The AAAI 2017 Spring Symposium on
Computational Construction Grammar and Natural Language Understanding
Technical Report SS-17-02

Modelling the Meaning of Argument Constructions with Distributional Semantics

Gianluca E. Lebani and **Alessandro Lenci**

Computational Linguistics Laboratory
Department of Philology, Literature, and Linguistics
University of Pisa, Italy
gianluca.lebani@for.unipi.it, alessandro.lenci@unipi.it

Abstract

Current computational models of argument constructions typically represent their semantic content with hand-made formal structures. Here we present a distributional model implementing the idea that the meaning of a construction is intimately related to the semantics of its typical verbs. First, we identify the typical verbs occurring with a given syntactic construction and build their distributional vectors. We then calculate the weighted centroid of these vectors in order to derive the distributional signature of a construction. In order to assess the goodness of our approach, we replicated the priming effect described by Johnson and Golberg (2013) as a function of the semantic distance between a construction and its prototypical verbs. Additional support for our view comes from a regression analysis showing that our distributional information can be used to model behavioral data collected with a crowdsourced elicitation experiment.

Argument constructions and verb meaning

One of the main tenets of Constructionist approaches (Goldberg 2006; Hoffmann and Trousdale 2013) is that grammar contains argument constructions (Cxns) associated with abstract meanings independently of the verbs occurring in them. For instance, the Ditransitive Cxn "Subj V Obj_1 Obj_2" (*She gave him a book*) is associated with a semantic content that can be paraphrased as X CAUSES Y TO RECEIVE Z.

This represents a major breaking point between Constructionist models and so-called "projectionist" approaches stemming from the formal generative tradition (Chomsky 2000) and arguing for the the pivotal role played by the main verb in the interpretation of the sentence. According to this latter view, grammar includes only formal combinatorial principles, while the semantic properties of a sentence are determined by the syntactic and semantic properties projected from the main verb. For instance, the syntactic configurations in the sentences in (1) are projections of the main verb *to slice*:

(1) a. He sliced the bread.
     b. Pat sliced the carrots into the salad.
     c. Pat sliced Chris a piece of pie.
     d. Pat sliced the box open.
         (examples from Goldberg (2006))

This view has been criticized in the Constructionist tradition, following theoretical considerations as well as psycholinguistics arguments (Goldberg 1995; 2006). From the theoretical side, it has been pointed out that, if argument structure is projected solely by the meaning of the verb, than it is necessary to stipulate a different meaning for each occurrence of a given verb in various argument structures. As for the sentences in (1), this view would lead to postulate special senses of this verb roughly meaning (1a) *to cut something with a sharp instrument*; (1b) *to cut something with a sharp instrument so as to move it*; (1c) *to cut something for someone else with a sharp instrument*; (1d) *to cut something with a sharp instrument so as to change its state*. In a Constructionist perspective, the verb "to slice" is always used with the intuitive meaning of *to cut with a sharp instrument*, the additional meaning coming from the construction in which it occurs, whose semantics can be paraphrased as (1a) *something acting on something else*; (1b) *something causing something else to move*; (1c) *someone intending to cause someone to receive something*; (1d) *someone causing something to change state* (Goldberg 1995).

Psycholinguistic evidence, on the other side, mostly originates from research on language comprehension and language acquisition. As for the former, studies like Bencini and Goldberg (2000), Kaschak and Glenberg (2000), Kako (2006), Goldwater and Markman (2009) and Johnson and Goldberg (2013) support the idea that the construction of a sentence (rather than the verb only) plays a role in its interpretation. In a sorting experiment, Bencini and Goldberg (2000) showed that, when asked to sort sentences on the basis of their overall meaning, subjects were as likely to rely on the verb as on the construction. Kaschak and Glenberg (2000) and Goldwater and Markman (2009) tapped into the semantic content of different syntactic frames by using novel denominal verbs in a comprehension task. Likewise, Kako (2006) investigated the meaning of six syntactic frames by collecting linguistic judgments over phrases whose content words were replaced by nonsense words (a.k.a. "Jabberwocky"sentences like *The grack mecked the zarg*). While all these works exploited off-line tasks or explicit judgments, Johnson and Goldberg (2013) demonstrated that the constructional meaning is accessed quickly by asking their participants to perform a speeded lexical decision task on a target verb, after being exposed to Jab-

berwocky prime sentences. From an acquisition perspective, studies supporting the so-called 'Syntactic Bootstrapping" hypothesis show that speakers use their knowledge about the meaning of syntactic pattern in order to infer the semantics of a novel verb (Landau and Gleitman 1985; Gleitman and Gillette 1995; Gillette et al. 1999), thus endorsing the idea that argument structures have an abstract semantics that dynamically interacts with the semantics of the main verb.

The fact that Cxns have independent semantic content raises the question of how their meaning is acquired. Goldberg (2006) has argued that the learning of the semantic content of argument Cxn heavily relies on the meaning of high frequency verbs used with them. For instance, the most frequent verb occurring in an intransitive motion Cxn in a corpus of children's early speech is *to go*, which roughly corresponds to the meaning of this Cxn. The same goes for the ditransitive and the caused-motion Cxns and their most frequent verbs, i.e. *to give* and *to put*, respectively (Goldberg 1999). The skewed distribution of verbs and Cxns, with a small number of "general purpose" verbs accounting for most of Cxn tokens, is therefore argued to play a key role in the acquisition of construction meaning. Among the others, Kidd et al. (2010) showed that 4- to 6-years old children were better able to recall finite sentential complement Cxn instances when these contained high frequency verbs, as opposed to when they contained low frequency verbs. Experimenting with artificial languages, Casenhiser and Goldberg (2005) not only showed that 5- to 7-year-old children are able to associate an abstract meaning to a phrasal form, but also that this process is facilitated when a verb occurs in a Cxn with a disproportionately high frequency. Barak et al. (2013) provide further support by exploiting a probabilistic computational model to investigate the acquisition of the English sentential complement Cxns. The obtained results suggest that the learning of an argument Cxn is influenced by a series of distributional properties of the input, among which verb frequency, co-occurrence frequency of a verb with the Cxn, and the frequency of each semantic verb class with the Cxn.

In this paper, we bring support to such hypothesis with a simple corpus-based method apt to infer the semantic content of a syntactic Cxn. Our proposal transposes into distributional terms the idea that the meaning of a Cxn is related to that of the verbs that most frequently appear in it. Traditionally the meaning of a Cxn has always been described in symbolic terms (see Table 1). Alishai and Stevenson (2008) proposed a computational model to learn argument constructions from corpora, whose semantic layer is represented in terms of hand-coded formal features. We instead introduce a fully unsupervised distributional semantic representation of argument constructions. Our representation allows for the measurement of the semantic similiarity between a Cxn and other Cxns and/or lexical elements.

In the next section we will present a distributional semantic model to represent the semantic content of syntactic Cxn. We validate this model on two test beds. In the first experiment, described in section (1), we test the ability of our approach to model the Cxn-verb priming effect reported by Johnson and Goldberg (2013). Section (1) reports a second study in which we investigated whether our distributional model is able to account for behavioral data concerning the intimate semantic link between a Cxns and its prototypical verbs. Final remarks and possible improvements are reported in section (1).

The distributional signature of a syntactic construction

Distributional Semantic Models (Sahlgren 2006; Lenci 2008; Turney and Pantel 2010, DSMs) are unsupervised corpus-based models of semantic representation implementing the so-called "Distributional Hypothesis" (Harris 1954; Miller and Charles 1991), that takes the similarity of the contexts in which two linguistic expressions occur as a proxy to their similarity in meaning. DSMs are typically built by searching all the occurrences of a target expression in a corpus, identifying its contexts of occurrence and representing the target-by-contexts frequencies as a matrix. Contexts can be words, syntactic relations, lexicalized patterns, documents and so on, while the vectors composing the final matrix are assumed to be the distributional representation of the semantics of the target elements. Distributional vectors can be used to evaluate the semantic distance between lexical elements by means of geometric methods (Bullinaria and Levy 2007; 2012; Lapesa and Evert 2014) or manipulated to represent more complex linguistic entities (Baroni 2013).

Our model implements the idea that the meaning of a syntactic Cxn is intimately related to the semantics of its typical verbs. It is a two step process, that starts by identifying the typical verbs that occur in our target syntactic Cxn and building their distributional $\overrightarrow{v}$ vectors. We calculated the weighted centroid of these verb vectors in order to build a $\overrightarrow{\text{CXN}}$ vector encoding the distributional properties of Cxn. The notion of centroid is the generalization of the notion of mean to multidimensional spaces. In a DSM it can be intuitively pictured as the prototype of a set of lexical elements, that is as a representation of the characteristics that are common to the verbs associated with our target Cxn. A positive by-product of a centroid-based representation is that it allows to soften the influence of the idiosyncratic or non-relevant properties of the verbs, as well as the influence of the noise produced by verb polysemy. Given the role of the skewed verb-Cxn frequency distribution, we weighted the salience of each verb in the calculation of the centroid on the basis of its co-occurrence frequency with the target Cxn. Coherently, then, we calculated our weighted centroids as:

$$\overrightarrow{\text{CXN}} = \frac{1}{|V|} \sum_{v \in V} f_{rel}(v, \text{CXN}) \cdot \overrightarrow{v} \qquad (1)$$

where CXN is our target construction, V the set of its top-associated verbs v and $f_{rel}(v, \text{CXN})$ the relative frequency of occurrence of a verb in a construction. For instance, given a Ditransitive target Cxn, whose associated verbs are *to give* ($f_{rel} = 0.75$) and *to hand* ($f_{rel} = 0.25$), its distributional signature would be estimated as:

$$\overrightarrow{\text{DITRANSITIVE}} = \frac{0.75 \cdot \overrightarrow{give} + 0.25 \cdot \overrightarrow{hand}}{2} \quad (2)$$

Weighting verb vectors with their co-occurrence frequency captures the role played by token frequency in determining the construction meaning. In fact, given equation (1) the vectors of high frequency verbs co-occurring with a construction have a greater weight in the resulting centroid, consistently with the evidence about the importance of such verbs for construction learning (Goldberg 2006).

It is worth noticing that our proposal shares a "family resemblance" with the "collostructional analysis" techniques that have been extensively exploited to study the relationship between a verb and the constructions encoding argument structures, tense/aspect, mood and modality, both from a theoretical as well as from a psycholinguistic perspective (Stefanowitsch 2013). The aim of our proposal is, however, radically different: while the collostructional paradigm has been developed to model the strength of association between a Cxn and the grammatical structures it occurs in, our primary intent is to derive the meaning of argument Cxns from the distributional semantic representations of the verbs co-occurring with them.

Implementing the model

We tested the psycholinguistic plausibility of our model by simulating the behavioral data reported by Johnson and Goldberg (2013), further reviewed in the first part of section (1). The requirement for our model is to account for the association between a Cxn and a target verb as a function of their geometric distance in the distributional semantic space. Given the exploratory nature of the work presented in these pages, we did not tune all the possible settings and hyperparameters of our DSM. Rather, whenever possible we relied on what on best practice in the literature.

To implement our proposal we need two kinds of information: the distributional signature of a set of verbs and their relative frequency with a set of syntactic Cxns. We extracted the latter from VALEX (Korhonen, Krymolowski, and Briscoe 2006), an automatically built subcategorization lexicon that encodes information for 6,397 English verbs. From this list we selected, for each of the four Cxns used by Johnson and Goldberg (2013) reported in Table 1, the set of 75 top associated verbs.

To model the distributional behavior of our verbs we built a syntax-based DSM (Grefenstette 1994; Lin 1998; Padó and Lapata 2007; Baroni and Lenci 2010), that is a space in which linguistic expressions are represented in terms of the syntactic dependency paths they co-occur with. For instance, given the sentence *The cat ate my homework*, in a syntax-based model the distributional entry for the verb $\overrightarrow{eat}$ is represented with the `dependency:filler` patterns `subj:cat`, `obj:homework`. We extracted the raw co-occurrence statistics from the extended arcs of the American English section of the Google Books Syntactic Ngrams corpus (Goldberg and Orwant 2013), a 146.2B tokens corpus built from 1.4M books. Verbs failing to reach the minimal threshold of 500 occurrences were discarded.

The raw co-occurrence matrix has been weighted with *Positive Local Mutual Information* (Evert 2008, PLMI) to calculate the strength of association between a verb and a syntactic pattern. PLMI is defined as the log ratio between the joint probability of a target v and a context c and their marginal probabilities, multiplied by their joint frequency, setting to zero all the negative results:

$$PLMI(c,v) = max\left(0, f(c,v) \cdot \log_2 \frac{p(c,v)}{p(c) \cdot p(v)}\right) \quad (3)$$

PLMI corresponds to the Positive Pointwise Mutual Information score (Church and Hanks 1991) between the verb and the context, weighted by their joint frequency, and differs from PPMI in avoiding the bias towards low-frequency events. To ignore unwanted variance and to reduce the processing cost we adopted the context selection strategy proposed by Polajnar and Clark (2014) and limited the distributional characterization of each verb to its 240 top-associated contexts. Finally, we fed equation (1) with all the previously collected statistics on each group of 75 top-associated verbs, thus obtaining the distributional signature of our target Cxns that will be tested in the remaining of the paper.

Jabberwocky sentences prime associated verbs

The starting point of the reflections by Johnson and Goldberg (2013, henceforth JG) is the psycholinguistic literature showing that speakers associate semantic knowledge to argument structures, independently of the linguistic properties of the verb governing it. Moving further, these authors tested the possibility that this knowledge is used automatically, that is quickly and instinctively, in sentence comprehension.

To this end, they submitted 40 speakers with a lexical decision task in which they were required to read a Jabberwocky sentence (i.e., a sentence whose content words have been replaced by meaningless strings) and then to judge as quickly as possible if a target verb was a real lexical element or a non-word. Table 1 reports the four syntactic constructions investigated by JG, along with an informal representation of their meaning and the Jabberwocky sentence.

Half of the target words seen by each participant were non-words, while the other half were the target verbs reported in Table 2, that were further classified into three classes: "High Frequency associate" (HF), i.e. a verb that most frequently occurs in a given Cxn; "Low Frequency associates" (LF), i.e. a verb that frequently occurs in a given Cxn, albeit significantly less than the relevant HF; "Semantically Related nonassociate" (SR), i.e. a verb whose meaning is related to the semantics of the Cxn, but that does not occur in it. Frequencies were estimated from the 400M words COCA corpus (Davies 2009). Each target verb could be presented either in a congruent context, i.e. after a Jabberwocky sentence instantiating the Cxn to which it is associated with (e.g., *Gave* preceded by a Ditransitive prime), or in an incongruent condition (e.g., *Gave* preceded by a Removal prime). In order to simplify the experimental design, the congruency-incongruency conditions were obtained by opposing either the Ditransitive and the Removal Cxns, or the Caused-motion and the Resultative Cxns.

Construction	Structure	Meaning	Jabberwocky Prime
Ditransitive	Subj V Obj$_1$ Obj$_2$	X CAUSES Y TO RECEIVE Z	*he daxed the norp*
Resultative	Subj V Obj Pred	X CAUSES Y TO BECOME Z	*she jorped it miggy*
Caused-motion	Subj V Obj Obl$_{path}$	X CAUSES Y TO MOVE Z	*he lorked it on the molp*
Removal	Subj V Obj Obl$_{source}$	X CAUSES Y TO MOVE FROM Z	*she vakoed it from her*

Table 1: JG's experimental constructions. Adapted from Johnson and Goldberg (2013, Tables 1,3).

Construction	High Frequency associate	Low Frequency associate	Semantically Related nonassociate
Ditransitive	*Gave*	*Handed*	*Transferred*
Resultative	*Made*	*Turned*	*Transformed*
Caused-motion	*Put*	*Placed*	*Decorated*
Removal construction	*Took*	*Removed*	*Ousted*

Table 2: JG's experimental target verbs. Adapted from Johnson and Goldberg (2013, Table 4).

The extent of priming was computed for each target verb as the difference between the reaction times in the congruent condition and the reaction time after the incongruent sentence. JG report a main effect of congruency, according to which each verb was recognized faster after a related Cxn. HF and LF associates were recognized faster in a congruent condition, both by-subject and by-item. SR verbs, on the other side, were recognized faster only in a by-subject analysis, a fact that can be attributed to the well-known weakness of semantic priming with respect to associative priming. Finally, the priming effect was recorded for all classes of verbs but those associated with the Resultative Cxn, a null effect that the authors ascribed to the plausibility of a metaphorical Caused-motion interpretation of these verbs ($^{??}$*She made/turned/transformed into the room*).

All in all, by recording a priming effect of the Jabberwocky sentences instantiating the Cxns in Table 1 over their associated verbs, JG showed not only that argument structures have an inherent abstract meaning independently of their main verb semantics, but also that this knowledge is accessed quickly and implicitly in the process of sentence comprehension.

Modeling the priming effect

The effect reported by JG is a viable and significant testing ground for our model. Replicating the same results with distributional semantic methods allows us to draw conclusions concerning the psycholinguistic plausibility of distributional representations, at the same time supporting the hypothesis that construction meaning is the result of a usage-based process of abstraction from the meaning of co-occurring verbs.

In our DSM, verb and Cxn vectors lie in the same distributional space, that is, they are described by means of the same contexts. This allows us to model the "semantic congruency" of a verb and a Cxn as a measure of the geometric distance between the $\overrightarrow{\text{CXN}}$ and the $\overrightarrow{verb}$ vectors. Following a common practice in the literature, we opted to calculate vectors similarity by measuring the co-

sine of the angle between them (Bullinaria and Levy 2007; Lapesa and Evert 2014).

JG see their priming effect as a proof of the fact that the constructions presented in Jabberwocky sentences have a meaning strongly associated with the one of the congruent target verbs. Accordingly, we expect higher similarity scores between the $\overrightarrow{\text{CXN}}$ and the congruent $\overrightarrow{verb}$ vectors, as opposed to the similarity scores between the $\overrightarrow{\text{CXN}}$ and the incongruent $\overrightarrow{verb}$ vectors. A major difference between JG's analysis and ours, however, concerns the number of oppositions in the incongruent condition. While in JG the congruency-incongruency conditions for each Cxn were obtained by opposing either the Ditransitive and the Removal Cxns, or the Caused-motion and the Resultative Cxns, we opted for a one-vs-all design, in which an incongruent condition is simply a Cxn-verb pairing inconsistent with the pattern in Table 2. We adopted this solution mainly in order to collect more data points for our analysis.

Coherently with JG, moreover, we expect an effect of the frequency class. That is, we expect higher similarity scores between the $\overrightarrow{\text{CXN}}$ and its High Frequency $\overrightarrow{verb}$ vectors, as opposed to the similarity scores between the $\overrightarrow{\text{CXN}}$ and the Semantically Related $\overrightarrow{verb}$ vectors, with the case of the Low Frequency $\overrightarrow{verb}$ vectors falling somehow in the middle.

Results and discussion

A two-ways ANOVA was conducted to compare the effect of the condition (congruent vs. incongruent) and of the frequency class (HF, LF and SR) on the similarity between each verb and the centroid of its class. Following JG, we expected weaker effects due to the relatively low number of items.

We found a significant main effect both for condition $F(1, 42) = 15.91$, $p < .001$, and frequency class $F(2, 42) = 4.86$, $p < .05$. Overall, our verbs are more similar to their congruent construction ($m = 0.32$, $sd = 0.32$) than to their incongruent construction ($m = 0.13$, $sd = 0.09$). Post-hoc analysis using Tukey Honest Significant Dif-
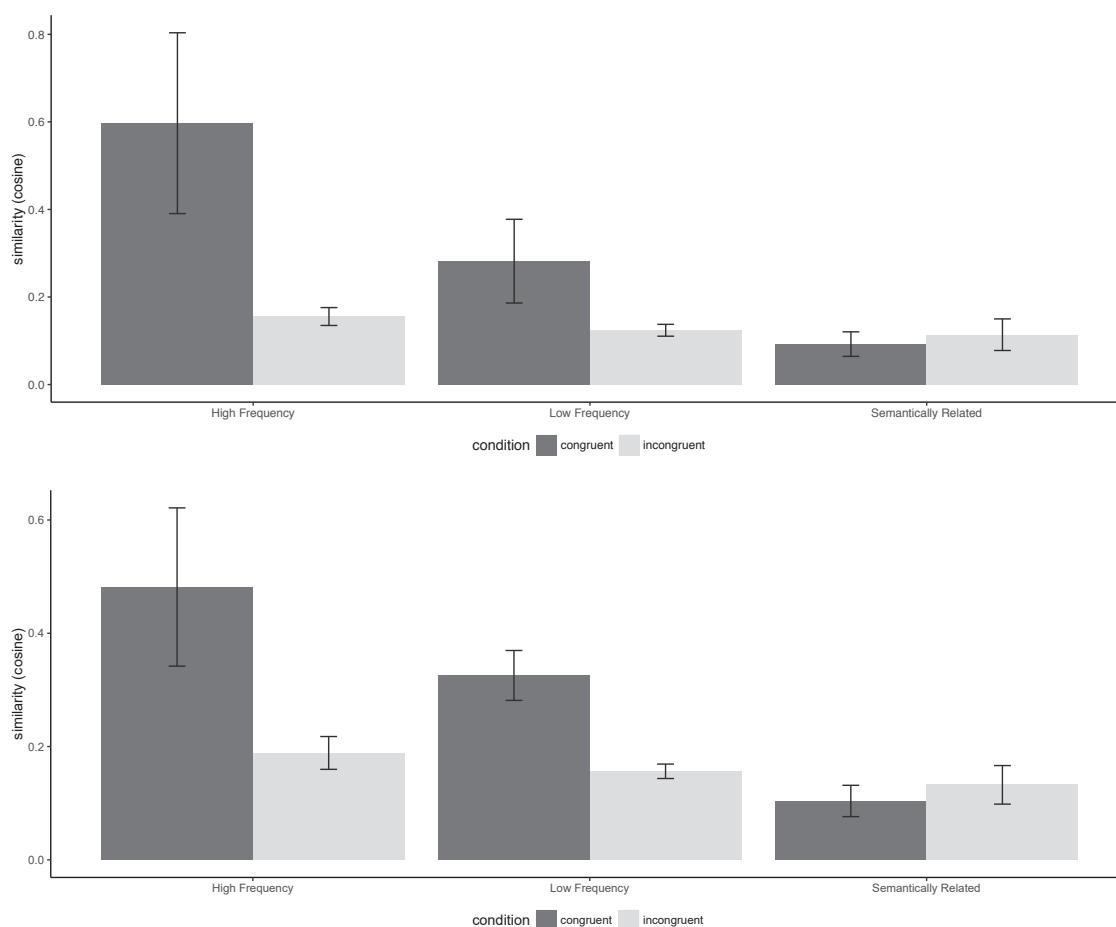
Figure 1: Mean cosine similarity scores as a function of frequency class (High Frequency, Low Frequency and Semantically Related) and experimental condition (congruent vs. incongruent). Vertical capped lines atop bars indicate standard error of the mean. Top: Constructions have been represented with frequency-weighted centroid vectors, as in (1). Bottom: Constructions have been represented with centroid vectors without frequency weighting.

ferences indicated a significant overall difference only between HF ($m = 0.27$, $sd = 0.27$) and SR ($m = 0.11$, $sd = 0.11$) cosines ($p < .05$), but no significant difference involving the LF verbs ($m = 0.16$, $sd = 0.12$).

A significant interaction between the two conditions has been found as well $F(2, 42) = 7.79$, $p < .01$ (see Figure 1). Post-hoc analysis using Tukey Honest Significant Differences indicated a significant difference between congruent ($m = 0.6$, $sd = 0.41$) and incongruent ($m = 0.155$, $sd = 0.07$) condition for HF verbs ($p < .001$), between HF verbs and SR ($m = 0.09$, $sd = 0.06$) verbs in their congruent conditions ($p < .001$), and between HF and LF ($m = 0.28$, $sd = 0.19$) verbs in their congruent conditions ($p < .05$), but no other meaningful contrast reaches statistical significance.

A one-way ANOVA was conducted to compare the effect of the Cxn type on the cosine similarity between each verb and the centroid of its Cxn. We were interested in assessing whether there was a significant difference in how similar

each Cxn vector is to its 75 most associated verbs, i.e. in how dense is the semantic space around each Cxn vector. The answer was affirmative: we found a significant main effect of the Cxn on the cosine similarity for all the four conditions $F(3, 277) = 0.0012$, $p < .01$. Post-hoc analysis using the Bonferroni correction for multiple comparisons indicated a significant ($p < .01$) difference in the densities of the removal ($m = 0.19$, $sd = 0.12$) and of the resultative constructions ($m = 0.11$, $sd = 0.13$), a significant ($p < .05$) difference in the difference in the densities of the removal and of the ditransitive constructions ($m = 0.13$, $sd = 0.12$), and a marginally significant ($p < .1$) difference in the densities of the removal and of the caused motion constructions ($m = 0.14$, $sd = 0.12$). No significant difference in densities has been found for all the other comparisons. This is coherent with the null effect on Resultative Cxn that puzzled JG. But while these ascribed it to a design flaw, i.e. to the fact that Resultative verbs could have a metaphorical Caused-motion interpretation, our results suggests a different inter-

pretation. The fact that in our design we implemented all the possible pairwise oppositions, indeed, suggests that the null effect on the Resultative Cxn is due to the low density of this group of vectors. This is in turn related to the fact that the verbs co-ccurring with the Resultative construction are less semantically homogenous. An in-depth study of the reasons behind the higher distance between the prototypical Resultative verbs and the Cxn is left for further investigation.

The effect of frequency weighting In our model, HF verbs are more similar to their congruent construction than to the incongruent one. One might argue that this is simply the effect of using frequency to weight the vectors in the centroid, because HF verbs have the highest weight in generating the construction vector. In order to exclude this possibility, we carried out a follow-up analysis using unweighted centroids. Differently from equation (1), now each verb type contributes the same to the resulting construction vector.

As can be seen in Figure 1, the patterns of verb-construction cosines in the unweighted model are strikingly similar to the ones in the original model. This similarity is confirmed by a two-ways ANOVA that was conducted to compare the effect of the condition (congruent vs. incongruent) and of the frequency class (HF, LF and SN) on the similarity between each verb and the centroid of its class. We found a significant main effect both for condition $F(1,42) = 14.31, p < .001$, and frequency class $F(2,42) = 5.7, p < .01$. Overall, our verbs are more similar to their congruent construction ($m = 0.30, sd = 0.22$) than to their incongruent construction ($m = 0.16, sd = 0.09$). Post-hoc analysis using Tukey Honest Significant Differences indicated a significant overall difference only between HF ($m = 0.26, sd = 0.2$) and SN ($m = 0.13, sd = 0.1$) cosines ($p < .01$), but no significant difference involving the LF verbs ($m = 0.19, sd = 0.09$). This proves that our distributional model of JS priming is not an artifact of verb frequency weighting.

All in all, we found a pattern that mirrors the priming effect reported by JG. In our DSM, the congruency condition, that in JG leads to faster reaction times, is associated with significantly higher similarity scores. Besides bringing further support to the idea of the link between the meaning of a Cxn and that of its typical verbs, these results confirm the psycholinguistic plausibility of our distributional model.

Isn't frequency enough? Analyzing crowdsourced production data

Works investigating the acquisition of Cxns usually stress the role played by the top-frequent verbs. Psycholinguistic findings (Casenhiser and Goldberg 2005; Kidd, Lieven, and Tomasello 2010) as well as computational simulations (Barak, Fazly, and Stevenson 2013) stress the importance of many frequency-related characteristics, such as the marginal frequency for the verb and the relative frequencies of the verb and of the verb semantic class. One may wonder if the semantic resemblance between a Cxn and its most-associated verbs may be explained simply as a function of frequency, rather than the distributional similarity between

verb and Cxn vectors. We tested this hypothesis by collecting production data from native speakers and assessing whether the inclusion of semantic similarity in a frequency-based model would result in a significant increase in fit.

Data collection

Behavioral data were collected from English speakers through the Crowdflower crowdsourcing platform. 40 certified "highest quality" contributors from the U.K., the U.S.A. or Canada were recruited. Each participant was allowed to complete only a hit (i.e., a "Human Intelligent Task"). In each hit the workers were required to generate, for each of the Jabberwocky prime tested by JG (see Table 1), five verbs that could replace the nonsense main verb of the sentence. Workers were also required to complete, for each Jabberwocky sentence, a language comprehension question of the form "is ghase an English word?". Participants failing to provide 5 descriptions for all the Cxns were not allowed to complete the hit, while participants that did not answer correctly to all the test questions were rejected. On the average, workers needed approximately 6 minutes ($m = 364.075$", $sd = 256.23$") to complete a valid hit. The data collection process took approximately 18 hours.

The workers accepted by the system submitted a total of 800 Cxn-verb pairings, that were subsequently manually filtered and formatted. This processing phase lead to the removal of the verbs submitted by one scammer and to the identification of 376 unique Cxn-verb pairings.

Modeling production frequency

We ran a linear regression analysis on the crowdflower-collected data with production frequency as dependent variable and the joint frequency $f(verb, \text{CXN})$ estimated from VALEX and the verb-Cxn cosine similarity calculated with our model as predictors. We were interested in assessing whether the frequency of production of a verb-Cxn in our crowdsourced data could be modeled on the basis of its relative frequency alone or whether the semantic similarity between the Cxn and the verb plays a role as well.

In a preprocessing phase we removed from the crowdsourced data all those data points corresponding to verb-Cxn pairings that occurred in VALEX less than 100 times. This reduced our dataset to 73 Cxn-verb pairings. Moreover, the raw frequency extracted from VALEX were log-transformed to approximate a normal distribution. Collinearity in the data matrix was evaluated by calculating the Variance Inflation Factors ($VIF = 1.27$) and the Condition Number ($\kappa = 20.76$). While a $VIF < 5$ value is undoubtedly reassuring, the κ value may be cause for concerns, even if it well below the critical threshold of 30 that is commonly taken as an indication of the risk of high collinearity (Cohen et al. 2003; Baayen 2008).

We defined the simplest model as the one in which the only predictor is the log-transformed joint frequency estimated from the corpus. As shown by Table 3, this model looks significantly better that the intercept-only model. We then enriched this model by adding the cosine similarity between each verb and the construction centroid, obtaining significant improvement in the goodness-of-fit. Finally, we

Model 1	Model 2	Δ AIC	Δ BIC	RSS	F
intercept only	frequency	-7.19	-4.9	216.1	9.53 **
frequency	frequency + similarity	-4.34	-2.05	134	6.35 *
frequency + similarity	frequency * similarity	-9.8	-7.51	220.3	12.1 ***

Table 3: Results of the production frequency models comparisons. AIC: Akaike Information Criterion; BIC: Bayesian Information Criterion; RSS: reduction of residual sum of squares; F: F-test statistics and significance values (* $= p < .05$; ** $= p < .01$; *** $= p < .001$).

	Estimate	SE	t
(intercept)	7.06	5.85	1.21
frequency	-0.73	0.99	-0.74
similarity	-64.44	21.39	-3.01 **
frequency:similarity	10.97	3.15	3.48 ***

Table 4: Parameters included in the final model and relevant statistics (** $= p < .01$; *** $= p < .001$).

added the interaction between corpus frequency and cosine similarity, thus obtaining our best fitting model ($F(3, 69) = 10.45$, $p < .001$, $R^2 = 0.312$, $R^2_{adj} = 0.282$). The low R^2 values were not unexpected due to the fact that crucial sources of variance has not been controlled or taken into consideration for the present study, such as the socio-cultural background of the speakers, the different varieties of the English language they were proficient in, the time spent in completing the micro-task and so forth. In this model the significant predictors are the semantic similarity and its interaction with the joint frequency, as reported in the Table 4.

All in all, we interpret these results as proving that the distributional information encoded in the distributional semantic representation of Cxns we have tested in this paper is able to model the linguistic behavior of adult native speaker over and above the variance that can be explained by the joint frequency of the single verbs in a given Cxn. The analysis of its possible theoretical implications are outside the scope of this paper, but we take this result as an additional confirmation of the goodness of our proposal.

Conclusion

We proposed a simple unsupervised corpus-based model that represents the meaning of a syntactic construction as the weighted centroid of the vectors encoding the distributional behavior of its prototypical verbs. Given the exploratory nature of this work, we did not explore the full parameter space of our model, an issue that follow-up studies could investigate, e.g. by comparing the alternative DSM implementations ability to model the priming effect magnitude (Ettinger and Linzen 2016).

Our model and experimental results show that distributional semantics is able to provide a usage-based representation of the semantic content of argument constructions, which is consistent with the available evidence concerning the psycholinguistic reality of construction semantics (Bencini and Goldberg 2000; Kaschak and Glenberg 2000; Kako 2006; Goldwater and Markman 2009; Johnson and Goldberg 2013) and how this knowledge is acquired (Gold-

berg 1999; Casenhiser and Goldberg 2005; Kidd, Lieven, and Tomasello 2010). At the same time, the increment in descriptive and explanatory power obtained by moving from a simple frequency-based measurement to a more complex frequency-based approach like ours shows the importance of developing a more articulate account of the relationship between a syntactic construction and its prototypical verbs.

References

Alishahi, A., and Stevenson, S. 2008. A Computational Model of Early Argument Structure Acquisition. *Cognitive Science* 32(5):789–834.

Baayen, R. H. 2008. *Analyzing linguistic data. A Practical Introduction to Statistics Using R.* Cambridge, UK: Cambridge University Press.

Barak, L.; Fazly, A.; and Stevenson, S. 2013. Modeling the Emergence of an Exemplar Verb in Construction Learning. In *Proceedings of the 35th Annual Meeting of the Cognitive Science Society (CogSci 2013)*, 1815–1820.

Baroni, M., and Lenci, A. 2010. Distributional Memory: A General Framework for Corpus-based Semantics. *Computational Linguistics* 36(4):673–721.

Baroni, M. 2013. Composition in Distributional Semantics. *Language and Linguistics Compass* 7(10):511–522.

Bencini, G. M. L., and Goldberg, A. E. 2000. The Contribution of Argument Structure Constructions to Sentence Meaning. *Journal of Memory and Language* 43(4):640–651.

Bullinaria, J. A., and Levy, J. P. 2007. Extracting semantic representations from word co-occurrence statistics: a computational study. *Behavior Research Methods* 39(3):510–526.

Bullinaria, J. A., and Levy, J. P. 2012. Extracting semantic representations from word co-occurrence statistics: stop-lists, stemming, and SVD. *Behavior Research Methods* 44(3):890–907.

Casenhiser, D., and Goldberg, A. E. 2005. Fast mapping be-

tween a phrasal form and meaning. *Developmental Science* 8(6):500–508.

Chomsky, N. 2000. *New Horizons in the Study of Language and Mind*. Cambridge, UK: Cambridge University Press.

Church, K. W., and Hanks, P. 1991. Word Association Norms, Mutual Information, and Lexicography. *Computational Linguistics* 16(1):22–29.

Cohen, J.; Cohen, P.; West, S. G.; and Aiken, L. S. 2003. *Applied Multiple Regression/correlation Analysis for the Behavioral Science*. Mahwah, NJ: Lawrence Erlbaum Associates, 3rd edition.

Davies, M. 2009. The 385+ million word Corpus of Contemporary American English (1990-2008+): Design, architecture, and linguistic insights. *International Journal of Corpus Linguistics* 14(2):159–190.

Ettinger, A., and Linzen, T. 2016. Evaluating vector space models using human semantic priming results. In *Proceedings of the 1st Workshop on Evaluating Vector Space Representations for NLP*, 72–77.

Evert, S. 2008. Corpora and collocations. In Lüdeling, A., and Kytö, M., eds., *Corpus Linguistics. An International Handbook*, volume 2. Berlin, GE: Mouton de Gruyter. 1212–1248.

Gillette, J.; Gleitman, H.; Gleitman, L. R.; and Lederer, A. 1999. Human simulations of vocabulary learning. *Cognition* 73(2):135–76.

Gleitman, L. R., and Gillette. 1995. The Role of Syntax in Verb Learning. In Ritchie, W. C., and Bhatia, T. K., eds., *Handbook of Child Language Acquisition*. Lndon, UK: Academic Press. 279–295.

Goldberg, Y., and Orwant, J. 2013. A Dataset of Syntactic-Ngrams over Time from a Very Large Corpus of English Books. In *Second Joint Conference on Lexical and Computational Semantics (*SEM)*, 241–247.

Goldberg, A. E. 1995. *Constructions. A construction Grammar Approach to Argument Structure*. Chicago, IL and London, UK: The University of Chicago Press.

Goldberg, A. E. 1999. The Emergence of the Semantics of Argument Structure Constructions. In MacWhinney, B., ed., *The Emergence of Language*. Mahwah, NJ: Lawrence Erlbaum Associates. 197–212.

Goldberg, A. E. 2006. *Constructions at work. The nature of generalization in language*. Oxford, UK: Oxford University Press.

Goldwater, M. B., and Markman, A. B. 2009. Constructional Sources of Implicit Agents in Sentence Comprehension. *Cognitive Linguistics* 20(4):675–702.

Grefenstette, G. 1994. *Explorations in Automatic Thesaurus Discovery*. Kluwer Academic Publishers.

Harris, Z. S. 1954. Distributional structure. *Word* 10:146–162.

Hoffmann, T., and Trousdale, G., eds. 2013. *The Oxford Handbook of Construction Grammar*. Oxford, UK: Oxford University Press.

Johnson, M. A., and Goldberg, A. E. 2013. Evidence for automatic accessing of constructional meaning: Jabberwocky sentences prime associated verbs. *Language and Cognitive Processes* 28(10):1439–1452.

Kako, E. 2006. The semantics of syntactic structures. *Language and Cognitive Processes* 21(5):562–575.

Kaschak, M. P., and Glenberg, A. M. 2000. Constructing Meaning: The Role of Affordances and Grammatical Constructions in Sentence Comprehension. *Journal of Memory and Language* 43(3):508–529.

Kidd, E.; Lieven, E. V.; and Tomasello, M. 2010. Lexical frequency and exemplar-based learning effects in language acquisition: evidence from sentential complements. *Language Sciences* 32(1):132–142.

Korhonen, A.; Krymolowski, Y.; and Briscoe, T. 2006. A Large Subcategorization Lexicon for Natural Language Processing Applications. In *Proceedings of the 5th Edition of the Language, Resources and Evaluation Conference (LREC 2006)*, 1015–1020.

Landau, B., and Gleitman, L. R. 1985. *Language and Experience: Evidence from the Blind Child*. Cambridge, MA: Harvard University Press.

Lapesa, G., and Evert, S. 2014. A Large Scale Evaluation of Distributional Semantic Models: Parameters, Interactions and Model Selection. *Transactions of the Association for Computational Linguistics* 2:531–545.

Lenci, A. 2008. Distributional semantics in linguistic and cognitive research. A foreword. *Italian Journal of Linguistics* 20(1):1–30.

Lin, D. 1998. Automatic Retrieval and Clustering of Similar Words. In *Proceedings of the 36th Annual Meeting of the Association for Computational Linguistics and 17th International Conference on Computational Linguistics (COLING - ACL 1998)*, 768–774.

Miller, G. A., and Charles, W. G. 1991. Contextual correlates of semantic similarity. *Language and Cognitive Processes* 6(1):1–28.

Padó, S., and Lapata, M. 2007. Dependency-Based Construction of Semantic Space Models. *Computational Linguistics* 33(2):161–199.

Polajnar, T., and Clark, S. 2014. Improving Distributional Semantic Vectors through Context Selection and Normalisation. In *Proceedings of the 14th Conference of the European Chapter of the Association for Computational Linguistics (ACL 2014)*, 230–238.

Sahlgren, M. 2006. *The word-space model: Using distributional analysis to represent syntagmatic and paradigmatic relations between words in highdimensional vector spaces*. Phd thesis, Stockholm University.

Stefanowitsch, A. 2013. Collostructional Analysis. In Hoffmann, T., and Trousdale, G., eds., *The Oxford Handbook of Construction Grammar*. Oxford, UK: Oxford University Press. 290–306.

Turney, P. D., and Pantel, P. 2010. From Frequency to Meaning: Vector Space Models of Semantics. *Journal of Artificial Intelligence Research* 37:141–188.