption that the statistical distribution of linguistic items in context plays a key role in characterizing their semantic behaviour (*Distributional Hypothesis* (Harris, 1954)). More precisely, Distributional Semantic Models (DSMs) represent the lexicon in terms of vector spaces, where a lexical target is described in terms of a vector (also known as *embedding*) built by identifying in a corpus its syntactic and lexical contexts (Lenci, 2018). Lately, neural models to learn distributional vectors have gained massive popularity: these algorithms build low-dimensional vector representations by learning to optimally predict the contexts of the target words (Mikolov et al., 2013). On the negative side, DS lacks a clear

connection with usage-based theoretical frameworks. To the best of our knowledge, existing attempts of linking DS with models of grammar have rather targeted formal theories like Montague Grammar and Categorical Grammar (Baroni et al., 2014; Grefenstette and Sadrzadeh, 2015).

To sum up, both CxG and DS share the assumption that linguistic structures naturally emerge from language usage, and that a representation of both form and meaning of any linguistic item can be modeled through its distributional statistics, and more generally, with the quantitative information derived from corpus data. However, these two models still live in parallel worlds. On the one hand, CxG is a model of grammar in search for a consistent usage-based model of meaning, and, conversely, DS is a computational framework to build semantic representations in search for an empirically adequate theory of grammar.

As we illustrate in Section 2, occasional encounters between DS and CxG have already happened, but we believe that new fruitful advances could come from the exploitation of the mutual synergies between CxG and DS, and by letting these two worlds finally meet and interact in a more systematic way. Following this direction of research, we introduce a new representation framework called **Distributional Construction Grammar**, which aims at bringing together these two theoretical paradigms. Our goal is to integrate distributional information into constructions by completing their semantic structures with distributional vectors extracted from large textual corpora, as samples of language usage.

These pages are structured as follows: after reviewing existing literature on CxG and related computational studies, in Section 3 we outline the key characteristics of our theoretical proposal, while Section 4 provides a general discussion about what contributions DSMs can provide to CxG representation from a linguistic and cognitive perspective. Although this is essentially a theoretical contribution, we outline ongoing work focusing on its computational implementation and empirical validation. We conclude by reporting future perspectives of research.

2 Related Work

Despite the popularity of the constructional approach in corpus linguistics (Gries and Stefanowitsch, 2004), computational semantics research

has never formulated a systematic proposal for deriving representations of constructional meaning from corpus data. Previous literature has mostly focused either on the automatic identification of constructions on the basis of their formal features, or on modeling the meaning of a specific CxG.

For the former approach, we should mention the works of Dunn (2017, 2019) that aim at automatically inducing a set of grammatical units (Cxs) from a large corpus. On the one hand, Dunn's contributions provide a method for extracting Cxs from corpora, but on the other hand they are mainly concerned with the formal side of the constructions, and especially with the problem of how syntactic constraints are learned. Some sort of semantic representation is included, in the form of semantic cluster of word embeddings to which the word forms appearing in the constructions are assigned. However, these works do not present any evaluation of the construction representations in terms of semantic tasks.

Another line of research has focused in using constructions for building computational models of language acquisition. Alishahi and Stevenson (2008) propose a model for the representation, acquisition and use of verb argument structure by formulating constructions as probabilistic associations between syntactic and semantic properties of verbs and their arguments. This probabilistic association emerges over time through a Bayesian acquisition process in which similar verb usages are detected and grouped together to form general constructions, based on their syntactic and semantic properties. Despite the success of this model, the semantic representation of argument structure is still symbolic and each semantic category of input constructions are manually compiled, in contrast with the usage-based nature of constructions.

Other studies used DSMs to model constructional meaning, by focusing on a specific type of Cx rather than on the entire grammar. For example, Levshina and Heylen (2014) build a vector space to study Dutch causative constructions with *doen* ('do') and *laten* ('let'). They compute several vector spaces with different context types, both for the nouns that fill the Causer and Causee slot and for the verbs that fill the Effected Predicate slot. Then, they cluster these nouns and verbs at different levels of granularity and test which classification better predicts the use of *laten* and *doen*.

A recent trend in diachronic linguistics investi-

gates linguistic change as a sequence of gradual changes in distributional patterns of usage (Bybee, 2010). For instance, Perek (2016) investigates the productivity of the *V the hell out of NP* construction (e.g., *You scared the hell out of me*) from 1930 to 2009. On one side, he clusters the vectors of verbs occurring in this construction to pin point the preferred semantic domains of the Cx in its diachronic evolution. Secondly, he computes the density of the semantic space of the construction around a given word in a certain period to be predictive of that word joining the construction in the subsequent period. A similar approach is applied to study changes in the productivity of the *Way*-construction over the period 1830-2009 (Perek, 2018). Perek’s analysis also proves that distributional similarity and neighbourhood density in the vector space can be predictive of the usage of a construction with a new lexical item. Other works have followed this approach, demonstrating the validity of DSMs to model the semantic change of constructions in diachrony. Amato and Lenci (2017) examine the Italian Gerundival Periphrases *stare* (to stay) *andare* (to go), *venire* (to come) followed by a gerund. As in previous works, they use DSMs to i) identify similarities and differences among Cxs clustering the vectors of verbs occurring in each Cx, and ii) investigate the changes undergone by the semantic space of the verbs occurring in the Cxs throughout a very long period (from 1550 to 2009).

Lebani and Lenci (2017) present an unsupervised distributional semantic representation of argument constructions. Following the assumption that constructional meanings for argument Cxs arise from the meaning of high frequency verbs that co-occur with them (Goldberg, 1999; Casenhiser and Goldberg, 2005; Barak and Goldberg, 2017), they compute distributional vectors for CxS as the centroids of the vectors of their typical verbs, and use them to model the psycholinguistic data about construction priming in Johnson and Goldberg (2013). This representation of construction meaning has also been applied to study valency coercion by Busso et al. (2018).

Following a parallel research line on probing tasks for distributed vectors, Kann et al. (2019) investigate whether word and sentence embeddings encode the grammatical distinctions necessary for inferring the idiosyncratic frame-selectional properties of verbs. Their findings show that, at least

for some alternations, verb embeddings encode sufficient information for distinguishing between acceptable and unacceptable combinations.

3 Distributional CxG Framework

We introduce a new framework aimed at integrating the computational representation derived from distributional methods into the explicit formalization of Construction Grammars, called **Distributional Construction Grammar** (DisCxG).

DisCxG is based on three components:

- **Constructions:** stored pairings of form and function, including morphemes, words, idioms, partially lexically filled and fully general linguistic patterns (Goldberg, 2003);
- **Frames:** schematic semantic knowledge describing scenes and situations in terms of their semantic roles;
- **Events:** semantic information concerning particular event instances with their specific participants. The introduction of this component, which is a novelty with respect to traditional CxG frameworks, has been inspired by cognitive models such as the *Generalized Event Knowledge* (McRae and Matsuki, 2009) and the *Words-as-Cues* hypothesis (Elman, 2014).

The peculiarity of DisCxG is that we distinguish two layers of semantic representation, referring to two different and yet complementary aspects of semantic knowledge. Specifically, frames define a prototypical semantic representation based on the different semantic roles (the *frame elements*) defining argument structures, while events provide a specialization of the frame by taking into account information about specific participants and relations between them. Crucially, we assume that *both these layers have a DS representation in terms of distributional vectors learned from corpus co-occurrences*.

Following the central tenet of CxGs, according to which linguistic information is encoded in similar way for lexical items as well as for more abstract Cxs (e.g., *covariational-conditional* Cx, *ditransitive* Cx etc.), the three components of DisCxG are modeled using the same type of formal representation with recursive feature-structures, which is inspired by Sign-Based Construction Grammar (SBCG) (Sag, 2012; Michaelis, 2013).

3.1 Constructions

In DisCxG, a construction is represented by form and semantic features. The following list presents the set of main features of Cxs adapting the formalization in SBCG:

- The FORM feature contains the basic formal characteristics of constructions. It includes the (i) PHONological/SURFACE form, (ii) the (morpho)syntactic features (SYN), i.e. part-of-speech (TYPE), CASE (nominal, accusative), the set of elements subcategorized (VAL), and (iii) PROPERTIES representing explicitly the syntactic relations among the elements of the Cx.
- The ARGument-STructure implements the interface between syntactic and semantic roles. The arguments are in order of their accessibility hierarchy (subj < d-obj < obl...), encoding the syntactic role. Each argument specifies the case, related to the grammatical function, and links to the thematic role.¹
- The SEMantic feature specifies the properties of Cx's meaning (Section 3.2).

Unlike SGBG or other CxG theories, we include inside FORM a new feature called PROPERTIES, borrowed from *Property Grammars* (Blache, 2005). Properties encode syntactic information about the components of a Cx, and they play an important role in its recognition. However, the discussion of this linguistic aspect is not presented here, as the focus of this paper is on the semantic side of constructions.²

As said above, a Cx can describe linguistic objects of various levels of complexity and schematicity: words, phrases, fully lexicalized idiomatic patterns, partially lexicalized schemas, etc. Thus, the attribute-value matrix can be applied to lexical entries, as the verb *read* in Figure 1, as well as to abstract constructions that do not involve lexical material. Figure 2 depicts the *ditransitive* Cx. The semantic particularity of this construction is that whatever the lexicalization of the verb, this

¹SGCG distinguishes between valence and argument structure: the ARG-ST encodes overt and covert arguments, including extracted (non-local) and unexpressed elements, while VAL in the form description represents only realized elements. When no covert arguments occur, these features are identical.

²For more details on the Property Grammar framework, see Blache (2016).

construction always involve a possession interpretation (more precisely the transfer of something to somebody), represented in the TRANSFER frame.

Differently from standard SBCG formalization of Cxs, we add the distributional feature DS-VECTOR into the semantic layer in order to integrate lexical distributional representations. The semantic structure of a lexical item can be associated with its distributional vector (e.g., the embedding of *read*), but we can also include a distributional representation of abstract syntactic constructions following the approach of Lebani and Lenci (2017) we have illustrated in Section 2.

3.2 Frames

A frame is a schematic representation of an event or scenario together with the participating actors/objects/locations and their (semantic) role (Fillmore, 1982). For instance, the sentences

1. (a) *Mary bought a car from John (for 5000\$).*
(b) *John sold a car to Mary (for 5000\$).*

activate the same COMMERCIAL_TRANSACTION frame, consisting of a SELLER (*John*), a BUYER (*Mary*), a GOOD which is sold (*car*), and the MONEY used in the transaction (*5000\$*).

Semantic frames are the standard meaning representation in CxG, which represent them as symbolic structures. The source of this information is typically FrameNet (Ruppenhofer et al., 2016), a lexical database of English containing more than 1,200 semantic frames linked to more than 200,000 manually annotated sentences. The not negligible problem of FrameNet is that entries must be created by expert lexicographers. This has led to a widely recognized coverage problem in its lexical units (Baker, 2012).

In DisCxG, semantic frames are still represented as structures, but the value of semantic roles consists of distributional vectors. As for the COMMERCIAL_TRANSACTION frame in Figure 3, each frame element has associated a specific embedding. It is worth noting that in this first version of the DisCxG model, frame representations are still based on predefined lists of semantic roles, as defined in FrameNet (e.g., BUYER, SELLER, etc.). However, some works have recently attempted to automatically infer frames (and their roles) from distributional information³. Woodsend and Lap-

³Lately, SemEval 2019 proposed a task on unsupervised

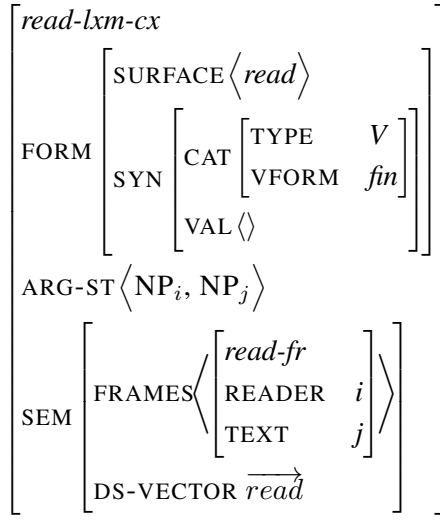


Figure 1: Description of *read* verb

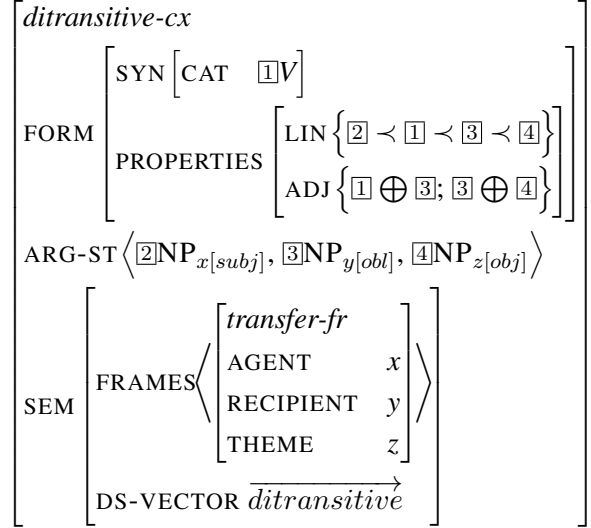


Figure 2: Description of *ditransitive Cx*

ata (2015) use distributional representations to induce embeddings for predicates and their arguments. Ustalov et al. (2018) propose a different methodology for unsupervised semantic frame induction. They build embeddings as the concatenations of subject-verb-object triples and identify frames as clustered triples. Of course, a limit of this approach is that it only uses subject and object arguments, while frames are generally associated with a wider variety of roles. Lebani and Lenci (2018) instead provide a distributional representation of verb-specific semantic roles as clusters of features automatically induced from corpora.

In this paper, we assume that at least some aspects of semantic roles can be derived from combining (e.g., with summation) the distributional vectors of their most prototypical fillers, following an approach widely explored in DS (Baroni and Lenci, 2010; Erk et al., 2010; Sayeed et al., 2016; Santus et al., 2017). For instance, the *buyer* role in the COMMERCIAL_TRANSACTION frame can be taken as a vector encoding the properties of the typical nouns filling this role. We are aware that this solution is just an approximation of the content of frames elements. How to satisfactorily characterize semantic frames and roles using DS is in fact still an open research question.

3.3 Events

Neurocognitive research has brought extensive evidence that stored world knowledge plays a key role in online language production and compre-

lexical semantic frame induction (<http://alt.qcri.org/semeval2019/index.php?id=tasks>)

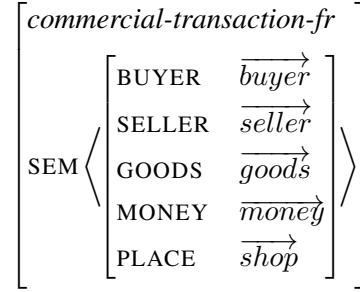


Figure 3: The COMMERCIAL_TRANSACTION frame containing the distributional representation of the semantic roles

hension. An important aspect of such knowledge consists of the events and situations that we experience under different modalities, including the linguistic input. McRae and Matsuki (2009) call it Generalized Event Knowledge (GEK), because it contains information about prototypical event structures. Language comprehension has been characterized as a largely predictive process (Kuperberg and Jaeger, 2015). Predictions are memory-based, and experiences about events and their participants are used to generate expectations about the upcoming linguistic input, thereby minimizing the processing effort (Elman, 2014; McRae and Matsuki, 2009). For instance, argument combinations that are more ‘coherent’ with the event scenarios activated by the previous words are read faster in self-paced reading tasks and elicited smaller N400 amplitudes in ERP experiments (Paczynski and Kuperberg, 2012).

In DisCxG, events have a crucial role: they

bridge the gap between the concrete instantiation of a Cx in context and its conceptualized meaning (conveyed from frames). For example, let's consider the verb *read*. We know that this verb subcategorizes for two noun phrases (form) and involves a generic READING frame in which there is someone who reads (READER) and something that is read (TEXT). This frame only provides an abstract, context-independent representation of the verb meaning, and the two roles can be generally defined as clusters of properties derived from singular subjects and objects of *read*. However, the semantic representation comprehenders build during sentence processing is influenced by the specific fillers that instantiate the frame elements. If the input is *A student reads...*, the fact that the word *student* appears as the subject of the verb activates a specific scenario, together with a series of expectations about the prototypicality of other lexical items. Consequently, the object of the previous sentence is more likely to be *book* rather than *magazine* (Chersoni et al., 2019). Accordingly, in DisCxG events are considered as functions that specialize the semantic meaning encoded in frames. The word *student* specializes the READING frame into a specific event, triggering expectations about the most likely participants of the other roles: the READER is encoded as a lexical unit vector, and the distributional restriction applied to the TEXT is represented by a subset of possible objects ordered by their degree of typicality in the event. Figure 4 gives a simple example of the specialization brought out by event knowledge.

$$\left[\begin{array}{l} \text{student-read-event} \\ \text{SPECIALIZE reading-fr} \\ \text{SEM} \left\langle \left[\begin{array}{ll} \text{READER} & \overrightarrow{\text{student}} \\ \text{TEXT} & \{\overrightarrow{\text{book}}, \overrightarrow{\text{paper}}..\} \end{array} \right] \right\rangle \end{array} \right]$$

Figure 4: *Student-read* event as the specialization of READING frame

In a similar way, events can instantiate an abstract construction dynamically, according to the context. The different lexicalization of the AGENT and the RECIPIENT in the ditransitive construction causes a different selection of the THEME. For example, the fact that the sentence fragment *The teacher gives students ...* could be completed as in (2) expresses a distributional restriction that can be encoded as an event capturing the co-occurrences

$$\left[\begin{array}{l} \text{SPECIALIZE ditransitive-cx} \\ \text{ARG-ST} \langle \text{NP}_x, \text{NP}_y, \text{NP}_z \rangle \\ \text{SEM} \left[\begin{array}{l} \text{FRAMES} \left\langle \left[\begin{array}{l} \text{transfer-fr} \\ \text{AGENT } x_{[\overrightarrow{\text{teacher}}]} \\ \text{RECIPIENT } y_{[\overrightarrow{\text{students}}]} \\ \text{THEME } z_{\langle \overrightarrow{\text{exercise}}, \overrightarrow{\text{homework}}.. \rangle} \end{array} \right] \right\rangle \end{array} \right] \end{array} \right]$$

Figure 5: *Ditransitive* event specialization

teacher/student/exercises (Figure 5).

2. The teacher gives students ... → The teacher gives students exercises

Any lexical item activate a portion of event knowledge (Elman, 2014): in fact, if verbs evoke events, nouns evoke entities that participate into events. Thus, events and entities are themselves interlinked: there is not a specific feature EVENT in the description of the lexical entry *teacher*, but events are activated by the lexical entry, generating a network of expectations about upcoming words in the sentence (McRae and Matsuki, 2009).

Given this assumption, Chersoni et al. (2019) represent event knowledge in terms of a Distributional Event Graph (DEG) automatically built from parsed corpora. In this graph, nodes are embeddings and edges are labeled with syntactic relations and weighted using statistic association measures (Figure 6). Each event is a a path in DEG. Thus, given a lexical cue w , it is possible to identify the events it activates (together with the strength of its activation, defined as a function of the graph weights) and generate expectations about incoming inputs on both paradigmatic and syntagmatic axes. With this graph-based approach, Chersoni et al. (2019) model sentence comprehension as the dynamic and incremental creation of a semantic representation integrated into a semantically coherent structure contributing to the sentence interpretation.

We propose to include in our framework the information encoded in DEG. Each lexical entry contains a pointer to its corresponding node in the graph. Therefore, the *frame specialization* we have described above corresponds to an event encoded with a specific path in the DEG. Event information represents a way to unify the schematic descriptions contained in the grammar with the

world knowledge and contextual information progressively activated by lexical items and integrated during language processing.

4 Some Other Arguments in Favor of a Distributional CxG

As we said in Section 2, few works have tried to use distributional semantic representations of constructions and existing studies have focused more on applying DS to a particular construction type, instead of providing a general model to represent the semantic content of Cxs. We argue that DSMs could give an important contribution in designing representations of constructional meaning. In what follows, we briefly discuss some specific issues related to Construction Grammars that could be addressed by combining them with Distributional Semantics.

Measuring similarity among constructions and frames The dominant approaches like frame semantics and traditional CxGs tend to represent entities and their relations in a formal (hand-made) way. A potential limitation of these methods is that it is hard to assess the similarity between frames or constructions, while one advantage of distributional vectors is that one can easily compute the degree of similarity between linguistic items represented in a vector space. For example, [Busso et al. \(2018\)](#) built a semantic space for several Italian argument constructions and then computed the similarity of their vectors, observing that some Cxs have similar distributional behaviour like Caused-Motion and Dative.

As for frames, there has been some work on using distributional similarity between vectors for their unsupervised induction ([Ustalov et al., 2018](#)), for comparing frames across languages ([Sikos and Padó, 2018](#)), and even for the automatic identification of the semantic relations holding between them ([Botschen et al., 2017](#)).

Identifying idiomatic meaning Many studies in theoretical, descriptive and experimental linguistics have recently questioned the *fregean* principle of compositionality, which assumes that the meaning of an expression is the result of the incremental composition of its sub-constituents. There is a large number of linguistic phenomena whose meaning is accessed directly from the whole linguistic structure: this is typically the case with idioms or multi-word expressions, where the figura-

tive meaning cannot be decomposed. In computational semantics, a large literature has been aiming at modeling idiomaticity using DSMs. [Senaldi et al. \(2016\)](#) carried out an idiom type identification task representing Italian V-NP and V-PP Cxs as vectors. They observed that the vectors of VN and AN idioms are less similar to the vectors of lexical variants of these expressions with respect to the vectors of compositional constructions. ([Cordeiro et al., 2019](#)) realized a framework for predict compound compositionality using DSMs, evaluating to what extent they capture idiomaticity compared to human judgments. Results revealed a high agreement between the models and human predictions, suggesting that they are able to incorporate information about idiomaticity.

In future works, it would be interesting to see if DSMs-based approaches can be used in combination with methods for the identification of the formal features of constructions ([Dunn, 2017, 2019](#)), in order to tackle the task of compositionality prediction simultaneously with syntactic and semantic features.

Modeling sentence comprehension A trend in computational semantics regards the application of DSMs to sentence processing ([Mitchell et al., 2010](#); [Lenci, 2011](#); [Sayeed et al., 2015](#); [Johns and Jones, 2015, i.a.](#)).

[Chersoni et al. \(2016, 2017\)](#) propose a Distributional Model of sentence comprehension inspired by the general principles of the Memory, Unification and Control framework ([Hagoort, 2013, 2015](#)). The memory component includes events in GEK with feature structures containing information directly extracted from parsed sentences in corpora: attributes are syntactic dependencies, while values are distributional vectors of dependent lexemes. Then, they model semantic composition as an event construction and update function F , whose aim is to build a coherent semantic representation by integrating the GEK cued by the linguistic elements.

The framework has been applied to the logical metonymy phenomenon (e.g. *The student begins the book*), using the semantic complexity function to model the processing costs of metonymic sentences, which was shown to be higher compared to non-coercion sentences ([McElree et al., 2001](#); [Traxler et al., 2002](#)). Evaluation against psycholinguistic datasets proves the linguistic and

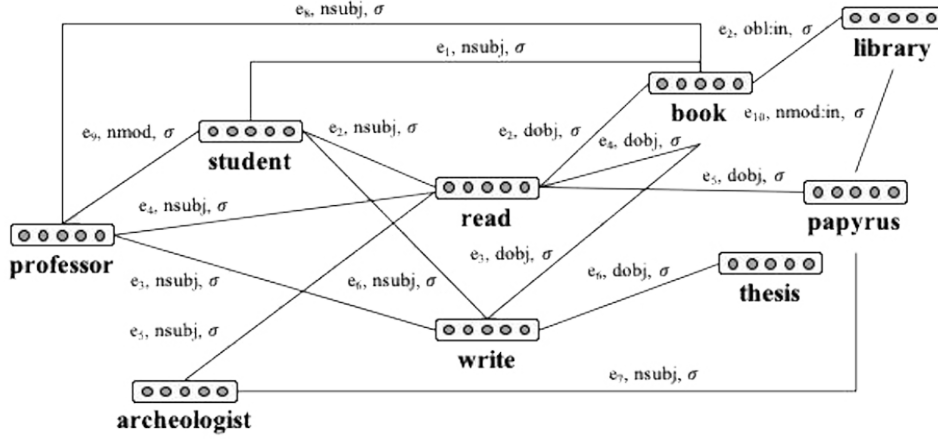


Figure 6: An extract of DEG showing several instances of events (Chersoni et al., 2019)

psycholinguistic validity of using embeddings to represent events and including them in incremental model of sentence comprehension.

Evaluations based on experimental evidence

DSMs have proved to be very useful in modeling human performance in psycholinguistic tasks (Mandera et al., 2017). This is an important finding, since it allows to test the predictions of Construction Grammar theories against data derived from behavioral experiments.

To cite an example from the DS literature, the models proposed by Lebani and Lenci (2017) replicated the priming effect of the lexical decision task by Johnson and Goldberg (2013), where the participants were asked to judge whether a given verb was a real word or not, after being exposed to an argument structure construction in the form of a Jabberwocky sentence. The authors of the study created distributional representations of constructions as combinations of the vectors of their typical verbs, and measured their cosine similarity with the verbs of the original experiment, showing that their model can accurately reproduce the results reported by Johnson and Goldberg (2013).

5 Conclusion

In this paper, we investigated the potential contribution of DSMs to the semantic representation of constructions, and we presented a theoretical proposal bringing together vector spaces and constructions into a unique framework. It is worth highlighting our main contributions:

- We built a unified representation of grammar and meaning based on the assumption

that language structure and properties emerge from language use.

- We integrated information about events to build a semantic representation of an input as an incremental and predictive process.

Converging different layers of meaning representation into a unique framework is not a trivial problem, and in our future work we will need to find optimal ways to balance these two components: semantic vectors derived from corpus data on the one hand, and a possibly accurate formalization of the internal structure of the constructions on the other hand. In this contribution, we hoped to show that merging the two frameworks would be worth the efforts, as they share many theoretical assumptions and complement themselves on the basis of their respective strengths.

Our future goal is the automatic building and inclusion of a distributional representation of frames and event in DisCxG; our aim is to exploit the final formalism to build for the first time a *Distributional Construction Treebank*. Moreover, we are planning to apply this framework in a predictive model of language comprehension, defining how a Cx is activated by the combination of syntactic, lexical and distributional cues occurring in DisCxG. We believe this framework could be a starting point for applications in NLP such as Knowledge representation and reasoning, Natural Language Understanding and Generation, but also a potential term of comparison for psycholinguistic models of human language comprehension.

References

- Afra Alishahi and Suzanne Stevenson. 2008. A Computational Model of Early Argument Structure Acquisition. *Cognitive Science*, 32(5):789–834.
- Irene Amato and Alessandro Lenci. 2017. Story of a Construction: Statistical and Distributional Analysis of the Development of the Italian Gerundival Construction. In G. Marotta and F. S. Lievers, editors, *Strutture Linguistiche e Dati Empirici in Diacronia e Sincronia*, pages 135–158. Pisa University Press.
- Collin F Baker. 2012. FrameNet, Current Collaborations and Future Goals. *Language, Resources and Evaluation*, 46(2):269–286.
- Collin F. Baker, Charles J. Fillmore, and John B. Lowe. 1998. The Berkeley FrameNet project. In *Proceedings of COLING-ACL*, pages 86–90, Montreal, Canada. ACL.
- Libby Barak and Adele Goldberg. 2017. Modeling the Partial Productivity of Constructions. In *Proceedings of the 2017 AAAI Spring Symposium Series on Computational Construction Grammar and Natural Language Understanding*, pages 131–138.
- Marco Baroni, Raffaella Bernardi, and Roberto Zamparelli. 2014. Frege in Space: A Program of Compositional Distributional Semantics. *Linguistic Issues in Language Technology*, 9(6):5–110.
- Marco Baroni and Alessandro Lenci. 2010. Distributional Memory: A General Framework for Corpus-Based Semantics. *Computational Linguistics*, 36(4):673–721.
- Philippe Blache. 2005. Property Grammars: A Fully Constraint-Based Theory. In Skadhauge P. R. Christiansen, H. and J. Villadsen, editors, *Constraint Solving and Language Processing. CSLP 2004*, volume 3438 of *LNAI*, pages 1–16. Springer, Berlin, Heidelberg.
- Philippe Blache. 2016. Representing Syntax by Means of Properties: A Formal Framework for Descriptive Approaches. *Journal of Language Modelling*, 4(2):183–224.
- Teresa Botschen, Hatem Mousselly Sergieh, and Iryna Gurevych. 2017. Prediction of Frame-to-Frame Relations in the FrameNet Hierarchy with Frame Embeddings. In *Proceedings of the ACL Workshop on Representation Learning for NLP*, pages 146–156.
- Joyce Tang Boyland. 2009. Usage-Based Models of Language. In D. Eddington, editor, *Experimental and Quantitative Linguistics*, pages 351–419. Lincom, Munich.
- Lucia Busso, Ludovica Pannitto, and Alessandro Lenci. 2018. Modelling Italian Construction Flexibility with Distributional Semantics: Are Constructions Enough? In *Proceedings of the Fifth Italian Conference on Computational Linguistics (CLiC-it 2018)*, pages 68–74.
- J. Bybee. 2010. *Language, Usage and Cognition*. Cambridge University Press.
- Devin Casenhiser and Adele E Goldberg. 2005. Fast Mapping Between a Phrasal Form and Meaning. *Developmental science*, 8(6):500–508.
- Emmanuele Chersoni, Philippe Blache, and Alessandro Lenci. 2016. Towards a Distributional Model of Semantic Complexity. In *Proceedings of the COLING Workshop on Computational Linguistics for Linguistic Complexity (CL4LC)*, pages 12–22.
- Emmanuele Chersoni, Alessandro Lenci, and Philippe Blache. 2017. Logical Metonymy in a Distributional Model of Sentence Comprehension. In *Proceedings of *SEM*, pages 168–177.
- Emmanuele Chersoni, Enrico Santus, Ludovica Pannitto, Alessandro Lenci, Philippe Blache, and Churen Huang. 2019. A Structured Distributional Model of Sentence Meaning and Processing. *Journal of Natural Language Engineering*. To Appear.
- Silvio Cordeiro, Aline Villavicencio, Marco Idiart, and Carlos Ramisch. 2019. Unsupervised Compositionality Prediction of Nominal Compounds. *Computational Linguistics*, 45:1–57.
- Jonathan Dunn. 2017. Computational Learning of Construction Grammars. *Language and Cognition*, 9(2):254–292.
- Jonathan Dunn. 2019. Frequency vs. Association for Constraint Selection in Usage-Based Construction Grammar. In *Proceedings of the NAACL Workshop on Cognitive Modeling and Computational Linguistics*.
- Jeffrey L. Elman. 2014. Systematicity in the Lexicon: On Having Your Cake and Eating It Too. In Paco Calvo and John Symons, editors, *The Architecture of Cognition*, pages 115–145. The MIT Press.
- Katrin Erk, Sebastian Padó, and Ulrike Padó. 2010. A Flexible, Corpus-Driven Model of Regular and Inverse Selectional Preferences. *Computational Linguistics*, 36(4):723–763.
- Charles J. Fillmore. 1982. Frame Semantics. *Linguistics in the Morning Calm*, pages 111–37.
- Charles J Fillmore. 1988. The Mechanisms of Construction Grammar. *Annual Meeting of the Berkeley Linguistics Society*, 14:35–55.
- Adele E. Goldberg. 1999. The Emergence of the Semantics of Argument Structure Constructions. In B. MacWhinney, editor, *The Emergence of Language*, pages 197–212. Lawrence Erlbaum Publications, Hillsdale, NJ.
- Adele E. Goldberg. 2003. Constructions: A New Theoretical Approach to Language. *Trends in Cognitive Sciences*, 7(5):219–224.

- Adele E Goldberg. 2006. *Constructions at Work: The Nature of Generalization in Language*. Oxford University Press on Demand.
- Edward Grefenstette and Mehrnoosh Sadrzadeh. 2015. Concrete Models and Empirical Evaluations for the Categorical Compositional Distributional Model of Meaning. *Computational Linguistics*, 41(1):71–118.
- Stefan Th. Gries and Anatol Stefanowitsch. 2004. Extending Collostructional Analysis. A Corpus-Based Perspective on 'Alternations'. *International Journal of Corpus Linguistics*, 9(1):97–129.
- Peter Hagoort. 2013. MUC (Memory, Unification, Control) and Beyond. *Frontiers in Psychology*, 4:1–13.
- Peter Hagoort. 2015. MUC (Memory, Unification, Control): A Model on the Neurobiology of Language Beyond Single Word Processing. In *Neurobiology of language*, pages 339–347. Elsevier.
- Zellig S. Harris. 1954. Distributional Structure. *Word*, 10:146–62.
- Thomas Hoffman and Graeme Trousdale, editors. 2013. *The Oxford Handbook of Construction Grammar*. Oxford University Press, Oxford.
- Brendan T Johns and Michael N Jones. 2015. Generating Structure from Experience: A Retrieval-Based Model of Language Processing. *Canadian Journal of Experimental Psychology*, 69(3):233–251.
- Matt A. Johnson and Adele E. Goldberg. 2013. Evidence for Automatic Accessing of Constructional Meaning: Jabberwocky Sentences Prime Associated Verbs. *Language and Cognitive Processes*, 28(10):1439–1452.
- Katharina Kann, Alex Warstadt, Adina Williams, and Samuel R Bowman. 2019. Verb Argument Structure Alternations in Word and Sentence Embeddings. In *Proceedings of the Society for Computation in Linguistics (SCiL) 2019*, pages 287–297.
- Gina R. Kuperberg and T. Florian Jaeger. 2015. What Do We Mean by Prediction in Language Comprehension? *Language Cognition & Neuroscience*, 3798.
- Gianluca E Lebani and Alessandro Lenci. 2017. Modelling the Meaning of Argument Constructions with Distributional Semantics. In *Proceedings of the AAAI 2017 Spring Symposium on Computational Construction Grammar and Natural Language Understanding*, pages 197–204.
- Gianluca E. Lebani and Alessandro Lenci. 2018. A Distributional Model of Verb-Specific Semantic Roles Inferences. In Thierry Poibeau and Aline Villavicencio, editors, *Language, Cognition, and Computational Models*, pages 118–158. Cambridge University Press, Cambridge.
- Alessandro Lenci. 2011. Composing and Updating Verb Argument Expectations: A Distributional Semantic Model. In *Proceedings of the ACL Workshop on Cognitive Modeling and Computational Linguistics*, pages 58–66.
- Alessandro Lenci. 2018. Distributional Models of Word Meaning. *Annual Review of Linguistics*, 4(1):151–171.
- Natalia Levshina and Kris Heylen. 2014. A Radically Data-Driven Construction Grammar: Experiments with Dutch Causative Constructions. In R. Boogaart, T. Colleman, and Rutten G., editors, *Extending the Scope of Construction Grammar*, pages 17–46. Mouton de Gruyter, Berlin.
- Paweł Mandera, Emmanuel Keuleers, and Marc Brysbaert. 2017. Explaining Human Performance in Psycholinguistic Tasks with Models of Semantic Similarity Based on Prediction and Counting: A Review and Empirical Validation. *Journal of Memory and Language*, 92:57–78.
- Brian McElree, Matthew J Traxler, Martin J Pickering, Rachel E Seely, and Ray Jackendoff. 2001. Reading Time Evidence for Enriched Composition. *Cognition*, 78(1):B17–B25.
- Ken McRae and Kazunaga Matsuki. 2009. People Use their Knowledge of Common Events to Understand Language, and Do So as Quickly as Possible. *Language and Linguistics Compass*, 3(6):1417–1429.
- Laura A Michaelis. 2013. Sign-Based Construction Grammar. In *The Oxford Handbook of Construction Grammar*, pages 133–152, Oxford. Oxford University Press.
- Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. Distributed Representations of Words and Phrases and their Compositionality. In *Advances in Neural Information Processing Systems (NIPS 2013)*, pages 3111–3119.
- Jeff Mitchell, Mirella Lapata, Vera Demberg, and Frank Keller. 2010. Syntactic and Semantic Factors in Processing Difficulty: An Integrated Measure. In *Proceedings of ACL*, pages 196–206, Uppsala, Sweden. ACL.
- Martin Paczynski and Gina R. Kuperberg. 2012. Multiple Influences of Semantic Memory on Sentence Processing: Distinct Effects of Semantic Relatedness on Violations of Real-World Event/State Knowledge and Animacy Selection Restrictions. *J Memory and Language*, 67(4).
- Florent Perek. 2016. Using Distributional Semantics to Study Syntactic Productivity in Diachrony: A Case Study. *Linguistics*, 54(1):149–188.
- Florent Perek. 2018. Recent Change in the Productivity and Schematicity of the Way-Construction: A Distributional Semantic Analysis. *Corpus Linguistics and Linguistic Theory*, 14(1):65–97.

- Josef Ruppenhofer, Michael Ellsworth, Myriam Schwarzer-Petruck, Christopher R Johnson, and Jan Scheffczyk. 2016. *FrameNet II: Extended Theory and Practice*.
- Ivan A Sag. 2012. Sign-Based Construction Grammar: An Informal Synopsis. In *Sign-based construction grammar*, volume 193, pages 69–202. CSLI: CSLI Publications.
- Enrico Santus, Emmanuele Chersoni, Alessandro Lenci, and Philippe Blache. 2017. Measuring Thematic Fit with Distributional Feature Overlap. In *Proceedings of EMNLP*, pages 648–658.
- Asad Sayeed, Stefan Fischer, and Vera Demberg. 2015. Vector-Space Calculation of Semantic Surprisal for Predicting Word Pronunciation Duration. In *Proceedings of ACL-IJCNLP*, pages 763–773, Beijing, China. ACL.
- Asad Sayeed, Clayton Greenberg, and Vera Demberg. 2016. Thematic Fit Evaluation: An Aspect of Selectional Preferences. In *Proceedings of the ACL Workshop on Evaluating Vector-Space Representations for NLP*, pages 99–105.
- Marco Silvio Giuseppe Senaldi, Gianluca E Lebani, and Alessandro Lenci. 2016. Lexical Variability and Compositionality: Investigating Idiomaticity with Distributional Semantic Models. In *Proceedings of the ACL Workshop on Multiword Expressions*, pages 21–31.
- Jennifer Sikos and Sebastian Padó. 2018. Using Embeddings to Compare FrameNet Frames Across Languages. In *Proceedings of the COLING Workshop on Linguistic Resources for Natural Language Processing*, pages 91–101.
- Michael Tommasello. 2003. *Constructing a Language: A Usage-Based Theory of Language Acquisition*. Harvard University Press, Cambridge, MA.
- Matthew J Traxler, Martin J Pickering, and Brian McElree. 2002. Coercion in Sentence Processing: Evidence from Eye-Movements and Self-Paced Reading. *Journal of Memory and Language*, 47(4):530–547.
- Dmitry Ustalov, Alexander Panchenko, Andrei Kutuzov, Chris Biemann, and Simone Paolo Ponzetto. 2018. Unsupervised Semantic Frame Induction Using Triclustering. In *Proceedings of ACL*, pages 55–62.
- Kristian Woodsend and Mirella Lapata. 2015. Distributed Representations for Unsupervised Semantic Role Labeling. In *Proceedings of EMNLP*, pages 2482–2491.