# Deriving the Structure of Variation from the Structure of Non-variation in the English Dative Alternation

## Max Bane
**University of Chicago**

## 1. Introduction

The question of how to accommodate probabilistic variability in models of grammatical knowledge has been the subject of much recent theoretical work, especially with the growing use of experimental and observational (e.g., corpus-based) data. We can distinguish two general approaches to deriving probabilistic outcomes. The first, which I will call the *gradient grammar* approach, incorporates gradient, real-valued parameters directly into the grammar as objects whose values must be fitted or learned. Examples include variable rule models (Labov 1969, Cedergren & Sankoff 1974), stochastic OT (Boersma 1997), maximum entropy models (Hayes & Wilson 2008), and noisy harmonic grammar (Boersma & Pater 2008). In contrast, a second approach, *grammar sampling*, posits random sampling from a set of parameter values for some categorical grammar, with membership in the set determined by learned categorical restrictions. Anttila's (1997, 2007) partially ordered constraints and Adger's (2006; also Adger & Smith 2010) combinatorial minimalism are examples of this approach.

This paper describes an application of grammar sampling to predict English speakers' nondeterministic variability in arranging the arguments of a certain class of dative, ditransitive verbs, using a version of Anttila's (1997, 2007) partially ordered constraints model. An otherwise categorical, non-variable OT model of the alternation allows us to make variable predictions by supposing that speakers select a grammar (total constraint ordering) uniformly at random from some restricted set of possible grammars; thus the probability of an outcome corresponds to the relative number of possible grammars that generate it. The model deviates from Anttila's original proposals in that the form of the restriction on possible grammars is not a stratified hierarchy or partial order, but a set of elementary ranking conditions (ERCs; Prince 2002), which describe exactly the information that a learner can infer from observed input-output mappings in OT. Though ERCs are more complicated than stratified hierarchies or partial orders, this approach retains the essential simplicity of Anttila's grammar sampling model: in the move from a categorical framework (i.e., standard OT) to a variable one, we need introduce no new representational entities such as probabilities or other real-valued parameters; we merely put to new use objects that were already necessary for the analysis of categorical behavior. Furthermore, in the case considered here, we have the remarkable result that the small subset of the data that shows *no* variation (~7% of tokens) yields a restriction on possible rankings, in the language of ERCs, that is sufficient to predict the variable data with high accuracy.

We wish to predict the distribution of usage of three alternating syntactic constructions, the *dative alternation*:

(1)    The DOUBLE OBJECT construction, V $NP_{goal}$ $NP_{theme}$: *give [my sister] [the old book]*.

(2)    The PREPOSITIONAL construction, V $NP_{theme}$ [$_{PP}$P $NP_{goal}$]: *give [the old book] [to my sister]*.

(3)    The HEAVY NP SHIFT construction, V [$_{PP}$P $NP_{goal}$] $NP_{theme}$: *give [to my sister] [the old book]*.

In each case we have a ditransitive verb and its two (non-subject) argument NPs, one of which, the *theme* (also "patient" or "direct object"), denotes some entity being transferred (perhaps in an

abstract sense, depending on the verb and context), and the other of which, the *goal* (also "recipient" or "indirect object"), denotes the recipient or destination.[1] In the *double object construction* (1), the goal NP immediately follows the verb, and is in turn immediately followed by the theme NP. The *prepositional construction* (2), by contrast, places the theme NP immediately after the verb, while the goal NP follows as the complement of a prepositional phrase headed by an appropriate preposition. The third construction, the so-called *heavy NP shift* (3), has the same structure, but with the order of the PP and theme NP reversed. A fourth logically possible construction would be of the form V $NP_{theme}$ $NP_{goal}$, like the double object but with theme and goal reversed, e.g., *give it me*. This ordering has historical relevance (Curme 1928, Cassidy 1937), and appears in restricted contexts in some non-standard dialects (Murphy 2007); nonetheless it remains quite rare or ill-formed overall, and does not occur in the corpus used here. Its absence will be held as a fact to be derived, rather than assumed *a priori*.

An extensive literature aims to explain the observation that for some combinations of verbs and arguments, constructions (1–3) alternate freely as apparent paraphrases, while with other verbs or arguments such alternation is restricted or outright ungrammatical. In general, the prepositional construction is available for any verb-argument combination. Some verbs, however, resist the double object construction altogether, e.g.:

(4)    * They returned/donated/revealed [their friends] [the books].

Other verbs permit the heavy NP shift construction with some arguments (typically phonologically "large" or "heavy" themes), but not with others:

(5)    a.  I'm going to reveal [to you] [everything I've learned in this business].

       b.  * I'm going to reveal [to you] [it].

Furthermore, in large corpora of spontaneous speech or writing (cf. Collins 1995, Gries 2003, Bresnan et al. 2007, Bresnan & Nikitina 2007), the usage facts are highly gradient, with verb-argument combinations exhibiting one, two, or all three constructions in systematically different proportions of frequency. It is such usage-frequency data that we seek to address.

In the remainder of this paper, section 2 describes the corpus from which the present data derive, and section 3 follows with a brief presentation of Anttila's (2008) prosodic OT analysis of the alternation, which forms the underlying categorical model from which possible grammars are sampled. Section 4 then shows how the variable usage frequencies are predicted, and quantifies the overall fit of the model. Section 5 offers some concluding remarks.

## 2. The Blogspot corpus

The usage frequency data come from 1,601 sentences of informal written English sampled from free, online weblog and homepage hosting sites (about 80% from `blogspot.com` and 20% from `livejournal.com`, `geocities.com`, and other sources). The corpus is reported on by Anttila (2008) and Anttila et al. (2010). It consists of approximately 100 finite dative constructions found for each of the following sixteen ditransitive verbs:

(6)    SAMPLED VERBS:
       assign, award, bring, give, offer, promise, administer, bequeath, concede, convey, deliver, donate, explain, guarantee, recommend, reveal

Anttila et al. report that the sampling procedure was to include the first 100 hits containing the relevant ditransitive context from the results of a search for each verb. Several auxiliary frames (*will give, have given,* etc.) were also included in the search queries, in order to better restrict the results to verbal contexts. Each verb usage sampled in this way was classified as an instance of either the double object, prepositional, or heavy NP shift construction. Table 1 shows the overall corpus frequencies, across all verbs, of the three construction types. It can be seen that at this gross level, the prepositional construction is by far the most common.

---

[1]An agentive subject of the verb is of course also usually present, but will not be relevant here.

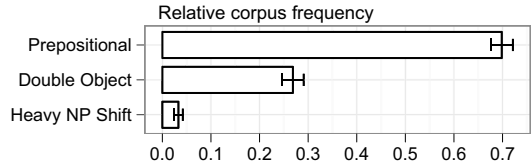| Construction | Count | Rel. freq. | Std. err. |
|---|---|---|---|
| Prepositional | 1118 | 0.698 | 0.0114 |
| Double object | 430 | 0.269 | 0.0110 |
| Heavy NP shift | 53 | 0.033 | 0.0045 |



**Table 1:** The overall corpus counts, relative frequencies, and multinomial standard errors in the relative frequencies of the three dative constructions. Error bars indicate twice the standard error.

The usual advantages and disadvantages of internet corpora apply: the broad coverage afforded by the internet is useful for uncovering rare forms like the heavy NP shift, but it also introduces the potential for contamination by non-native speakers and systematic differences between dialects.[2] Still, our concern is that the data contain robust and systematic patterns, the successful prediction of which is nontrivial.

## 3. Anttila et al.'s prosodic OT model

A number of semantic and pragmatic factors have been proposed to account for the categorical facts of alternation and non-alternation (e.g., by Green (1974), Oehrle (1976), Goldsmith (1980), Levin (1993), Krifka (1999), and others). In addition, though more rarely, phonological variables have also been advanced as relevant, either beside other semantic considerations, or by themselves as the primary determinants of the alternation (Green 1974, Grimshaw 2005). This paper starts from the latter sort of analysis, in the form of an optimality theoretic (OT; Prince & Smolensky 1993) model proposed by Anttila (2008, et al. 2010) in which the acceptability of a construction for a given choice of verb and arguments is determined by the interaction of a ranked set of constraints sensitive to various prosodic properties of the linguistic forms in question.

This section reviews this OT model of the dative alternation, including: its representation of the input (section 3.1) and output (section 3.2) forms that it establishes a mapping between, and how they correspond to the available corpus observations; the set of constraints and a proposed *a priori* restriction on their ranking (section 3.3).

### 3.1. Representation of inputs

A model input is the specification of three variables: the number of prosodic feet in the verb (1 or 2+), and the number of lexical, word-level, primary stresses in the goal and theme (each 0, 1, or 2+). We will follow Anttila (2008) in using a mnemonic representation, summarized in (7).

(7)   VERB(GOAL, THEME), where:

    a. VERB ∈ {give, donate},

    b. GOAL ∈ {her, my sister, my little sister},

    c. THEME ∈ {it, the book, the old book}.

The convention will be that *give* stands in for any 1-foot verb, and *donate* for any 2+-foot verb (according to Grimshaw's (2005) analysis; see below). For the postverbal arguments, *her* stands in for any stressless goal, *my síster* for any goal with one lexical word-level stress, and *my líttle síster* for any with two or more such stresses. Similarly, *it*, *the bóok*, and *the óld bóok* represent any theme with 0, 1, or 2+ lexical stresses. So for example *give(her, the book)* describes any token in the corpus in which a dative construction was formed with a 1-foot verb, a stressless goal, and a theme with one lexical stress.

Figure 1 shows the corpus distribution of usage of each construction type as a function of the input, encoded in this way. For example, it was found that a two-foot verb with a 1-stress theme and stressless goal (*donate(her, the book)*) occurred in prepositional constructions 48.7% of the time, in double object constructions 38.7% of the time, and in the heavy NP shift 12.6% of the time. It is these conditional
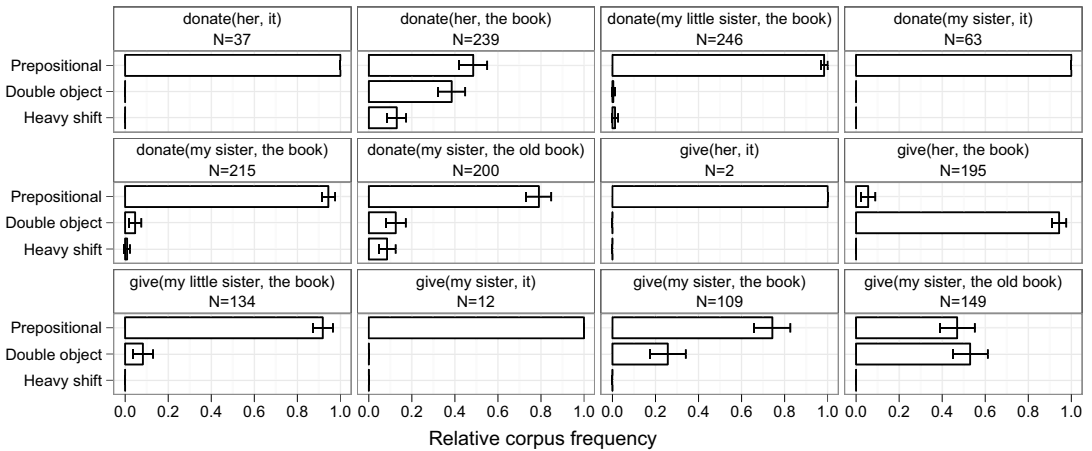
**Figure 1:** The empirical proportions of each construction type, conditioned on input. Error bars indicate twice the multinomial standard error in proportions.

frequency data that we aim to account for—to predict, for the given prosodic characteristics of a verb and its arguments, how often each construction will be used.

In determining the number of feet in a verb, Grimshaw's (2005) prosodic constituency analysis is adopted. Anttila et al. (2010) provide the details of how this analysis is applied to the sixteen corpus verbs (6), which break down for footedness as follows:

(8)  a.  ONE-FOOT VERBS ($n = 6$):
        assign, award, bring, give, offer, promise

     b.  TWO-FOOT VERBS ($n = 10$):
        administer, bequeath, concede, convey, deliver, donate, explain, guarantee, recommend, reveal

### 3.2. Representation of outputs

For each input, the model entertains eight possible output forms. These correspond to four possible linearizations of the post-verbal arguments crossed with two possible prosodic phrasings of the VP. The linearizations considered are those of the three attested constructions—prepositional, double object, and heavy NP shift—plus the fourth, unattested (in this corpus) *give it me* ordering. These constructions will be represented as $P$, $D$, $H$, and $\#$, respectively, in notational contexts. For each linearization, the model distinguishes between two possible phonological phrasings: one in which the verb is prosodically parsed together with the linearly following argument, and one in which the verb occupies its own prosodic phrase. It is assumed that the final argument must always have its own phrase. Thus the choice is between a phonological rendering in which there are two distinct phonological phrases (that containing the verb plus first argument, and that of the final argument), and one in which there are three phrases (one each for the verb, theme, and goal). Notationally, the two-phrase output will be represented with a subscript 2 (e.g., $P_2$, $D_2$, $H_2$), and the three-phrase output with a subscript 3 ($P_3$, etc.). For example, the eight possible outputs for the input *give(my sister, the book)* are listed in (9).

(9)  CANDIDATE OUTPUTS FOR INPUT *give(my sister, the book):*
     $P_3$:  (give) (the bóok) (to my síster)     $P_2$:  (give the bóok) (to my síster)
     $D_3$:  (give) (my síster) (the bóok)        $D_2$:  (give my síster) (the bóok)
     $H_3$:  (give) (to my síster) (the bóok)     $H_2$:  (give to my síster) (the bóok)
     $\#_3$:  (give) (the bóok) (my síster)       $\#_2$:  (give the bóok) (my síster)

Due to the textual nature of the Blogspot corpus, it is not possible to observe speakers' intended phonological phrasings. Thus the usage frequencies reported (e.g., in Figure 1) are in fact collapsed across the possible phrasings, as are the frequency predictions developed below.[3] That is, the model

---

[3]It is even possible that, in textual media, speakers *don't have* intended phonological phrasings, and that whatever procedure they use to generate utterances collapses the possibilities.

| Constraint | Penalizes... |
|---|---|
| SYNTAX | a goal NP not adjacent to its head (verb or preposition); $\#$ linearization. |
| WRAP-XP | an XP not contained in a phonological phrase; $D_3$. |
| STRESS | each stressless phonological phrase. |
| *CLASH | each lexical stress beyond the first in a phonological phrase. |
| *TERNARY | a phonological phrase containing 3 or more feet; $P_2$, $D_2$, $H_2$, $\#_2$ for *donate*. |
| *TO | prepositions; $P_2$ and $P_3$. |
| STRESSTOSTRESS | (STS) each lexical stress in a non-final phonological phrase. |
| *PHRASE | each phonological phrase. |
| FOCUS(GOAL) | a nonfinal goal. |
| FOCUS(THEME) | a nonfinal theme. |

**Table 2:** Summary of constraints.

will make its predictions partly on the basis of a representational distinction that is not present in the data against which the predictions are compared, by pooling the predictions across that distinction. This "hidden layer" of prosodic information that the model's internal computations are sensitive to, but which is not present in the results that the model returns is a part of the model's strictly internal structure, and the mapping from this structure to its predictions (i.e., the pooling of its apportioned frequency predictions across different prosodic phrasings) is nontrivial. When referring to a linearization collapsed across phrasings in this way, a bare, unsubscripted symbol ($P$, $D$, $H$, $\#$) will be used.

### 3.3. Constraints

Table 2 briefly describes the output environment or input-output relationship penalized by each constraint. For a more detailed presentation, with motivating arguments for each constraint, see Anttila 2008 and Anttila et al. 2010. Evaluation of the SYNTAX constraint assumes that the goal NP forms a constituent with the verb (in $D$, $\#$) or preposition (in $P$, $H$), with the latter being the head of that constituent (cf. Larson 1988). WRAP-XP and *PHRASE have been proposed by Truckenbrodt (2007). FOCUS(GOAL) and FOCUS(THEME) reflect an assumption that English marks a focused construction by placing it under phrasal stress, which is final. Similarly, STRESSTOSTRESS requires that lexical stress coincide with final phrasal stress.

The tableau in (10) demonstrates how the eight candidate outputs are evaluated for the input *donate(my síster, it)*. In this case, one of the outputs, $H_2$, is harmonically bounded—its constraint violations are a strict superset of those of $D_2$, and thus it cannot be optimal under any constraint ranking.

(10)

| Input: donate(my síster, it) | STRESS | SYNTAX | WRAP-XP | *CLASH | FOC(THEME) | FOC(GOAL) | *TERNARY | STS | *TO | *PHRASE |
|---|---|---|---|---|---|---|---|---|---|---|
| $P_3$: (donate) (it) (to my síster) | * | | | * | * | | | | * | *** |
| $H_3$: (donate) (to my síster) (it) | * | | | * | | * | | * | * | *** |
| $\#_3$: (donate) (it) (my síster) | * | * | | * | * | | | | | *** |
| $D_3$: (donate) (my síster) (it) | * | | * | * | | * | | * | | *** |
| $P_2$: (donate it) (to my síster) | | | | * | * | | * | | * | ** |
| × $H_2$: (donate to my síster) (it) | * | | | ** | | * | * | * | * | ** |
| $\#_2$: (donate it) (my síster) | | * | | * | * | | * | | | ** |
| $D_2$: (donate my síster) (it) | * | | | ** | | * | * | * | | ** |

Furthermore, in order to rule out the possibility of $\#$ (*give it me*) orderings, Anttila (2008) proposes an *a priori* restriction on possible constraint rankings, in the form of a stratified hierarchy:

(11)  {SYNTAX, WRAP-XP, STRESS}  $\gg$  {*CLASH, FOC(THEME), FOC(GOAL), *TERNARY, STS, *TO, *PHRASE}

| Model | Log likelihood | Class. acc. | Freq. corr. | $R_M^2$ | $R_N^2$ |
|---|---|---|---|---|---|
| *Non-grammatical* | | | | | |
| Fully saturated | -25.40 | 81.20% | 100% | 98.37% | 99.46% |
| Overall probabilities | -486.79 | 69.83% | 73.99% | 68.77% | 86.07% |
| Uniform probabilities | -1558.88 | 12.90% | — | — | — |
| *Grammar-sampling* | | | | | |
| Unrestricted | -1277.57 | 29.11% | 37.94% | — | — |
| Anttila's strata | -612.33 | 68.08% | 73.42% | 52.07% | 70.79% |
| Categorical ERCs (2-phrases) | -273.62 | 79.14% | 92.21% | 78.58% | 89.64% |
| Categorical ERCs (3-phrases) | $-\infty$ | 69.83% | 84.49% | — | — |
| Categorical ERCs (disjunction) | **-264.30** | **80.64**% | **92.40**% | **79.31**% | **90.05**% |

**Table 3:** Model statistics: log likelihood of the corpus, classification accuracy, correlation of maximal independent subsets of predicted and observed frequencies, and pseudo-$R^2$ values. The latter are with respect to the uniform probability model as a base for the non-grammatical models, and to the empty restriction as a base for the grammar-sampling models.

That is, SYNTAX, WRAP-XP, and STRESS must outrank all other constraints, but may be ranked freely with respect to each other. Subject to this restriction, the only viable output in (10) would be $P_2$, and indeed, all corpus instances of *donate(my sister, it)* are prepositional, shown in Figure 1. While this ranking restriction will *not* be assumed here, it makes a useful point of comparison against the grammatical restrictions derived below from the non-variable data.

## 4. Modeling variation in the Blogspot corpus

Recall that the data consist of counts of observed outcomes, from three possibilities ($P$, $D$, $H$), conditioned on an input form (cf. Figure 1). Explanatory completeness compels us to assume that there is a fourth possible outcome as well ($\#$), which has zero observed counts in all cases. We assume that, for each input, the observed output constructions are multinomially distributed, so that there is some true, underlying probability of observing each construction, conditional on the input. The job of any model is to generate, for each input, a conditional distribution over the outputs (i.e., four probabilities that sum to unity, one for each output). The model thus assigns a likelihood to the whole corpus of observations, as the product of multinomial likelihoods for each input.

### 4.1. Non-grammatical models

It is useful first to consider some non-grammatical models of the corpus observations as points of comparison for the grammatical models. The simplest, vacuous model assigns uniform $\frac{1}{4}$ probability to each construction for all inputs. A slightly more complex model might predict that each construction occur at its corpus-wide, overall rate in all cases. The maximally complex, or "fully saturated" model simply predicts all constructions to occur with exactly their empirical frequencies, conditional on each input. Given the present classification of inputs and outcomes, it assigns the highest possible probability to the corpus, but generalizes nothing. Thus we have three models to compare grammar-sampling against: two lower bounds (uniform and overall probabilities) and an upper bound (fully saturated).

We compare models through five statistics. First is the *log likelihood*, which is simply the natural logarithm of the probability assigned by the model to the corpus. Second is the *classification accuracy* of the model, which is the proportion of corpus tokens that would be correctly classified if each input always resulted in the construction that the model considers most probable for that input (n.b., the most probable of four outcomes may have probability $< 0.5$).[4] Third is the *correlation between predicted and observed* relative frequencies of conditional outcomes.[5] The remaining two statistics are varieties of *pseudo-$R^2$* values sometimes used in the statistics literature for categorical data: McFadden's $R^2$ (here,

---

[4]Ties are resolved randomly, and reported figures are the mean result.

[5]In order to avoid trivial inflation of the correlation, it is calculated only on outcomes whose empirical frequencies are logically independent of each other. For example, the empirical construction frequencies for

$R_M^2$) and Nagelkerke's $R^2$ (here, $R_N^2$). Both are meant to be roughly analogous to the $R^2$ of a linear regression—a value between 0 and 1 indicating the proportion of variance explained—but the notion of explained variance is not directly translatable to count data, hence the term *pseudo-$R^2$*; both metrics compare a model's corpus likelihood to that of some "null," base model. In this case, we may compare the above non-grammatical models to the simple uniform probability model. The first half of Table 3 gives these five statistics for these models.

## 4.2. Grammar-sampling models

We model the corpus as having been generated by a grammar-sampling process: for each input (choice of verb and arguments), the corresponding output construction is produced by randomly choosing a total constraint ranking with uniform probability from a delimited set of rankings $G$. Each potential input-output mapping $i \mapsto o$ is consistent with (i.e., optimal under) some finite number $r(i \mapsto o)$ of the $k!$ possible rankings, for $k$ constraints.[6] Let us call the proportion of such rankings the *unconditional ranking volume* of that mapping: $v(i \mapsto o) = r(i \mapsto o)/k!$. If we restrict our count to only include rankings *in $G$* that are consistent with $i \mapsto o$, and we call this restricted count $r(i \mapsto o|G)$, then the *conditional ranking volume given $G$* is $v(i \mapsto o|G) = r(i \mapsto o|G)/|G|$. Note that for any input, $\sum_o v(i \mapsto o) = 1$, where the summation is over all four possible outputs; and for fixed $G$, $\sum_o v(i \mapsto o|G) = 1$. If a ranking is selected uniformly at random from $G$ (with replacement), then it is clear that for any given input the conditional probability of each output is simply $v(i \mapsto o|G)$. To compute $r(i \mapsto o|G)$ and $|G|$ without brute-force enumeration of factorially many rankings, we use algorithms developed by Riggle (2010).

We consider several choices of $G$, the set of sampled rankings. The most basic, "null" grammar-sampling model imposes no restrictions on $G$, so that all $k!$ rankings are possible (here $k = 10$, so $k! =$ 3,628,800). As Table 3 shows, this is a poor model of the corpus, giving a very low log likelihood, and a classification accuracy much lower than if all observations were simply predicted to be prepositional. This is not surprising, since, for instance, the unattested # *give it me* construction may frequently occur with no restriction on rankings. We can instead sample from just those rankings that are consistent with Anttila's proposed stratified hierarchy (11), which, among other things, excludes all grammars that generate #. This model's prediction of the data improves significantly, but remains somewhat worse than giving each outcome a fixed, overall probability.

A much better restriction on rankings is found by considering the information implied by those cases that show *no* variation in the corpus. There are four non-variable inputs, all of which appear only in the prepositional construction (cf. Figure 1): *donate(her, it), donate(my sister, it), give(her, it), give(my sister, it)*, together accounting for 114 tokens, or about 7% of the corpus. In a classical OT framework, these are the only observations that could possibly be self-consistent, since variation on a single input necessarily implies at least two contradictory rankings. From a grammar sampling point of view, such contradictory implications are fully expected, and are the reason for modeling grammatical knowledge as a *set* of rankings, rather than as any single ranking. But it remains the case that any non-variable inputs will tend to be more informative than variable ones, by virtue of having fewer possible outcomes (i.e., just one) and therefore being consistent with stronger restrictions on possible rankings. For example, if input $i$ shows variation between outputs $o_1$ and $o_2$, then we must admit *both* the rankings consistent with $o_1$ *and* those consistent with $o_2$ as possibilities; if, on the other hand, we were confident that $o_2$ never occurred, it would be sufficient to consider just the rankings implied by $o_1$.

Prince (2002) describes a formalism for reasoning about the rankings implied by observed input-output mappings. Consider the tableau in (10). What could we infer about the possible ranking of constraints if $P_3$ were the observed output? If we compare the violation profile of $P_3$ to that of $H_3$, we see that these two candidates differ in their evaluation by three constraints: FOCUS(THEME), FOCUS(GOAL), and STS. $P_3$ is preferred by FOCUS(GOAL) and STS, while $H_3$ is preferred by FOCUS(THEME). Thus, according to the rules of OT, observing $P_3$ as the optimal output implies that at

---

*donate(her, the book)* are $P = 48.7\%$, $D = 38.7\%$, $H = 12.6\%$, $\# = 0\%$, so that any choice of three fully determines the fourth; thus one of them (arbitrarily chosen), together with the corresponding predicted frequency, must be excluded from the correlation.

[6]Outputs are collapsed for phonological phrasing. That is, $r(i \mapsto P) = r(i \mapsto P_2) + r(i \mapsto P_3)$, etc.

least one of FOCUS(GOAL) and STS must outrank FOCUS(THEME). This inference can be represented by an *elementary ranking condition* (ERC): $erc(P_3 \succ H_3) = \langle e, e, e, e, \text{L}, \text{W}, e, \text{W}, e, e \rangle$. This is a length-10 vector of symbols $\{\text{W}, \text{L}, e\}$ in which the symbol at the $i$th component indicates the preference of the $i$th constraint (as ordered in tableau (10)): a W means that the constraint favors the "winner" (here, $P_3$), an L means that the "loser" ($H_3$) is favored, and $e$ means that neither candidate is preferred. The interpretation of any ERC is that it describes the set of all rankings in which *at least one* of the constraints marked by W outranks *all* of those marked by L. If there are $n$ candidates in a tableau, the observed output generates $n - 1$ ERCs, one for each winner-loser pair (e.g., $erc(P_3 \succ \#_3)$, $erc(P_3 \succ D_3)$, etc.). The entailed set of possible rankings is then the intersection of the sets implied by each of these ERCs. If we observe the optimal outputs for several inputs (i.e., several tableaux), then the set of possible rankings is simply the intersection of the sets described by all generated ERCs.

This method of grammatical inference assumes that there is exactly one optimal (i.e., observed) output form for each input. If an input ever has multiple observed outputs whose violation profiles are distinct, then the set of rankings consistent with one must necessarily be disjoint from the set of rankings consistent with another, so that their intersection is empty. Thus the non-variable inputs, which have just one observed output each, are a natural first place to look for a restriction on possible grammars, since they permit the most straightforward application of standard methods of grammatical inference in OT. The present case is complicated, however, by the fact that even these non-variable observations are ambiguous for phonological phrasing; an observed prepositional construction could be either $P_2$ (two phrases) or $P_3$ (three phrases), each yielding different inferences (ERCs). We could simply assume that all constructions are 2-phrase, or all 3-phrase, though this is rather unsatisfying, as we have no independent reason to believe so. Interestingly, though, the 2-phrase assumption does assign high probability to the corpus, much higher than using the overall construction frequencies as constant predicted probabilities; see Table 3. The 3-phrase assumption, on the other hand, yields ERCs that describe a set of rankings under which some corpus observations are actually impossible (*donate(my sister, the book)* $\mapsto H_3$ and *donate(my little sister, the book)* $\mapsto D_3, H_3$). That is, assuming all 3-phrase outputs is actually inconsistent with some of the data, and thus assigns zero probability to the corpus.

Fortunately, the logic of ERCs gives us a more principled way to accommodate the ambiguity of our observations. We can apply the *ERC-disjunction operator* (see Prince (2002) for details) to generate "weakened" inferences that specify the most restrictive ranking conditions that would be consistent with either of the ambiguous outputs. Briefly, the disjunction of two ERCs is a third ERC than contains a W wherever *either* of the two ERCs have a W, an L wherever *both* ERCs have an L, and $e$ everywhere else. As Table 3 shows, sampling from the disjunctive set of grammars consistent with either phrasing of the non-variable outputs approaches the theoretical maximum classification accuracy of the fully saturated model, and gives the highest corpus likelihood of the grammatical models considered. We can take this to make two qualitative predictions about phonological phrasing in the dative alternation: first, since the disjunctive approach does better than assuming all 2- or 3-phrases, we expect there to be variation in the phrasing of outputs for at least some inputs; and second, since it does only slightly better than the all 2-phrase assumption, we might expect most outputs to be realized with two phonological phrases.

## 5. Conclusion

The primary thrust of this paper has been to demonstrate that in a grammar sampling model of the English dative alternation, the categorical facts contain the seed of the variable facts. That is, the grammatical implications of what *doesn't* vary restrict the set of possible grammars to one that, when sampled from randomly, corresponds closely to the frequency structure of what *does* vary. The theoretical import of this correspondence, and its suggestion of a possible general relation between variable and categorical linguistic competence deserve further exploration. Note, though, that the set of grammars derived here from the non-variable data is by no means necessarily the best set of grammars available for modeling the corpus. It remains to establish general model-fitting and learning procedures that allow us to infer the most probable set of sampled grammars given the data (which in general may or may not contain any strictly non-variable subsets). For instance, there exists a partial order of Anttila's constraints (discoverable by a local search procedure; Bane to appear) that describes a set of rankings within which grammar-sampling assigns the corpus a probability much nearer to that of the

fully saturated model (log likelihood $\approx -80$). Nonetheless it is intriguing that so much of the necessary restriction on possible rankings can be inferred from the small non-variable subset of the alternation.

# References

Adger, David (2006). Combinatorial variability. *Journal of Linguistics* 42, 503–530.

Adger, David & Jennifer Smith (2010). Variation in agreement: A lexical feature-based approach. *Lingua* 120, 1109–1134.

Anttila, Arto (1997). Deriving variation from grammar. Hinskens, Frans, Roeland van Hout & Leo Wetzels (eds.), *Variation, Change and Phonological Theory*, John Benjamins Press, Amsterdam/Philadelphia, 35–68.

Anttila, Arto (2007). Variation and optionality. de Lacy, Paul (ed.), *The Cambridge Handbook of Phonology*, Cambridge University Press, Cambridge.

Anttila, Arto (2008). Phonological constraints on constituent ordering. Chang, Charles & Hannah Haynie (eds.), *Proceedings of the 26th West Coast Conference on Formal Linguistics*, Cascadilla Proceedings Project, Somerville, MA, 51–59.

Anttila, Arto, Matthew Adams & Michael Speriosu (2010). The role of prosody in the English dative alternation. *Language and Cognitive Processes* 25, first published on: 14 April 2010. DOI: 10.1080/01690960903525481.

Bane, Max (to appear). A grammar sampling model of variation in the English dative alternation. Cleary-Kemp, Jessica, Clara Cohen, Stephanie Farmer, Laura Kassner, John Sylak, & Melinda Woodley (eds.), *Proceedings of the 36th Annual Meeting of the Berkeley Linguistics Society*, BLS, Berkeley, CA.

Boersma, Paul (1997). How we learn variation, optionality, and probability. *Proceedings of the Institute of Phonetic Sciences of the University of Amsterdam*, vol. 21, 43–58. ROA 221.

Boersma, Paul & Joe Pater (2008). Convergence properties of a gradual learner in harmonic grammar. Manuscript. Amsterdam and Amherst, MA: Meertens Institute and University of Massachusetts Amherst. ROA 970.

Bresnan, Joan & Tatiana Nikitina (2007). The gradience of the dative alternation. Uyechi, Linda & Lian Hee Wee (eds.), *Reality Exploration and Discovery: Pattern Interaction in Language and Life*, CSLI Publications, Stanford.

Bresnan, Joan, Anna Cueni, Tatiana Nikitina & Harald Baayen (2007). Predicting the dative alternation. Boume, G., I. Kraemer & J. Zwarts (eds.), *Cognitive Foundations of Interpretation*, Royal Netherlands Academy of Science, Amsterdam, 69–94.

Cassidy, Frederic G. (1937). 'Case' in modern English. *Language* 13, 240–245.

Cedergren, Henrietta J. & David Sankoff (1974). Variable rules: Performance as a statistical reflection of competence. *Language* 50, 333–355.

Collins, Peter (1995). The indirect object construction in English: An informational approach. *Linguistics* 33, 35–49.

Curme, George Oliver (1928). Review of Jesperson *A modern English grammar on historical principles*. *Language* 4:2, 135–148.

Goldsmith, John (1980). Meaning and mechanism in grammar. Kuno, Susumo (ed.), *Harvard Studies in Syntax and Semantics*, Harvard University Department of Linguistics, Cambridge, 423–448.

Green, Georgia (1974). *Semantics and Syntactic Regularity*. Indiana University Press, Bloomington.

Gries, Stefan Th. (2003). Towards a corpus-based identification of prototypical instances of constructions. *Annual Review of Cognitive Linguistics* 1, 1–27.

Grimshaw, Jane (2005). *Words and Structure*. University of Chicago Press, Chicago.

Hayes, Bruce & Colin Wilson (2008). A maximum entropy model of phonotactics and phonotactic learning. *Linguistic Inquiry* 39, 379–440.

Krifka, Manfred (1999). Manner in dative alternation. *Proceedings of the 18th West Coast Conference on Formal Linguistics*, Cascadilla Press, Somerville, MA.

Labov, William (1969). Contraction, deletion and inherent variability of the English copula. *Language* 45, 715–762.

Larson, R.K. (1988). On the double object construction. *Linguistic Inquiry* 19, 335–391.

Levin, Beth (1993). *English Verb Classes and Alternations*. University of Chicago Press, Chicago.

Murphy, Lynne (2007). She gave it me. `http://separatedbyacommonlanguage.blogspot.com/2007/12/she-gave-it-me.html` (accessed 28 Apr, 2009).

Oehrle, Richard (1976). *The Grammar of the English Dative Alternation*. Ph.D. thesis, MIT, Cambridge.

Prince, Alan (2002). Entailed ranking arguments. *Rutgers Optimality Archive* ROA-500.

Prince, Alan & Paul Smolensky (1993). Optimality theory: Constraint interaction in generative grammar. Ms., Rutgers University and University of Colorado, Boulder.

Riggle, Jason (2010). Counting rankings. Manuscript, University of Chicago. ROA 1075.

Truckenbrodt, H. (2007). The syntax-phonology interface. de Lacy, Paul (ed.), *The Cambridge Handbook of Phonology*, Mouton de Gruyter, Berlin, 155–175.

ROA = Rutgers Optimality Archive, `roa.rutgers.edu`.

# Proceedings of the 28th West Coast Conference on Formal Linguistics

## edited by Mary Byram Washburn, Katherine McKinney-Bock, Erika Varis, Ann Sawyer, and Barbara Tomaszewicz

Cascadilla Proceedings Project     Somerville, MA     2011

## Ordering information

Orders for the library binding edition are handled by Cascadilla Press.
To place an order, go to www.lingref.com or contact:

Cascadilla Press, P.O. Box 440355, Somerville, MA 02144, USA
phone: 1-617-776-2370, fax: 1-617-776-2271, sales@cascadilla.com

## Web access and citation information

This entire proceedings can also be viewed on the web at www.lingref.com. Each paper has a unique document # which can be added to citations to facilitate access. The document # should not replace the full citation.

This paper can be cited as:

Bane, Max. 2011. Deriving the Structure of Variation from the Structure of Non-variation in the English Dative Alternation. In *Proceedings of the 28th West Coast Conference on Formal Linguistics*, ed. Mary Byram Washburn et al., 42-50. Somerville, MA: Cascadilla Proceedings Project. www.lingref.com, document #2434.