Lexical fillers influence the dative alternation: Estimating Constructional Saliency Using Web Document Frequencies

Gard B. Jenset, Bergen University College, gbj@hib.no, telephone: +47 55 58 59 65, fax: +47 55 58 58 09, English section, Faculty of Education, Bergen University College, P.O. box 7030, No-5020 Bergen, Norway

Christer Johansson, University of Bergen, christer.johansson@uib.no, telephone: +47 55 58 22 62, fax: +47 55 58 96 60, Department of Linguistic, Literary and Aesthetic Studies, P.O. box 7805, No-5020 Bergen, Norway

Abstract

We propose a new measure of constructional saliency for use with Web-data, which corrects for infrequent forms. The measure attempts to incorporate both collocational information as well as frequency of use for the whole construction. We report on results for a case study of the so-called dative alternation in English, and show that our measure of saliency indicates that not only do specific verbs have different preferences for the two forms of the alternation, as shown in previous research, but that there are also specific preferences with respect to the filler items. We interpret this as supporting the view that the dative alternation is governed by a rich web of syntactic, semantic, and pragmatic factors.
1 Introduction

Legend has it that the original basis for an inch was the knuckle length of King Edgar’s thumb, with the unit being further subdivided into barleycorns for a lack of suitable smaller parts of the monarch’s anatomy (Gould, 1993, p. 138). The story illustrates that measurement, itself a potentially complicated process, is inextricably intertwined with the choice of and motivation for the unit of measurement, none more so than the measurement of word associations. Measuring associations among words is a challenge since it is not clear exactly what the appropriate scale of measurement is, and different measures can yield different results (Schmitt, 2010, p. 124). A further complicating factor presents itself in that the status of what is being measured is uncertain: word association might be accidental properties of language production factors, or they might be a psychologically real property of language (Durrant & Doherty, 2010) with possible ramifications for language acquisition (Redington, Crater, & Finch, 1998). Nor is it entirely clear what the scope of the association ought to be (words, phrases, utterance fragments, etc.). The source of the linguistic material matters as well. A finite corpus, however large, offers a different situation than frequencies from the World Wide Web. Although the former does offer many advantages (Kilgariff, 2007), the Web is de facto an important source of linguistic material, despite the increasing availability of large corpora. Furthermore, a metric of association would ideally capture not only associations between words, but also the associations between both words and larger syntactic units, covering subcategorization frames or constructions (Croft & Cruse, 2004, pp. 247–256). The aim of the present article is to present a new metric for measuring associations between words and syntactic constructions that have been retrieved from the Web. As a test case, we look at the English dative alternation and its association with various lexical fillers. We suggest that this association captures a form of “saliency” of the assembled units, although we remain agnostic about the precise psychological implications it might have.

1.1 Measures of association

There exist a wide variety of measures of word association (Schmitt, 2010, pp. 120–132). A basic distinction can be drawn between those that are based on statistical null hypothesis tests (e.g. t-score, z-score and chi-square) and those based on mutual information (MI). The former can be said to indicate our confidence in some association (conditional on the assumptions of the test statistic), while the latter indicates how much the words have in common (Schmitt, 2010, p. 130). An extension of the basic measurement of association between words is presented in Stefanowitsch & Gries (2003) and Gries & Stefanowitsch (2004) who use the log transformed p-value of Fisher’s exact test as an ordinal measure of the association between words and the construction in which they occur, which they call a collostructional analysis. Further measures of association include conditional probability (Bilisoly, 2008, pp. 115–117) and the related measure minimum sensitivity, or MS (Evert, 2005, p. 85). For a more comprehensive technical comparison of such association measures, see Evert (2005).

An important question concerning measures of word association is their psychological status. Wiechmann (2008) discusses a number of measures of collocational strength based on a combination of corpus and eye-tracking studies. Although no measure performs statistically significantly better than the others, he argues that MS, defined as the minimum of \( \Pr(word1 \mid word2) \) and \( \Pr(word2 \mid word1) \), is preferable due to its lack of dependence on sample size and
distributional assumptions, its ease of computation, and the fact that it to some extent outperformed other measures in his study (Wiechmann, 2008, p. 282). Since Wiechmann’s study did not result in a significant difference between the measures, it is premature to conclude that one measure has a greater psychological plausibility than others. However, based on the results reported in Durrant and Doherty (2010), a priming experiment studying high frequency collocations, it seems reasonable to conclude that collocations express some kind of psychological reality. Interestingly, their results suggest that collocations which score high on any association measure (in their case either t-score or MI) exhibited priming effects (Durrant & Doherty, 2010, p. 145). Hence, various metrics of word association appear, once a certain threshold is reached, to capture a psychological reality of sorts. However, this psychological reality seems to depend more on the size of the association and less on the specific metric chosen.

TABLE 1 ABOUT HERE

A further consideration is the question of how a word is related not only to neighboring words, but also to its syntactic environment proper. Stefanowitsch and Gries (2003) propose to test this using collostructional analysis, an approach that makes use of Fisher’s exact test for 2 x 2 contingency tables. Thus, to test the association between a construction C (Goldberg, 1995) and a verb v, the observed frequencies of their co-occurrence as well as their independent occurrences would be tabulated and tested (cf. table 1). The resulting log-transformed p-value can be compared to the value for other verbs, generating a ranked list of specific verb-construction preferences. While later developments of the method allows for testing multiple relationships within the same construction (Stefanowitsch & Gries, 2005), the method is nevertheless better adapted for data drawn from a finite corpus than for Web data. Fisher’s exact test assumes fixed row and column totals (Pedersen, 1996), which is a fair assumption to make in a finite corpus, but a more problematic one with respect to Web frequencies. Consequently, we believe there to be room for a metric which can measure (possibly multiple) word-construction associations, and which is appropriate for Web frequencies.

1.2 Data

The present study deals with English verb argument structure, specifically the so-called dative construction and its alternative (Bresnan, Cueni, Nikitina, & Baayen, 2007). It is well established that a number of English verbs can express their arguments in two alternative ways, either with a NP–PP structure or a NP–NP structure, as exemplified in (1) and (2):

(1) I gave the book to her (V–NP–PP)
(2) I gave her the book (V–NP–NP)

This will be referred to as the dative alternation (Levin, 1993); (Bresnan et al., 2007). We will largely follow the terminology of Bresnan et al. (2007), illustrated in table 2. The terms Recipient and Theme will collectively be referred to as fillers. The specific instantiations of constructions with fillers will be referred to as constructs.

TABLE 2 ABOUT HERE
The data for the study was collected from the Web, and we decided to collect only instances of the dative alternation where the recipient was animate and expressed as a pronoun. This simplified the data collection process since the data was collected in raw, un-annotated form. For practical reasons, this was further reduced to *her* and *him*. To avoid confounding the pronominal use of *her* with the determiner *her*, the search was limited to cases where *her* was followed by a punctuation mark. The use of pronominal recipients should ensure a potentially wide range of data, since the dative alternation tends to allow greater flexibility with such recipients, cf. Bresnan et al. (2007).

Since un-annotated data do not allow for lemma searches, it was decided to limit the verb forms to a single unambiguous form, the simple past tense, for the sake of consistency. The following verbs, selected on the basis that they can potentially exhibit the dative alternation, were included in the study:

- blew,
- brought,
- designed,
- faxed,
- gave,
- handed,
- kicked,
- left,
- made,
- painted,
- promised,
- sent,
- showed,
- sold,
- struck,
- threw,
- told.

The frequencies were collected from *MSN Search* with a number of different fillers. The reason for choosing *MSN Search* rather than the leading search engine *Google* was that the former’s API interface allows more searches and returns more hits. For each verb, the search was carried out with the two pronominal recipients: *her* and *him*. A number of different topic NPs were included, however, it was not always possible to find examples of all fillers with all verbs. This results in an unbalanced data set, which can cause problems for some statistical analyses. To counter this problem, we employed a mixed effects model to assess statistical significance and determine effect sizes. The statistical modeling is further discussed in section 4.

2 Measuring saliency

In the discussion below, we use the term *saliency* in a technical sense denoting degree of association. The choice of terminology is meant to underline the basic insights from Cognitive Linguistics that grammar should be considered in light of language use, and that language is a form of conceptualization (Croft & Cruse, 2004, pp. 2–3). From such a perspective, it is not unnatural to think of the choice of one argument construct (including fillers) as opposed to another one, as partially a result of semantics expressing different conceptualizations, cf. also Goldberg (1995). However, we do wish to make it clear that this is our interpretation, and that the measure to be proposed in itself is a straightforward association metric.

2.1 Defining saliency

To measure the saliency of constructs, in this case the combinations of specific dative construction alternations with specific fillers, we propose a new measure which relies on Web document frequencies. We first define a Mutual Information (MI) type of measure, based on the frequency of specific verb-noun patterns divided by the total number of occurrences of the words in any order:

\[
MI_D = \log \frac{N \times f_d(pattern)}{f_d(words)};
\]
where \( f_d(\text{pattern}) \) is the document frequency of the specific pattern, i.e. a Web search with quotation marks, which returns the documents containing that exact phrase. \( f_d(\text{words}) \) is the document frequency of the words in any configuration, i.e. a Web search without quotation marks, which returns documents containing all of the words of the pattern irrespective of their order or proximity.

\( N \) is the size of the corpus for conventional mutual information, i.e. the number of Web documents if the units are documents, approximated by a large constant. In our measure, \( N \) is used to scale the measure to appropriate size, since the real size of the corpus is open ended and growing. Our measure relates frequencies of documents with the full pattern, to frequencies of documents containing the same words as in the pattern but in other configurations. We expect that the ratio is a near constant, as the proportion of documents containing the pattern is a property of the population of documents containing the individual words in the pattern, just like the mean of a distribution is a property of the measured population. What may vary is that some patterns have sudden bursts in popularity, which is interesting to study in its own right (Church, 2000). Note that documents that contain repetitions are only counted once which ensures that each document has an equal weight and also makes the measure less sensitive to individual quirks of various authors. Our measure is roughly based on a measure proposed in (Modjeska, Markert, & Nissim, 2003) for a machine learning approach to anaphora resolution. However, our measure differs in some aspects, most noticeably in the inclusion of a scaling constant and a step to correct for infrequent patterns (see below). Our measure can be viewed as the degree of association between the elements of the construct (i.e. lexical fillers and the construction), where a higher score indicates greater association. Point-wise (or maximum likelihood) association measures are known to be overly sensitive to small frequencies (Evert, 2005, p. 86), and we would like a saliency measure for the Web to be more robust. A common remedy is to simply exclude frequencies lower than 5 (or some other suitable constant). We chose to grade MI by the logarithm of the pattern frequency; using the base 10 excludes all patterns seen only once and not until 10 examples are seen will the measure start to get higher than without this correction factor. It gives a weight of 1 to patterns seen 10 times, and gradually increases the weight as the pattern frequency goes up, although at a fairly slow pace.

To define the saliency \( SA_c \) of the construct, i.e. the specific configuration of verbs and arguments, as well as their lexical fillers, we take

\[
SA_c = MI_D \times \log f_d(\text{pattern}),
\]

where \( MI_D \) is multiplied by the base-ten logarithm of the document frequency of the pattern.

As explained above this amounts to controlling for very infrequent patterns, which might otherwise receive an unduly large saliency measure. The frequency adjusted log scale provides a potentially psychologically plausible scale for human perception. Thus, we assume that saliency (in our technical definition) involves both high mutual information and frequent recurrence. Although this is not a definition of the psychological term saliency, Durrant and Doherty (2010) found that high association leads to priming effects, as mentioned above. Thus, we consider our measure an attempt to quantify the essence of the psychological notion salience as manifested in priming effects, namely attention and selection, from observed Web frequencies.

Although there is reason to believe that the measure proposed above could have some psychological validity as a measure of salience, this is ultimately an empirical question reserved for further research. However, the general assumption behind the measure is that cognitive
(including linguistic) processing is probabilistic and sensitive to frequency effects (Bod, 2003; Baayen, 2003); an assumption which we consider fairly uncontroversial. For the moment we remain agnostic as to the specific psychological validity of the specific, logarithmic scale provided by the measure, but propose that the relative ordering it provides has heuristic value and – judging from other studies referred to above – probably some psychological relevance.

2.2 Distributional properties

In the current data set, the saliency values range from −0.98 to 22.39, with a mean of 9.24 and a standard deviation of 7.04. As the density plot in the top left panel of figure 1 shows, the measure follows a bimodal distribution with a pronounced dip in the lower tail. Given the uneven saliency scores of the verbs for the two alternations, it is reasonable to ask whether this is a result of the two constructions involved or if it is merely a reflection of the document frequencies of the verbs.

The middle and rightmost columns of panels in figure 1 show saliency plotted by construction. As the plots show, the saliency for the double object construction is reasonably close to a normal distribution, while that of the prepositional dative is more problematic. A formal test of normality rejects the normality hypothesis for all the current data; however, these tests can be unduly influenced by sample size, and we prefer instead to rely on graphical evaluations such as quantile-quantile plots, while at the same time acknowledging that a model such as the normal distribution is a theoretical model that is fitted to data using only a few parameters derived from all the observations. In practice the observations are very rarely independent of each other, and in our case the population of documents is changing, which adds a time dimension that we have excluded from investigation for simplicity, although we think it could be interesting in relation to, for example, novelty detection.

FIGURE 1 ABOUT HERE

Below, we will treat the saliency score as a normally distributed continuous variable for modeling purposes, while taking care to investigate whether this assumption causes severe problems for the modeling. This, we believe, is a justified empirical approach, which is required irrespective of whether a test or plot indicates statistical normality, since such a distribution always represents an idealization.

FIGURE 2 ABOUT HERE

Before we can address the question of saliency and the dative alternation, it is pertinent to consider our proposed metric’s performance as a measure of association between verbs and themes. If there is an association between the filler items in the alternation constructs and the constructions, and if our proposed metric captures some of this association, then it should also capture some of the verb–theme association. To investigate this, we constructed a matrix where each row represented a verb, and each cell was the mean saliency score for that verb with each NP theme in our dataset. Based on this matrix, a principal component analysis (PCA) was conducted, the output of which can be seen in figure 2. Together the two plots display three dimensions or principal components, covering 65.5% of the variance, which we consider an acceptable result for our purposes. A close scrutiny of the plots reveals that we find expected verb–NP theme associations: money is clustered with gave, house is clustered with left and sold, ball is clustered
with *threw* and *kicked*, to take a few examples. However, in many cases no specific associations can be seen and the fillers are clustered together near the center. This qualitative assessment of the PCA plots suggests that our measure is able to capture conventional and expected word associations, at least in the case of strong or frequent collocation pairs. Based on this observation, we would again like to point out that the results reported in Durrant and Doherty (2010) indicate that strong association, regardless of the specific metric used, is a prerequisite for psychological plausibility. Our measure has in other words passed an initial performance test which strengthens our confidence in its usefulness. To reiterate our position so far: we suggest that the measure should be treated as an approximation to saliency, that it possesses a necessary feature for psychological relevance (capturing frequency association), but that it does not measure “reality” directly. Rather, it can be seen as a proxy to the psychological reality that we are interested in capturing, and this must be further tested in appropriate experiments.

2.3 Document frequencies

It is necessary to comment on the frequencies employed in the present study. Document frequencies, the frequencies returned by a Web search engine, are not frequencies of occurrence proper, but rather an indication of dispersion since they count web pages, that is, documents with one or more occurrences of the term, cf. Kilgarriff (2007). Document frequency is a term closely linked to *polytextuality*, i.e. the number of contexts a word or phrase may occur in (Köhler, 2005, p. 767). However, for the present investigation we prefer the term “document frequency” which serves to highlight the role of the search engine algorithm in the retrieval of Web documents. Although we believe there are certain qualitative differences between such Web documents and the parts of a finite (and possibly balanced) electronic corpus, we would like to emphasize that the choice of terminology is motivated by a desire for precise communication and not part of some theoretically founded terminological dispute.

Perhaps the greatest conceptual problems with document frequencies, as pointed out by Kilgarriff, are that they are not constant, may contain duplicates within the document, several versions of a document may exist on the web, and we do not know exactly how the search engines obtain them (to these can be added a number of practical problems). These are problems that need to be acknowledged. However, the size of the Web and the coverage of search engines make them attractive to linguists in spite of these problems. We also think that even though some documents may be duplicated, it will not hurt our measure, since we are comparing proportions of documents with the pattern to documents with the same word. Additionally, search engines often smooth document frequencies and very often cluster very similar documents together to count less in the frequency counts. A worry is, of course, that we do not know exactly how the search engines handle frequencies, and the algorithms used by the search engines may change without notice. We agree with Kilgarriff (2007) that the linguistics community should strive towards making larger linguistically processed corpora, but we are of the opinion that this is not in conflict with work to exploit the currently available information from the Web more efficiently; rather, the two approaches should be seen as complimentary. On a more pragmatic note, it should be noted that Modjeska, Markert, and Nissim (2003) found that their Web based approach greatly improved their system, which gives at least some heuristic credibility to a Web based approach. The
measure proposed in the present paper compensates to some degree for the shortcomings mentioned by Kilgarriff (2007), since it is scaled by multiplication by a large constant.

3 Saliency and the dative alternation

Returning to the English dative alternation, we will now consider in more detail the association between construction, verb, and fillers. The question we want to address is whether the verbs have different saliency scores with different alternations, different themes, and different recipients.

3.1 Saliency by construction type

As a first step, let us consider whether there are any overall differences between verbs with respect to saliency. Since the saliency measure encodes the interaction of the specific verb with specific themes and recipients for each alternation, the scores give an indication of the preference that a verb has for the alternation patterns with different recipients and themes. As all recipients here are pronominal, we would not expect any skew related to the choice of recipient for grammatical reasons (but possibly for social and pragmatic reasons). This expectation is supported by a Welch t-test for unequal samples, which indicates that there is no significant difference in saliency for her and him ($t = 0.83$, $df = 112.29$, $p$-value = 0.41).

Based on this, we can predict that if all verbs have similar preferences for the two alternations with the specific recipients and themes, they should follow a monotonous trend so that verbs which score low for the prepositional dative construction should score low for the double object construction, and vice versa. This follows from the fact that the saliency score is not only an expression of verb preferences, but also corrects for frequency. However, this begs the question of whether the saliency score expresses the same information as verb frequency, that is, can the constructional saliency measure be predicted by verb frequency? To account for this possibility, we collected raw frequencies of the (past tense) verbs in question from the Corpus of Contemporary American English (Davies, 2008) or COCA, and performed a Spearman rank correlation test on the frequencies and the mean saliency of each verb. Spearman’s rank correlation test is a robust ordinal measure of rank association, and an association between verb frequency and saliency was expected to manifest itself as a significant departure from the null hypothesis of no correlation. Since the result of the test was not significant at the 0.05 level ($S = 858$, $p = 0.85$, rho = 0.05), we can be reasonably confident that there is no overlap between the two. Put differently: although constructional saliency incorporates information about verb frequency, constructional saliency clearly measures something beyond, and cannot be reduced to, verb frequency.

Thus, although some verbs might have a stronger preference for one of the alternations overall, we would nevertheless expect it to score relatively low both for the prepositional dative construction and the double object construction, given the choice of recipient. We strongly suspect the verbs to behave differently with respect to the alternation patterns. Essentially, a monotonous relationship would imply that the scores for the two alternations would measure the same thing, possibly some general saliency/frequency aspect of the verbs’ usage.

FIGURE 3 ABOUT HERE
Figure 3 shows the mean saliency for the two alternations per verb. For the sake of comparison, a linear regression line is included in the plot. As the plot clearly shows, the saliency of the two alternations with respect to the different verbs are very different. Some of the verbs are reasonably closely aligned with the regression line, whereas others are further away, indicating that they have different preferences for the two alternations. We interpret this as indicating that the saliency score is able to capture real differences between the verbs.

3.2 Saliency by verb

In the previous section we plotted mean saliency scores. We next turn to the full variation in saliency per construction type and verb. Figure 4 shows that the situation varies: some verbs have quite similar saliency scores for either construction, while others display large differences, with *gave* being an extreme case. There are a number of potential confounding factors that must be considered here, but which are difficult to control for. An obvious candidate is the unbalanced data mentioned above. Another is the likely difference in the number of senses of these verbs, which is difficult to take into account without manual interventions in the data.

The overall impression is thus that the choice of verb is likely to influence the saliency of the construct. However, we find that there is a great deal of variation. Nevertheless, it seems likely that the choice of verb will have an impact on the saliency of the construct. This will be explored further in section 4. Before that, we will consider the case of *gave* in more detail, since this is a verb which is both frequent and exhibits a clear difference between the double NP and the prepositional dative constructions.

FIGURE 4 ABOUT HERE

3.3 The case of *gave*

In the previous section we saw that *gave* displayed a large variation regarding saliency, depending on construction type. Figure 5 shows mean saliency by construction type, with 95% confidence intervals. As the plot shows, the double object construction has a much higher mean saliency than the prepositional dative construction, for *gave*. Interestingly, some outliers are visible in the plot, which suggests that within the *gave* dative constructions there are filler-specific lexical preferences influencing the final construct. Although *gave* seems to prefer the double object construction, we find some constructs with the prepositional dative that have a saliency that are on a par with the double object construction. The themes are “money” and “flowers”, and, interestingly “money” is highly salient with both male and female recipients, while “flowers” is salient only with female recipients.

The results reported here are consistent with the results found in a large corpus such as COCA. Searching COCA for frequencies of two partial instantiations of the constructions resulted in the following: *gave + him + the* [noun] (n = 545) and *gave + the* [noun] + *to him* (n = 4), revealing a strong preference for the double object construction (the frequencies with *her* are double object = 335, prepositional dative = 18). However, even a corpus as a large as COCA with over 400 million words is not able to capture the full variation seen in our Web data. This variation has both empirical and theoretical implications. The empirical implication is that although *gave* overwhelmingly prefers the double object construction, there are filler-specific lexical effects, which, judging from the items themselves, seem motivated in pragmatics and wider cultural notions. Thus, even a very large corpus such as COCA will leave out interesting information.
concerning the associations between constructions and fillers, highlighting the usefulness of the Web as a supplement to corpora. This has implications for corpus linguistics as a methodological subject, as well as for linguistic theorizing that aims at descriptive adequacy.

FIGURE 5 ABOUT HERE

In brief, looking at the saliency of different fillers within the dative constructions nuances the picture from COCA. Although, on the whole, gave prefers the double object dative, our measure of saliency highlights specific fillers that occur as themes and recipients more often than the generalization based on corpus frequencies would suggest. This is clearly a usage effect, stemming from social conventions. It illustrates that the saliency metric is capable of capturing usage effects which lie at the intersection between syntactic, lexical, and pragmatic factors.

4 Modeling contributing factors

The next question we want to address is: which variables are in fact expressed through our measure. In other words, how much of the variation in constructional saliency can be explained by the verb, how much can be explained by the construction, and how much can be explained by other factors such as the theme or syntactic complexity? These questions will allow us to better assess the usefulness of our saliency measure compared to the information that can be gleaned from frequencies, lexical preferences, and syntactic measures of complexity. Obviously, frequency information about verb and construction co-occurrence is an integrated part of the saliency measure, cf. above. However, it is of interest to assess how much of the variation in saliency is due to the factors listed above. For instance, if the verb explains all, or almost all of the variation, with little explanatory force left for the construction type, we would have a result which suggested that verbs have little or no variability in their choice of the dative alternation; at least as far as Web usage is concerned. Conversely, if variables such as construction, verb and frequency only account for a small part of the variation in saliency, the hypothesis that our saliency measure provides information above and beyond these variables will be strengthened.

4.1 Linear mixed effects modeling

As a first approximation to pinpointing the factors influencing saliency, a simple linear regression model was attempted, with saliency score as the dependent variable and “Construction” and “Verb” as predictors. Although the fit of the model was acceptable, there were severe problems with underdispersion, that is, there was less variation in the data than expected by the model. This is a result of the fact that we have multiple observations for each verb, that is, the independence assumption is false. This suggests that there were systematic structures in the data that the model was not able to capture. In addition to this comes the unbalanced data set pointed out above. To correct these problems, a second model was set up in the statistics environment R (R Development Core Team, 2010), using a linear mixed effects model (Baayen, 2008a) with the lmer function in R (Bates, Maechler, & Dai, 2008).
As mentioned above, a benefit from using the \textit{lmer} function is that it handles unbalanced data in an adequate way (Pinheiro & Bates, 2000, pp. 25–27), although it is nevertheless important to check that the confidence intervals are reasonable. This should not be seen as a way of circumventing bad research design with bloated statistical machinery as Gorard (2003) warns against. Rather, we approach both the data and the statistical model by considering them as uncertain and requiring an exploratory approach, as well as theoretically informed judgment.

Venables (2000) suggests that mixed effect models can be viewed in two ways: either as a set of fixed and random effects, both of which we want to estimate. This is consistent with a hypothesis-testing approach. Alternatively, the random effects are seen as a kind of model or template which might be appropriate, given the dependent nature of observations and groups. This approach is more exploratory than hypothesis-testing oriented in nature, and it is consonant with the pragmatic attitude towards random effect modeling advocated in e.g. Gelman & Hill (2007). In the present article, we follow this pragmatic approach. The aim of the mixed effects model presented here is not so much to identify statistical significance as to identify an empirically adequate overall model of the relationship between saliency and the variables at hand.

Several models were attempted in the modeling process. Based on the discussion above, it seemed reasonable to include the type of pronoun in the model. However, the effect of this predictor was minute, and it was not found significant based on an $F$-test comparison of models. The final form of the model, with a response, or dependent variable, (“ConstructionalSaliency”), an overall mean $\alpha_0$, a fixed effect $\beta$ for the “Construction” factor (levels: NP-NP and NP-PP), a random effect $\alpha_i$ for the “Verb” factor, an adjusted mean $\alpha_j$ and an error term $\varepsilon$, can be written as

\[
\text{ConstructionalSaliency} = \alpha_j + \beta \text{Construction} + \varepsilon
\]

\[
\varepsilon \sim N(0, \sigma_\varepsilon^2),
\]

\[
\alpha_j = \alpha_0 + \alpha_i, \quad \alpha_i \sim N(0, \sigma_i^2).
\]

In R code, the model can be expressed as follows:

\[
\text{lmer(ConstructionalSaliency} \sim \text{Construction + (1|Verb), family = gaussian).}
\]

Both notations specify that for any given construct, its constructional saliency is a linear function of the construction to which it belongs as well as a certain amount of individual noise $\varepsilon$, with adjustments for systematic differences in constructional saliency between verbs. The by-verb adjustments themselves are expected to follow a normal distribution with a center of 0.

**FIGURE 6 ABOUT HERE**

The adjustment, or random effect, added for Verb allows the model to take into account the fact that the observations are made in groups for each verb. The residuals showed no signs of patterning, indicating that the fit to the data is good. Furthermore, the diagnostic plots in figure 6 show that both the residuals and the random effects are acceptably close to a random distribution. An additional random effect for the theme was not found significantly different from the simpler
model described above, and hence omitted for reasons of parsimony.

**TABLE 2 ABOUT HERE**

Tables 2 and 3 provide the estimates for the model, alongside estimated $p$-values and confidence intervals. Table 2 shows that there is a large difference in saliency for the two alternations; the difference is much greater than the uncertainties as shown by the lack of overlap in the confidence intervals.

The group level structure, “Verb”, is shown in Table 3. The standard deviations show the amount of variation explained by the “Verb” factor and how much is left unexplained (i.e. the residual within-verb variation). The within-verb variation is much greater than the between-verb variation, and the difference is again greater than the uncertainties represented by the estimated confidence intervals. To put the difference into perspective, the standard deviations can be compared with the 7.04 overall standard deviation reported above for the *ConstructionalSaliency* variable. The within-verb variation amounts to about 47% of one standard deviation, while the between-verb standard deviation amounts to about 82% of one standard deviation.

**TABLE 3 ABOUT HERE**

There is no $R^2$ measure for modeling the goodness-of-fit in linear mixed effect models, since these models are not based on reducing variance in the way classical regression models are. However, it is possible to get a similar measure by comparing the fitted (that is, the estimated) response values from the model, with those of the observed data (i.e. *Saliency*), cf. Johnson (2008, pp. 237–239); Baayen (2008a, pp. 258–259). For the model above, the square correlation between estimated and observed data is 0.41. This score is not necessarily interesting on its own, but it can be compared with a model which only uses the information about the verbs, not the differences in construction. Such a model has a square correlation of only 0.26. That is, a model with only a random effect for “Verb” explains 63% of the estimated – observed correlation found in the full model. In plain terms, this corroborates the view that information about both verb and construction play an important and substantial role in our measure, which is precisely what we wanted to achieve. Put differently, this result documents that our proposed saliency measure provides information about both verbs and constructions, packed into a single score.

To furthermore assess the magnitude of influences, it is instructive to compare the standard deviations of the random effects for these two models (with or without a predictor for construction). The full model, which takes includes the differences between verbs as well as the alternation, has a 0.9% higher between-verb standard deviation, but a 13% lower within-verb standard-deviation. In other words, if we look only at verbs we lose a substantial amount of information since we miss the variation within each verb with respect to the dative alternation. For both random and fixed effects the confidence intervals are small, which increases our belief in the model, since this suggests that the program had enough data to produce sensible estimates.

**4.2 Logistic mixed effects modeling**
Above we established that our measure captures information about both constructions and verbs. To estimate the influence of theme on the dative alternation, we fitted the following mixed effects binomial logistic regression model, again using the lmer function:

\[
\Pr(\text{DoubleObject}) = \log \frac{1}{1 + \exp(-\alpha_i + \beta \text{ConstructionalSaliency} + \beta \text{Theme} + \varepsilon)}
\]

\[
\varepsilon \sim N(0, \sigma^2_{\varepsilon}),
\]

\[
\alpha_i = \alpha_0 + \alpha_i, \quad \alpha_i \sim N(0, \sigma^2_{\alpha_i}).
\]

That is, the probability of finding the double object construction rather than the prepositional dative is estimated as an inverse logistic function of “ConstructionalSaliency” and “Theme”, with a random effect for “Verb”. Although the predictor for “Theme” is redundant from a hypothesis testing perspective, it was kept for modeling purposes, in line with the exploratory approach discussed above. Figure 7 plots the effects of the “Saliency” predictor against the estimated probability of finding a double object construction. Individual (clusters of) constructs are plotted as hexagons around the S-curve. Since the random effect adjusts for differences among verbs, the deviations from the S-curve are in large part due to the different constructional preferences for themes. The plot shows that the “Theme” predictor has different effects at different points of the S-curve. Near the bottom of the curve, i.e. with the prepositional dative, we find a few cases far above the curve, which indicates that these points have an unexpectedly high saliency. We take this to mean that verbs differ in their sensitivity to filler effects, but that the sensitivity is construction-specific (rather than verb-specific). If this interpretation is correct it partially accounts for why the theme did not have a large effect in the models explored above, since verbs and constructions account for such a large portion of the overall variation.

**FIGURE 7 ABOUT HERE**

Furthermore, the results corroborate the results for gave in section 3.3 and demonstrate that these results are not specific to that verb but can to some extent be generalized to the class of verbs exhibiting the dative alternation. Figure 7 shows that, even when we take the effects of different verb preferences into account, there are still theme-specific effects in operation, and that these theme-specific effects appear to be greater when the probability of the double object construction is low, supporting the view that themes also have an independent role to play among the factors influencing the dative alternation.

### 4.3 The role of complexity and processing

Finally we will consider the relationship between our measure and linguistic complexity. The dative alternation is a question of constituent ordering, and Arnold, Losongco, Wasow, & Ginstrom (2000) document that it is sensitive to both complexity constraints (such as heaviness) and pragmatic constraints (such as newness), with the former possibly accounting for more than the latter. This begs the question of whether there is a systematic relation between our saliency measure and complexity. Such a hypothetical relationship would be interesting in its own right, but would also prove useful in the framework of *synergetic linguistics* which seeks to construct explanations for linguistic phenomena based on established general laws and hypotheses (Köhler, 2005). In our case, Behaghel’s law with its predictions of a short-before-long ordering of
constituents (Köhler, 1999, p. 49) would be particularly relevant. However, processing complexity with respect to constituent structure is also of interest (Hawkins, 1994); (Hawkins, 2004). Finally, information (again from a processing, or information theory viewpoint) can be considered relevant since this can be used as an approximation to the amount memory needed for processing (Köhler, 2005, p. 770).

From a processing point of view, we would expect an overall preference for the double object construction over the prepositional dative for two reasons pertinent to the question of complexity. Firstly, following the hypothesis proposed by Hawkins (1994, p. 76) that the hearer can more efficiently process the syntactic structure of an utterance when the ratio between words (in the linear input) and immediate syntactic constituents is as high as possible, we can predict a preference for the double object construction over the prepositional dative. Figure 8 illustrates this point. The prepositional dative example in 8 a. minimally requires four words, gave, the, money, to in order to identify the three constituents of the VP (V, NP, PP), yielding a ratio of 3/4 or 75%. In comparison, the double object construction in 8 c. requires only three words, gave, me, the to establish the structure (V, NP, NP) resulting in a higher ratio (3/3 or 100%). Since the prepositional dative has a slightly worse immediate constituent ranking compared to the double object construction, we would expect the latter to be the preferred construction. This prediction is borne out in our material. Secondly, if we consider complexity from the vantage point of information, it could be assumed that higher entropy entails greater degree of uncertainty or unpredictability when it comes to processing. Since greater uncertainty means that more variants must be considered simultaneously in processing, we can posit that less uncertainty leads to less complexity from a processing point of view (Köhler, 2005). If we consider the linear (surface) order of constituents, it becomes clear that the pattern verb NP to can realize at least two different syntactic structures (figure 8 a. and b.), whereas the pattern verb NP NP, i.e. the double object construction, has only a single structural interpretation (figure 8 c.). Thus, there are at least two general hypotheses regarding complexity and processing that predict the results presented above. However, these are general hypotheses about the constructions that do not relate specifically to our measure, even if they predict the overall distribution seen in figure 4. Next we will specifically consider the relation between our measure and other operational measures of complexity.

**FIGURE 8 ABOUT HERE**

It is not straightforward to investigate the relationship between our saliency measure and syntactic complexity using the data that was gathered from the Web. The reason is that our MSN data is un-annotated and predicate-centered. Since it is un-annotated it is difficult to take a phrase-structure/node-counting approach to complexity. Furthermore, since it is centered on specific predicates (rather than sentences) it might be problematic to take a word-count approach to complexity. Figure 8 illustrates that the number of immediate constituents is the same for the prepositional dative (a) and for the double object construction (c), making it difficult to use this as a numeric proxy for complexity.
Although a full-scale investigation falls outside the scope of the present paper, we offer some predictions and preliminary results. Further studies might benefit from more data and other approaches; we believe there is a potential for using machine learning and Bayesian networks for further investigation. Following the principles of synergetic linguistics, we start by outlining in some detail some predictions based on general principles for constituent ordering.

i. In accordance with Behaghel’s law (Köhler, 1999, p. 49) we would expect a “short-before-long” ordering of arguments in both constructions, irrespective of which construction they appear in.

ii. As a consequence of (i), we would expect the double object construction (NP-NP) to have a greater tendency toward expressing recipients as pronouns, because the recipient appears at an earlier position in the linear ordering of constituents in the double object construction.

iii. As a consequence of (i) and (ii), we would expect that longer, more complex themes are found with the prepositional dative, where the longer, more complex PP recipient in final position would ensure adherence to Behaghel’s law. Conversely, we would expect to find shorter themes with the double object construction.

iv. As a consequence of (i – iii), we would expect constructs with:
   a. recipients expressed as pronouns and short themes to have a higher saliency with the ditransitive construction (NP-NP), and
   b. recipients expressed as full noun phrases and longer themes to have a higher saliency with the prepositional dative construction (NP-PP).
   c. As a consequence of (a and b) we would expect that when recipients (in both constructions) are expressed as pronouns, saliency for the NP-NP construction will go up and saliency for the NP-PP construction will go down as the length of the theme increases. The reasoning behind this would be that such an arrangement would go against Behaghel’s law.

v. As a consequence of (i – iv), we would expect constructs with highly complex elements (e.g. themes and recipients realized by NPs with center embedded relative clauses, or very long phrases in general) to have an overall low saliency for either construction. The reason is that such examples are bound to be rare, and our measure incorporates explicit steps to tone down the influence of rare events. We will not test this last prediction, since it follows both from the steps outlined above, the general tendency for linguistic units to be Zipf-distributed, the rareness of highly complex linguistic phenomena (e.g. multiple center embedding), and finally the fact that the design of our measure would ensure that such rare occurrences would receive a low saliency score.

In short, we expect our measure to be sensitive to trends in usage that arise for a number of reasons, including conventionality, lexical semantics, constructional semantics, and complexity and processing constraints. However, we expect this relationship to come about because our measure is based on usage (or rather: dispersion, which in itself is a product of usage) frequencies, which themselves are partially shaped by the factors mentioned above. Here lies some of the reason for naming the measure “saliency”: the idea is that a high score on our measure translates into something that, hypothetically, stands out as a noticeable combination of linguistic units. Before looking at the predictions above in more detail, we need to consider the issue of measuring complexity.
As mentioned above, measuring complexity requires some deliberation, perhaps more so in this case than usual, because of the nature of our data (summary data of untagged, predicate centered patterns). Here we will focus on two types of complexity, namely syllabic complexity (number of syllables) and morphological complexity (whether or not the noun displays a derivational affix).

Turning next to the predictions in (i – iv) above, we gathered some new material from COCA to supplement the material in our paper. We will first deal with the predictions relating to recipients (ii), then themes (iii), finally we will turn to complexity as outlined in prediction (iv).

To test prediction (ii), that is, the realization of recipients, we searched COCA for two patterns:

\[
gave * \text{the [noun]}
gave \text{the [noun]} \text{to *}
\]

where the asterisk denotes a wildcard character to match any potential recipient and [noun] denotes any noun. Since this is a preliminary study, we limited ourselves to the past tense of the verb give. For each of the two searches we took the first 100 hits returned by COCA and identified the relevant cases, manually excluding constructions like gave up. This left us with 171 relevant hits, the distribution of which is shown in table 5 below. It is not surprising that such a clearly visible difference is statistically significant ($\chi^2 = 151.4$, df = 1, $p << 0.001$, $\varphi = 0.95$). In our data, the double object construction massively prefers a pronoun as its recipient, whereas the opposite is true for the prepositional dative.

TABLE 5 ABOUT HERE

Thus, we can confidently say that prediction (ii) was confirmed. This in itself is not surprising, based both on Behagel’s law and previous research. However, it illustrates that the realization of arguments within these constructions is far from arbitrary.

Prediction (iii) dealt with the complexity of themes, that is, the entity being affected, or the direct object. Again drawing on COCA, we searched for two patterns:

\[
gave \text{PERS.PRN the [noun]},
gave \text{the [noun]} \text{ to PERS.PRN.}
\]

Since we wanted to focus on the theme, we kept the recipient role constant across both conditions by limiting ourselves to only personal pronouns. This would also ensure that the data more closely resembled the data in our original paper. We chose to use all personal pronouns rather than just him and her to obtain more data for the prepositional dative. For the prepositional dative construction we found 21 cases; for the double object construction we obtained 1,277 hits. However, for convenience we only used the first 100 hits provided by COCA. For each of the 121 cases the number of syllables in the theme was recorded and we also noted whether or not the theme had a derivational affix. Using a Welch two-sample $t$-test, we found a statistically significant, medium-sized difference in the number of syllables exhibited by the nouns acting as
themes, which confirmed our predictions ($t = 3.22, df = 54.83, p = 0.002, d = 0.52$). Nouns acting as themes in the NP-NP construction had fewer syllables (mean = 1.33, sd = 0.58) than the NP-PP construction (mean = 1.87, sd = 1.09). Our predictions were not borne out with respect to derivational morphology, however. A chi-square test of independence showed no significant difference in the number of derivational affixes for themes in the two constructions ($\chi^2 = 1.41, df = 1, p = 0.24$). Thus, for one of the measures of complexity we proposed, the prediction was supported: the prepositional dative typically occurs with somewhat longer themes than those appearing in the double object construction.

For prediction (iv), we returned to our original MSN data to look at the direction relationship between complexity and saliency. In our material, all recipients were expressed as pronouns, which meant that we could not investigate predictions (iv a.) and (iv b.) directly. However, using the number of syllables in the theme noun as a measure of complexity, we could test prediction (iv c.). We would a priori expect a higher saliency for the NP-NP construction and a lower saliency for the NP-PP construction as the number of syllables in the theme goes up, for the special case when the recipient is expressed as a pronoun. The reason is that the NP-NP construction places the theme in the final position while the NP-PP construction places the recipient in the final position. Thus, if the recipient is a short pronoun, we would (following Behaghel’s law) expect a lower saliency for the NP-PP construction and an increased saliency for the NP-NP construction, as the length of the theme increases. Using Spearman’s rank correlation test we investigated – for each verb (gave, promised, faxed, etc.) and each filler in the theme slot (money, house, car, etc.) – the correlation between the saliency in each of the constructions with the number of syllables in the theme filler. A portion of the data is exemplified in table 6 below.

**TABLE 6 ABOUT HERE**

The results showed a significant negative correlation between saliency for the prepositional dative and the number syllables in the filler ($S = 44948.45, p = 0.003, \rho = -0.38$). There was no statistically significant correlation for the double object construction ($S = 28486.43, p = 0.35, \rho = 0.12$).

Prediction (iv c.) was in other words partially confirmed, in accordance with Behaghel’s law: increasing the length of the theme NP lowers the saliency score for the prepositional dative construction when the recipient is expressed as a pronoun. However, there was no corresponding increase in saliency for the double object construction. The reason for this result might be that the overall saliency for the double object construction with a pronoun recipient is relatively high to start with, resulting in a ceiling-effect.

In summary, our preliminary results illustrate that there might be some kind of relationship between our measure, complexity, and general hypotheses such as Behaghel’s law, but that the relationship is indirect and moderated by other factors. We believe that the measure occupies a niche which is more closely related to pragmatics than to processing-complexity in the strict sense. Although it is an open question whether such issues can really be separated (especially in a construction-grammar framework), we think that our measure sheds light on an interesting aspect
of grammar precisely because it deals with issues that lie on the border between syntax, lexicon and cultural conventions.

5 Discussion of results

The dative alternation has received extensive attention in linguistic theorizing and analysis; see e.g. Levin (1993); Goldberg (1995); Arnold et al. (2000); Stefanowitsch and Gries (2003); Gries and Stefanowitsch, (2004); Bresnan et al. (2007); Rappaport Hovav & Levin (2008); for further references see e.g. Levin (1993, p. 45). Much of the debate has focused on which verbs exhibit this alternation (Levin, 1993, pp. 47–48). Goldberg (1995) shifts the focus to the interaction between verbs and constructions, something which Stefanowitsch and Gries follow up on with their collostructional analysis. The basic conclusion in Goldberg (p. 151) is that the construction’s semantics is associated with a successful transfer between willing participants. Bresnan et al. (2007) note that double object constructions which are not expected based on linguistic theorizing are found on the Web, but typically with pronominal recipients. Using a range of different types of data, they show that in addition to syntax and semantics, factors such as animacy (of recipient and theme) and discourse accessibility are also involved in the dative alternation. Since pronouns are typically given elements, Bresnan et al. (2007) conclude that pragmatic factors and usage lead to a greater syntactic flexibility with the dative alternation than can be predicted from intuitions and theory alone. Based on this, they criticize what they see as reductionist tendencies, whereby attempts are made to explain the dative alternation by a single factor.

The results presented in the present paper corroborate and extend those reported in Bresnan et al. (2007). We have shown that in addition to the verb and the recipient, the theme also plays a role in deciding the type of construction the dative alternation. Our measure of constructional saliency highlighted the fact that we find interesting variations within each construction and within each verb with respect to how salient specific fillers are, a fact that is not readily observed with corpus frequencies. Furthermore, it should be pointed out that since we have used un-annotated data, only verb forms have been used, something which probably makes the task of distinguishing verb-construction-filler saliency harder, since we would expect subtle sense-specific preferences. Despite such difficulties, our measure was able to capture associations in the data that are entirely expected from a pragmatic point of view. That the theme variable was not kept in the final mixed effects model discussed in 4 can probably be explained by the large explanatory potential offered by the combination of verb and construction. Recall that the mixed effects model indicated that there was substantial variation both between and within each verb, i.e. the random effect. The fixed effects could only explain what variation was left over or beyond that which was accounted for by the individual verbs. Thus, what the mixed effects model told us was that there was a consistent effect of the construction variable across all verbs, even after the individual differences between verbs had been adjusted for. This no doubt left less explanatory potential for the theme variable, especially since some of the more frequent and general verbs typically occur with a wider range of theme NPs. Also, the saliency measure itself incorporates frequency information about themes as part of specific constructs, which further reduces the explanatory potential for the theme variable. Since our results are closely aligned with those found in previous studies on the psychological status of collocations, it seems warranted to conclude that our measure has some psychological plausibility, and that our results have some real bearing on the interpretation of the English dative alternation. We interpret the results above as supporting the arguments in Bresnan et al. (2007) that pragmatics play an important role in the dative alternation. We extend their findings by including information about the type of theme as well as verb and construction,
finding that all themes, in various degrees, interact with the choice of construction. As such, our results modify somewhat the claims made by Rappaport Hovav & Levin (2008) who place a heavy emphasis on verb semantics. Here it should once again be pointed out that our results are based on an empirical association measure which draws on a vast collection of text, thus providing a very good coverage of the full range of usage variation, something we consider important for any usage-based approach to linguistic theorizing.

6 Conclusions

We have presented a new measure of association between constructions, verbs, and arguments intended for Web data. Based on previous studies, it was argued that this measure can be interpreted as the saliency of the specific configuration of fillers given the construction. As a case study we have investigated the saliency of specific verb-filler configurations in the English dative construction, finding that our measure highlights a wider range of pragmatic influences on the dative alternation, and that it reveals that the lexical fillers themselves seem to have an influence on the choice of construction. We began the article by discussing the problems of measuring associations in linguistic data drawn from the Web. For all its complications, we believe that the Web holds information of relevance to linguists, and hope to have shown that our proposed measure for such data have brought some deeper insights about a much-studied grammatical phenomenon, even if the deepening might not be more than a barleycorn.

References


### Tables and figures

**Table 1** Terminology used for the dative alternation.

<table>
<thead>
<tr>
<th>Prepositional dative construction</th>
<th>… gave [the book] [to her]</th>
<th>V NP PP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Double object construction</td>
<td>… gave [him] [the book]</td>
<td>V NP NP</td>
</tr>
<tr>
<td>Recipient</td>
<td>… gave [the book] [to her]</td>
<td>V NP PP</td>
</tr>
<tr>
<td>Theme</td>
<td>… gave [him] [the book]</td>
<td>V NP NP</td>
</tr>
<tr>
<td>Construct</td>
<td>The librarian gave her the book</td>
<td>Specific instantiation</td>
</tr>
</tbody>
</table>
Table 2: Example 2 by 2 table for collostructional analysis. $O$ refers to observed frequencies and subscripts to cell numbers.

<table>
<thead>
<tr>
<th>Construction</th>
<th>Verb $v$</th>
<th>$\neg$ Verb $v$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$C$</td>
<td>$O_{11}$</td>
<td>$O_{12}$</td>
</tr>
<tr>
<td>$\neg C$</td>
<td>$O_{21}$</td>
<td>$O_{22}$</td>
</tr>
</tbody>
</table>

Table 3: Fixed effects from the mixed effect model, with estimated $p$-values and 95% confidence intervals. The prepositional dative construction is associated with a far lower saliency than the double object construction. Note that the difference between the constructions is much greater than the uncertainty in the estimates. The estimated $p$-values were created with the pvals.fnc function from (Baayen, 2008b).

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Est. coef.</th>
<th>SE</th>
<th>$t$-value</th>
<th>$p$MCMC</th>
<th>95%HPD C.I.</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>11.68</td>
<td>1.13</td>
<td>10.33</td>
<td>0.0001</td>
<td>9.83, 13.94</td>
</tr>
<tr>
<td>Construction: PP</td>
<td>-5.75</td>
<td>1.07</td>
<td>-5.36</td>
<td>0.0001</td>
<td>-7.98, -3.51</td>
</tr>
</tbody>
</table>

Table 4: Random effects from the mixed effect model, with 95% confidence intervals. Note that the within-verb standard deviation (5.77) is larger than the between-verb standard deviation (3.28) by a good margin.

<table>
<thead>
<tr>
<th>Groups</th>
<th>Std.Dev.</th>
<th>95%HPD C.I.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Verb (Intercept)</td>
<td>3.28</td>
<td>0.00, 3.76</td>
</tr>
<tr>
<td>Residual</td>
<td>5.77</td>
<td>5.21, 6.95</td>
</tr>
</tbody>
</table>

Table 5: Raw and relative (in parentheses) frequencies of types of recipient realization with the prepositional dative and the double object construction for the verb gave in COCA. The observed difference is statistically significant with a high effect size.

<table>
<thead>
<tr>
<th>Construction</th>
<th>Pronoun</th>
<th>Full NP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prepositional dative (NP-PP)</td>
<td>2 (0.03)</td>
<td>75 (0.97)</td>
</tr>
<tr>
<td>Double object (NP-NP)</td>
<td>92 (0.98)</td>
<td>2 (0.02)</td>
</tr>
</tbody>
</table>
Table 6 Table showing excerpt of the data used to measure the associations between themes, verbs, saliency for the two different constructions, and complexity as represented by the number of syllables in the theme.

<table>
<thead>
<tr>
<th>Theme (filler)</th>
<th>Verb</th>
<th>Saliency:NP-NP</th>
<th>Saliency:NP-PP</th>
<th>Theme:syllables</th>
</tr>
</thead>
<tbody>
<tr>
<td>disease</td>
<td>gave</td>
<td>20.50</td>
<td>0.08</td>
<td>2</td>
</tr>
<tr>
<td>ball</td>
<td>kicked</td>
<td>13.03</td>
<td>13.21</td>
<td>1</td>
</tr>
<tr>
<td>ball</td>
<td>threw</td>
<td>16.16</td>
<td>18.02</td>
<td>1</td>
</tr>
</tbody>
</table>

Figure 1 Plots showing the overall distribution of saliency, as well as by construction type.
Figure 2 The association between verbs and themes as measured by mean saliency. The figure shows PCA biplots for dimensions 1-3, accounting for a total of 65.5% of the variation. Items clustered near the center of the plot do not display any specific, strong associations.
Figure 3 Mean verb saliency for the double object and prepositional dative constructions. The dashed gray line is a linear regression line of the vertical axis on the horizontal axis. As the plot shows, there are clear differences between the saliency scores of the two alternations.
Figure 4 Constructional saliency for the dative alternation by verb with added nonparametric smoothing lines. For most verbs, there are small or no differences between the alternations. However, gave, threw, showed, and sent stand out as having a noticeable preference for the double object construction.
Figure 5 Box and whiskers plot for gave with the two alternation patterns. Outliers for the prepositional dative construction have been identified. The verb give normally prefers a double object construction; except when the theme is money, or when the theme is flowers and the recipient female.
Figure 6 Diagnostic plots for the mixed effect model. The top left panel shows residuals vs. fitted, which indicate that the fit assumption of constant variance is reasonable. The top right and lower right panels show normal quantile-quantile plots for the residuals and for the random effect, none of which indicate any serious problems. The lower left panel shows observed vs. fitted data (“Saliency”), with a non-parametric regression line. Although there is considerable variation, it is evident that the predictions are not unreasonable.
Figure 7 Plot showing the partial effect of saliency on the predicted probability of the double object construction for a mixed binomial model with saliency and theme as predictors. Hexagons are cases, with density indicated by color intensity. Note the outliers near the bottom of the curve, mainly due to the influence of NP themes. The R code for the function used to generate the plot was developed by T. Florian Jaeger.

Figure 8 Syntactic trees showing an instance of the prepositional dative (a), a sentence partially sharing the prepositional dative’s surface structure (b), and an instance of the double object construction. All cases share the same number of immediate constituents, but the shared surface order of verb + NP + to in (a) and (b) could imply that (a) holds more information than (c) and hence represents increased processing complexity.