

Semantic similarity to high-frequency verbs affects syntactic frame selection

Abstract: This paper investigates the effect of the high frequency of occurrence of a verb in a syntactic frame on speakers' selection of that syntactic frame for other verbs. We hypothesize that the frequent co-occurrence of a syntactic frame and a particular verb (what we call an *anchor* verb) leads to a strong association between the verb and the frame analogous to the relationship between a category and its best exemplar. Our Verb Anchor Hypothesis claims that verbs that are more semantically similar to the anchor are more likely to occur in that syntactic frame than verbs that are less semantically similar to the anchor. We tested the Verb Anchor Hypothesis on the dative alternation. A corpus study determined that *give* was the anchor verb for the ditransitive frame. We then examined whether high semantic similarity to *give* increases the likelihood of an alternating verb occurring in the ditransitive frame. The results of several logistic regression analyses show that semantic similarity to *give* makes a unique contribution to predicting the choice of the ditransitive frame aside other known factors that affect syntactic frame selection. Additional analyses suggest that the Verb Anchor Hypothesis might also hold for narrow classes of alternating verbs (in the sense of Pinker 1989).

Keywords: frequency, verb similarity, argument structure, syntactic frame selection, dative alternation, corpus study, exemplar, categorization, sentence production

1 Introduction

Most English verbs can occur in more than one *syntactic frame*. The active forms of the verb *give*, for example, can occur in at least two distinct syntactic frames. In the syntactic frame in (1a), the Prepositional Dative or Prepositional frame, the recipient of the gift is described via a PP, typically following the NP that describes the gift. In the frame in (1b), the Ditransitive or Double Object frame, the recipient argument is realized as an NP and precedes the gift argument

realized also as NP. The contrast in syntactic frames illustrated in (1) is often referred to as the *dative alternation*, a tradition we follow.

- (1) a. Mary gave a book to the boy.
 b. Mary gave the boy a book.

Since Green (1974), many researchers have observed that the set of verbs that can occur in a particular syntactic frame or participate in alternate syntactic frames tend to share semantic properties. For example, many of the verbs participating in the dative alternation (e.g. *give*, *promise*, *lend*, *toss* and *offer*) are semantically similar in that they describe situations that involve some ‘transfer.’ But even when verbs meet the semantic criteria that allow them to alternate between two syntactic frames, they do not occur in each frame with equal frequency, as most verbs have a syntactic preference for one frame over the other (see Ford, Bresnan and Kaplan 1982 among others). For example, some verbs like *give* occur very frequently in one variant, here, the ditransitive frame, and much more rarely in the other frame, here, the prepositional frame. In other words, speakers exhibit a syntactic bias associated with a given alternating verb, although both frames are available to speakers and express roughly the same meaning. In fact, *syntactic preferences* vary drastically with verbs. For example, for the verb *throw* in the sense of ‘give by throwing’ speakers choose the prepositional frame far more frequently than the ditransitive frame. While a body of lexical semantics literature (e.g. Gruber 1965; Fillmore 1968; Green 1974; Pinker 1989; Koenig and Davis 2006; Rappaport and Levin 2008) has shown that verbs’ semantic features are crucial determinants of their *possibility* of occurring in alternate frames, less is known about whether and how *individual verb meaning* affects a verb’s syntactic preferences or speakers’ *choice* of one syntactic frame over another. This paper is primarily concerned with the role that the meaning of individual verbs plays in speakers’ preference and choice of syntactic frame.

Many previous studies of the dative alternation have shown that properties of the post-verbal arguments (the *gift* and *recipient* arguments) play an important role in predicting the choice of syntactic frame (Collins 1995; Thompson 1990; Wasow 2002; Bresnan et al. 2007; Theijssen et al. 2013). Summarizing this research, the ditransitive frame tends to be preferred if the situation described by the verb includes a participant that is categorized as a recipient and is

semantically animate and definite; it is also preferred if the syntactic expression of this recipient argument is pronominal and /or morphophonologically short and conveys pragmatically given information. Bresnan and her colleagues (2007) showed, based on a corpus study, that all those factors can help predict a speaker's choice of the ditransitive frame or prepositional dative frame. It is important to note, though, that, to date, the majority of the factors shown to modulate the choice of syntactic frame for the dative alternation pertains to properties of the arguments that co-occur with the verb, not properties of the individual verb chosen by the speaker, in particular, its meaning.

Another robust finding to note, and one particularly relevant to this paper, is that a few verbs occur in a syntactic frame far more frequently than others and often a single verb occurs frequently enough to be considered semantically representative of that frame, e.g. *go, do, get, take, put, give*, etc. (Carroll et al. 1971; Gropen et al. 1989; Goldberg 1995, 1999, 2006; Stefanowitsch and Gries 2003; Ellis et al. 2014). This small number of verbs tend to be *general-purpose* verbs and, as a result, are used in a wide variety of contexts and situations. They usually are the first verbs to be learned by children (Clark 1978) and considered to play a significant role in second language acquisition as they serve as prototypical verbs for particular syntactic frames. They can thus be considered *pathbreaking* verbs (Ninio 1999; Ellis and Ferreira-Junior 2009; Ellis and O'Donnell 2011, 2012). Furthermore, Goldberg and her colleagues (Goldberg, Casenhiser and Sethuraman 2004; Casenhiser and Goldberg 2005) provide evidence that the fact that the frequency of occurrence of verbs in a syntactic frame is highly skewed (some verbs occur in the frame far more frequently than others) actually facilitates the acquisition of a construction. They report that participants acquire artificial grammars more easily and more accurately when they are exposed to syntactic structures with skewed verb frequencies than when they are exposed to syntactic structures with evenly-distributed verb frequencies.

In this study, we aim to explore the possibility that the verb most frequently used in a syntactic frame attracts other verbs to that frame and thus serves as an *anchor* for other verbs. More specifically, we hypothesize that semantic similarity to a verb that occurs very frequently in a syntactic frame modulates speakers' choice of syntactic frame for other verbs. We call our hypothesis the *Verb Anchor Hypothesis*. We call a verb that occurs highly frequently in a frame and that modulates speakers' choice of frame for other verbs an *anchor* for the frame. The syntactic frames we focus on in this paper are the ditransitive and prepositional dative frames.

We investigate whether some verbs can serve as anchor of either frame and whether semantic similarity between these anchors and other alternating verbs modulates speakers' choice of frame, i.e. whether the more similar a verb is to an anchor, the more likely it is to occur in the same frame that the anchor is biased towards.

The motivation behind our hypothesis is the assumption that the high frequency of occurrence of a verb in a syntactic frame leads to a strong cognitive association between that verb and that frame in the same way that a highly frequent exemplar of a category is strongly associated with the category (see Hintzman 1986; Komatsu 1992; Medin and Schaffer 1978; Medin 1989; Nosofsky 1988, among others on categorization in general and Lakoff 1982, 1986; Langacker 1987 for the role of categorization in language). Research on categorization shows that the more similar an entity is to the most frequent member of a category, the more likely it is to be considered a member of that category. Several previous studies support our assumption that the relationship between a verb and a syntactic frame is cognitively analogous to the relationship between an exemplar and a category (e.g. Johnson and Goldberg 2013; Snider 2008). First, syntactic frames are not explicitly taught but are abstracted from repeated exposure to sentences with various verbs in the same way categories are extracted from exemplars (Goldberg et al. 2007; Tomasello 1992). Second, syntactic frames often display typicality effects. For example, speakers tend to think of the most typical verb when asked to name a verb that can occur in a particular syntactic frame. If asked to provide an example verb that would fit the word string '*A man ___ a kid a toy,*' people are most likely to come up with the verb *give* (Goldberg 1995; Ellis et al. 2014). Importantly for our concerns, these typicality effects seem to be a consequence of the fact that verbs occur in syntactic frames with different frequencies.

We assume two joint causes for the putative effect the meaning of verb anchors may have on production. First, verbs overlap in meaning and the concept that a speaker wishes to express activates the meaning of verbs it overlaps with (see, among others, Dell 1986). For example, McRae and Boisvert (1986), among many others, showed that words are primed to the degree to which two words are semantically similar to each other. Second, verbs whose meanings are activated activate in turn the syntactic frame(s) they occur in. Although we talk for most of this paper as if there was only one anchor for a syntactic frame, there are in fact multiple putative anchors, with more or less effect on production, as each verb is more or less semantically similar to all other verbs and each verb is more or less associated with a syntactic

frame. Some verbs play a particularly important role, though, as they occur particularly frequently in a frame and this paper focuses on the impact these verbs have on syntactic frame selection, both because the literature on argument structure has stressed the important role these verbs play and because it is initially most appropriate to test our Verb Anchor Hypothesis on verbs whose putative effect on syntactic frame selection is the largest.

Finally, note that the Verb Anchor Hypothesis only depends on one situation category (a verb meaning) activating similar situation categories (verbs with similar meanings) and the strength of association between verbs and the alternating frames they occur in. It does not depend on and is agnostic about whether syntactic frames themselves are associated with a meaning as scholars such as Pinker (1989), Goldberg (1995), or Ambridge et al. (2014) have argued. We return to this issue in the General Discussion.

As just mentioned above, we investigate the Verb Anchor Hypothesis on the set of verbs that participate in the dative alternation (Levin 1993). We first examine the frequency distribution of alternating verbs in the ditransitive and prepositional frames and identify a verb that may constitute the anchor of either frame. Once the anchor is identified, we measure the semantic similarity between the anchor and each of the other alternating verbs and enter this similarity into logistic regressions as a predictor of the choice of the ditransitive or prepositional frame, following Bresnan et al.'s (2007) methodology. The Verb Anchor Hypothesis will be confirmed if an alternating verb's semantic similarity to the anchor verb is a significant predictor of the syntactic frames in which the verb occurs in our corpus. To test our hypothesis we collected from the British National Corpus sentences that instantiate either the ditransitive or the prepositional frame for each alternating verb. Each sentence token in this dataset served as an individual data point in our logistic regression models.

The organization of this paper is as follows. Section 2 presents the results of an extensive corpus study that investigates the frequency distribution of verbs in the ditransitive and prepositional frames and introduces measures of the strength of cognitive association between a verb and a syntactic frame. Section 3 provides details on how we measured semantic similarity between verbs and presents a simple logistic regression model of our corpus data that supports our Verb Anchor Hypothesis. Sections 4 and 5 investigate whether semantic similarity to the anchor verb makes a unique contribution to the prediction of syntactic frame selection above and beyond other factors known to affect the choice of syntactic frame. Section 6 examines

whether semantic similarity to an anchor verb affects syntactic frame selection within narrower semantic classes than the entire class of alternating verbs. Section 7 concludes the paper.

2 The dative alternation in the British National Corpus

Previous researchers have shown that the frequency distribution of verbs in the ditransitive and prepositional frames is highly skewed (Gropen et al. 1989; Goldberg et al. 2004). In this section, we corroborate previous research with a study of all verbs listed as alternating in Levin (1993). Our corpus was a version of the British National Corpus (BNC) syntactically annotated by an automatic parser (Charniak 1997).¹ The results we report provide an overview of the frequency distribution of verbs that occur in the two frames in that corpus.²

We first retrieved from the parsed BNC the verb phrases that instantiate the ditransitive or prepositional frame, i.e. [V NP NP] or [V NP PP]. We then discarded sentences whose main verb was not one of the 122 verbs that Levin (1993) listed as participating in the dative alternation. Levin’s original list includes 127 verbs. The number of verbs we started with, however, was 122, as five verbs are listed twice in Levin’s list but with two different senses and our parsed BNC cannot discriminate between verb senses. 13 verbs from Levin’s list occurred in neither frame in our corpus (*schlep, tote, bus, truck, modem, netmail, satellite, semaphore, telecast, telex, wireless, bunt, and punt*). We excluded *render* and *vote* by hand because most of their tokens did not instantiate the meaning ‘transfer.’ Finally, we had to remove *pass* and *relay* as well because they have two distinct senses, according to Levin, one of which entails caused possession irrespective of syntactic frame and one of which does not (Rappaport and Levin 2008). Since entailing (or not entailing) caused possession irrespective of syntactic frame is a semantic factor in several of our models and our automatic annotation of the BNC cannot

¹ The *precision*, *recall*, and *F measure* for prepositional phrase structures in the BNC annotated via the Charniak parser are 82%, 92%, and 87%, respectively; the *precision*, *recall*, and *F measure* for the ditransitive frame are 81%, 85%, and 83% (Roland et al. 2007).

² As is well-known, some verbs of transfer of possession can occur only in one of the frames in the dative alternation (e.g. *bet* and *transfer*) and some recent verbs that can alternate are not included in Levin’s list (e.g. *tweet*). We considered only verbs that Levin (1993) listed as participating in the dative alternation for two reasons. First, since our hypothesis pertains to the *choice* of syntactic frame, we can only use verbs that alternate. Second, there are individual as well as inter-dialectal differences in the acceptability of the occurrence of new verbs in either syntactic frame. We thus conservatively restricted our analysis to verbs where variation is less pronounced. Nothing critical, we believe, hinges on this latter point.

distinguish between verb senses, we had to omit these two verbs from our study. This process left us with 105 distinct verbs and 62,713 sentences or tokens of either the ditransitive or prepositional frame.

Table 1. Frequency distribution of *give* vs. the other 104 verbs

Verb	Tokens					Proportions
	Ditransitive (D)		Prepositional (P)		D+P	D:P
<i>give</i>	15,311	58%	8,402	22%	23,713	65:35
other 104 verbs	10,732	42%	28,268	78%	39,000	28:72
Total	26,043	100%	36,670	100%	62,713	42:58

As shown in Table 1, not only does our corpus study confirm the extremely high frequency of *give* in the ditransitive frame, but it also shows that *give* differs from other alternating verbs in its relative frequency of occurrence in the ditransitive and prepositional frames. *Give* is overwhelmingly more frequent in the ditransitive frame than any other verb. *Give* accounts for 58% of all ditransitive tokens while the 104 other verbs account for only 42%. The second most frequent verb in the ditransitive frame is *tell*, which only accounts for 10% of the tokens of the ditransitive frame. Additionally, *give* occurs more frequently in the ditransitive frame than in the prepositional frame (D:P = 65:35) whereas the vast majority of other verbs occur more frequently in the prepositional frame (D:P = 28:72 is the mean distribution for the other 104 verbs). *Give*'s strong preference for the ditransitive frame stands out as most alternating verbs have a preference for the prepositional frame. These two distributional facts (relative and absolute frequency of *give* in the ditransitive frame) suggest that there may exist a strong cognitive association between *give* and the ditransitive frame and that the relation between the ditransitive frame and *give* may be analogous to the relation between a category and its most typical exemplar.

A few verbs (*email*, *ask*, and *tell*) show an even stronger bias for the ditransitive frame than *give*. But, none of them, we argue, constitutes a better exemplar of the ditransitive frame than *give* as they occur much less frequently in the ditransitive frame. What increases the strength of association between a verb and a syntactic frame is, we claim, not only the

preference the verb exhibits for that frame (the *relative* frequency of occurrence of a verb in a syntactic frame), but also the token frequency of occurrence of the verb in that syntactic frame (its *absolute* frequency of occurrence in that frame). In other words, a verb is the best exemplar or anchor of a syntactic frame when (1) it occurs very frequently in that frame and (2) it strongly prefers that frame over the alternation's competing frame. When both conditions are met, not only should the verb evoke the syntactic frame but the frame should conversely evoke the verb. Such is the case for the verb *give* and the ditransitive frame, whereas verbs like *email*, *ask* or *tell* do evoke (or activate) the ditransitive frame, but the ditransitive frame does not strongly evoke (or activate) either *ask* or *tell* as their total number of occurrences in the frame is very low.

There are various ways to quantify the strength of association between a verb and a frame. We discuss three possible measures here: Gries and Stefanowitsch's (2004) collocational strength, Ellis and Ferreira-Junior's (2009) ΔP (more precisely, ΔP Attraction or ΔP Construction \rightarrow Word),³ and a measure based on Hebb's (1949) learning rule (see Proulx and H  lie 2005 and McClelland 2006 for recent discussion of the role Hebbian learning may play in human cognition). We used several measures of association, as each measure has advantages and drawbacks (see Schmid and K  cenhoff 2013 for a discussion and comparison of various verb-syntactic frame association measures). The use of several convergent measures helps ensure the robustness of our models of speakers' choice of syntactic frame, as despite important conceptual differences between these three measures, they agree on all the anchors we make use of in this paper.

To define *collocational strength* between a verb and a frame, one needs to build for each verb a contingency table that compares the number of occurrences of a verb in a frame (vs. its occurrences in competing frames) to the number of occurrence of all other verbs in the frame (vs. their occurrences in competing frames). Collocational strength is then defined as the *p* value of the Fisher exact test run on each contingency table. The ΔP *Attraction* measure is defined as

³ Ellis and Ferreira-Junior (2009) also considered the converse ΔP Word \rightarrow Construction or " ΔP Reliance" measure. This measure indicates how good a cue of a syntactic frame a word is. In the context of our study, a drawback of the ΔP Word \rightarrow Construction measure is that it does not properly reflect the raw frequency of occurrence of verbs in the construction. For example, in our dataset, *email* (D:P = 100:0, total counts = 1) and *tell* (D:P = 89:11, total counts = 3,041) showed higher ΔP Reliance than *give* (D:P = 58:42, total counts = 23,713), due to their high *relative* frequency of occurrence in the ditransitive frame despite their extremely low *absolute* frequency of occurrence in the ditransitive frame. This over-sensitivity to low-frequency words that are highly biased towards a particular syntactic frame is why, Ellis and Ferreira-Junior (2009) surmise, the ΔP Word \rightarrow Construction is less predictive of the acquisition of syntactic frames by non-native speakers.

the difference between the probability of the verb given the frame and the probability of the verb given competing frames. Finally, Hebbian association strength (w_{verbD}) is proportional to the difference between how many times a verb occurs in a frame vs. how many times it occurs in competing frames (see the appendix for more details of this measure). For purposes of this paper whose main concern is speakers' choice between the ditransitive and prepositional frames, competing frames for each of these measures is simply the prepositional frame when measuring the association strength of a verb to the ditransitive frame, and the ditransitive frame when measuring the association strength of a verb to the prepositional frame.

When applied to our corpus data, all three measures of association strength select *give* as the verb most strongly associated with the ditransitive frame and *tell* is ranked second. A look at the magnitude of the strength of association furthermore shows that *give* is about three times more strongly associated with the ditransitive frame than *tell* if one adopts a Hebbian rule based measure ($w_{giveD} = 6909$ vs. $w_{tellD} = 2363$) and similar results obtain if the ΔP Attraction measure is used ($\Delta P_{D \rightarrow give} = 0.359$ and $\Delta P_{D \rightarrow tell} = 0.095$). We do not provide quantitative comparison of strengths of association for Gries and Stefanowitsch's measure, as p values cannot be used to estimate effect size, as Schmid and Kücenhoff (2003) point out, but we note that their rank-ordered list of the verb-ditransitive association strengths matched the other two measures well particularly for high-frequency verbs. Despite differences in theoretical assumptions, then, all measures we considered confirm that the verb *give* is most strongly associated with the ditransitive frame and is far more so than any other verb (all three measures also agree on the other anchors we consider in Section 6). The full list of verbs and estimates of their connection strengths as per the three measures we just discussed are provided in the appendix.

We selected our anchor verb from the rank-ordered list of association strengths. We surmise that the magnitude of association strength matters (both absolute strength and, possibly, the difference between the top-ranked verb's strength and the second ranked verb's strength). Unfortunately, we do not know of any quantitative criteria by which we can decide what magnitude of association strength indicates a strong enough association for an anchor effect to occur. Furthermore, magnitude of strength of association may depend on the particular measure of association strength one chooses (e.g. the association strength of *give* with the ditransitive frame is three times higher than that of *tell* by the Hebbian rule based measure but

four times higher by ΔP Attraction). We therefore only consider rank in selecting anchors in this paper.

3 Semantic similarity to *give* predicts the choice of syntactic frame

Our Verb Anchor Hypothesis predicts that high semantic similarity to a frame's anchor facilitates the occurrence of other verbs in the same frame. Since *give* is the ditransitive frame's anchor, the Verb Anchor Hypothesis predicts that the more semantically similar a verb is to *give*, the more likely it will occur in the ditransitive frame. We use logistic regression to test whether a verb's semantic similarity to an anchor verb, e.g. *give*, is a significant predictor of the dative alternation. We test our Verb Anchor Hypothesis by entering in a logistic regression model the semantic similarity between the anchor (*give*) and an alternating verb as predictor and the syntactic frame of any sentence that includes that alternating verb as an outcome variable. In other words, our predictor is the semantic similarity score and what is predicted is whether the speaker used the ditransitive frame or not.

This section presents the simplest of our models, where we test whether a main verb's semantic similarity to *give* is on its own a predictor of the choice of syntactic frame. We will report more and more complex models in the following sections. We introduce another verb semantic predictor we hypothesize to influence syntactic frame selection in Section 4. We report logistic regression and correlation analyses that study the effect of both verb-internal semantic predictors on the choice of syntactic frame. We then verify the effects of the two verb-internal predictors we introduced in this section and the next in the context of the well-known effects of verb-external factors on syntactic frame selection in Section 5, where we examine whether our two semantic verb-internal factors make a unique contribution to predicting the dative alternation above and beyond the effects of previously known predictors. To prepare the outcome or dependent variable for all logistic regression analyses, we coded as either ditransitive (1) or prepositional (0) frame every sentence collected from the parsed British National Corpus. We also identified the verb in every sentence so that we could measure its semantic similarity to *give*. Since *give* was chosen as the anchor of the ditransitive frame and similarity to *give* was our predictor, sentences whose main verb was *give* were excluded from the model. Including sentences that include the anchor verb would artificially boost the effect of

semantic similarity since the anchor is, of course, the verb that is semantically most similar to the anchor and the anchor, by definition, occurs very frequently in the ditransitive frame.

To estimate semantic similarity between alternating verbs and *give* (our predictor), we used Latent Semantic Analysis (LSA hereafter; Landauer et al. 2007), which is one of the distributional models of semantic memory developed in computational linguistics. The goal of distributional models in general is to build “a formal cognitive mechanism to learn semantics from repeated episodic experience in the linguistic environment, typically from a text corpus” (Jones et al. 2015, p.239). The tenet of this approach is often described by Firth’s (1957) wording “you shall know a word by the company it keeps.” LSA computationally and statistically simulates the *contextual* usage or overlap of words and computes their similarities or relatedness using natural language corpora that are meant to reflect our *experience* with language (see Landauer et al. 1998, for more technical details).

LSA is one of many currently available measures of semantic similarity between words (see Jones et al. 2015 for a review) and each similarity measure has advantages as well as drawbacks for different research purposes. Unfortunately, as semantic similarity measures have been evaluated mostly on their performance in noun-to-noun comparisons (e.g. Maki et al. 2004), it is difficult to determine which measure is the most accurate estimate of semantic similarity between verbs. We will use LSA cosines as our primary measure of verb similarity, but we will also report some additional analyses that use a WordNet-based similarity (Pedersen et al. 2004) and GloVe vectors (Pennington et al. 2014) to ensure that our findings are not driven by biases in LSA.⁴

The intuition behind LSA is that the similarity in meaning of two words or sets of words can be estimated by their co-occurrence and contextual overlap: Do the two words occur in the same documents? Do the two words occur with the same set of words even if they do not co-occur directly? (See Deerwester et al. 1990, Landauer and Dumais 1997 and Kwantes 2005 for technical details.) Taking an example in Jones et al. (2015), *robin* and *egg* may be related because

⁴ The most serious drawback of using a non-distributional measure of semantic similarity like WordNet for the purposes of our study is that it requires hand-picking senses from dictionary-style verb entries. Which senses to include is not always easy. Even when the relevant senses can be selected for a particular pair of verbs, we must deal with the fact that sense-to-sense similarities can greatly vary across senses. The pairwise semantic similarity of the three dative senses of *give* and seven senses of *grant* vary from .087 (*give* 1 - *grant* 7) to .810 (*give* 1 - *grant* 2). The mean, the median, or the maximum value of all pairwise values may be chosen as *the* pairwise semantic similarity (e.g. Ellis et al. 2014), but we do not know of any basis for any of these choices.

they often directly co-occur with each other. But, *robin* and *sparrow* may also be related because they frequently occur in similar contexts or with the same set of words, although they rarely co-occur directly.

When applied to two expressions or sets of expressions, LSA produces cosines ranging from 0 to 1 as a measure of similarity (but artifacts of LSA computations sometimes produce slightly negative values). Cosines closer to 1 indicate high semantic similarity. Crucially, LSA does *not* take into account syntactic differences in computing semantic similarity. That is, when LSA compiles the corpus term-by-document frequency matrix, the document is treated as a ‘bag of words’ and does not include any transitional or syntactic information. For example, the LSA cosine of the sentences *she gave the boy the book* and *she gave the book to the boy* is 1, a reflection of the irrelevance of syntactic frames in LSA estimates of semantic similarity. This insensitivity to syntactic differences is critical for our purposes as it means that LSA does not produce high estimates of the semantic similarity between two verbs just because they often occur in the same syntactic frame. The insensitivity of LSA to syntactic context is particularly appropriate for our study, as we investigate the role of verb similarity on syntax, not the role of syntax on verb similarity.

Our LSA was trained on the British National Corpus, used 400 dimensions and treated paragraphs as documents. Using BNC-based spaces allows us to alleviate concerns that may be raised by the relatively small size and somewhat idiosyncratic nature of the spaces used by the on-line LSA (school textbook samples) (<http://lsa.colorado.edu>). In applying LSA, we used the past tense form of *give* (i.e. *gave*) and the past tense form of each of the 104 alternating verbs. Since LSA uses word forms not lemmas to compute cosines, measuring a verb’s similarity to *gave* (i.e. *gave* vs. *sent*) or to multiple forms of the same verb (e.g. *give gives gave* vs. *send sent*) produces slightly different results. We ran LSA both ways, using the online version of LSA and the default settings (i.e. pairwise comparison with a topic space of ‘General Reading up to 1st year college (300 factors)’ and term-to-term comparison). Although the results are highly correlated (Pearson’s $r = .946$, $p < .001$), a close review of individual pairwise comparisons revealed that cosines tend to be slightly higher for verbs whose non-third singular present tense / bare infinitive forms are identical to their noun forms, e.g. *kick* and *offer*. In order to minimize inconsistencies across verbs, we chose to use past tense forms in our pairwise verb comparisons. The LSA cosines for our 104 pairwise comparisons ranged from 0.047 to 0.946.

While preparing the data, we observed a noticeable trend: Cosines tend to be bigger for high frequency words than low frequency words. Taking the number of occurrences of a verb in our data (summing across ditransitive and prepositional uses) as an estimate of a verb's frequency, we found that verb frequencies significantly correlate with LSA cosines (Pearson's $r = .443$, $p < .001$). As our hypothesis is not meant to take frequency into account in our semantic similarity measure, we residualized LSA cosines over verb frequency, statistically removing the portion of the cosine that is contributed by verb frequency. Using residualized cosines allows us to see the effect of semantic similarity while controlling the effect of verb frequency. We observed that residualization lowered the LSA cosines of high-frequency verbs. For example, *handed* and *brought* have similar non-residualized LSA cosines to *gave* (.53 and .56) but when LSA cosines are corrected for frequency of occurrence (567 and 5507 for *handed* and *brought*, respectively), their LSA cosines to *give* are more dissimilar (0.247 and 0.098). It should be noted, though, that there was a high correlation between unresidualized and residualized LSA cosines (Pearson's $r = .896$, $p < .001$), indicating that residualization did not dramatically alter, overall, the relative sizes of LSA cosines.

Overall, post-hoc analyses revealed that our particular choices did not have much of an effect: The same effects of semantic similarity we report below are found whether a BNC-based or online-based LSA matrix is chosen as a basis of semantic similarity, whether past tense verb forms alone or multiple verb forms are used in verb similarity comparisons, or whether cosines are residualized over frequency or not, suggesting the results of our analyses are robust. The choices we made simply contribute to refining and improving the accuracy of our estimate of the independent variable (verb-to-*give* semantic similarity), but do not drive the results, as the same effects hold when other choices are made.

After estimating our predictor and outcome variables, we entered into a logistic regression model residualized LSA cosines between *give* and each sentence's main verb as predictor variable and syntactic frame as outcome variable. For this model as well as all subsequent models fitted to our BNC data, we centered all predictor variables to minimize any problems that multicollinearity between variables may cause (Aiken and West 1991). In reporting the models, we present the *coefficients* (b) and the *z-statistic* of individual predictors to show the direction as well as significance of each predictor (i.e. whether similarity to *give* makes a significant *positive* or *negative* contribution to predicting the choice of ditransitive frame).

When model selection is discussed, we also report Akaike Information Criterion (AIC) estimates, i.e. ΔAIC and weight (Wagenmakers and Farrell 2004).

The results of our first model showed that, as predicted by the Verb Anchor Hypothesis, a verb being semantically similar to *give* contributes to predicting the use of the ditransitive frame ($b = 0.48, z = 35.41, p < .001$). This result suggests that the frequency-driven association between an anchor verb like *give* and the ditransitive frame plays a significant role in predicting speakers' choice of the ditransitive frame. To address the concern that the result was unexpectedly driven by residualized LSA cosines over frequency, we also tested the effect of non-residualized LSA cosines for the same model. The results of this model confirm that this was not the case: We found a significant effect of verb similarity to *give* ($b = 0.27, z = 22.53, p < .001$). We also tested the Verb Anchor Hypothesis with a different estimate of semantic similarity, the WordNet-based vector measure (Patwardhan 2003), to a subset of our data (22,787 sentences, $D = 8,031$; $P = 14,756$) that included the 49 alternating verbs that occur in four narrow classes in Pinker's (1989) sense, namely *give* verbs (but excluding *give*), message transfer verbs, future having verbs, and *send* verbs.⁵ We chose these four classes because they constitute the main narrow classes of alternating verbs and choosing the 'right' senses proved difficult for many verbs in other classes (often, none of their senses in the WordNet dictionary was appropriate for their ditransitive use). We used the mean of the vectors between verb senses as an estimate of verb similarity. This model replicated the effects of similarity to *give* on syntactic frame selection ($b = 0.12, z = 8.37, p < .001$) that we found when using LSA as our measure of semantic similarity.

The study we just reported is the first, to our knowledge, to quantitatively investigate the role that the meaning of the most typical exemplar of a syntactic frame (what we call an anchor verb) plays in the process of syntactic frame selection. Highly frequent and "typical" verbs have received much attention in language acquisition, as several studies have shown language learners are sensitive to verbs' frequency distributions and biased frequencies can facilitate their grammar learning (Goldberg et al. 2004; Ellis and Ferreira-Junior 2009; Ambridge et al. 2014, among others). The present study provides convergent evidence on the importance

⁵ We also compared LSA cosines and human ratings for 34 alternating verbs. LSA cosines and human ratings were found to significantly correlate (Pearson's $r = .534, p < .01$).

of these most frequent and typical verbs (or anchor verbs) by showing the influence their meanings may have on adult speakers' syntactic choices.

For the current simple model where only one predictor variable is included, the Verb Anchor Hypothesis can be tested via a correlation analysis. The Verb Anchor Hypothesis is confirmed if there is a significant positive correlation between an alternating verb's semantic similarity to *give* and its proportion of uses in the ditransitive frame. And indeed there is a significant correlation between a verb's semantic similarity to *give* and its proportion of uses in the ditransitive frame (Pearson's $r = .295$, $p < .001$).

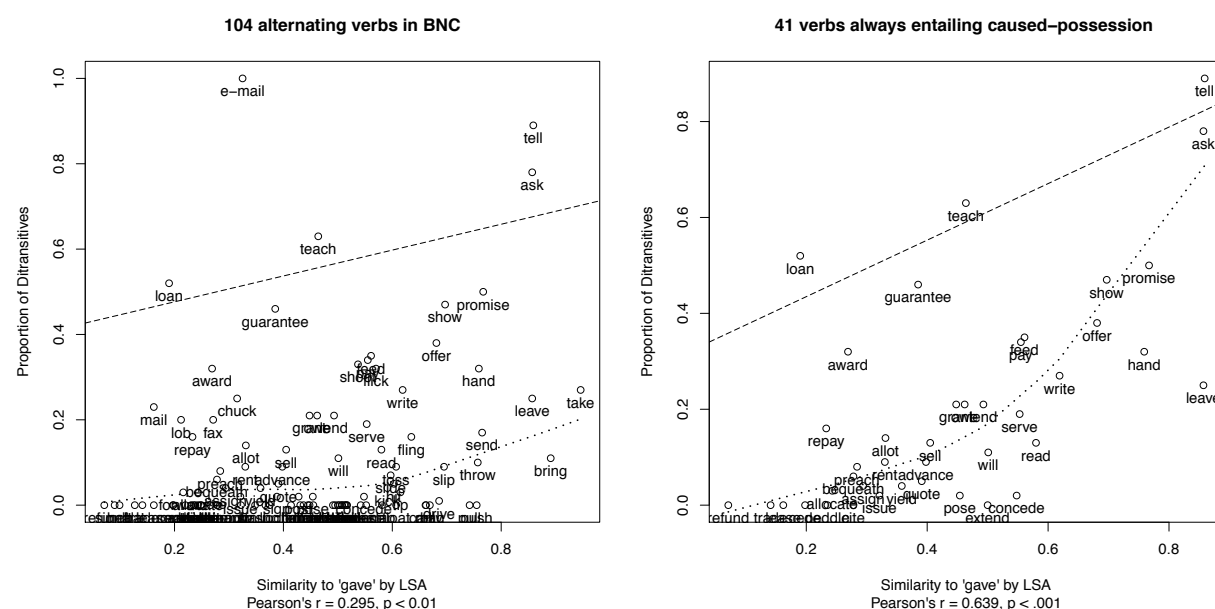


Figure 1. Plots of the correlation between verbs' semantic similarity to *give* (x -axis) and their proportions of occurrence in the ditransitive frame (y -axis) with regressional (dashed) and lowess (dotted) lines

As illustrated in Figure 1, however, many of the putative alternating verbs occur in the ditransitive frame very rarely. In fact, 47 verbs never occurred in the ditransitive frame in our dataset. The correlation between semantic similarity to *give* and proportion of occurrence in the ditransitive frame is stronger when only the 41 verbs that entail caused possession irrespective of syntactic frame are included in the analysis (Pearson's $r = .639$, $p < .001$) (see Rappaport and

Levin 2008 and Section 4 for more details on verbs that entail caused possession irrespective of the syntactic frame they occur in).

Although logistic regression and correlation analyses both support the Verb Anchor Hypothesis, logistic regression better serves the overall purposes of our study. Firstly, as each verb's proportion of ditransitive uses constitutes a single data point in a correlation analysis, results are significantly affected by a few obvious outliers, an issue that is particularly problematic for narrow-class analyses we discuss in Section 6. Secondly, in contrast to logistic regressions, a correlation analysis does not allow us to simultaneously evaluate multiple predictors of choice of syntactic frame (see Sections 4 and 5) and compare our results with those of previous studies, in particular Bresnan et al.'s (2007). As already mentioned, previous literature has shown that various factors modulate the choice of syntactic frame. In the next two sections, we test whether semantic similarity to the anchor remains a significant predictor of syntactic frame selection when other factors are included in a model. We start with another factor related to verb meaning which we alluded to in our description of Figure 1. We then discuss more complex models that include the non-verbal factors known to influence syntactic frame selection in the dative alternation (Bresnan et al. 2007).

4 Combining lexical semantic similarity and lexical entailment

We have assumed so far that both the ditransitive and prepositional frames convey more or less the same meaning. However, many researchers have argued that the two frames are associated with different conceptual or semantic structures (Pinker 1989, among others): The ditransitive frame entails caused possession while the prepositional frame does not. Rappaport and Levin (2008) have further argued that the identity of the main verb modulates this difference in meaning between the two frames. For some verbs, both frames entail caused possession (at least, if we abstract away potential sublexical modal components that restrict this entailment to a subset of possible worlds, see Koenig and Davis 2001). For other verbs, only the ditransitive frame is guaranteed to entail caused possession. This contrast between the two classes of verbs is illustrated in (2).

- (2) a. A woman *gave/throw* her friend a ball.
 b. A woman *gave/throw* a ball to her friend.

Both (2a) and (2b) entail caused possession when the main verb is *give*, whereas only (2a) entails caused possession when *throw* is the main verb. The verb *throw* can be used in the prepositional frame in (2b) not only when a speaker intends to describe an event of transferring a ball to a friend so that he possesses it, but also when, say, a woman throws a ball to hit a man, but not necessarily to cause him to possess it. In other words, the prepositional frame is a meaning-preserving alternative to the ditransitive frame only for *give*-like verbs. For *throw*-like verbs, only the ditransitive frame conventionally conveys caused possession.

The contrast between *give*-like and *throw*-like verbs suggests that the verbs we considered in Section 3 actually consist of two semantically distinct subclasses. Now, intuitively, verbs that carry the same entailment in both frames can be expected to be semantically more similar to each other than verbs that carry different entailments in each frame. Since our verb anchor is *give*, which entails caused possession in both syntactic frames, this means that we can expect *give*-like verbs to be overall more semantically similar to *give* than *throw*-like verbs. To make sure that semantic similarity to *give* affects syntactic frame selection above and beyond the effect of whether or not a verb entails caused-possession in both frames, we conducted a second logistic regression analysis that included both semantic predictors, the binary ‘caused-possession entailment’ predictor (always entailing (1) vs. not always entailing (0) caused possession) and the continuous predictor ‘semantic similarity to *give*,’ and the outcome variable, the syntactic frame of the sentence (ditransitive (1) and prepositional (0) frame). Our goal was to determine whether both predictors make an independent contribution to speakers’ choice of the ditransitive frame or one predictor is reducible to the other. Except for the addition of the new ‘entailment’ predictor, all other settings were kept the same as in our first logistic regression model reported in Section 3.

The results of our analysis showed that both binary and continuous semantic predictors make distinct significant contributions to the prediction of speakers’ choice of syntactic frame and that their effect goes in the expected direction (always entailing caused possession, $b = 0.74$, $z = 58.25$, $p < .001$; similarity to *give*, $b = 0.62$, $z = 47.37$, $p < .001$). This second model ($\Delta AIC = 0$, $weight = 1$) is better at predicting which syntactic frame is selected than our first model ($\Delta AIC =$

3684, $weight < 0.001$). Our results suggest that the likelihood of a speaker choosing the ditransitive frame increases when the verb always entails caused possession and is semantically similar to *give*. They also suggest that the caused-possession entailment plays an additional predictive role in explaining the choice of syntactic frame. Most importantly, our second model shows that our Verb Anchor Hypothesis cannot be reduced to whether a verb always entails caused possession or not.

We then attempted to further clarify the relationship between the two semantic predictors. Since only one of the two semantic classes of verbs shares the entailment of caused possession with our ditransitive anchor *give* (*give*-type verbs *vs.* *throw*-type verbs), we expect that semantic similarity to *give* plays a more important role for verbs that entail caused possession irrespective of syntactic frame than for verbs that do not. In fact, semantic similarity to *give* may only affect ditransitive frame selection for *give*-type verbs. To quantitatively examine the relationship between the two predictors, we fitted to our data a new regression model that included an interaction term between the two semantic predictors and found the interaction is significant ($b = 0.24$, $z = 10.40$, $p < .001$). An examination of this interaction using the simple slopes procedure recommended in Cohen, Cohen, West, and Aiken (2003) revealed that high semantic similarity to *give* is a better predictor of choice of syntactic frame for verbs that invariably entail caused possession than for verbs that do not. However, a post hoc examination of interaction effects revealed that this interaction is not as robust as their first-order main effects, as selection of dataset modulates the effect of this interaction. See Section 5.1 for more details.

5 Combining verb-internal and verb-external factors

The logistic regression models we reported in the previous two sections investigated how factors related to *verb meaning* help predict choice of syntactic frame. Most previous studies have paid more attention to factors only tangentially related to verb meaning, for example, whether the recipient argument is pronominal or definite, whether the theme argument represents given information, or what the difference in length is between the two post-verbal arguments. Bresnan et al. (2007) conducted a comprehensive analysis of a natural language corpus and showed that these verb-external factors *do* predict, together and individually, the choice of ditransitive or

prepositional frame. In this section, we are concerned with the question of whether the predictor this paper is particularly concerned with, a verb’s semantic similarity to the ditransitive frame’s anchor (*give*), plays a role in choice of syntactic frame that is independent of the verb-external factors that Bresnan et al.’s study showed to be important.

5.1. Combining verb-external and verb-internal factors with a ditransitive frame biased verb as anchor

To determine whether semantic similarity to *give* affects choice of syntactic frame independently of the verb-external factors Bresnan and her colleagues studied, we conducted two separate sets of analyses with distinct datasets. In the first set of analyses, we used the data we collected from the BNC, i.e. the same dataset we used for our previous models, and added some verb-external factors to the list of verb-meaning predictors we discussed in Sections 3 and 4. We kept constant all other settings. We coded each sentence with three more factors, namely, (non-)pronominality of the recipient argument, (non-)pronominality of the theme argument and difference in length between phrases that express the recipient and theme arguments, for a total of five predictors. We chose these three verb-external factors from Bresnan et al.’s list of predictors because they can be relatively easily automatically annotated and are known to significantly modulate the choice of syntactic frame for the dative alternation. Given the large size of our data set, we could not perform a manual annotation of all verb-external factors Bresnan et al. considered. But the three verb-external factors we included in our model are known to be highly correlated with factors that require manual annotations. For example, the pronoun *him* in *Jean gave him a book* is definite, animate, constitutes given information, and is short. Although our first model does not include all of Bresnan et al.’s predictors, it provides a useful point of comparison with previous models we reported as all models try to predict the same outcome for the same dataset.

In order to compensate for the limitations of this model, though, we ran another set of models using Bresnan et al.’ dataset collected from the Switchboard and the Wall Street Journal corpus. This dataset is much smaller in size and number of verbs involved but it is manually annotated with nearly the full set of verb-external factors known to affect the dative alternation. These two sets of models complement each other. The BNC-based model includes fewer predictors than the Bresnan et al.-corpus-based model but benefits from an increase in statistical reliability because of the BNC’s large size (i.e. 100 million words vs. 3 million words). The

Bresnan et al.-corpus-based model includes a more comprehensive set of verb-external predictors, but makes use of a small number of observations for a smaller number of verbs when compared to our BNC-based model.

The results of the BNC-based model show that (i) all five predictors are independently significant predictors of the sentences' syntactic frame, i.e. verb similarity to *give* ($b = 0.50, z = 27.79, p < .001$), entailing caused possession irrespective of syntactic frame ($b = 0.43, z = 24.89, p < .001$), pronominal recipient ($b = 0.43, z = 78.07, p < .001$), pronominal theme ($b = -1.09, z = -42.03, p < .001$), and recipient-theme length difference ($b = 0.82, z = 30.49, p < .001$) and that (ii) the addition of the three verb-external predictors makes the model a far better fit to the data ($\Delta AIC = 0, weight = 1$) than the previous model that included only two verb-internal predictors ($\Delta AIC = 16273, weight < 0.001$). These results show that verb-external factors make an independent contribution above and beyond that of verb-internal factors as well as confirm our previous finding: The likelihood of a sentence occurring in the ditransitive frame increases (and the likelihood of a sentence occurring in the prepositional frame decreases) when the verb in the sentence is semantically similar to *give* and/or always entails caused possession. Our model also replicates Bresnan et al.'s results on a larger data set, namely the choice of ditransitive frame is facilitated when the recipient argument is, but the theme argument is not, expressed by a pronoun, as well as when the recipient is shorter in length than the theme.

To ensure that the results of this first model are not unduly influenced by our choice of semantic similarity estimate (LSA cosines), we ran the same model but with a different estimate of semantic similarity, one based on GloVe (Pennington et al. 2014), with all other settings constant. Note that since GloVe estimates semantic distance (i.e. the smaller the value, the closer the meanings), the Verb Anchor Hypothesis predicts that a larger GloVe semantic distance predicts occurrence in the prepositional frame (rather than occurrence in the ditransitive frame); in other words we predict a significant *negative* coefficient for GloVe semantic distance. The GloVe measure we used in this analysis was trained on a 6-billion-word corpus (Wikipedia and Gigaword 5) and used 300 dimensions. All five predictors remained significant predictors of the choice of syntactic frame when GloVe measure of semantic distance is used rather than LSA cosine similarity. Verb dissimilarity to *give*, as measured by GloVe ($b = -0.65, z = -29.29, p < .001$), entailing caused possession irrespective of syntactic frame ($b = 0.61, z = 30.01, p < .001$), pronominality of the recipient ($b = 1.42, z = 82.80, p < .001$), pronominality of the theme ($b =$

-1.07, $z = -41.34$, $p < .001$) and recipient-theme length difference ($b = 0.79$, $z = 29.60$, $p < .001$) were all significant predictors of choice of syntactic frame. We also found a strong negative correlation between GloVe semantic distance and LSA cosines (Pearson's $r = -0.669$, $p < .001$).

Not only did we ensure that the effect of semantic similarity to *give* is robust across distinct semantic similarity measures, we also made sure our findings were not driven by the particular dataset we used but holds across datasets that include different subsets of alternating verbs that vary in their frequencies of occurrences. It is important to ensure robustness of the effect of semantic similarity to *give* across sets of verbs that differ in frequency of occurrence, as the entire dataset has a skewed frequency distribution, i.e. high-frequency verbs constitute a very large proportion of data points in the regression analysis. In particular, we examined the possibility that the results might have been unduly affected by high-frequency verbs such as *tell* which is similar to *give* in many respects (see Section 2 for its frequency distribution and Section 6 for its semantic properties). We therefore constructed four different datasets by removing sentences (or data points) that included the four verbs with the highest frequency of occurrence one by one from our original dataset (which included 104 verbs and 39,000 ditransitive sentences). We removed, first, sentences whose main verb is *tell* (subset 1: 103 verbs; 35,959 sentences), then additionally removed sentences whose main verb is *take* (subset 2: 102 verbs; 28,295 sentences), followed by additionally removing sentences that included *bring* as main verb (subset 3: 101 verbs; 22,788 sentences) and, finally, additionally removing sentences that included *send* as main verb (subset 4: 100 verb; 18,996 sentences). We then created two additional datasets using a pseudo-random sampling procedure. We first kept all 104 alternating verbs but reduced the bias introduced by high frequency verbs by randomly selecting 100 sentences for the 31 verbs that occurred more than 100 times in the original dataset (subset 5: 104 verbs; 4,816 sentences). Finally, we further removed from subset 5 sentences with verbs that never occurred in the ditransitive frame in our data (subset 6: 57 verbs; 3,751 sentences). The skewedness of the dataset decreased substantially through these procedures, as shown by the fact that the mean and standard deviation of the number of occurrences per verb dropped from 375 and 1072 in our original dataset to 66 and 35.7 in subset 6. We then fitted the same model to these six additional datasets as was fitted to our original dataset.

Table 2. The effect of the verb similarity to *give* across seven datasets (** $p < .001$)

Dataset	Entire data	Subsets					
		[1]	[2]	[3]	[4]	[5]	[6]
<i>b</i> coefficient & significance	0.50 ***	0.31 ***	0.33 ***	0.36 ***	0.40 ***	0.63 ***	0.55 ***
Odds ratio	1.65	1.37	1.39	1.44	1.49	1.88	1.73
95% C.I.	1.59~1.71	1.32~1.42	1.33~1.45	1.37~1.51	1.41~1.56	1.66~2.13	1.53~1.96

All five factors (semantic similarity to *give*, caused possession entailment irrespective of syntactic frame, pronominality of the recipient, length difference between the recipient and theme, and pronominality of the theme) significantly facilitate the choice of the ditransitive frame (first four factors) or the prepositional frame (last factor) across all datasets, as summarized in Table 2. The coefficients (*b*) of semantic similarity to *give* ranged from 0.31 to 0.63 and the odd ratios from 1.37 to 1.88 across the 7 datasets (all $p < .001$). The fact that all factors remain significant predictors of the choice of syntactic frame across all datasets suggests that the results of our original model were not simply due to the high frequency of occurrence of a few verbs, but reflected properties of all alternating verbs we examined.

The models of syntactic frame selection we have considered until now did not include any interaction term, as the Verb Anchor Hypothesis does not make any prediction regarding interactions between semantic similarity to *give* and other factors and followed Bresnan et al.'s (2007) analysis, as they too did not discuss interactions between their predictors of interest. Still, as one reviewer pointed out, it is potentially interesting to explore interactions between the various predictors we included in our model. Since unpredicted interactions can be driven by unknown properties or noise in particular datasets and make it more difficult to interpret predicted main effects, we took the conservative approach to focus on interactions whose effects were not limited to any particular dataset. When we fitted a full model with all possible interaction terms (up to a 5-way interaction) to our original dataset as well as to the six different datasets we just described, only one interaction was significant across all models, the three-way interaction between semantic similarity to *give*, pronominal recipient, and pronominal theme (see the appendix for tables of the results of all models fitted to all seven datasets). When we

examined this interaction using a simple slopes procedure, we found out that non-pronominal themes increase the likelihood of the use of the ditransitive frame more for verbs with high semantic similarity to *give* than for verbs with low semantic similarity to *give* and that, additionally, high semantic similarity to *give* increased the likelihood of the choice of the ditransitive frame only when the theme was not pronominal. The interaction between similarity to *give* and caused-possession we observed and reported in Section 4 was not as robust as we expected. We then fitted a new model including this three-way interaction and all its component two-way interactions to the seven different datasets. Only one two-way interaction was significant irrespective of dataset, the interaction between pronominal recipient and pronominal theme. The same two-way interaction was also the only interaction to be significant across all datasets when we fitted a model that included all possible two-way interactions (ten all in all). An examination of the pronominal-recipient-by-pronominal-theme interaction using a simple slopes procedure revealed that a pronominal recipient facilitates the selection of the ditransitive frame in the presence of a non-pronominal theme but has no such effect when the theme is pronominal. This interaction reflects the grammatical constraint that when both recipient and theme are pronominal, only the prepositional frame is felicitous (*gave it to me* vs. **gave me it*). Critically, *all* five main effects of relevance to the purpose of this paper remained significant in all models with interaction terms fitted to all datasets, i.e. models with all possible interaction terms, models with the three-way and its component two-way interactions and models with all two-way interaction terms.

The final model we report on in this section is an attempt to replicate our results on Bresnan et al.'s corpus of alternating sentences using most of the verb-external factors they considered.⁶ This corpus is manually annotated with all the verb-external properties we mentioned but also includes information on a few verbs classes (*transfer*, *future having* and *communication* classes). The manual annotation means it is relatively noise-free, but it is small in size, as it consists of 3,265 sentences drawn from the Switchboard and Wall Street Journal corpora and only 75 different verbs are observed in these sentences (compared to 62,713 sentences and 105 verbs in our BNC dataset). A further difference between the two datasets is

⁶ The data is publicly available online at the website of the publisher of *Quantitative methods in linguistics* by Keith Johnson (2008), Blackwell. A couple of factors from the original study are omitted in the on-line version, i.e. *person*, *number*, *structural parallelism* and *concreteness of theme*. Conversely, *inanimate theme* is not present in Bresnan et al.'s study but is available in the online data. We used all variables available online in our analysis.

that whereas we collected from the BNC only sentences that contained verbs that Levin (1993) listed as alternating, 34 verbs from Bresnan et al.’s dataset are not on Levin’s list. These 34 verbs account for about 10% of the total tokens (329 sentences) and three of them (*cost*, *charge*, and *do*) make up three quarters of those tokens (242 ditransitive and 2 prepositional sentences). We excluded sentences that contained these 34 verbs from our analysis to allow for a better comparison of the results of this model with those of the previous models. As with all other models, sentences that contained the anchor verb *give* were excluded to prevent an artificial boost in the effect of semantic similarity to *give*. Table 3 presents an overview of the distribution of verbs in the Bresnan et al.-corpus-based dataset.

Table 3. *Give* and other alternating verbs in the ditransitive and prepositional frames from the corpus of Bresnan et al. (2007)

Verb	Tokens					Proportions
	Ditransitive (D)		Prepositional (P)		D+P	D:P
<i>give</i>	1,411	63%	256	31%	1,667	85:15
other 40 verbs	704	33%	565	69%	1,269	55:45
Total	2,115	100%	821	100%	2,936	72:28

As in the BNC, the verb *give* is by far the most frequent verb in the ditransitive frame in Bresnan et al.’s corpus (cf. Table 1 in Section 2). It accounts for 63% of all ditransitive tokens (cf. 58% in our data). Also, *give* seems to occur even more frequently in the ditransitive frame (D:P = 85:15) in Bresnan et al.’s corpus (cf. D:P = 65:35 in Table 1) and, in contrast to our corpus study, other verbs also show a (slight) preference for the ditransitive frame (D:P = 55:45, cf. D:P = 27:73 in Table 1). These differences may be a consequence of the larger proportion of spoken language in Bresnan et al.’s corpus, which tends to contain more pronominal expressions for animate entities (*you*, *me*, *him*, *them*, etc.) than written language corpora. Pronominal recipients are known to favor the choice of ditransitive frame, e.g. *Someone gave me the big book*.

All settings were kept the same as in Bresnan et al.’s (2007) study. To assess the contribution of our verb-internal predictors (a verb’s semantic similarity to *give* and entailment of caused possession) more precisely, we first replicated Bresnan et al.’s model using only their predictors. This model (A) provides us with a baseline to which we can compare our two new

models. One model (B) includes both of our lexical semantic predictors, i.e. semantic similarity to *give* and caused-possession entailment, as additional predictors. The other model (C) additionally includes an interaction term between (non-)pronominal recipient and theme, whose consistent effect on syntactic frame selection on our BNC-based datasets we discussed above. Bresnan et al. coded the ditransitive frame as 0 and the prepositional frame as 1, the opposite of what we did in our previous models. For ease of comparison we follow Bresnan et al.'s coding and the predictor variables are not centered, just as in their study. When interpreting the results of the models we report in this section, it should be borne in mind that a positive coefficient indicates that the predictor decreases the likelihood of the selection of the ditransitive frame (and increases the likelihood of prepositional frame selection), whereas a negative coefficient indicates that the predictor increases the likelihood of the selection of the ditransitive frame (and decreases the likelihood of prepositional frame selection). The coefficients and the significance levels of each predictor are summarized in Table 4.

Table 4. A comparison of the relative magnitude of predictors (coefficients (*b*))

Predictors	A Replication of Bresnan et al.'s	B ⊕ sim-to- <i>give</i> ⊕ csd-poss	C ⊕ prorec-by-prothm interaction
inanimate recipient	3.59 ***	3.64 ***	3.82 ***
inanimate theme	-1.20 *	-1.26 *	-1.28 *
nonpronominal recipient	1.21 ***	1.33 ***	-0.31 ns
nonpronominal theme	-0.70 *	-0.88 **	-1.32 ***
nongiven recipient	1.38 ***	1.34 ***	1.34 ***
nongiven theme	-1.14 ***	-1.18 ***	-1.14 ***
indefinite recipient	0.56 *	0.70 **	0.65 **
indefinite theme	-1.25 ***	-1.19 ***	-1.20 ***
transfer semantic class	0.05 ns	-0.30 ns	-0.29 ns
communication semantic class	-2.62 ***	-2.15 ***	-2.05 ***
future having semantic class	-1.36 **	-1.32 **	-1.29 **
length difference	-0.91 ***	-0.91 ***	-0.93 ***
verb similarity to <i>give</i> (LSA cosines)	-	-3.54 ***	-3.58 ***
verb caused-possession entailment	-	-1.20 ***	-1.28 ***
nonpron recipient : nonpron theme	-	-	1.88 ***

Significance: ns $p > .05$, * $p < .05$, ** $p < .01$, *** $p < .001$, [Ditransitive = 0, Prepositional = 1]

The results show that semantic similarity to *give* and the entailment of caused possession (in shade in Table 4) make an independent contribution above and beyond the verb-external predictors Bresnan et al. discussed. A higher degree of semantic similarity to *give* and the invariable presence of a caused-possession entailment increase the likelihood of occurrence in the ditransitive frame and decrease the likelihood of occurrence in the prepositional frame. Comparing A and B, all the coefficients and significances of verb-external predictors remain essentially the same whether or not semantic similarity to *give* and an invariable caused-possession entailment are included as predictors. Comparing B and C, the addition of an interaction effect made the main effect of the (non-)pronominal recipient no longer significant while it does not impact the effect of our semantic predictors, suggesting that semantic similarity to *give* is orthogonal to the effect of verb-external factors. Model comparison showed that the most complex model (C) is the best-fitting model ($\Delta AIC = 0$, $weight = 1$), when compared to the replication (A) of Bresnan et al. ($\Delta AIC = 29.3$, $weight < 0.001$) or the model with two semantic predictors (B) ($\Delta AIC = 21.5$, $weight < 0.001$).⁷

Summarizing the results of all models reported on in this section, the effect of semantic similarity to *give*, our hypothesized anchor for the ditransitive frame, makes a unique contribution to the prediction of choice of syntactic frame, a contribution that is independent of the effects of (verb-external) properties of post-verbal arguments reported in previous literature.

5.2. Combining verb-external and verb-internal factors with a prepositional frame biased anchor

Sections 3, 4, and 5.1 tested the Verb Anchor Hypothesis when *give* is the anchor to the ditransitive frame. But, of course, the Verb Anchor Hypothesis predicts that an anchor to the prepositional frame, if any, should conversely attracts verbs similar to it to the prepositional

⁷ We did not replicate the effects of semantic classes Bresnan et al. reported. Note that Bresnan et al.'s semantic classes are not based on verb meaning, but rather the interpretation of each verb in each sentence context. They are thus quite distinct from the semantic classes we consider in this paper which center on the meaning of verbs and correspond to the notion of semantic classes discussed in Pinker (1989), Levin (1993), and Rappaport and Levin (2008). In Bresnan et al.'s dataset, a single verb is annotated differently depending on its sense in a particular sentence. In our model A, the *future having class* predictor has a negative coefficient (increasing the likelihood of the choice of the ditransitive frame) rather than the positive coefficient Bresnan et al. reported. This results is actually what we expected since future having verbs always entail caused possession, which we showed promotes ditransitive frame use. Second, the effect of their *transfer semantic class* turned out not to be significant in our model. This difference seems to be due to the fact that our model excludes sentences with the anchor verb *give* as main verb. Post hoc analyses revealed that the *transfer semantic class* predictor is a significant predictor for sentences that contain *give* ($b = 1.48$, $z = 5.02$, $p < .001$), while it is not for other verbs.

frame. It is this converse prediction we now test. Since we limit the choice of frames to the ditransitive and prepositional frames (to ensure that the verb sense involved stays relatively constant across frames), the verb least associated with the ditransitive frame (i.e. the lowest ranked verb in the association strength table in the appendix) is also the verb most strongly associated with the prepositional frame. This is the verb *bring* in our ΔP Attraction and Hebbian measures ($\Delta P_{D \rightarrow \text{bring}} = -0.112$; $w_{\text{bring}D} = -4347$). Before discussing our study of *bring* as a putative anchor for the prepositional frame, though, we should point out that there are important differences in verb frequency distributions between the ditransitive and prepositional frames.

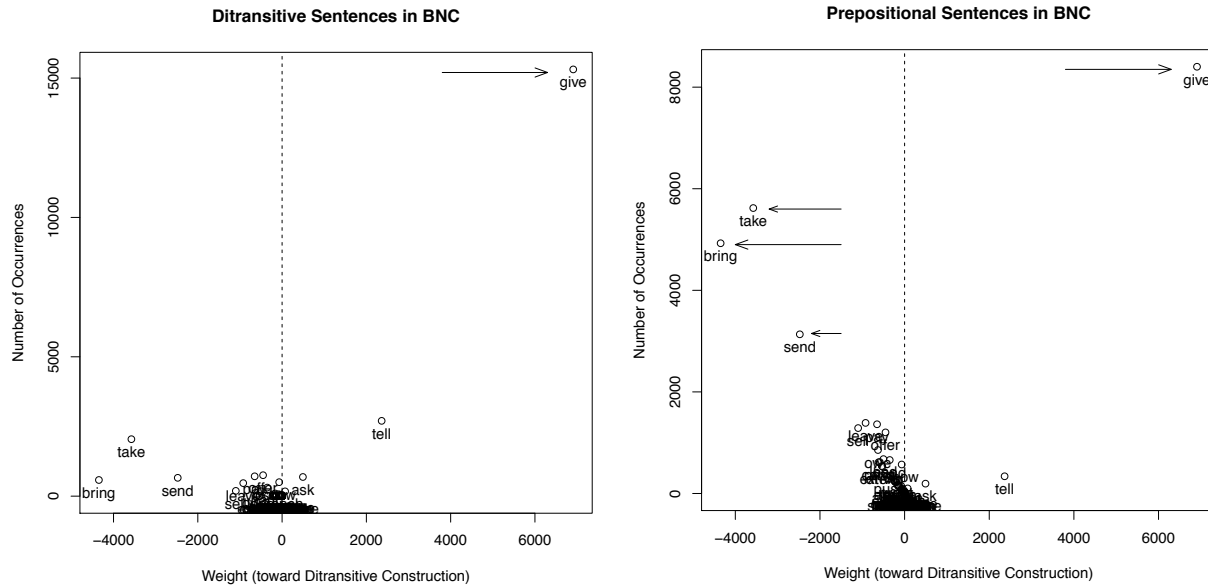


Figure 2. Plots of Hebbian-based measures of association strength between verbs and syntactic frame; (left) exemplar verbs in the ditransitive frame; (right) exemplar verbs in the prepositional frame.

As illustrated in Figure 2, the verb *give* stands out as by far the verb most strongly associated with the ditransitive frame as there are no other competing verbs that frequent and strongly biased towards the ditransitive frame. In contrast, the association between *take* ($\Delta P_{D \rightarrow \text{take}} = -0.075$; $w_{\text{take}D} = -3576$) or *send* ($\Delta P_{D \rightarrow \text{send}} = -0.060$; $w_{\text{send}D} = -2476$) and the prepositional frame is not that much lower than that of *bring*. Given the similarity of *bring*, *take*, and *send* in

their degrees of association to the prepositional frame, all three verbs, separately or together, may serve as anchors for the prepositional frame (see the General Discussion).

We then tested whether the more semantically similar to *bring* a verb is, the more likely it is to occur in the prepositional frame. Given the effect of *give* as an anchor to the ditransitive frame, we fitted a model to our BNC data that included these two semantic similarity predictors, i.e. verb similarity to *bring* (just as with our previous models, we used the past tense form *brought* in the computation of LSA cosines and LSA cosines were residualized over frequency) and verb similarity to *give*, and the four other predictors we discussed in Section 5.1, i.e. caused-possession entailment irrespective of syntactic frame, pronominality of recipient, pronominality of theme, and recipient-theme length difference. The results of this new regression analysis show that all six predictors are independently significant predictors of the choice of syntactic frame, i.e. verb similarity to *bring* ($b = -0.71, z = -15.47, p < .001$), verb similarity to *give* ($b = 1.24, z = 26.26, p < .001$), entailment of caused possession irrespective of syntactic frame ($b = 0.32, z = 17.36, p < .001$), pronominality of recipient ($b = 1.29, z = 69.62, p < .001$), pronominality of theme ($b = -1.14, z = -42.39, p < .001$), and recipient-theme length difference ($b = 0.85, z = 29.86, p < .001$). Importantly, in line with the predictions of the Verb Anchor Hypothesis, semantic similarity to *bring* decreases the likelihood of a sentence occurring in the ditransitive frame and increases the likelihood of a sentence occurring in the prepositional frame (see the *negative* coefficient estimate), while semantic similarity to *give*, as was shown previously, does the opposite. We further examined whether there is an interaction between the two verb similarity predictors and found a significant interaction ($b = -0.15, z = -9.66, p < .001$): Semantic similarity to *bring* increases the likelihood of speakers choosing the prepositional frame only when the main verb is not highly similar to *give*. By model comparison, this more complex model better fits the data than a model without the interaction term ($\Delta AIC = 98.9, weight < 0.001$). Although the results of this model confirm our predictions, we think they should be treated with caution since the frequency distribution of verbs in the prepositional frame is more complex than the distribution of verbs in the ditransitive frame: There exist several plausible anchors that may attract semantically similar verbs towards the prepositional frame.

6 Testing the Verb Anchor Hypothesis within narrow semantic classes

Until now, we have investigated our Verb Anchor Hypothesis for the entire set of verbs listed in Levin (1993) as alternating between the ditransitive and prepositional frames. In doing so, we treated the verb *give* as an anchor for verbs that rather loosely share the semantic notion of ‘transfer’ when they occur in those frames. But previous research has shown that when trying to predict which verb can or cannot alternate between two syntactic frames, the “right” level of semantic abstraction may be smaller than the entire class of alternating verbs. Pinker (1989), in particular, has argued that a verb’s ability to alternatively occur in the ditransitive and prepositional frames is conditioned by whether its meaning instantiates a meaning common to a *narrow* class of verbs. It is within these semantically coherent narrow classes, Pinker argues, that the dative alternation is productive and it is the nature of those narrow classes that children must acquire. Similarly, Goldberg (1995) has argued that the ditransitive construction is associated with closely related meanings (or is polysemous) and that these various meanings arise out of the interaction between the construction and the verbs that occur in it. It has also been argued that some argument structure constructions, e.g. the conative construction (*The horse pulled at the cart*), cannot be defined at the highest possible level of abstraction but must be defined at relatively lower levels of schematization that are assigned more specific meanings (Perek and Lemmens 2010; Perek 2014). These studies suggest that narrower and more semantically coherent subclasses of alternating verbs play an important role in the mental representation of alternations.

If narrow classes play a role in the representation of alternations, our Verb Anchor Hypothesis should also apply to narrow classes. Verbs within each narrow class share more specific semantic information than a rather loosely-defined notion of ‘transfer.’ For example, verbs such as *tell*, *teach* and *write*, and other verbs in the *message transfer* class all describe events in which an abstract message is metaphorically transferred. Verbs such as *offer*, *promise* and *bequeath*, and other verbs in the *future having* class share the semantic property that the transfer will occur in the future. Just as we argued *give* serves as the ditransitive anchor for the entire broad class of alternating verbs, there may be a highly frequent and typical verb within each narrow class that is strongly associated with either frame. This possibility is particularly likely if the ditransitive frame is associated with a slightly different meaning or sense across narrow

classes, as Goldberg argues. To draw an analogy from natural categories, there can be a typical exemplar of the bird category and another for the subordinate eagle category. We call such verbs, if they exist, *narrow class verb anchors*. Our hypothesis is that, as was the case for *give* and the entire class of alternating verbs, the degree of semantic similarity to a narrow class anchor affects which syntactic frame other narrow class members occur in. We test this hypothesis on three narrow classes, the *give* class, the *message transfer* class, and the *future having* class.

The first step in testing the Verb Anchor Hypothesis on narrow verb classes is to select a narrow class anchor for each of the three narrow classes. We examined the frequency distribution of verbs within each narrow class, where membership in the class was based on Levin (1993). (We manually changed the membership for the verb *hand*. It is listed as a *send* verb in Levin 1993, but its meaning as well as its usage in the corpus indicate that it is more appropriate to assign it to the *give* class than to the *send* class.) The distributions are illustrated in Figure 3.

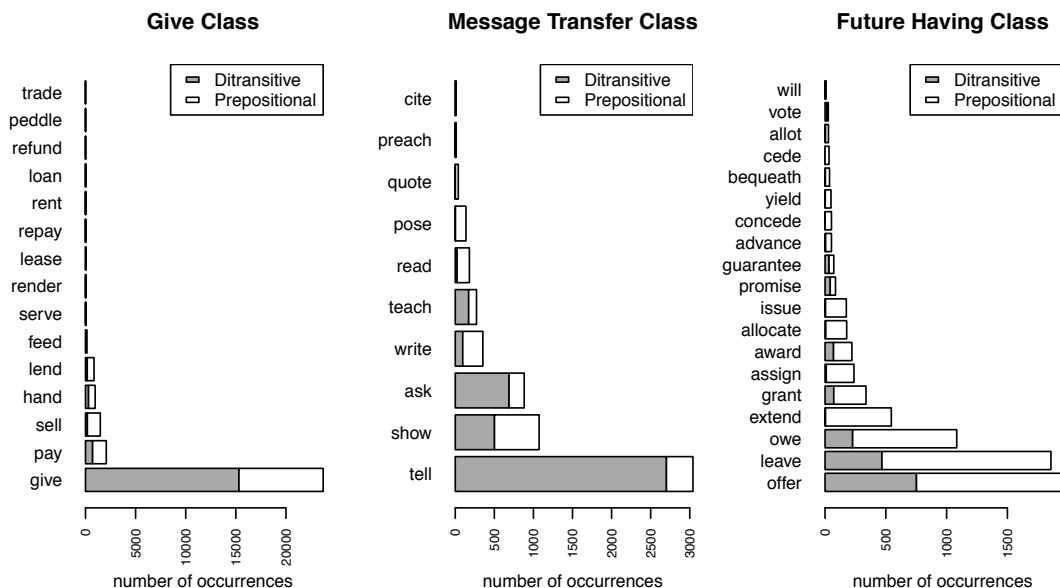


Figure 3. Verb frequencies in three narrow classes

Within the *give* class, as one can expect, *give* has the highest strength of association with the ditransitive frame and was selected as anchor. For this narrow class, we tested how a verb's semantic similarity to *give* predicts the choice of ditransitive frame for a smaller set of verbs than

in our previous models. In the *message transfer* class, the verb *tell* has the highest strength of association with the ditransitive frame and was selected as anchor. In the *future having* class, none of the frequent verbs favors the ditransitive frame and most are highly biased towards the prepositional frame. It is interesting to speculate on why verbs in the future having class favor the prepositional frame, particularly since they entail caused possession in both frames (Rappaport and Levin 2008). Our tentative answer is that the ‘goal’ meaning associated with the preposition *to* may better match the future component of the verbs’ meaning, i.e. the fact that the transfer occurs in the future. Be that as it may, we chose as anchor the verb with the highest strength of association with the prepositional frame, namely the verb *leave*. Our prediction, here, is that high semantic similarity to *leave* increases the likelihood of a verb occurring in the prepositional frame.

Having selected anchor verbs, we fitted the same kind of logistic regression models used in previous sections to each of the three narrow classes. We sorted the entire set of sentences extracted from the BNC into three separate datasets, one for each set of narrow class verbs. We then measured semantic similarity between narrow class members and their respective narrow class anchors using the same procedure as before (see Section 3 for details), i.e. *give* vs. *give* verbs, *tell* vs. message transfer verbs, and *leave* vs. future having verbs. This measure of semantic similarity was added as predictor to our former predictors, namely semantic similarity to *give* as broad-range or ‘global’ anchor and the three verb-external predictors included in the BNC-based models discussed in Section 5, i.e. pronominality of the recipient, pronominality of the theme, and length difference between the phrases that express the recipient and theme arguments. The entailment of caused possession irrespective of syntactic frame cannot be used as predictor since all the verbs in these narrow classes analyses entail caused possession irrespective of syntactic frame (Rappaport and Levin 2008).

The goal of our analyses is to determine whether semantic similarity to narrow class anchors affects the choice of syntactic frame independently of the contribution of those other predictors. We can test only three predictors for the *give* class, however, as the verb *give* is the narrow as well as broad class anchor. But for the message transfer and future having classes, we basically fitted two logistic regression models to the data, one with four predictors (a replication of our results reported in Section 5 within each narrow class as a baseline) and the other with semantic similarity to the narrow class anchor as an additional predictor. When necessary, we

also fitted more models to test interactions between the two similarity predictors and performed standard model selection procedures. The verb-external predictors, *pronominality of recipient*, *pronominality of theme*, and *theme-recipient length difference*, are included in all regression models reported in this section, and they all turned out to be significant predictors. We are not reporting numbers for these predictors as they are not crucial for our purposes here.

The *give* class data set consists of 5,731 sentences. 13 distinct verbs occur in these sentences: *pay*, *sell*, *hand*, *lend*, *feed*, *serve*, *lease*, *repay*, *loan*, *rent*, *refund*, *peddle*, and *trade*. As before, sentences that contain the anchor verb *give* were excluded from the analysis. For this class, we can only test whether previous findings can be replicated within this smaller set of verbs as the narrow class anchor is the same as the anchor for the entire set of alternating verbs. This model showed that all four predictors make independent contributions to predicting the choice of syntactic frame, i.e. verb similarity to *give* ($b = 0.19$, $z = 4.67$, $p < .001$). Not too surprisingly, and in line with our Narrow Verb Anchor Hypothesis, the results show that *give* serves as anchor to the narrow *give* class just as it did for the broad class of alternating verbs.

The data set for the *message transfer* class includes 2,960 sentences. 9 non-*tell* verbs occur in these sentences: *show*, *ask*, *write*, *teach*, *read*, *pose*, *quote*, *preach*, and *cite*. We ran two models, one with and the other without semantic similarity between *tell* and other class members as a predictor. When both similarity to *give* and similarity to *tell* are entered in the analysis, the similarity to the narrow class anchor *tell* turns out not to be a significant predictor ($b = 0.24$, $z = 1.35$, $p = .18$) while similarity to *give* is significant ($b = 0.85$, $z = 4.70$, $p < .001$). Moreover, adding similarity to *tell* to the model ($\Delta AIC = 0.2$, $weight = 0.48$) does not improve the model performance compared to a model without this predictor ($\Delta AIC = 0$, $weight = 0.52$). This suggests that what we defined as the narrow class anchor for message transfer verbs, namely *tell*, does not make a unique contribution to predicting the choice of syntactic frame for message transfer verbs when semantic similarity to *give* is also a predictor. To investigate the reasons for this outcome, we fitted another model with similarity to *tell* rather than similarity to *give* as the sole semantic similarity predictor. In this case, similarity to *tell* is a significant predictor of choice of syntactic frame ($b = 1.03$, $z = 14.80$, $p < .001$). Semantic similarity to *tell* is, thus, only predictive of syntactic frame selection when semantic similarity to *give* is not included in the model. The cause of the absence of a unique contribution of semantic similarity to *tell* seems to be the extremely high correlation between semantic similarity to *tell* and semantic similarity to *give*.

(Pearson's $r = .94, p < .001$), as indicated by higher VIFs (VIF = 7.18 for similarity to *give* and 6.74 for similarity to *tell*) than when one similarity predictor is included in the model (VIF = 1.17 for similarity to *give* and 1.43 for similarity to *tell*).⁸ Moreover, *give* and *tell* have very similar frequency profiles: Both occur far more frequently than other class members and both are highly associated with the ditransitive frame. This is why semantic similarity to *give* and semantic similarity to *tell* play a similar predictive role for message transfer verbs. Semantic similarity to *give* is a better predictor when both similarities compete, though, as shown by the relative magnitude of the coefficients between similarity to *give* and similarity to *tell* in models where they are considered as sole semantic similarity predictors. This is expected as *give* is more strongly associated with the ditransitive frame than *tell* as it occurs in the ditransitive frame almost six times more often than *tell*.

The future having class data set consists of 5,149 sentences. 17 different verbs occur in these sentences, when the narrow class anchor *leave* is excluded: *offer, owe, extend, grant, assign, award, allocate, issue, promise, guarantee, advance, concede, yield, bequeath, cede, allot, and will*. We fitted a model with both semantic similarity to *give* and semantic similarity to *leave* as predictors as well as the other three verb-external predictors we included in other models. *Give* and *leave* are expected to pull speakers in opposite directions. We replicated the positive coefficient of the semantic similarity to *give* predictor ($b = 0.79, z = 3.72, p < .001$) but the negative coefficient of the semantic similarity to *leave* predictor was not significant ($b = -0.23, z = -1.09, p = .28$). When their interaction was additionally taken into account, we found the effect of similarity to *give* to be robust and to significantly interact with similarity to *leave* ($b = 0.18, z = 0.07, p < .01$). Model comparison shows that this complex model ($\Delta AIC = 0, weight = 0.93$) is better fitting the data than the previous model ($\Delta AIC = 5, weight = 0.07$). A post hoc examination of this interaction revealed that high similarity to *leave* slightly increases the likelihood of selecting the prepositional frame for verbs of low-similarity to *give* but slightly increases the likelihood of the ditransitive frame for verbs of high-similarity to *give*. The results suggest that what we expected from an anchor of the prepositional frame, *leave*, may take effect only for verbs that are less

⁸ Variance inflation factor (VIF) indicates how much multicollinearity there is between predictors in a regression analysis and is standardly used to determine whether there is too high a correlation between predictors. A less than 10 VIF value is often considered tolerable, while some argue that 5 or 6 is the maximum. It is best for all predictors to have VIFs close to 1. Note that since all our predictors were centered, any multicollinearity between our predictors in this analysis and in analyses reported in Sections 5 and 6 is essential in the sense of Cohen et al. (2003).

similar to *give*. In this case, our Narrow Verb Anchor Hypothesis is borne out only when its relationship with the effect of the broad-class anchor is taken into account.

We investigated in this section whether there can be anchors that are specific to more semantically cohesive classes of verbs than the broad class of verbs that alternate between the ditransitive and prepositional frames. We showed that *tell* which we tested as the narrow-class ditransitive anchor for message transfer verbs plays a role in predicting syntactic frame selection only when the broad-class anchor *give* is not included in the model as its effect highly overlaps with that of *give*, as evidenced by the extremely high correlation between semantic similarity to *give* and semantic similarity to *tell*. We also showed that the verb *leave* which we tested as an anchor of the prepositional frame for future having verbs plays a role that counteracts the role of *give* for verbs that are less semantically similar to *give* and thus less affected by the strong effect of *give*. Our results suggest that small-scale narrow class anchors can play a role in syntactic frame selection independently of or cooperatively with broad class anchors.

7 General Discussion

This paper investigated the consequences on sentence production of the highly skewed distribution of verbs that occur in a particular syntactic frame. Drawing insight from the categorization literature and the notion of exemplars, we treated syntactic frames or constructions as categories and sentences occurring in those frames, more specifically verbs that occur in sentences, as exemplars. We hypothesized that frequent experience with exemplar sentences where a particular verb occurs in a particular syntactic frame may lead to a strong association between the verb and the frame, analogous to the association between a category and its best exemplar (e.g. Medin and Schaffer 1978). Categorization research showed that similarity to the best exemplar is a critical factor for an entity to be considered a member of the category the best exemplar belongs to. We hypothesized that the “best exemplar” verb of a syntactic frame constitutes a lexical anchor for that frame, so that how semantically similar another verb is to the anchor affects speakers’ choice of syntactic frame. Our Verb Anchor Hypothesis claims that high semantic similarity to a syntactic frame’s anchor tends to lead speakers to choose the same syntactic frame.

We tested the Verb Anchor Hypothesis by investigating sentences that instantiate the dative alternation, i.e. sentences that include, as main verbs, verbs that can occur either in the ditransitive frame or in the prepositional frame. Since the choice of syntactic frame in the dative alternation is known to be influenced by a variety of factors, not only did we investigate the role of a highly frequent and typical verb in syntactic frame selection, we also made sure any effect of a syntactic frame's anchor was independent from the effect of other factors known to influence syntactic frame selection. We conducted a series of logistic regression analyses on ditransitive or prepositional sentences collected from the British National Corpus. Our main predictor was semantic similarity between the ditransitive anchor verb *give* and other alternating verbs, as measured by Latent Semantic Analysis. To summarize our results, semantic similarity to the anchor *give* is a significant predictor of syntactic frame selection and the effect of this predictor is not reducible to that of other factors such as the presence of a caused possession entailment in both frames or pronominality of postverbal arguments. Overall, our results confirm our hypothesis that semantic similarity to the verb most highly associated with a syntactic frame modulates syntactic frame selection in sentence production.

We also tested our Verb Anchor Hypothesis on semantically more cohesive narrow classes like those discussed in Pinker (1989) and showed that narrow class anchors can also play a role in predicting syntactic frame selection. Of particular note is the fact that narrow class anchors may counteract the effect of broad class anchors: *Leave* pulls future having verbs in the opposite direction than *give* does, i.e. *leave* biases speakers towards choosing the prepositional frame rather than the ditransitive frame.

Many researchers have pointed out that verbs occurring particularly frequently in a certain syntactic frame tend to have a very general meaning (e.g. *go*, *give*, and *put*) that may be quite similar to the putative abstract meaning of the syntactic frame. Pinker (1989, p.212), for example, suggests that "[*give*'s semantic] representations are virtually identical to the double-object thematic core" and Goldberg (1997, p. 386) claims that "[*give*] codes an elaboration of the meaning of the [ditransitive] construction." One may therefore argue that the results of the present study are due to the similarity between the meaning of *give* and that of the ditransitive frame rather than their high co-occurrence frequency. Since *give* occurs both highly frequently in the ditransitive frame and its meaning is indeed quite general and not much more than the notion of transfer of possession many assign to the ditransitive frame, teasing apart these two

hypotheses is hard. The results of our analysis of the anchoring role of *bring* and of our narrow class analyses, though, suggest that the mechanisms underlying the effect of verb anchors amount to more than the semantic generality of verbs like *give* and provide support for the effect of frequency of occurrence of verbs in particular frames. Both *bring* and *leave* pull class members towards the prepositional frame, not towards the ditransitive frame as *give* does. Since it is not the case that “the meaning of the prepositional frame” is roughly the same as that of either *bring* or *leave*, verb anchoring does not seem to reduce, in this case, to semantic similarity between the meaning of the anchors and the meaning of the frame or to the fact that verbs similar in meaning to the anchors would be verbs whose meanings are most compatible with that of the frame. Rather, we surmise, the effect of verb anchors arises from the lexical similarity between alternating verbs and the verb most strongly associated with a syntactic frame because of the latter’s high frequency of occurrence in that frame.

The results of our narrow class models also suggest that our focus on a *single* anchor for the ditransitive frame might be an oversimplification of the role of semantic similarity in sentence production. Verbs overlap in meaning with many verbs, not just the verb that occurs most frequently in a syntactic frame: The verb that best expresses the situation category that a speaker wishes to communicate activates many other verbs. Each of these verbs activates the syntactic frames they occur in. The way in which semantic overlap increases the activation of a syntactic frame is thus probably the result of a complex interaction between the weighted activation of other verb meanings and the weighted association of each of these verbs with the relevant syntactic frames (see Chang et al. 2006 for an approach to syntactic frame selection broadly compatible with the view we are articulating). Talking of a frame’s anchor as we have done throughout this paper is nonetheless justified, we believe, for two reasons. It is justified theoretically by previous claims about the importance of verbs that occur most frequently in syntactic frames in the acquisition (e.g. Goldberg et al. 2004; Ellis et al. 2014) or representation (e.g. Pinker 1989; Goldberg 1995; Yoon 2013) of argument structure constructions. Second, and more practically, the effect of most verbs other than anchors may be too small to be measurable, as the strength of their association with a syntactic frame may be quite weak. Starting the investigation of the influence of semantic similarity on syntactic frame selection with most frequently occurring verbs is, we believe, the wisest course of action.

The claim that the activation of a syntactic frame results from the activation of all verbs that overlap in meaning with the verb that best expresses the speaker's message resembles the view put forth in Ambridge et al. (2014) that the meaning associated with the ditransitive construction incorporates the meaning of all the verbs that occur in the construction weighted by their frequency of occurrence in the construction. Ambridge et al. (2014, p. 238) suggest that "the properties of a particular slot [in a syntactic frame] are a weighted average of the properties shared by the items that have appeared in this position in the input utterances that gave rise to the construction." Similarly, Ellis et al. (2014, p. 58) proposed "a VAC [verb-argument structure construction] inherits its schematic meaning from the constituency of all of the verb exemplars experienced within it, weighted according to their frequency of experience."

There are some important differences, however, between these claims and the view we just articulated: Both the behavior Ambridge et al. are trying to explain and the predictors of that behavior differ from the focus of our study. Ambridge et al.'s models are meant to study the effect of preemption on argument structure learning and explain the relative acceptability on a five-point scale of sentences that contain alternating and non-alternating verbs. Our study is meant to explain speakers' *choice* of syntactic frame in naturally occurring English sentences. So, whereas Ambridge et al.'s empirical study focuses on metalinguistic behavior (acceptability judgments), our study focuses on syntactic frame selection in sentence production. Not only is the behavior to be explained different, so are the predictors. Frequency of occurrence is a critical *predictor* of behavior in their study (since they focus on preemption as an explanation for the absence of overgeneralization); frequency of occurrence in one frame or the other is what we are trying to predict (either directly through a correlation analysis or indirectly through logistic regressions explaining the choice of frame for each of our 39,000 dative sentences, excluding *give* sentences). Their semantic predictor of interest and our semantic predictors of interest are also critically distinct. Our semantic predictor of interest is the semantic *similarity* between verb meanings (in particular the meaning of the main verb used in a sentence and the meaning of the verbs that occur the most frequently in a syntactic frame). Ambridge et al.'s semantic predictors of interest are the semantic features Pinker (1989) thought important for the dative alternation. Finally, our model, as mentioned in the introduction, does not assume that syntactic frames have meanings. Nor does it rely on a theory of meaning that probabilistically include features from the meaning of all verbs that occur in the syntactic frame. Our hypothesis relies on the

more conservative view that lexical meaning overlap is what leads to the activation of other concepts, an assumption critical to explaining semantic priming effects (McRae and Boisvert 1986). Despite those differences, however, our two main claims are consonant with what we take to be Ambridge et al.'s main intuition: (1) Each verb meaning is associated with a syntactic frame as a monotonically increasing function of their frequency of occurrence in that frame, (2) the overlap in meaning between the situation category a speaker wishes to express and the meaning of other verbs occurring in that frame mediates the effect those associations have on the choice of syntactic frame.

The last issue we would like to tackle is whether the effect of anchor verbs on syntactic frame selection is a peculiarity of the dative alternation or can be generalized to other syntactic frames. As one reviewer pointed out, other verbs have been argued to be typical exemplars of argument structure constructions, e.g. *put* for the 'caused-motion' construction. The Verb Anchor Hypothesis predicts that such verbs should attract semantically similar verbs to occur in the construction they are typical exemplars of. The results of a model we ran on the 45 verbs listed in Levin (1993) as participating in the locative alternation suggest that this prediction is borne out. After retrieving from the BNC all locative (*on(to)*, *in(to)*, *around*, etc.) and *with* PP tokens of the 45 alternating verbs, we ran a logistic regression similar to the first model we presented in Section 3 where the locative PP frame was coded as 0 and the *with* PP frame as 1 (e.g. *John sprayed the powder on the wall* vs. *John sprayed the wall with the powder*). Semantic similarity to *put* was indeed a significant predictor of the occurrence of verbs in the locative PP variant ($b = -1.658$; $p < .001$). Despite this positive result, whether the Verb Anchor Hypothesis applies to most alternations or just a few depends on several factors whose effect we cannot at present ascertain and is thus a matter for further study. As Sun and Koenig (2017) and Sun (2018) point out, the dative-alternation verb frequency distribution is rather unique among English verb alternations. It is the only alternation (or construction) where a single verb, *give*, accounts for such a large proportion of the syntactic frame's tokens. *Give* might be uniquely strongly associated with the ditransitive frame and the weaker association of putative anchors to other frames might decrease the likelihood of finding an effect of semantic similarity to the anchor verb's meaning.

Appendix

A.1 Verb frequencies, association strengths, and residualized LSA cosines

The table below lists *give* and the 104 verbs we used in our analyses and provides the following information for each verb: (i) number of occurrences in the ditransitive and prepositional frame in the British National Corpus (D:P), (ii) verb class defined by Levin (1993), (iii) Hebbian association strength with the ditransitive frame (W_{verbD}), (iv) ΔP Attraction ($\Delta P_{D \rightarrow V}$) value, (v) p values from the Fisher Exact Tests when the target construction is the ditransitive frame (D) and the competing construction is the prepositional frame (P), (vi) semantic similarity to *give* estimated by cosines from the Latent Semantic Analysis (LSA) and (vii) LSA cosines residualized over frequency (D+P) (see Section 2 for a discussion of these measures and Section 3 for a discussion of LSA and residualization). The Hebbian association strength is computed by subtracting the number of a verb's occurrences in the prepositional frame from its occurrences of the ditransitive frame, as the constant learning rate is the same across all verbs. Below the table we provide the assumptions we are making in computing this association strength measure. See Ellis and Ferreira-Junior (2009) and Stefanowitsch and Gries (2003) for more details about the computation and application of ΔP and Fisher Exact test (see Schmid and Küchenhoff 2013 for a review of those and other measures).

No	Verb (D:P)	Verb class	W_{verbD}	$\Delta P_{D \rightarrow V}$	p of Fisher's	LSA cosine	LSA cosine (residualized)
1	give (15311:8402)	<i>Give</i>	6909	0.35879	0.00E+00	1.000	-0.311
2	tell (2702:339)	<i>TrsMsg</i>	2363	0.09451	0.00E+00	0.859	0.311
3	ask (688:194)	<i>TrsMsg</i>	494	0.02113	1.61E-109	0.857	0.389
4	teach (172:100)	<i>TrsMsg</i>	72	0.00388	6.02E-13	0.464	0.018
5	loan (12:11)	<i>Give</i>	1	0.00016	4.62E-04	0.190	-0.247
6	e-mail (1:0)	<i>Instr</i>	1	0.00004	1.25E-01	0.325	-0.111
7	promise (43:43)	<i>FutHav</i>	0	0.00048	4.74E-01	0.767	0.328
8	barge (0:1)	<i>Drive</i>	-1	-0.00003	3.00E-01	0.431	-0.005
9	bash (0:1)	<i>Throw</i>	-1	-0.00003	4.15E-01	0.348	-0.088
10	bat (0:1)	<i>Throw</i>	-1	-0.00003	1.00E+00	0.127	-0.309
11	cable (0:1)	<i>Instr</i>	-1	-0.00003	1.00E+00	0.249	-0.187
12	heft (0:1)	<i>Carry</i>	-1	-0.00003	1.00E+00	0.099	-0.337
13	slap (0:1)	<i>Throw</i>	-1	-0.00003	1.00E+00	0.581	0.145

No	Verb (D:P)	Verb class	W_{verbD}	ΔP D->V	p of Fisher's	LSA cosine	LSA cosine (residualized)
14	sneak (0:1)	<i>Send</i>	-1	-0.00003	1.00E+00	0.514	0.078
15	chuck (1:3)	<i>Throw</i>	-2	-0.00004	1.00E+00	0.315	-0.121
16	radio (0:2)	<i>Instr</i>	-2	-0.00005	1.00E+00	0.220	-0.216
17	shuttle (0:2)	<i>Drive</i>	-2	-0.00005	6.46E-01	0.243	-0.193
18	sling (0:2)	<i>Throw</i>	-2	-0.00005	5.14E-01	0.503	0.067
19	lob (1:4)	<i>Throw</i>	-3	-0.00007	5.14E-01	0.212	-0.224
20	telegraph (0:3)	<i>Instr</i>	-3	-0.00008	5.14E-01	0.279	-0.157
21	bounce (0:4)	<i>Slide</i>	-4	-0.00011	4.1E-01	0.542	0.106
22	lug (0:4)	<i>Carry</i>	-4	-0.00011	2.72E-01	0.375	-0.061
23	trade (0:4)	<i>Give</i>	-4	-0.00011	1.47E-01	0.141	-0.295
24	phone (0:5)	<i>Instr</i>	-5	-0.00014	1.47E-01	0.455	0.019
25	row (0:5)	<i>Drive</i>	-5	-0.00014	1.47E-01	0.322	-0.114
26	guarantee (33:39)	<i>FutHav</i>	-6	0.00020	8.07E-02	0.385	-0.054
27	fax (2:8)	<i>Instr</i>	-6	-0.00014	8.07E-02	0.271	-0.165
28	slam (0:6)	<i>Throw</i>	-6	-0.00016	2.11E-01	0.552	0.116
29	tow (0:6)	<i>Carry</i>	-6	-0.00016	4.19E-01	0.231	-0.205
30	wire (0:6)	<i>Drive/Instr</i>	-6	-0.00016	2.61E-01	0.238	-0.198
31	mail (3:10)	<i>Send</i>	-7	-0.00016	4.51E-02	0.162	-0.274
32	will (1:8)	<i>FutHav</i>	-7	-0.00018	4.51E-02	0.501	0.065
33	cart (0:8)	<i>Drive</i>	-8	-0.00022	4.51E-02	0.197	-0.239
34	catapult (0:8)	<i>Throw</i>	-8	-0.00022	9.03E-02	0.288	-0.148
35	float (0:8)	<i>Slide</i>	-8	-0.00022	2.45E-02	0.607	0.171
36	smuggle (0:8)	<i>Send</i>	-8	-0.00022	2.45E-02	0.367	-0.069
37	wheel (0:8)	<i>Drive</i>	-8	-0.00022	2.45E-02	0.495	0.059
38	flick (8:17)	<i>Throw</i>	-9	-0.00016	2.45E-02	0.570	0.133
39	telephone (0:9)	<i>Instr</i>	-9	-0.00025	2.45E-02	0.507	0.071
40	preach (1:11)	<i>TrsMsg</i>	-10	-0.00026	2.66E-01	0.284	-0.152
41	peddle (0:10)	<i>Give</i>	-10	-0.00027	1.30E-02	0.235	-0.201
42	pitch (0:10)	<i>Throw</i>	-10	-0.00027	1.91E-02	0.437	0.001
43	cite (0:11)	<i>TrsMsg</i>	-11	-0.00030	6.97E-03	0.278	-0.158
44	refund (0:11)	<i>Give</i>	-11	-0.00030	6.97E-03	0.071	-0.365
45	signal (0:11)	<i>Instr</i>	-11	-0.00030	3.77E-03	0.090	-0.346
46	tug (0:11)	<i>Carry</i>	-11	-0.00030	3.77E-03	0.516	0.080
47	flip (0:12)	<i>Throw</i>	-12	-0.00033	3.77E-03	0.414	-0.022
48	shoot (13:27)	<i>Throw</i>	-14	-0.00024	3.77E-03	0.537	0.100
49	hurl (0:14)	<i>Throw</i>	-14	-0.00038	2.07E-03	0.448	0.012
50	shove (0:15)	<i>Carry/Throw</i>	-15	-0.00041	6.39E-04	0.491	0.055
51	repay (4:21)	<i>Give</i>	-17	-0.00042	3.61E-04	0.233	-0.204
52	heave (0:18)	<i>Carry</i>	-18	-0.00049	1.31E-02	0.517	0.080
53	rent (2:21)	<i>Give</i>	-19	-0.00050	1.08E-04	0.330	-0.107
54	allot (4:24)	<i>FutHav</i>	-20	-0.00050	1.01E-03	0.331	-0.106
55	hoist (0:22)	<i>Carry</i>	-22	-0.00060	3.29E-03	0.508	0.071
56	kick (1:27)	<i>Carry/Throw</i>	-26	-0.00070	9.40E-06	0.591	0.154

No	Verb (D:P)	Verb class	W_{verbD}	$\Delta P D \rightarrow V$	p of Fisher's	LSA cosine	LSA cosine (residualized)
57	tip (1:30)	Throw	-29	-0.00078	1.33E-01	0.615	0.178
58	fling (7:38)	Throw	-31	-0.00077	8.06E-06	0.635	0.197
59	bequeath (2:34)	FutHav	-32	-0.00085	3.67E-04	0.278	-0.159
60	forward (1:34)	Send	-33	-0.00089	1.57E-06	0.216	-0.221
61	cede (0:33)	FutHav	-33	-0.00090	1.71E-06	0.198	-0.239
62	hit (2:36)	Throw	-34	-0.00090	2.67E-07	0.603	0.166
63	quote (2:37)	TrsMsg	-35	-0.00093	2.44E-08	0.391	-0.046
64	serve (12:52)	Give	-40	-0.00096	1.04E-06	0.553	0.115
65	advance (5:48)	FutHav	-43	-0.00112	5.82E-07	0.398	-0.040
66	roll (0:43)	Slide	-43	-0.00117	1.80E-04	0.661	0.224
67	feed (52:96)	Give	-44	-0.00062	3.84E-07	0.561	0.120
68	yield (2:47)	FutHav	-45	-0.00120	1.14E-10	0.358	-0.080
69	slide (4:50)	Slide	-46	-0.00121	3.94E-09	0.597	0.159
70	ferry (0:48)	Drive	-48	-0.00131	3.53E-08	0.248	-0.190
71	sign (1:50)	Instr	-49	-0.00133	1.13E-11	0.388	-0.050
72	lease (0:49)	Give	-49	-0.00134	9.59E-11	0.162	-0.276
73	concede (1:52)	FutHav	-51	-0.00138	5.99E-12	0.548	0.110
74	post (1:55)	Send	-54	-0.00146	2.79E-11	0.428	-0.010
75	toss (6:63)	Throw	-57	-0.00149	3.19E-03	0.607	0.169
76	slip (7:67)	Send	-60	-0.00156	4.74E-12	0.695	0.256
77	show (502:571)	TrsMsg	-69	0.00370	1.98E-09	0.697	0.222
78	ship (0:69)	Send	-69	-0.00188	1.61E-09	0.266	-0.172
79	haul (0:73)	Carry	-73	-0.00199	1.38E-16	0.445	0.006
80	award (70:151)	FutHav	-81	-0.00143	1.20E-17	0.269	-0.175
81	pose (3:134)	TrsMsg	-131	-0.00354	2.64E-08	0.454	0.013
82	read (24:157)	TrsMsg	-133	-0.00336	1.43E-16	0.580	0.137
83	fly (0:143)	Drive	-143	-0.00390	2.00E-27	0.512	0.071
84	drag (0:158)	Carry	-158	-0.00431	6.82E-34	0.669	0.227
85	write (96:257)	TrsMsg	-161	-0.00332	5.52E-03	0.619	0.170
86	allocate (6:172)	FutHav	-166	-0.00446	1.59E-37	0.243	-0.199
87	issue (4:171)	FutHav	-167	-0.00451	2.29E-32	0.321	-0.121
88	grant (72:265)	FutHav	-193	-0.00446	5.44E-15	0.448	0.000
89	throw (25:222)	Throw	-197	-0.00509	2.55E-34	0.757	0.312
90	assign (10:228)	FutHav	-218	-0.00583	7.96E-28	0.294	-0.151
91	pull (0:231)	Carry	-231	-0.00630	1.89E-40	0.742	0.298
92	push (0:344)	Carry	-344	-0.00938	4.58E-10	0.755	0.306
93	hand (308:659)	Give	-351	-0.00614	1.37E-54	0.759	0.287
94	offer (752:1203)	FutHav	-451	-0.00393	5.56E-81	0.681	0.173
95	lend (177:677)	Give	-500	-0.01167	8.47E-12	0.493	0.026
96	drive (4:530)	Drive	-526	-0.01430	2.75E-38	0.686	0.230
97	extend (2:541)	FutHav	-539	-0.01468	9.03E-117	0.500	0.044
98	carry (0:615)	Carry	-615	-0.01677	4.18E-47	0.662	0.203
99	owe (227:856)	FutHav	-629	-0.01463	5.51E-123	0.462	-0.014

No	Verb (D:P)	Verb class	W_{verbD}	$\Delta P D \rightarrow V$	p of Fisher's	LSA cosine	LSA cosine (residualized)
100	pay (712:1363)	<i>Give</i>	-651	-0.00983	6.11E-145	0.555	0.043
101	leave (468:1390)	<i>FutHav</i>	-922	-0.01994	2.00E-50	0.857	0.353
102	sell (190:1288)	<i>Give</i>	-1098	-0.02783	4.49E-131	0.405	-0.085
103	send (658:3134)	<i>Send</i>	-2476	-0.06020	2.44E-236	0.765	0.189
104	take (2044:5620)	<i>Bring&take</i>	-3576	-0.07477	1.08E-182	0.946	0.227
105	bring (580:4927)	<i>Bring&take</i>	-4347	-0.11209	0.00E+00	0.891	0.252

Our Hebbian measure of association strength is based on two assumptions. First, we assume that the association between a verb and a syntactic frame is learned in a Hebbian manner (Hebb 1949), that is, that the change in the weight of the connection between a verb and a syntactic frame is proportional to the activations of the verb and the frame ($\Delta w = \eta a_v a_F$; a_v and a_F refer to the activation of the verb and syntactic frame, respectively, and η to a learning rate constant). Second, we assume that the two alternating frames are competing for selection (a verb cannot simultaneously occur in both frames), but that verbs are not competing with each other for selection (since we are measuring the bias of verbs towards each frame, once a message/meaning is selected, other verbs are irrelevant). The consequence of this second assumption is that once a verb and a frame are selected, the other frame is inhibited, but other verbs are merely inactive. Under these two assumptions, the weight of the connection between a verb and a frame is simply proportional to the difference between the number of times the verb occurs in the frame and the number of times it occurs in the competing frame ($w_{vF}(t) = \eta(t_1(v, +F) - t_2(v, -F))$), where F and $-F$ are the alternative frames, and F and $-F$ are encountered t_1 times and t_2 times, respectively ($t = t_1 + t_2$)).

A.2 Corpus data and model results tables

The table below provides the details of seven BNC-based datasets we used to verify our hypothesis, i.e. our original dataset [1] and six additional subsets [2~7]. It summarizes the number of ditransitive and prepositional sentences and the mean and standard deviation of number of sentences per verb in each dataset. Below the table, we also provide each model's result tables, i.e. the models including all possible interactions in A.2.1, the models including all

possible two-way interactions in A.2.2, and the models with five predictors but with no interactions in A.2.3, each fitted to the seven different datasets.

Dataset	Entire data	Subsets					
		[1]	[2]	[3]	[4]	[5]	[6]
# of verbs	104 verbs	103 verbs (- tell)	102 verbs (- tell, take)	101 verbs (- tell, take, bring)	100 verbs (- tell, take, bring, send)	104 verbs (max freq. = 100/verb)	57 verbs (max freq. = 100 & excl. D=0 verbs)
D sentences	10,732	8,030	5,986	5,406	4,748	851	851
P sentences	28,268	27,929	22,309	17,382	14,248	3,965	2,900
D + P total	39,000	35,959	28,295	22,788	18,996	4,816	3,751
D:P	28:72	22:78	21:79	24:76	25:75	18:82	23:77
# of sentences per verb mean sd	375 1072.5	350 1044.6	277 752.9	226 544.4	190 411.8	46 40.1	66 35.7

A.2.1 Result tables of all full-interaction models

Formula = str ~ crlsagave * cposs * cprorec * cprothm * cldiff
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

A.2.1.1 - Dataset: Entire data

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)	
(Intercept)	-1.820784	0.040825	-44.600	< 2e-16	***
crlsagave	0.413170	0.062914	6.567	5.13e-11	***
cposs	0.599246	0.039838	15.042	< 2e-16	***
cprorec	0.630950	0.063292	9.969	< 2e-16	***
cprothm	-1.168698	0.059569	-19.619	< 2e-16	***
cldiff	2.232568	0.169006	13.210	< 2e-16	***
crlsagave:cposs	0.044812	0.060758	0.738	0.460790	
crlsagave:cprorec	-0.005253	0.092533	-0.057	0.954725	
cposs:cprorec	0.289320	0.061495	4.705	2.54e-06	***
crlsagave:cprothm	-0.323442	0.074085	-4.366	1.27e-05	***
cposs:cprothm	0.176084	0.058018	3.035	0.002406	**
cprorec:cprothm	-0.934299	0.094481	-9.889	< 2e-16	***
crlsagave:cldiff	0.321915	0.236641	1.360	0.173718	
cposs:cldiff	-0.582639	0.163978	-3.553	0.000381	***
cprorec:cldiff	2.748211	0.296595	9.266	< 2e-16	***
cprothm:cldiff	1.947841	0.286786	6.792	1.11e-11	***
crlsagave:cposs:cprorec	-0.016013	0.089282	-0.179	0.857659	
crlsagave:cposs:cprothm	0.105644	0.071645	1.475	0.140335	
crlsagave:cprorec:cprothm	-0.430331	0.099695	-4.316	1.59e-05	***
cposs:cprorec:cprothm	0.130490	0.091657	1.424	0.154542	
crlsagave:cposs:cldiff	-0.203759	0.228281	-0.893	0.372083	
crlsagave:cprorec:cldiff	0.973250	0.401552	2.424	0.015363	*
cposs:cprorec:cldiff	-1.026519	0.287710	-3.568	0.000360	***
crlsagave:cprothm:cldiff	1.278512	0.376443	3.396	0.000683	***

cposs:cprothm:cldiff	-0.752836	0.278159	-2.706	0.006800	**
cprorec:cprothm:cldiff	3.744924	0.505061	7.415	1.22e-13	***
cr1sagave:cposs:cprorec:cprothm	0.297523	0.096398	3.086	0.002026	**
cr1sagave:cposs:cprorec:cldiff	-0.526827	0.387414	-1.360	0.173876	
cr1sagave:cposs:cprothm:cldiff	-0.822373	0.363247	-2.264	0.023577	*
cr1sagave:cprorec:cprothm:cldiff	2.584511	0.636567	4.060	4.91e-05	***
cposs:cprorec:cprothm:cldiff	-1.037716	0.489860	-2.118	0.034142	*
cr1sagave:cposs:cprorec:cprothm:cldiff	-1.743683	0.614421	-2.838	0.004541	**

A.2.1.2 - Dataset: Subset [1]

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)	
(Intercept)	-2.19495	0.04605	-47.668	< 2e-16	***
cr1sagave	0.30078	0.06604	4.555	5.25e-06	***
cposs	0.43414	0.04271	10.165	< 2e-16	***
cprorec	0.38158	0.07779	4.905	9.33e-07	***
cprothm	-1.13081	0.06588	-17.164	< 2e-16	***
cldiff	2.02201	0.16690	12.115	< 2e-16	***
cr1sagave:cposs	-0.05160	0.05901	-0.874	0.38189	
cr1sagave:cprorec	-0.10358	0.10339	-1.002	0.31641	
cposs:cprorec	0.22190	0.07116	3.118	0.00182	**
cr1sagave:cprothm	-0.32739	0.08117	-4.034	5.49e-05	***
cposs:cprothm	0.16898	0.06100	2.770	0.00560	**
cprorec:cprothm	-0.96001	0.11475	-8.366	< 2e-16	***
cr1sagave:cldiff	0.29107	0.22511	1.293	0.19600	
cposs:cldiff	-0.46938	0.15219	-3.084	0.00204	**
cprorec:cldiff	2.63851	0.32559	8.104	5.32e-16	***
cprothm:cldiff	1.46373	0.27422	5.338	9.40e-08	***
cr1sagave:cposs:cprorec	-0.05368	0.09210	-0.583	0.56002	
cr1sagave:cposs:cprothm	0.08857	0.07279	1.217	0.22367	
cr1sagave:cprorec:cprothm	-0.52570	0.11877	-4.426	9.59e-06	***
cposs:cprorec:cprothm	0.12803	0.10502	1.219	0.22278	
cr1sagave:cposs:cldiff	-0.12519	0.20018	-0.625	0.53173	
cr1sagave:cprorec:cldiff	1.05955	0.41782	2.536	0.01122	*
cposs:cprorec:cldiff	-0.94988	0.29714	-3.197	0.00139	**
cr1sagave:cprothm:cldiff	1.06607	0.35243	3.025	0.00249	**
cposs:cprothm:cldiff	-0.69399	0.24994	-2.777	0.00549	**
cprorec:cprothm:cldiff	3.29886	0.53728	6.140	8.26e-10	***
cr1sagave:cposs:cprorec:cprothm	0.31348	0.10654	2.943	0.00326	**
cr1sagave:cposs:cprorec:cldiff	-0.47743	0.37193	-1.284	0.19926	
cr1sagave:cposs:cprothm:cldiff	-0.65610	0.31359	-2.092	0.03642	*
cr1sagave:cprorec:cprothm:cldiff	2.60179	0.65071	3.998	6.38e-05	***
cposs:cprorec:cprothm:cldiff	-0.97268	0.49059	-1.983	0.04741	*
cr1sagave:cposs:cprorec:cprothm:cldiff	-1.61855	0.58002	-2.791	0.00526	**

A.2.1.3 - Dataset: Subset [2]

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)	
(Intercept)	-2.59103	0.06596	-39.284	< 2e-16	***
cr1sagave	0.29636	0.09474	3.128	0.00176	**
cposs	1.07865	0.07205	14.971	< 2e-16	***
cprorec	0.82943	0.07597	10.917	< 2e-16	***
cprothm	-1.02474	0.10026	-10.221	< 2e-16	***
cldiff	1.82526	0.17262	10.574	< 2e-16	***
cr1sagave:cposs	-0.02433	0.10499	-0.232	0.81672	
cr1sagave:cprorec	-0.07082	0.10308	-0.687	0.49202	
cposs:cprorec	-0.09959	0.08211	-1.213	0.22520	
cr1sagave:cprothm	-0.22961	0.13734	-1.672	0.09457	.
cposs:cprothm	0.12529	0.10977	1.141	0.25368	
cprorec:cprothm	-0.68239	0.11043	-6.180	6.43e-10	***
cr1sagave:cldiff	0.24863	0.22287	1.116	0.26460	
cposs:cldiff	-0.32820	0.18691	-1.756	0.07910	.
cprorec:cldiff	1.95979	0.30700	6.384	1.73e-10	***
cprothm:cldiff	0.75909	0.29652	2.560	0.01047	*
cr1sagave:cposs:cprorec	-0.11291	0.11372	-0.993	0.32075	

cr1sagave:cposs:cprothm	-0.01874	0.15229	-0.123	0.90207	
cr1sagave:cprorec:cprothm	-0.48329	0.12174	-3.970	7.19e-05	***
cposs:cprorec:cprothm	-0.10790	0.11899	-0.907	0.36454	
cr1sagave:cposs:cldiff	-0.08595	0.24597	-0.349	0.72678	
cr1sagave:cprorec:cldiff	0.83822	0.38735	2.164	0.03047	*
cposs:cprorec:cldiff	-0.56789	0.33135	-1.714	0.08655	.
cr1sagave:cprothm:cldiff	1.05378	0.36128	2.917	0.00354	**
cposs:cprothm:cldiff	-0.02939	0.32077	-0.092	0.92699	
cprorec:cprothm:cldiff	2.19878	0.53192	4.134	3.57e-05	***
cr1sagave:cposs:cprorec:cprothm	0.33850	0.13344	2.537	0.01119	*
cr1sagave:cposs:cprorec:cldiff	-0.32452	0.42688	-0.760	0.44713	
cr1sagave:cposs:cprothm:cldiff	-0.74285	0.39822	-1.865	0.06212	.
cr1sagave:cprorec:cprothm:cldiff	2.28122	0.62636	3.642	0.00027	***
cposs:cprorec:cprothm:cldiff	-0.05666	0.57319	-0.099	0.92126	
cr1sagave:cposs:cprorec:cprothm:cldiff	-1.59131	0.68871	-2.311	0.02086	*

A.2.1.4 - Dataset: Subset [3] Coefficients:

	Estimate	Std. Error	z value	Pr(> z)	
(Intercept)	-2.253055	0.061800	-36.457	< 2e-16	***
cr1sagave	0.282338	0.082032	3.442	0.000578	***
cposs	0.966691	0.083957	11.514	< 2e-16	***
cprorec	0.964829	0.074693	12.917	< 2e-16	***
cprothm	-0.940342	0.095638	-9.832	< 2e-16	***
cldiff	1.609127	0.200931	8.008	1.16e-15	***
cr1sagave:cposs	0.005799	0.117362	0.049	0.960593	
cr1sagave:cprorec	-0.026788	0.102348	-0.262	0.793529	
cposs:cprorec	-0.220072	0.098658	-2.231	0.025705	*
cr1sagave:cprothm	-0.261596	0.112032	-2.335	0.019543	*
cposs:cprothm	0.066569	0.131363	0.507	0.612327	
cprorec:cprothm	-0.492028	0.113452	-4.337	1.45e-05	***
cr1sagave:cldiff	0.107752	0.240499	0.448	0.654126	
cposs:cldiff	-0.031030	0.271587	-0.114	0.909037	
cprorec:cldiff	1.326606	0.331469	4.002	6.28e-05	***
cprothm:cldiff	0.325219	0.359076	0.906	0.365089	
cr1sagave:cposs:cprorec	-0.227645	0.145336	-1.566	0.117270	
cr1sagave:cposs:cprothm	0.001992	0.160008	-0.012	0.990069	
cr1sagave:cprorec:cprothm	-0.217259	0.117872	-1.843	0.065302	.
cposs:cprorec:cprothm	-0.385601	0.150024	-2.570	0.010162	*
cr1sagave:cposs:cldiff	0.147407	0.342028	0.431	0.666484	
cr1sagave:cprorec:cldiff	0.354029	0.371031	0.954	0.339995	
cposs:cprorec:cldiff	0.116903	0.440803	0.265	0.790852	
cr1sagave:cprothm:cldiff	0.478586	0.399311	1.199	0.230711	
cposs:cprothm:cldiff	0.673156	0.485935	1.385	0.165967	
cprorec:cprothm:cldiff	1.237834	0.593858	2.084	0.037124	*
cr1sagave:cposs:cprorec:cprothm	0.045542	0.162359	0.281	0.779093	
cr1sagave:cposs:cprorec:cldiff	0.357174	0.520155	0.687	0.492291	
cr1sagave:cposs:cprothm:cldiff	-0.030333	0.565291	-0.054	0.957206	
cr1sagave:cprorec:cprothm:cldiff	0.992783	0.596338	1.665	0.095953	.
cposs:cprorec:cprothm:cldiff	1.279557	0.788627	1.623	0.104694	
cr1sagave:cposs:cprorec:cprothm:cldiff	-0.077140	0.824257	-0.094	0.925437	

A.2.1.5 - Dataset: Subset [4] - model error due to multicollinearity Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-2.43126	4.04564	-0.601	0.548
cr1sagave	0.32004	4.71728	0.068	0.946
cposs	1.82165	8.82636	0.206	0.836
cprorec	1.23831	2.22890	0.556	0.579
cprothm	-1.30611	7.66451	-0.170	0.865
cldiff	1.16575	6.74699	0.173	0.863
cr1sagave:cposs	-0.02067	10.29177	-0.002	0.998
cr1sagave:cprorec	-0.04829	2.59647	-0.019	0.985
cposs:cprorec	-0.84606	4.86199	-0.174	0.862
cr1sagave:cprothm	-0.21612	8.93701	-0.024	0.981

cposs:cprothm	0.93447	16.72184	0.056	0.955
cprorec:cprothm	-0.08142	4.22180	-0.019	0.985
crlsagave:cldiff	-0.04331	8.63326	-0.005	0.996
cposs:cldiff	0.93146	14.71831	0.063	0.950
cprorec:cldiff	0.59679	3.79940	0.157	0.875
cprothm:cldiff	-0.38784	12.78239	-0.030	0.976
crlsagave:cposs:cprorec	-0.31112	5.66411	-0.055	0.956
crlsagave:cposs:cprothm	-0.14520	19.49823	-0.007	0.994
crlsagave:cprorec:cprothm	-0.12782	4.91808	-0.026	0.979
cposs:cprorec:cprothm	-1.41271	9.20982	-0.153	0.878
crlsagave:cposs:cldiff	0.59447	18.83444	0.032	0.975
crlsagave:cprorec:cldiff	0.13860	4.77368	0.029	0.977
cposs:cprorec:cldiff	1.67423	8.27837	0.202	0.840
crlsagave:cprothm:cldiff	0.16485	16.35621	0.010	0.992
cposs:cprothm:cldiff	2.52994	27.88472	0.091	0.928
cprorec:cprothm:cldiff	-0.03756	7.19509	-0.005	0.996
crlsagave:cposs:cprorec:cprothm	-0.15052	10.72926	-0.014	0.989
crlsagave:cposs:cprorec:cldiff	1.09923	10.40792	0.106	0.916
crlsagave:cposs:cprothm:cldiff	0.75017	35.68329	0.021	0.983
crlsagave:cprorec:cprothm:cldiff	0.36653	9.04082	0.041	0.968
cposs:cprorec:cprothm:cldiff	4.54684	15.67876	0.290	0.772
crlsagave:cposs:cprorec:cprothm:cldiff	1.43715	19.71288	0.073	0.942

A.2.1.6 - Dataset: Subset [5] Coefficients:

	Estimate	Std. Error	z value	Pr(> z)	
(Intercept)	-3.09412	0.21223	-14.579	< 2e-16	***
crlsagave	0.66103	0.21667	3.051	0.00228	**
cposs	1.21444	0.24169	5.025	5.04e-07	***
cprorec	0.90507	0.22375	4.045	5.23e-05	***
cprothm	-1.09238	0.33286	-3.282	0.00103	**
cldiff	1.98274	1.23291	1.608	0.10780	
crlsagave:cposs	-0.17000	0.24781	-0.686	0.49271	
crlsagave:cprorec	-0.22693	0.21356	-1.063	0.28797	
cposs:cprorec	0.12778	0.25281	0.505	0.61326	
crlsagave:cprothm	-0.43072	0.32820	-1.312	0.18940	
cposs:cprothm	0.19880	0.37879	0.525	0.59970	
cprorec:cprothm	-0.58366	0.34526	-1.690	0.09093	.
crlsagave:cldiff	0.43472	1.28380	0.339	0.73490	
cposs:cldiff	-1.08022	1.43461	-0.753	0.45147	
cprorec:cldiff	1.31798	1.03865	1.269	0.20446	
cprothm:cldiff	1.60172	2.19964	0.728	0.46651	
crlsagave:cposs:cprorec	-0.11978	0.24191	-0.495	0.62050	
crlsagave:cposs:cprothm	0.09299	0.37495	0.248	0.80414	
crlsagave:cprorec:cprothm	-0.71947	0.29522	-2.437	0.01481	*
cposs:cprorec:cprothm	0.11570	0.38824	0.298	0.76570	
crlsagave:cposs:cldiff	-0.08656	1.49839	-0.058	0.95393	
crlsagave:cprorec:cldiff	1.43260	0.99869	1.434	0.15144	
cposs:cprorec:cldiff	-0.97036	1.17534	-0.826	0.40903	
crlsagave:cprothm:cldiff	1.59072	2.28573	0.696	0.48647	
cposs:cprothm:cldiff	-1.56936	2.56006	-0.613	0.53987	
cprorec:cprothm:cldiff	0.97064	1.81864	0.534	0.59354	
crlsagave:cposs:cprorec:cprothm	0.34368	0.32997	1.042	0.29762	
crlsagave:cposs:cprorec:cldiff	-0.16652	1.13799	-0.146	0.88366	
crlsagave:cposs:cprothm:cldiff	-0.93016	2.66833	-0.349	0.72740	
crlsagave:cprorec:cprothm:cldiff	3.31058	1.72295	1.921	0.05467	.
cposs:cprorec:cprothm:cldiff	-0.23494	2.05566	-0.114	0.90901	
crlsagave:cposs:cprorec:cprothm:cldiff	-1.34200	1.95993	-0.685	0.49352	

A.2.1.7 - Dataset: Subset [6] Coefficients:

	Estimate	Std. Error	z value	Pr(> z)	
(Intercept)	-2.2458371	0.1473772	-15.239	< 2e-16	***
crlsagave	0.5268896	0.1542435	3.416	0.000636	***
cposs	0.7993994	0.1987521	4.022	5.77e-05	***

cprorec	1.1273218	0.1758177	6.412	1.44e-10	***
cprothm	-0.8544808	0.2331206	-3.665	0.000247	***
cldiff	1.4615399	0.8239052	1.774	0.076077	.
crlsagave:cposs	-0.1805868	0.2126258	-0.849	0.395705	
crlsagave:cprorec	-0.1940481	0.1809816	-1.072	0.283631	
cposs:cprorec	0.1371790	0.2359519	0.581	0.560981	
crlsagave:cprothm	-0.4242653	0.2263180	-1.875	0.060842	.
cposs:cprothm	0.0303092	0.3051455	0.099	0.920879	
cprorec:cprothm	-0.3240876	0.2742948	-1.182	0.237392	
crlsagave:cldiff	0.6256561	0.8420518	0.743	0.457473	
cposs:cldiff	-0.7898275	1.2119917	-0.652	0.514609	
cprorec:cldiff	0.5631215	0.9497430	0.593	0.553235	
cprothm:cldiff	0.5374923	1.6612806	0.324	0.746286	
crlsagave:cposs:cprorec	-0.2651019	0.2456542	-1.079	0.280513	
crlsagave:cposs:cprothm	-0.0006267	0.2959711	-0.002	0.998310	
crlsagave:cprorec:cprothm	-0.6189950	0.2512237	-2.464	0.013743	*
cposs:cprorec:cprothm	-0.0662045	0.3479967	-0.190	0.849117	
crlsagave:cposs:cldiff	0.0393953	1.2564211	0.031	0.974986	
crlsagave:cprorec:cldiff	1.5537075	0.9305155	1.670	0.094973	.
cposs:cprorec:cldiff	-0.6743074	1.2825185	-0.526	0.599049	
crlsagave:cprothm:cldiff	2.0561307	1.6880310	1.218	0.223199	
cposs:cprothm:cldiff	-0.8749628	2.4451418	-0.358	0.720465	
cprorec:cprothm:cldiff	-0.3395955	1.8753959	-0.181	0.856305	
crlsagave:cposs:cprorec:cprothm	0.1580924	0.2982209	0.530	0.596030	
crlsagave:cposs:cprorec:cldiff	0.1543466	1.2737455	0.121	0.903552	
crlsagave:cposs:cprothm:cldiff	-0.9002879	2.5184728	-0.357	0.720737	
crlsagave:cprorec:cprothm:cldiff	3.8063599	1.8058535	2.108	0.035049	*
cposs:cprorec:cprothm:cldiff	0.7784788	2.5080826	0.310	0.756266	
crlsagave:cposs:cprorec:cprothm:cldiff	-1.0505132	2.4340909	-0.432	0.666044	

A.2.2 Result tables of all 2-way interaction models

Formula = str ~ crlsagave * cprorec + crlsagave * cprothm + crlsagave * cposs + crlsagave *
 cldiff + cprorec * cprothm + cprorec * cposs + cprorec * cldiff + cprothm * cposs +
 cprothm * cldiff + cposs * cldiff
 Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

A.2.2.1 - Dataset: Entire data

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)	
(Intercept)	-1.58323	0.02633	-60.134	< 2e-16	***
crlsagave	0.43575	0.03580	12.173	< 2e-16	***
cprorec	1.04889	0.03076	34.101	< 2e-16	***
cprothm	-0.80181	0.02740	-29.262	< 2e-16	***
cposs	0.45211	0.02205	20.504	< 2e-16	***
cldiff	0.96789	0.05119	18.909	< 2e-16	***
crlsagave:cprorec	0.04938	0.01794	2.752	0.00592	**
crlsagave:cprothm	-0.11112	0.02553	-4.353	1.34e-05	***
crlsagave:cposs	0.07850	0.03323	2.363	0.01814	*
crlsagave:cldiff	-0.13432	0.03070	-4.375	1.21e-05	***
cprorec:cprothm	-0.30010	0.02358	-12.726	< 2e-16	***
cprorec:cposs	0.02603	0.01837	1.417	0.15646	
cprorec:cldiff	0.49853	0.07271	6.856	7.08e-12	***
cprothm:cposs	0.06269	0.02525	2.483	0.01303	*
cprothm:cldiff	-0.15253	0.05285	-2.886	0.00390	**
cposs:cldiff	0.04590	0.02829	1.622	0.10474	

A.2.2.2 - Dataset: Subset [1]

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)	
(Intercept)	-1.95388	0.02920	-66.922	< 2e-16	***
crlsagave	0.33195	0.03977	8.346	< 2e-16	***
cprorec	0.88020	0.03166	27.799	< 2e-16	***
cprothm	-0.79023	0.03000	-26.341	< 2e-16	***

cposs	0.30069	0.02527	11.897	< 2e-16	***
cldiff	0.93215	0.04807	19.391	< 2e-16	***
crlsagave:cprorec	-0.03896	0.01731	-2.251	0.024383	*
crlsagave:cprothm	-0.14171	0.02820	-5.026	5.01e-07	***
crlsagave:cposs	-0.01693	0.03395	-0.499	0.617997	
crlsagave:cldiff	-0.09943	0.03341	-2.976	0.002924	**
cprorec:cprothm	-0.30627	0.02558	-11.971	< 2e-16	***
cprorec:cposs	-0.05514	0.01760	-3.132	0.001734	**
cprorec:cldiff	0.49145	0.06903	7.119	1.08e-12	***
cprothm:cposs	0.03425	0.02837	1.207	0.227282	
cprothm:cldiff	-0.20342	0.05489	-3.706	0.000211	***
cposs:cldiff	0.07554	0.03055	2.472	0.013423	*

A.2.2.3 - Dataset: Subset [2]

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)	
(Intercept)	-2.39894	0.04280	-56.048	< 2e-16	***
crlsagave	0.30901	0.05052	6.117	9.52e-10	***
cprorec	1.16234	0.03689	31.505	< 2e-16	***
cprothm	-0.69624	0.04409	-15.790	< 2e-16	***
cposs	0.93930	0.04327	21.707	< 2e-16	***
cldiff	1.08369	0.06026	17.985	< 2e-16	***
crlsagave:cprorec	-0.03963	0.02030	-1.953	0.0509	.
crlsagave:cprothm	-0.14912	0.03062	-4.870	1.12e-06	***
crlsagave:cposs	0.04154	0.05358	0.775	0.4381	
crlsagave:cldiff	-0.09648	0.03783	-2.550	0.0108	*
cprorec:cprothm	-0.33345	0.02961	-11.262	< 2e-16	***
cprorec:cposs	-0.30608	0.02359	-12.974	< 2e-16	***
cprorec:cldiff	0.49959	0.07980	6.260	3.84e-10	***
cprothm:cposs	-0.02648	0.04134	-0.641	0.5217	
cprothm:cldiff	-0.38740	0.05750	-6.737	1.62e-11	***
cposs:cldiff	-0.08176	0.05168	-1.582	0.1136	

A.2.2.4 - Dataset: Subset [3]

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)	
(Intercept)	-2.15194	0.04371	-49.231	< 2e-16	***
crlsagave	0.34178	0.04617	7.403	1.33e-13	***
cprorec	1.14489	0.04102	27.912	< 2e-16	***
cprothm	-0.76136	0.04916	-15.487	< 2e-16	***
cposs	0.92255	0.05372	17.172	< 2e-16	***
cldiff	1.17079	0.06788	17.248	< 2e-16	***
crlsagave:cprorec	-0.04425	0.02236	-1.979	0.04781	*
crlsagave:cprothm	-0.16612	0.03223	-5.154	2.56e-07	***
crlsagave:cposs	0.02672	0.06168	0.433	0.66488	
crlsagave:cldiff	-0.09468	0.04103	-2.308	0.02102	*
cprorec:cprothm	-0.35803	0.03253	-11.006	< 2e-16	***
cprorec:cposs	-0.28803	0.02935	-9.815	< 2e-16	***
cprorec:cldiff	0.46450	0.08962	5.183	2.18e-07	***
cprothm:cposs	0.08960	0.05741	1.561	0.11857	
cprothm:cldiff	-0.37418	0.06393	-5.852	4.84e-09	***
cposs:cldiff	-0.17586	0.06278	-2.801	0.00509	**

A.2.2.5 - Dataset: Subset [4]

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)	
(Intercept)	-2.14253	0.06971	-30.737	< 2e-16	***
crlsagave	0.37346	0.04225	8.839	< 2e-16	***
cprorec	1.19751	0.04903	24.422	< 2e-16	***
cprothm	-0.78678	0.08225	-9.566	< 2e-16	***
cposs	1.37544	0.13723	10.023	< 2e-16	***
cldiff	1.06702	0.07967	13.393	< 2e-16	***
crlsagave:cprorec	-0.05120	0.02480	-2.064	0.039019	*
crlsagave:cprothm	-0.17901	0.03508	-5.103	3.35e-07	***


```

cr|sagave:cposs      0.03779      0.07288      0.519 0.604046
cr|sagave:cldiff     -0.10731      0.04499     -2.385 0.017082 *
cprorec:cprothm      -0.38579      0.03414    -11.301 < 2e-16 ***
cprorec:cposs        -0.43239      0.06381     -6.776 1.24e-11 ***
cprorec:cldiff        0.36542      0.09431      3.875 0.000107 ***
cprothm:cposs         0.21330      0.16097      1.325 0.185135
cprothm:cldiff       -0.37135      0.06468     -5.741 9.40e-09 ***
cposs:cldiff         -0.08826      0.11228     -0.786 0.431850
---

```

A.2.2.6 - Dataset: Subset [5]

Coefficients:

```

              Estimate Std. Error z value Pr(>|z|)
(Intercept)   -2.86220    0.12664  -22.601 < 2e-16 ***
cr|sagave       0.72975    0.11423   6.388 1.68e-10 ***
cprorec        1.31963    0.09195  14.351 < 2e-16 ***
cprothm       -0.90688    0.15788   -5.744 9.24e-09 ***
cposs          0.98876    0.12642   7.821 5.23e-15 ***
cldiff         0.73081    0.17832   4.098 4.16e-05 ***
cr|sagave:cprorec -0.04739    0.05262   -0.901 0.367819
cr|sagave:cprothm -0.10947    0.11664   -0.939 0.347974
cr|sagave:cposs  -0.05692    0.10405   -0.547 0.584363
cr|sagave:cldiff -0.15440    0.09860   -1.566 0.117360
cprorec:cprothm -0.32001    0.08823   -3.627 0.000287 ***
cprorec:cposs  -0.31872    0.06810   -4.681 2.86e-06 ***
cprorec:cldiff  0.18944    0.14159    1.338 0.180903
cprothm:cposs   0.10359    0.12704    0.815 0.414843
cprothm:cldiff -0.21675    0.23374   -0.927 0.353752
cposs:cldiff    0.09864    0.15347    0.643 0.520372
---

```

A.2.2.7 - Dataset: Subset [6]

Coefficients:

```

              Estimate Std. Error z value Pr(>|z|)
(Intercept)   -2.12959    0.10109  -21.067 < 2e-16 ***
cr|sagave       0.64148    0.09729   6.594 4.29e-11 ***
cprorec        1.36062    0.08962  15.182 < 2e-16 ***
cprothm       -0.83485    0.13302   -6.276 3.47e-10 ***
cposs          0.60620    0.11264   5.382 7.37e-08 ***
cldiff         0.81870    0.15591   5.251 1.51e-07 ***
cr|sagave:cprorec -0.04785    0.05656   -0.846 0.397544
cr|sagave:cprothm -0.06961    0.10786   -0.645 0.518710
cr|sagave:cposs  -0.04223    0.10861   -0.389 0.697376
cr|sagave:cldiff -0.15263    0.10276   -1.485 0.137467
cprorec:cprothm -0.31833    0.08845   -3.599 0.000319 ***
cprorec:cposs  -0.27341    0.07242   -3.775 0.000160 ***
cprorec:cldiff  0.17428    0.15942    1.093 0.274298
cprothm:cposs   0.03559    0.10884    0.327 0.743640
cprothm:cldiff -0.20618    0.22657   -0.910 0.362817
cposs:cldiff    0.11219    0.14272    0.786 0.431812
---

```

A.2.3 Result tables of simple models with no interaction terms

Formula = cr|sagave + cposs + cprorec + cprothm + cldiff

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

A.2.3.1 - Dataset: Entire data

Coefficients:

```

              Estimate Std. Error z value Pr(>|z|)
(Intercept)  -1.67013    0.02033  -82.15 <2e-16 ***
cr|sagave     0.50227    0.01807   27.79 <2e-16 ***
cposs         0.43151    0.01734   24.89 <2e-16 ***
cprorec       1.31287    0.01682   78.07 <2e-16 ***
cprothm      -1.09498    0.02605  -42.03 <2e-16 ***
cldiff        0.81862    0.02685   30.49 <2e-16 ***

```

A.2.3.2 - Dataset: Subset [1]

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)	
(Intercept)	-2.05583	0.02325	-88.41	<2e-16	***
cr1sagave	0.31272	0.01909	16.38	<2e-16	***
cposs	0.25368	0.01853	13.69	<2e-16	***
cprorec	1.16617	0.01613	72.28	<2e-16	***
cprothm	-1.15553	0.03008	-38.42	<2e-16	***
cldiff	0.86342	0.02809	30.74	<2e-16	***

A.2.3.3 - Dataset: Subset [2]

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)	
(Intercept)	-2.47883	0.03136	-79.06	<2e-16	***
cr1sagave	0.33023	0.02284	14.46	<2e-16	***
cposs	0.73649	0.02523	29.19	<2e-16	***
cprorec	1.38283	0.02023	68.36	<2e-16	***
cprothm	-1.29997	0.03735	-34.80	<2e-16	***
cldiff	1.02543	0.03664	27.99	<2e-16	***

A.2.3.4 - Dataset: Subset [3]

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)	
(Intercept)	-2.20035	0.03183	-69.13	<2e-16	***
cr1sagave	0.36356	0.02394	15.19	<2e-16	***
cposs	0.64087	0.02808	22.82	<2e-16	***
cprorec	1.35685	0.02235	60.72	<2e-16	***
cprothm	-1.29389	0.03902	-33.16	<2e-16	***
cldiff	1.06379	0.03907	27.23	<2e-16	***

A.2.3.5 - Dataset: Subset [4]

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)	
(Intercept)	-2.11897	0.03557	-59.56	<2e-16	***
cr1sagave	0.39688	0.02555	15.54	<2e-16	***
cposs	0.88828	0.04683	18.97	<2e-16	***
cprorec	1.28226	0.02402	53.39	<2e-16	***
cprothm	-1.23793	0.04039	-30.65	<2e-16	***
cldiff	1.01149	0.04056	24.94	<2e-16	***

A.2.3.6 - Dataset: Subset [5]

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)	
(Intercept)	-2.99667	0.09333	-32.110	<2e-16	***
cr1sagave	0.62934	0.06404	9.828	<2e-16	***
cposs	0.74151	0.07068	10.491	<2e-16	***
cprorec	1.39810	0.05069	27.582	<2e-16	***
cprothm	-1.46411	0.10955	-13.364	<2e-16	***
cldiff	0.86493	0.09400	9.201	<2e-16	***

A.2.3.7 - Dataset: Subset [6]

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)	
(Intercept)	-2.25890	0.08000	-28.238	< 2e-16	***
cr1sagave	0.54783	0.06324	8.663	< 2e-16	***
cposs	0.47563	0.06861	6.932	4.14e-12	***
cprorec	1.46830	0.05527	26.565	< 2e-16	***
cprothm	-1.30629	0.10171	-12.843	< 2e-16	***
cldiff	0.90075	0.09925	9.075	< 2e-16	***

References

- Aiken, Leona S. & Stephen G. West. 1991. *Multiple regression: Testing and interpreting interactions*. Sage Publications.
- Ambridge, Ben, Pine, Julian M., Rowland, Caroline F., Freudenthal, Daniel, & Franklin Chang. 2014. Avoiding dative overgeneralisation errors: semantics, statistics or both? *Language, Cognition and Neuroscience* 29(2), 218-243.
- Bresnan, Joan, Cueni, Anna, Nikitina, Tatiana, & R. Harald Baayen. 2007. Predicting the dative alternation. In G. Boume, I. Kraemer, & J. Zwarts (eds.), *Cognitive foundations of interpretation*, 69-94. Amsterdam: Royal Netherlands Academy of Science.
- Casenhiser, Devin M. & Adele E. Goldberg. 2005. Fast mapping between a phrasal form and meaning. *Developmental Science* 8(6), 500-508.
- Carroll, John, Davies, Peter, & Barry Richman. 1971. *Word Frequency*. New York: Houghton Mifflin.
- Chang, Franklin, Dell, Gary S., & Kathryn Bock. 2006. Becoming syntactic. *Psychological Review* 113(2), 234-272.
- Charniak, Eugene. 1997. Statistical parsing with a context-free grammar and word statistics. In K. Ford (ed.), *Proceedings of the fourteenth national conference on artificial intelligence (AAAI-97)*, 598-603. American Association for Artificial Intelligence. Menlo Park: AAAI Press/MIT Press.
- Clark, Eve V. 1978. Discovering what words can do. *Parasession on the lexicon, Proceedings of the Chicago Linguistics Society*, 34-57.
- Cohen, Jacob, Cohen, Patricia, West, Stephen & Leona Aiken. 2003. *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences, 3rd Edition*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Collins, Peter. 1995. The indirect object construction in English: an informational approach. *Linguistics* 33, 35-49.
- Deerwester, Scott, Dumais, Susan T., Furnas, George W., Landauer, Thomas K., & Richard Harshman. 1990. *Journal of the American Society for Information Science* 41(6), 391-407.

- Dell, Gary S. 1986. A spreading activation theory of retrieval and sentence production. *Psychological Review* 93(3), 283-321.
- Ellis, Nick & Fernando Ferreira-Junior. 2009. Construction learning as a function of frequency, frequency distribution, and function. *The Modern Language Journal*, 93(3), 370-385.
- Ellis, Nick, O'Donnell, Matthew Brook, & Ute Römer. 2014. The processing of verb-argument constructions is sensitive to form, function, frequency, contingency and prototypicality. *Cognitive Linguistics* 25(1), 55–98.
- Ellis, Nick & Matthew Brook O'Donnell. 2011. Robust language acquisition - an emergent consequence of language as a complex adaptive system. In *Proceedings of the Annual Meeting of the Cognitive Science Society* 33, 3512-3517.
- Ellis, Nick & Matthew Brook O'Donnell, 2012. Statistical construction learning: Does a Zipfian problem space ensure robust language learning? In J. Rebuschat & J. Williams (eds.), *Statistical Learning and Language Acquisition*, 265–304. Berlin: Mouton de Gruyter.
- Fillmore, Charles J. 1968. The case for case. In E. Bach & R. T. Harms (eds.), *Universals in linguistic theory*, 1-90. Holt, Rinehart & Winston.
- Firth, John Rupert (1957). A synopsis of linguistic theory, 1930-1955. In Philological Society (Great Britain) (ed.), *Studies in linguistic analysis*. Oxford, England: Blackwell.
- Ford, Marilyn, Bresnan, Joan, & Ronald M. Kaplan. 1982. A competence based theory of syntactic closure. In J. Bresnan (ed.), *The mental representation of grammatical relations*, 727–796. Cambridge, MA: MIT Press.
- Goldberg, Adele E. 1997. The relationships between verbs and constructions. In M. H. Verspoor, K. D. Lee & E. Sweetser (eds.), *Lexical and syntactic constructions and the construction of meaning: Proceedings of the bi-annual ICLA meeting in Albuquerque, July 1995*, 383-398. John Benjamins Publishing Company.
- Goldberg, Adele E. 1995. *Constructions: A construction grammar approach to argument structure*. Chicago: The University of Chicago Press.
- Goldberg, Adele E. 1999. The emergence of the semantics of argument structure constructions. In B. MacWhinney (ed.), *The emergence of language*, 197-212. Lawrence Erlbaum Publications.

- Goldberg, Adele E., Casenhiser Devin M., & Nitya Sethuraman. 2004. Learning argument structure generalizations. *Cognitive Linguistics* 15, 289-316.
- Goldberg, Adele E. 2006. *Constructions at work: The nature of generalization in language*. Oxford University Press.
- Goldberg, Adele E., Casenhiser, Devin M., & Tiffani R. White. 2007. Constructions as categories of language. *New Ideas in Psychology* 25(2), 70-86.
- Green, Georgia. 1974. *Semantics and syntactic regularity*. Bloomington: Indiana University Press
- Gries, Stefan Th. & Anatol Stefanowitsch. 2004. Extending collocation analysis: A corpus-based perspective on 'alternations.' *International Journal of Corpus Linguistics* 9(1), 97-129.
- Gropen, Jess, Pinker, Steven, Hollander, Michelle, Goldberg, Richard, & Ronald Wilson. 1989. The learnability and acquisition of the dative alternation in English. *Language* 65, 203-257.
- Gruber, 1965. *Studies in Lexical Relations*. Doctoral dissertation. MIT.
- Hebb, Donald O. 1949. *The organization of behavior*. New York: Wiley & Sons.
- Hintzman, Douglas L. 1986. "Schema abstraction" in a multiple-trace memory model. *Psychological Review* 93, 411-428.
- Johnson, Matt A. & Adele E. Goldberg. 2013. Evidence for automatic accessing of constructional meaning: Jabberwocky sentences prime associated verbs. *Language and Cognitive Processes* 28(10), 1439-1452.
- Jones, Michael N., Willits, Jon, & Simon Dennis. 2015. Models of semantic memory. In J. R. Bussemeyer, Z. Wang, J. T. Townsend, & A. Eidels (eds.), *Oxford Handbook of Mathematical and Computational Psychology*, 232- 254. Oxford University Press..
- Koenig, Jean-Pierre & Anthony R. Davis. 2001. Sublexical modality and the structure of lexical semantic representations. *Linguistics and Philosophy* 24, 71-124.
- Koenig, Jean-Pierre & Anthony R. Davis. 2006. The KEY to lexical semantic representations. *Journal of Linguistics* 42, 71-108.
- Komatsu, Lloyd. K. 1992. Recent views of conceptual structure. *Psychological Bulletin* 112(3), 500-526.
- Kwantes, Peter J. 2005. Using context to build semantics. *Psychonomic Bulletin & Review* 12(4), 703-710.

- Lakoff, George. 1982. Categories: An essay in Cognitive Linguistics. In The Linguistic Society of Korea (ed.), *Linguistics in the morning calm: selected papers from SICOL-1981*, 139-193. Seoul: Hanshin Publishing Company.
- Lakoff, George. 1986. *Women, fire and dangerous things: What categories reveal about the mind*. Chicago: The University of Chicago Press.
- Langacker, Ronald W. 1987. *Foundations of Cognitive Grammar: Theoretical prerequisites*. Stanford 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.
- Landauer, Thomas K., Foltz, Peter W., & Carrell Laham. 1998. An introduction to Latent Semantic Analysis. *Discourse Processes* 25, 259-284.
- Landauer, Thomas K., McNamara, Danielle S., Dennis, Simon & Walter Kintsch. 2007. *Handbook of Latent Semantic Analysis*. Lawrence Erlbaum Associates, Inc.
- Levin, Beth. 1993. *English verb classes and alternations: A preliminary investigation*. Chicago: The University of Chicago Press.
- Maki, William S., McKinley, Lauren N., & Amber G. Thompson. 2004. Semantic distance norms computed from an electronic dictionary (WordNet). *Behavior Research Methods, Instruments, & Computers* 36(3), 421-431.
- McClelland, James L. 2006. How Far Can You Go with Hebbian Learning, and When Does it Lead you Astray? In Y. Munakata and M. H. Johnson (ed.), *Processes of Change in Brain and Cognitive Development: Attention and Performance XXI*, 33-69. Oxford: Oxford University Press.
- McRae, Ken & Steven Boisvert. 1986. Automatic semantic similarity priming. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 24, 558-572.
- Medin, Douglas L. 1989. Concepts and conceptual structure. *American Psychologist* 44, 1469-1481.
- Medin, Douglas L. & Marguerite M. Schaffer. 1978. Context theory of classification learning. *Psychological Review* 85, 207-238.
- Ninio, Anat. 1999. Pathbreaking verbs in syntactic development and the question of prototypical transitivity. *Journal of Child Language* 26, 619-653.

- Nosofsky, Robert M. 1988. Similarity, frequency, and category representations. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 14, 54-65.
- Patwardhan, Siddharth. 2003. *Incorporating dictionary and corpus information into a context vector measure of semantic relatedness*. Master's thesis, Univ. of Minnesota, Duluth.
- Pedersen, Ted, Patwardhan, Siddharth, & Jason Michelizzi. 2004. WordNet::Similarity - measuring the relatedness of concepts. In *Proceedings of the Nineteenth National Conference on Artificial Intelligence (AAAI-04)*, 1024-1025. American Association for Artificial Intelligence.
- Pennington, Jeffrey, Socher, Richard, & Christopher D. Manning. 2014. GloVe: Global vectors for word representation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 1532-1543. Doha, Qatar. Association for Computational Linguistics.
- Perek, Florent. 2014. Rethinking constructional polysemy: The case of the English conative construction. In D. Glynn & J. Robinson (eds.), *Corpus methods for semantics: quantitative studies in polysemy and synonymy*, 61-85. Amsterdam/Philadelphia: John Benjamins.
- Perek, Florent & Maarten Lemmens. 2010. Getting at the meaning of the English *at*-construction: the case of a constructional split. *CogniTextes* 5. URL: <http://cognitextes.revues.org/331>
- Pinker, Steven. 1989. *Learnability and cognition: the acquisition of argument structure*. The MIT Press.
- Proulx, Robert & Sébastien Hélie. 2005. Adaptive categorization and neural networks. In C. Lefebvre & H. Cohen (eds.), *Handbook of Categorization in Cognitive Science*, 793-815. Amsterdam: Elsevier
- Rappaport Hovav, Malka & Beth Levin. 2008. The English dative alternation: The case for verb sensitivity. *Journal of Linguistics* 44, 129-167.
- Roland, Douglas, Dick, Frederic, & Jeffrey L. Elman. 2007. Frequency of basic English grammatical structure: A corpus analysis. *Journal of Memory and Language* 57, 348-379.

- Schmid, Hans-Jörg & Helmut Küchenhoff. 2013. Collostructional analysis and other ways of measuring lexicogrammatical attraction: Theoretical premises, practical problems and cognitive underpinnings. *Cognitive Linguistics* 24(3), 531-577.
- Snider, Neal. 2008. *An exemplar model of syntactic priming*. Doctoral dissertation. Stanford University.
- Stefanowitsch, Anatol & Stefan Th. Gries. 2003. Collostructions: Investigating the interaction of words and constructions. *International Journal of Corpus Linguistics* 8(2), 209-243.
- Sun, Hao. 2018. *Association and Extension in Learning English Argument Structures*. Doctoral dissertation. University at Buffalo.
- Sun, Hao & Jean-Pierre Koenig. 2017. There are more valence alternations than the ditransitive. In *Proceedings of the Berkeley Linguistics Society* 43 (to appear).
- Theijssen, Daphne, Bosch, Louis ten, Boves, Lou, Cranen, Bert, & Hans van Halteren. 2013. Choosing alternatives: Using Bayesian Networks and memory-based learning to study the dative alternation. *Corpus Linguistics and Linguistic Theory* 9(2), 227-262.
- Thompson, Sandra A. 1990. Information flow and dative shift in English discourse. In J. A. Edmondson, C. Feagin & P. Mühlhäusler (eds.), *Development and diversity: Language variation across time and space: a festschrift for Charles-James N. Bailey*, 239–253. Summer Institute of Linguistics.
- Tomasello, Michael. 1992. *First verbs: a case study of early grammatical development*. Cambridge University Press.
- Wagenmakers, Eric-Jan & Simon Farrell. 2004. AIC model selection using Akaike weights. *Psychonomic Bulletin & Review* 11(1), 192-196.
- Wasow, Thomas. 2001. *Postverbal behavior*. CSLI Publications.
- Yoon, Soyeon. 2013. Correlation between semantic compatibility and frequency: A usage-based approach. *Linguistic Research* 30(2), 243-272.