Predicting Syntax: Processing Dative Constructions in American and Australian Varieties of English

Joan Bresnan & Marilyn Ford*

November 20, 2009

ABSTRACT

The present study uses probabilistic models of corpus data in a novel way, to measure and compare the syntactic predictive capacities of speakers of different varieties of the same language. The study finds that speakers' knowledge of probabilistic grammatical choices can vary across different varieties of the same language and can be detected psycholinguistically in the individual. In three pairs of experiments Australians and Americans responded reliably to corpus model probabilities in rating the naturalness of alternative dative constructions, their lexical decision latencies during reading varied inversely with the syntactic probabilities of the construction, and they showed subtle covariation in these tasks, which is in line with quantitative differences in the choices of datives produced in the same contexts.

*For assistance with data collection we thank Gabriel Recchia, Nick Romero, and Richard Futrell. For helpful comments we thank Victor Kuperman, Anette Rosenbach, and the *Language* reviewers. We gratefully acknowledge support from the Applied Cognitive Neuroscience Research Centre at Griffith University, Australia, and from Stanford University's Vice-Provost for Undergraduate Education. All modeling and graphics were done with open source software (R Development Core Team 2009). This material is based in part upon work supported by the National Science Foundation under Grant Number IIS-0624345 to Stanford University for the research project 'The Dynamics of Probabilistic Grammar' (PI Joan Bresnan). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

The ability to predict is necessary for survival; as an obvious example, predictions of motion and sequential action are continuously employed in human motor activities. In the past decade evidence has been growing that prediction also underpins linguistic perception and comprehension. For example, while listening to sentences unfold, people make anticipatory eye-movements to predicted semantic referents (Altmann and Kamide 1999, Kamide et al. 2003a, Kamide et al. 2003b). Event-related brain potential (ERP) changes show graded pre-activation of the word forms *a* or *an* as a function of their probability of occurrence in the context of reading a sentence (DeLong et al. 2005). Words that are less discourse-predictable evoke a greater positive deflection in the ERP waveform, and this effect diminishes when the predictive discourse context is eliminated; convergently, prediction-inconsistent adjectives slow readers down in a self-paced reading task (Van Berkum et al. 2005). People use language production predictively at all levels during comprehension, according to a functional model consistent with a wide range of evidence (Pickering and Garrod 2005). Language production is so intimately involved with language perception that listeners' auditory perception of words can be changed by robotic manipulation of their jaws and facial skin during pronunciation (Ito el al. 2009, Nasir and Ostry 2009). Predictive models can also explain many frequency effects in language acquisition, use, and historical change (see Diessel 2007 for a review). The logic common to many of these studies is that if people use language production to make predictions during comprehension, then probabilistic differences in production should be detectable in experiments on perception and comprehension, even with higher-level grammatical structures (syntax).

There is one potentially rich source of probabilistic production differences that has scarcely been tapped in this program—quantitative syntactic divergences between different varieties of the same language.[1] For an example of probabilistic differences between varieties of English, consider that for every one-word increase in the length of the possessum phrase (illustrated by *shadow* in (1)), the use of the Saxon genitive increases by 37% over the Norman genitive in American English, while there is no such effect in British English, according to a comparative study of corpora by Hinrichs and Szmrecsányi (2007: 466). The same study found that in both spoken and written productions, American English speakers differ from British in being more likely to produce the Saxon genitive with an inanimate possessor as illustrated in (1a), all else being equal (Hinrichs and Szmrecsányi 2007, Szmrecsányi

[1] But interest in comparative cross-varietal psycholinguistics is growing (Bender 2005, Bock et al. 2006, Trousdale and Clark 2008, Staum Casasanto 2008.

and Hinrichs 2008):

(1) a. *the building's shadow* ← Saxon (*'s*) genitive

b. *the shadow of the building* ← Norman (*of*) genitive

Supporting the link between probabilistic production and comprehension, a similar phenomenon was found in an earlier psycholinguistic study: Rosenbach (2002, 2003) devised a forced-choice experiment comparing British and American preferences for one or the other genitive construction after reading passages from a novel. She found major effects of animacy, topicality, and semantic relation in both groups, but younger American subjects had less effect of animacy compared to the British subjects.

Other examples of probabilistic production differences between varieties of English come from cross-corpus studies of the dative alternation, illustrated in (2).

(2) a. *Who gave that wonderful watch to you?* prepositional (*to*-)dative

b. *Who gave you that wonderful watch?* double object construction

In a study of *give* dative constructions in spoken New Zealand and US corpora, Bresnan and Hay (2008: 202) found that New Zealand English speakers are more likely to produce inanimate recipients in the double object construction (such as *who gave the school a distinctly scientific bias*) than Americans, all else being equal.

In general, these macro-regional varieties of English differ not in their grammatical rules for syntactic structures like those just illustrated, but in the probabilities of the structures occurring in spoken and written discourse (Schneider 2007; Rohdenburg and Schlüter 2009). Prior work in psycholinguistics has shown that there are important parallels between the comprehension and production of dative constructions in the use of information about their quantitative distribution (MacDonald 1999: 189; Stallings et al. 1998). Corpus probabilities of dative constructions can predict sentence ratings (Bresnan 2007b), and dative verb bias toward one or the other construction can predict anticipatory eye movements (Tily et al. 2008). Can we find evidence for differences in linguistic predictive behavior among speakers of different varieties of English? In the present study we use the dative alternation to compare syntactic predictive behavior of speakers of Australian and American (United States) varieties of English.

It might be thought that individual differences among speakers in experimental tasks would completely swamp any group effects that could reflect divergent probability distributions. To attack this problem head-on, we make use of *multilevel* (also

termed *mixed effect*) *regression models*, which are of growing importance in the language sciences (Baayen 2008, Johnson 2008, Baayen et al. 2008, Quené and van den Bergh 2008, Jaeger 2008). As explained below, these models provide an efficient way to adjust for random effects of the individual participants and items used in the study, so that any significant main effects or interactions with variety reliably hold across the particular samples of participants and items and can generalize beyond them.

It also might be thought that social differences between the speakers of the two varieties participating in the experiments would swamp any effects of macro-regional variety. Yet the studies of broad-coverage corpora cited above, as well as others (e.g. Grimm and Bresnan 2009) show that probabilistic grammatical differences exist between varieties at the macro-regional level. To determine the extent of sub-regional social influences on the dative alternation lies beyond the scope of the present study and remains a topic for future research. Given that speakers (and indeed, most linguists) are unaware of quantitative trends in the dative alternation, it is possible that there are few social influences on the variable (cf. Weiner and Labov 1983 on the English agentless passive).

Then there is the question of whether the two English dative constructions illustrated in (2a,b) are syntactic alternatives at all. The concept of syntactic alternations has been critically discussed in sociolinguistics (Lavandera 1978, Romaine 1984, Cheshire 1987, Silva-Corvalán 1986) and from the general position that any difference in word order also implies a difference in meaning (Bolinger 1977), although 'meaning' is construed differently in these discussions. There are nevertheless important studies explicitly proceeding from the assumption of semantic equivalence between syntactic variants (Weiner and Labov 1983, Jacobson 1980, Kroch 1989, 1994, Kroch and Taylor 1997, Pintzuk 1993, 1996, Pintzuk and Kroch 1989, Rosenbach 2002, Szmrecsányi 2006, Hinrichs and Szmrecsányi 2007, among many others), and the underlying concept of grammatical variation, fundamental in sociolinguistics, has been generalized and considerably refined there (Sankoff and Thibault 1981, Tagliamonte 2006: 70–76) and elsewhere (Rosenbach 2002: 22–25).

In syntax more generally it is true that alternative forms often have differing meanings (Pinker 1989, Levin 1993), frequently explained in terms of 'the principle of contrast' (Clark 1987). Syntactic studies of the dative alternation have differed on this point, with some assuming semantic equivalence (Aoun and Li 1989, Larson 1988a,b, den Dikken 1995), others arguing for lexical semantic differences (Green 1974, Oehrle 1976, Pinker 1989), and still others, for lexical equivalence and constructional semantic differences only (Goldberg 2002, Harley 2002, Gries and Stefanowitsch 2004). Nevertheless empirical studies of dative constructions have

shown that variants spontaneously occur as partial repetitions in discourse (Davidse 1996: 291, Bresnan and Nikitina 2009):

(3) 'You don't know how difficult it is to find something which will please everybody—especially the men.'

'Why not just *give them cheques*?' I asked.

'You can't *give cheques to people*. It would be insulting.'

(4) 'You *carrying a doughnut to your aunt* again this morning?' J.C. sneered. Shelton nodded and turned his attention to a tiny TV where 'Hawaii Five-O' flickered out into the darkness of the little booth. 'Looks like you *carry her some breakfast* every morning.'

Further, reported cases of non-alternation based on intuitive judgments of decontextualized examples are surprisingly inconsistent with actual usage (Fellbaum 2005, Bresnan et al. 2007a, Bresnan 2007a,b). Contrast, for example, the frequently reported ungrammaticality of *give a headache to, give the creeps to* with the examples harvested from usage by Bresnan et al. (2007a), Bresnan and Nikitina (2009):

(5) a. The spells that protected her identity also *gave a headache to* anyone trying to determine even her size

b. Design? Well, unless you take pride in *giving a headache to* your visitors with a flashing background? no.

(6) a. This life-sized prop will *give the creeps to* just about anyone! Guess he wasn't quite dead when we buried him!

b. Stories like these must *give the creeps to* people whose idea of heaven is a world without religion.

The purported ungrammaticality of such examples has been a primary justification for theories that semantics determine dative construction choice. Reported ungrammaticality can be overridden by manipulating the pronominality, definiteness, and information structure of the referring expressions (Green 1971, Kuno and Takami 1993, Polinsky 1998, Bresnan 2007a,b, Bresnan et al. 2007a). As an example, consider a commonly cited example of verbs that occur only in the double object construction, not the prepositional dative:

(7) a. Ted denied Kim the opportunity to march.
*Ted denied the opportunity to march to Kim.

b. The brass refused Tony the promotion.
*The brass refused the promotion to Tony.

Green (1981) showed that by looking at the interaction of the apparently ungrammatical examples with a strongly conflicting constraint against V NP Pronoun, fully grammatical cases of the prepositional dative can be constructed (8a). Similar examples can be found in use (8b):

(8) a. Ted gave Joey permission to march, but *he denied it to Kim*.

b. He extended it to everyone and *refused it to no one*.

The reportedly ungrammatical constructions are used grammatically to avoid the even worse constructions that would result from the violating of the constraint:[2]

(9) a. . . . he denied Kim it.

b. . . . he refused no one it.

That competing forms may be neutralized in discourse is also observed in the sociolinguistic literature (Sankoff 1988: 153; Tagliamonte 2006).

Lastly, the truth-conditional semantics of the dative alternation has been addressed in model-theoretic terms which assume truth-conditional equivalence for some alternating dative verbs (Krifka 2003). Although different word orders are associated with such truth-conditional semantic phenomena as quantifier scope and bound anaphora, these, too, are easily eliminated by careful examination and filtering of the variable context.

If it is not semantics alone, then what does determine the choice between alternative dative constructions? According to many previous studies, which alternative is used depends on multiple and often conflicting properties (McDonald et al. 1993, Arnold et al. 2000, Bock and Irwin 1980, Bock et al. 1992, Collins 1995, Gries 2005b, Hawkins 1994, Lapata 1999, Prat-Sala and Branigan 2000, Snyder 2003, Thompson 1990, Wasow 2002, Bresnan et al. 2007a). These include the accessibility of the referents in the context (has the possessor or recipient just been mentioned or is it new information to the hearer?), the complexity and pronominality of the

[2]Bresnan and Nikitina 2009 argue that this constraint itself is gradient.

descriptions of the referents (shorter before longer, pronouns adjacent to their governing head), the animacy of the referents, and the like. Previous studies have shown that the probability of a dative construction—either double-object or prepositional—is increased when the first of its two complements is a pronoun, is definite, refers to a highly accessible referent, has an animate referent, or is short. From these and other variables such as the previous occurrence of a parallel structure (Bock 1986, Gries 2005b, Pickering et al. 2002, Szmrecsányi 2005), it is possible to predict the choice of dative construction in spoken English with 94% classification accuracy on unseen data (against a baseline of 79%, Bresnan et al. 2007a). Prosodic information has also been found to contribute to dative construction choice (McDonald et al. 1993; Anttila 2008; cf. Shih et al. 2009).

The evidence thus suggests that the differences in the two constructions are preferences, not categorical regularities. This conclusion is further supported by historical and inter-variety divergences in the constructions. To cite just a few relevant findings, (i) the frequencies of double object constructions with the same set of verbs in British and American English in the 19th and early 20th centuries have been diverging (Rohdenburg 2007); (ii) Indian English has higher overall rates of prepositional dative than British English (Mukherjee and Hoffman 2006); (iii) in New Zealand English the overall probability of use of prepositional datives with the verb *give* has been significantly increasing from the early 1900s, after adjusting for other variables including verb semantics, discourse accessibility of referents, pronominality, and length (Bresnan and Hay 2008); (iv) in dative constructions found in British and American journalists' texts from the 1960's and 1990's there is a rise in the probability of the double object construction, according to a corpus study which controlled for verb lemma as well as length, pronominality, and text frequency of recipient and theme (Grimm and Bresnan 2009); and (v) the relative frequencies of prepositional datives are higher in the spoken and written Australian English dative data reported by Collins (1995) than in the combined spoken and written American English dataset of Bresnan et al. (2007a): 34.5% vs. 25%.[3]

[3]However, the selection criteria of the two datasets differ (for example, Collins included both *to* and *for* datives, while Bresnan et al. included only *to* datives), and there are many other possible unknown confounds. Additionally, corpus inputs may differ in a way which affects summary statistics without affecting the underlying probabilities of outputs (Bresnan et al. 2007a).

1 The corpus model

To measure predictive capacities of both Australian and US participants, we used an updated version of the Bresnan et al. (2007a) corpus model of American dative choices during spontaneous conversations. Both the original model and the original dataset have been extensively discussed elsewhere and are publicly available.[4] In this section we provide a summary of our updated version of the model and dataset in order to make the present study self-contained and methodologically transparent.

1.1 The Data

Bresnan et al. collected a database of 2360 instances of dative constructions from the three-million word Switchboard corpus of telephone conversations in English (Godfrey et al. 1992).[5] Based on findings from the previous literature on dative construction choice cited above and on hypotheses about markedness hierarchies in syntax (see Greenberg 1966, Silverstein 1976, Aissen 1999, O'Connor et al. 2004, Bresnan et al. 2007b, Bresnan and Nikitina 2009, among others), the data were coded for multiple variables or 'predictors', reviewed in the following paragraphs. Further information about the data sampling and annotation can be found in Bresnan et al. (2007a) and Bresnan and Hay (2008), both based on Cueni (2004).

For the present project we used a corrected version of the database created by Gabriel Recchia in 2006 by correlating the Bresnan et al. dataset with the time-aligned Switchboard corpus produced by the Mississippi State University Institute for Signal and Information Processing resegmentation project (Deshmukh et al. 1998).[6] The new dataset consists of 2349 observations of dative constructions, of which 499 are prepositional *to*-datives. Like the original dataset, it has a preponderance of double object constructions (79%).

[4]The original Bresnan et al. (2007a) paper itself is freely available from the Royal Netherlands Academy of Science (www.knaw.nl/publicaties/pdf/20051055.pdf). The corpus model and data are also incorporated into two recent textbooks on quantitative linguistic analysis (Baayen 2008a, Johnson 2008), where technical issues of model specification, validation, and interpretation are discussed in detail. The dative dataset of Bresnan et al. (2007a) is publicly available for download from the publisher of Johnson (2008) and in the CRAN internet archive (The Comprehensive R Archive Network, http://cran.r-project.org/) with the languageR package (Baayen 2008b).

[5]They also collected data from from the Penn Treebank *Wall Street Journal*, which are not included in the present study because of our interest in spoken varieties of English.

[6]This was a case study for Recchia (2007).

Verb The dative alternation has long been known in syntax to be governed by the verb (Levin 1993), in that only certain verbs can head alternative constructions like (2a,b) while preserving semantic equivalence. For example, *sent, brought, took, promised, offered* can replace *gave* in (2a,b), while *discussed, liked, knew, required, saw* cannot. Bresnan et al. (2007a) therefore sampled dative constructions by first finding a set of verbs that can appear in the alternative constructions in spoken English. For this they used the parsed (approximately one million-word) portion of the Switchboard corpus available in the Penn Treebank (Marcus et al. 1993) and the TGREP query language (Pito 1994) to extract all prepositional *to*-dative and double-object constructions. From these, thirty-eight alternating dative verb lemmas were identified by selecting those verbs for which at least five instances in each dative construction could be found on that portion of the Internet indexed by Google, excluding instances which were judged to be errors or which occurred on pages not based in the US, UK, or Australia; thus some verbs were included which alternate rarely (see Bresnan and Nikitina 2009 for discussion). All forms of these verbs were then searched for in the full (approximately three million-word) Switchboard corpus using character-string searches, and the examples were inspected and manually filtered, resulting in 2360 instances in the original dataset. Every instance was coded for one of the thirty-eight verb lemmas.

Verb Sense The sense in which a dative verb is used is predictive of the choice of dative construction it appears in: for example, 61% of verbs used in the transfer sense in the dataset occur in the double object construction, compared to 91% of verbs used in the communication sense. A Verb Sense predictor was created by crossing Verb with six broad semantic classes, creating subclassifications of each verb according to the senses in which it was used. The six semantic classes are 'transfer' of possession of an object (as with *give* in example (2a,b), 'future transfer' (as with *bet, offer, owe*), 'communication' of information (as with *tell, quote, show*), 'prevention of possession' (if it was an instance of a verb like *charge, cost, deny*), or 'abstract', for all other instances. There were 55 different verb senses in the (Switchboard) dataset.

Although semantic class was included as a main effect by Bresnan et al. (2007a), it was not included in the present study because its effects are included in the random effect of Verb Sense (see n. 17).

Structural Parallelism Structural parallelism or persistence is an important predictor of syntactic choice in corpus data including sociolinguistic interviews (see Weiner and Labov 1983 and Szmrecsányi 2005, among others). In the Bresnan

et al. dataset a possible effect of the presence of a parallel structure in the dative dataset was measured by a variable defined as the presence of the same syntactic construction type (prepositional *to*-dative or double object construction) in the same dialogue.

A preferable measure, following Szmrecsányi (2005, 2006), is the type of the nearest preceding dative structure in the dialogue. For the present study we adopted this measure of syntactic parallelism, created by Gabriel Recchia using the time-aligned and corrected dative dataset he developed (see Recchia 2007).[7]

Relative Syntactic Complexity of Theme and Recipient Another important predictor of word order and construction type is the relative syntactic complexity of complements (Hawkins 1994, Arnold et al. 2000). Measures of syntactic complexity are highly correlated and can be efficiently operationalized by counting the number of graphemic words (Wasow 2002, Szmrecsányi 2004). This was the metric used by Bresnan et al. In their model the complexity predictor is the signed logarithm of the absolute value of the difference between the theme and recipient lengths in words.[8] This measure is intended to capture the relative complexity of theme and recipient in one variable.

In the present study we use a simpler measure, the natural logarithm of the recipient length minus the natural logarithm of the theme length. This measure expresses the ratio of the two complements within a scale that also compresses extreme values.

Discourse Accessibility of Recipient, Theme Many previous researchers have presented evidence that discourse accessibility and/or focus placement influences the choice of alternative constructions (see Halliday 1970, Erteshik-Shir 1979, Givón 1984, Thompson 1995, among others). An important alternative hypothesis is that apparent discourse accessibility effects can be explained by ordering phrases so as to minimize syntactic complexity in comprehension (cf. Hawkins 1994, Arnold et al. 2000). By adjusting for the relative syntactic complexity of the complements, it is possible to determine whether discourse accessibility also makes a contribution to

[7] See Snider (2009) for a study of structural persistence with dative constructions using this augmented dataset.

[8] Length is transformed by the logarithm in order to compress outliers and bring the distribution more closely into the logistic regression model assumption of linearity in logit space. Because differences can have values ≤ 0, on which the logarithm is not defined, one unit was added to the absolute value of the length difference of theme minus recipient to shift it to a positive numerical domain on which the transformation is always defined; the original signs were then restored to the resulting values.

construction choice over and above syntactic complexity.

The dative data were coded for seven levels of discourse accessibility—'evoked', 'situationally evoked', 'frame inferrable', 'generic', 'containing inferrable', 'anchored', and 'new' (see Michaelis and Hartwell 2007, Prince 1981, Gundel et al. 1993). To overcome data sparseness in modeling, these seven categories were simplified to two. A theme or recipient phrase was defined as 'given' if (i) its *referent* was mentioned in the previous 10 lines of discourse ('evoked') or (ii) it was a first or second person pronoun (denoting a 'situationally evoked' referent). All others were 'non-given'.[9]

Definiteness of Recipient, Theme Semantic definiteness is known to interact with word order in a number of languages (see van Bergen and de Swart 2009 for a recent review and empirical study of Dutch). Recipients and themes were coded for semantic definiteness as operationalized by Garretson (2003). If placing the recipient or theme phrase in the context *There is/are* __ permits an existential interpretation (as opposed to a list or deictic reading), then the NP is coded as indefinite. Examples of indefinite NPs include *a little bit of flavor, one, some Amish bread, a puppy, different perspectives, more jobs, another box, something that I can create*; examples of definite NPs include *him, that chore, the pony, my photo album, Nolan Ryan, all my friends* (Cueni 2004).

Pronominality of Recipient, Theme Different nominal expression types (such as pronouns, proper names, common nouns) have been found to affect the choice of syntactic alternatives in a variety of constructions (Silverstein 1976, Aissen 1999, O'Connor et al. 2004). Themes and recipients in the dative data were coded for nominal expression type, the values of which were 'personal pronoun' (*him*), 'impersonal pronoun' (*someone*), 'demonstrative pronoun' (*that*), 'proper noun' (*Schwarzkopf*), 'common noun' (*a native Texan*), 'gerund' (*advancing their education*), and 'partitive' (*the rest of the family members*) (Cueni 2004). To avoid sparseness of data, Bresnan et al. (2007a) simplified the distinction. Specifically, pronominality was simplified to distinguish phrases headed by personal, demonstrative, indefinite, or reflexive pronouns from those headed by nonpronouns such as nouns and gerunds.

For the present study we redefined the pronoun category to bring it in line with the Treebank part-of-speech tagging guidelines ('PRP', 'PRP$' in the notation of Santorini 1990). 'Pronouns' are now personal (including *it, them* and generic *you*,

[9]The generic uses of *you* were counted as situationally evoked because they were considered to include the hearer semantically.

n. 9), demonstrative, or reflexive, and exclude indefinites.

Animacy of Recipient, Theme Animacy is an important cognitive category in humans with subtle effects on English word order, primarily showing up in variation (Ransom 1977, 1979, Thompson 1900, 1995, Bock et al. 1992, McDonald et al. 1993, Rosenbach 2002, 2005, 2008, Bresnan and Hay 2008, Lamers et al. 2008). For the dative dataset, animacy was coded in four categories—'human', 'organization', 'animal'[10], and 'inanimate,' derived from Garretson et al. (2004) by collapsing their 'concrete inanimate', 'non-concrete inanimate', 'place', and 'time' into a single 'inanimate' category (cf. Zaenen et al. 2004). In modeling, the animacy variable was further simplified because of data sparseness to a binary category of human or animal vs. other.[11]

Concreteness of Theme Theme arguments were coded for whether they referred to a concrete object, defined as a prototypically concrete inanimate object or substance perceivable by one of the five senses (Garretson 2003). The 'prototypical' restrictor was used to bring the category within ordinary conceptions of what a concrete object is: for example, it excludes currents of water, but includes plants. Concreteness of theme was added to compensate for the simplification of the original Garretson et al. animacy coding system with the four categories of inanimates.

Person of Recipient, Theme Person influences syntactic alternations categorically in some languages and variably in English (Bresnan et al. 2001). The data were coded for person distinguishing inclusive and specific uses of both first and second persons (Cueni 2004). For modeling, the more elaborate classification was simplified to a binary division between 'local' persons (first and second) and 'non-local' (third).[12]

[10] One instance of an intelligent machine occurred in the dataset and was classed with 'animal': 'you'll be able to just give it commands', said of a computer that responds to speech.

[11] 'Human' referents were individual humans and humanoid beings (such as gods, ghosts, or androids) and groups of humans which do not meet the criteria for organizations because of the lack of a collective voice and/or purpose. For example *people that come into this country, qualified students, their customers* refer to groups of humans and not to organizations.

[12] Although this person variable lacked significance in the original corpus model of Bresnan et al. (2007a), Bresnan and Nikitina (2009) subsequently found it to be a small significant influence on dative choice in a larger dataset combining dative constructions from the Penn Treebank *Wall Street Journal* (which Nikitina annotated for person) and the Switchboard corpus.

Number of Recipient, Theme Number is a typologically important category in grammar (Greenberg 1966), and plays a role in some types of morphosyntactic variation in English (Bresnan 2002, Bresnan et al. 2007b). In the dative dataset, words with formal plural marking (and one instance of fish which the context clearly indicated was plural) were coded 'plural'; other words were coded 'singular'.

Most of the variables described above have previously been observed to influence the dative alternation, and pervasive partial correlations are known to exist among them, as Bresnan et al. (2007a) emphasize. This is one motivation for using multiple regression modeling in a careful assessment of the contributions and interactions of multiple variables.

1.2 The Model

The corpus model is a mathematical formula which defines, for combinations of the predictor values enumerated above, the probability that a prepositional *to*-dative will be chosen from the two alternative dative constructions. The general structure of the model is shown in (10):

(10) **A Generalized Linear Model with a Single Random Intercept**

$$logit[Pr(\boldsymbol{Y_{ij}} = \boldsymbol{y_{ij}} \,|\boldsymbol{X_{ij}}, \boldsymbol{u_i}\,)] = \boldsymbol{X_{ij}}\boldsymbol{\beta} + \boldsymbol{u_i}$$

In this model the conditional probability of a response given a group i is systematically linked to a linear combination of fixed cross-group explanatory variables $\boldsymbol{X_{ij}}$ and a randomly varying normally distributed group effect.[13] The 'response' here is simply the choice between a prepositional or double-object dative construction as in (2), the 'groups' are the different Verb Senses as defined in Section 1.1, and the explanatory variables are the previously described predictors including relative syntactic complexity of theme and recipient, animacy of theme and recipient, and the like.

For the present project we re-fit the original Bresnan et al. (2007a) model (their 'Model B') to the corrected version of the database. We also took the opportunity to make several improvements to the model for the present study, already described in Section 1.1. First, we adopted the improved measure of structural parallelism based

[13]The $logit$ function gives the 'log odds'—the natural logarithm of the odds ratio, $\frac{P}{(1-P)}$—which is here the ratio of the probability of a V NP PP to the probability of the alternative, V NP NP.

on the corrected, time-aligned dataset (n. 6). Secondly, we re-defined the pronoun category to exclude indefinite pronouns in line with the Treebank part-of-speech tagging guidelines. Third, we simplified our measure of relative syntactic complexity, using the difference of log recipient length and log theme length. In addition we used currently available software for fitting glmms and tools for interpreting and validating them (Bates et al. 2009, Baayen 2008b) which were not available earlier. All references in subsequent sections to 'the (corpus) model' are to this new model of the corrected and time-aligned dataset. The fitted values of the present corpus model are highly correlated with the original (Spearman's $\rho = 0.949$). The present model substantially reduces the moderate collinearity of the original Bresnan et al. model.[14]

The new model formula is given in Figure 1. Like the older corpus model of Bresnan et al., it is of the general form in (10), with the inverse of the $logit$ function applied to both sides of the equation. This model was selected from a full model containing all predictors in the original corpus model (with the replacements described for parallelism, pronominality, and relative complexity of recipient and theme) by eliminating predictors where the magnitude of the estimated coefficient was less than the standard error. With the changes to the model predictors described above, discourse accessibility dropped out, although definiteness and (a narrower classification of) pronominality remain in the model.

The numerical coefficients of the predictors shown in Figure 1 are parameters estimated from the empirical distributions of the predictors in the data.[15] They can be likened to constraint weights in other optimization-based systems (Manning 2003). The numerical coefficients are multiplied by 1 if a predictor value is true and 0 if it is false. For example, line 5 of Figure 1 means that if the theme is a pronoun 4.2391 is added to the formula sum; otherwise, 0 is added. For numerical predictors like the recipient-theme length ratio in the penultimate line of the equation, the coefficient (here $+1.1819$) is multiplied by the numerical value of the predictor and added to the formula sum. As in ordinary logistic regression, positive coefficient signs favor the 1 response (V NP PP here) and negative coefficient signs favor the 0 response (V

[14]Using a stringent method of calculating multi-collinearity due to Belsley et al. (1980) and implemented by Baayen (2008a: 200), we found that the 'condition number'c of the original model is $c = 20.22$ (indicating medium collinearity), while that of the present model is $c = 8.97$ (slight collinearity). A condition number less than 6 indicates no collinearity to speak of, and $c > 30$ indicates harmful non-collinearity.

[15]Maximum likelihood is the criterion for fitting generalized mixed effect regression, while residual maximum likelihood (REML) is preferred for linear mixed effect regression (Pinheiro and Bates 2000; Bates et al. 2009).

Figure 1: The Corpus Model

$$\text{Probability}\{\text{Response} = \text{V NP PP}\,|X, u_i\} = \frac{1}{1 + e^{-(X\beta + u_i)}}, \text{ where}$$

$$\begin{aligned} X\hat{\beta} = \; & 1.1583 \\ & -3.3718\{\text{pronominality of recipient} = \text{pronoun}\} \\ & +4.2391\{\text{pronominality of theme} = \text{pronoun}\} \\ & +0.5412\{\text{definiteness of recipient} = \text{indefinite}\} \\ & -1.5075\{\text{definiteness of theme} = \text{indefinite}\} \\ & +1.7397\{\text{animacy of recipient} = \text{inanimate}\} \\ & +0.4592\{\text{number of theme} = \text{plural}\} \\ & +0.5516\{\text{previous} = \text{prepositional}\} \\ & -0.2237\{\text{previous} = \text{none}\} \\ & +1.1819 \cdot [\log(\text{length}(\text{recipient})) - \log(\text{length}(\text{theme}))] \end{aligned}$$

and $\hat{u_i} \sim N(0, 2.5246)$

NP NP here).

The final parameter $\hat{u_i}$ in Figure 1 refers to a random effect vector of normally distributed numerical values, each an individual adjustment to one of the Verb Senses representing its tendency to be expressed with the V NP PP construction—its prior 'bias' toward the construction, adjusting for all of the fixed effect predictors. Although there are 55 such values, the multilevel model only uses a single parameter for all of them, which is the standard deviation of their normal distribution around the mean of 0.[16] Table 1 lists these random effect values for the model; positive values represent a Verb Sense bias toward V NP PP, and negative values a bias toward V NP NP.[17]

[16]Technically the random effects are not estimated parameters for the statistical model, but 'best linear unbiased predictors (BLUPS)' (Pinheiro and Bates 2000: 71). However, they behave in some ways like intercepts and the term *random intercepts* is common parlance.

[17]Baayen (2008a: 181) observes of the dative dataset that broad semantic class as a separate predictor adds little that is not already explained by individual verb variability as a random effect. We have nevertheless used the broad semantic classes to subclassify the verbs, in order to capture striking differences in the usage of individual verbs: for example, *pay* in the abstract sense as in *pay attention* strongly favors the prepositional dative while *pay* in the transfer sense as in *pay money* favors the double object construction. In contrast, *hand* hardly differs in its abstract and transfer uses.

xxxx

Table 1: Random Effects of the Corpus Model

afford.a	0.9792	float.a	−0.0509	pay.a	5.2559
allot.a	−0.0387	give.a	−1.3000	pay.t	−1.9805
allot.f	−0.2157	give.c	0.0627	promise.a	−0.1252
allow.a	−2.3679	give.t	−0.1314	quote.c	−0.3027
assign.a	−1.8472	hand.a	1.9369	read.c	1.4258
assign.f	0.5526	hand.t	1.9365	sell.t	1.5342
award.f	−1.7434	leave.a	3.1092	send.a	−0.1944
bet.f	−0.2100	leave.c	1.3478	send.c	2.4383
bring.a	3.1927	leave.f	0.7237	send.t	1.3094
bring.t	2.0043	lend.a	0.1654	serve.t	−0.4671
cause.a	−0.3278	lend.t	−0.0182	show.c	−0.8993
charge.p	−1.1801	loan.t	0.4604	swap.t	−0.0509
cost.p	−3.4840	mail.a	−0.0760	take.a	0.8954
deny.p	−1.1051	mail.t	2.2962	take.t	2.7427
do.a	−0.7445	make.c	−0.2693	teach.c	−2.5230
feed.c	−0.1292	offer.a	1.3923	tell.c	−5.7870
feed.t	−1.3740	offer.f	0.7515	wish.a	−0.6953
flip.a	−0.1944	owe.a	−1.9362	write.t	3.3665
		owe.f	−2.5562		

Suffixes .a, .c, .f, .p, .t respectively designate Verb Sense values of ‘abstract’, ‘communication’, ‘future transfer of possession’, ‘prevention of possession’, and ‘transfer of possession’.

Verb Sense is treated as a random effect in this model because the study findings are not meant to be restricted to the thirty-eight verbs collected by Bresnan et al. (2007a), but to generalize across the entire population of possible dative verbs of English in their various broadly defined senses. Recall that the data in their study were sampled from the population of all possible dative constructions by collecting dative verbs in context from specific corpora. For this reason Verb Sense is a random variable and is treated as a random effect in the model. Speaker identity (an index of the individual speakers who produced the data) is also a random variable and could be included as another random effect, but the corpora contained many speakers each

with relatively few dative constructions, and Bresnan et al. (2007a) found that the fixed effects remain significant after taking speaker differences into account. The same holds for our present corpus model: we performed a likelihood ratio test on nested models, the model in Figure 1 with an added random intercept for speaker and without it; the added random effect of speaker does not significantly improve model log-likelihood, $\chi^2 = 0.1812, Df = 1, Pr(> \chi^2) = 0.6704$.

Notice that the model formula can be read off the standard table of the model parameters given in Table 2, which also shows the reliability of the model estimates for the fixed effects.

Table 2: Corpus Model parameters

Fixed Effects:

	Estimate	Standard Error	Z Value	Pr($>\|z\|$)
(intercept)	1.1583	0.5337	2.170	0.03
recipient = pronoun	−3.3718	0.3236	−10.420	0.0000
theme = pronoun	4.2391	0.4376	9.688	0.0000
recipient = indefinite	0.5412	0.3147	1.720	0.0001
theme = indefinite	−1.5075	0.2877	−5.239	0.0000
recipient = inanimate	1.7397	0.4595	3.787	0.0002
theme = plural	0.4592	0.2627	1.748	0.0805
previous *to*-dative	0.5516	0.3406	1.620	0.1053
no previous dative	−0.2237	0.2389	−0.936	0.3490
log rec-theme diff	1.1819	0.1686	7.008	0.0000

Random effects:

Group	Standard Deviation
verb sense (intercept)	2.5246

number of observations: 2349, groups: verb sense, 55

Tables 3–4 provide examples of the kind of data contained in the database (with preceding contexts edited for readability), together with the probabilities derived by the model for the prepositional dative choice. The italicized expressions are the ones actually observed. The alternative construction is given after the italicized one. The probabilities of such examples can be straightforwardly calculated by plugging into the formula the values of all of the predictors for that example, including the random

effects.

Table 3: Example of a very high-probability *to*-dative

Speaker:
About twenty-five, twenty-six years ago, my brother-in-law showed up in my front yard pulling a trailer. And in this trailer he had a pony, which I didn't know he was bringing. And so over the weekend I had to go out and find some wood and put up some kind of a structure to house that pony, because he *brought the pony to my children*/ brought my children the pony.

verb:	*bring*	=	transfer
theme:	*the pony*	=	non-pronoun, definite, length = 2
recipient:	*my children*	=	non-pronoun, definite, length = 2
previous:	none		

Probability of V NP PP $= 0.9497$

Figure 2 gives a worked example showing how the probability of the item in Table 3 is calculated from the predictor properties for which it was coded. The bolded lines are those that need to be used in the calculation of probability for the item in Table 3. Notice that the $\hat{u}_i$ parameter is for $bring.t$. The reader can verify the probabilities of the example in Table 4 by similar calculations, based on its differing linguistic properties.

1.3 Predictive Accuracy of the Model

The same model can be used on unseen data to predict the probabilities conditioned on the set of variables known to the model. For example, given a previously unencountered dative verb, together with required information about the recipient and theme arguments and the context of use, the model yields a predicted probability of syntactic realization as a prepositional *to*-dative vs. a double object construction.[18]

How accurate are the model's predictions? We divided our dataset of corpus dative constructions randomly 100 times into a training set of sufficient size to estimate the model parameters ($n = 2000$) and a testing set ($n = 349$). We fit the model to

[18]Model predictions for an unseen dative verb assume the mean random effect value of zero.

Table 4: Example of a very low probability *to*-dative

Speaker A:
We're like everybody else, you know, we've got several credit cards that sometimes, instead of paying them all off every month maybe you have to slip some and you pay part of it this month and part of it next month especially around Christmas time. You know, that's when everybody goes crazy on charging stuff.

Speaker B:
Well then, see, that's one of the points which I don't see. — Like I don't give a lot of gifts during Christmas and I, you know, don't like to give any at all because the thing is that, you know, it's like if I want something I'll ask somebody, you know. — Like, for Christmas my roommate goes, What do you want? And I said, I want a backpack. I told him if you want to give me a present for Christmas *give me a backpack*/ give a backpack to me.

verb:	*gave*	=	transfer
theme:	*a backpack*	=	non-pronoun, indefinite, length = 2
recipient:	*me*	=	pronoun, definite, length = 1
previous:	V NP NP		

Probability of V NP PP $= 0.0093$

Figure 2: Using the corpus model to calculate probabilities

$$\text{Probability}\{\text{Response} = \text{V NP PP} \mid X, u_i\} = \frac{1}{1 + e^{-(X\beta + u_i)}}, \text{ where}$$

$X\hat{\beta} =$

$\mathbf{1.1583}$

$-3.3718\{\text{pronominality of recipient} = \text{pronoun}\} = 0$

$+4.2391\{\text{pronominality of theme} = \text{pronoun}\} = 0$

$+0.5412\{\text{definiteness of recipient} = \text{indefinite}\} = 0$

$-1.5075\{\text{definiteness of theme} = \text{indefinite}\} = 0$

$+1.7397\{\text{animacy of recipient} = \text{inanimate}\} = 0$

$+0.4592\{\text{number of theme} = \text{plural}\} = 0$

$+0.5516\{\text{previous} = \text{prepositional}\} = 0$

$\mathbf{-0.2237\{previous = none\}} = 1$

$\mathbf{+1.1819 \cdot [\log(length(recipient)) - \log(length(theme))]}$

and $\hat{\mathbf{u}}_{\mathbf{bring.t}} = \mathbf{+2.0043}$

$= \frac{1}{1+e^{-(1.1583-0.2237+2.0043)}} = 0.9497$

each training set and generated its predictions for the corresponding unseen test set, scoring accuracy by concordance probability.[19] The mean concordance probability for the 100 test sets was $C = 0.945$. Note that these model predictions make no use of the random effect of Verb Sense, and so ignore known dative verb biases toward one or the other construction. For the corpus model fit on all of the data, $C = 0.984$, so on unseen data the model loses only 0.04 of its predictive accuracy, an indication that the model is not excessively overfitting the data. In other words, the high predictive accuracy of the model is not obtained at the cost of excessive model complexity.

We grouped the predicted probabilities for individual items into ten sets using equal cutoff points on the [0,1] interval. The mean probability for each group was compared to the observed proportions of the binary response 1 (= prepositional dative realization) in the same group. A plot of the observed proportions by the mean predicted probabilities for the ten groups is given in Figure 3, showing a very good fit of the model to the data.

The relative contribution of each predictor to the overall fit is displayed in Figure 4. Each bar represents the decrease in quality of fit caused by removing one of the predictors from the *full* model, as measured by likelihood ratio tests of the full and reduced-by-one models.[20] Each predictor contributes significantly to the model quality of fit, except for Definiteness of Recipient.

2 Quantitative harmonic alignment

One of the main findings of previous corpus work on the dative alternation is the existence of a statistical pattern in which, all else being equal, animate, definite, pronominal, discourse accessible and shorter arguments tend to precede inanimate, indefinite, non-pronominal, less discourse accessible or longer arguments in both dative constructions (2a,b). This pattern has been found in dative constructions in American, Australian, New Zealand, and British varieties of English and in both written and spoken modalities (Thompson 1995, Collins 1995, Bresnan et al. 2007, Bresnan and Hay 2008, Theijssen 2008, Grimm and Bresnan 2009). To illustrate,

[19] Bresnan et al. (2007a) used a similar validation method to test the classification accuracy of their model. Here concordance probability (as implemented by Harrell 2009) is used to measure how well the model discriminates all pairs of opposite responses; a value of $C = 0.5$ shows random discriminative accuracy, $C = 1$ is perfect, and a value 'greater than roughly $C = 0.8$ has some utility in predicting the responses of individual subjects' (Harrell 2001: 247). Concordance probability is equivalent to the area under the receiver operating characteristic (ROC) curve.

[20] Cf. n. 15. Although REML is the preferred criterion for optimizing the fit, maximum likelihood is an option and was used for calculating effect sizes of our linear model fits.

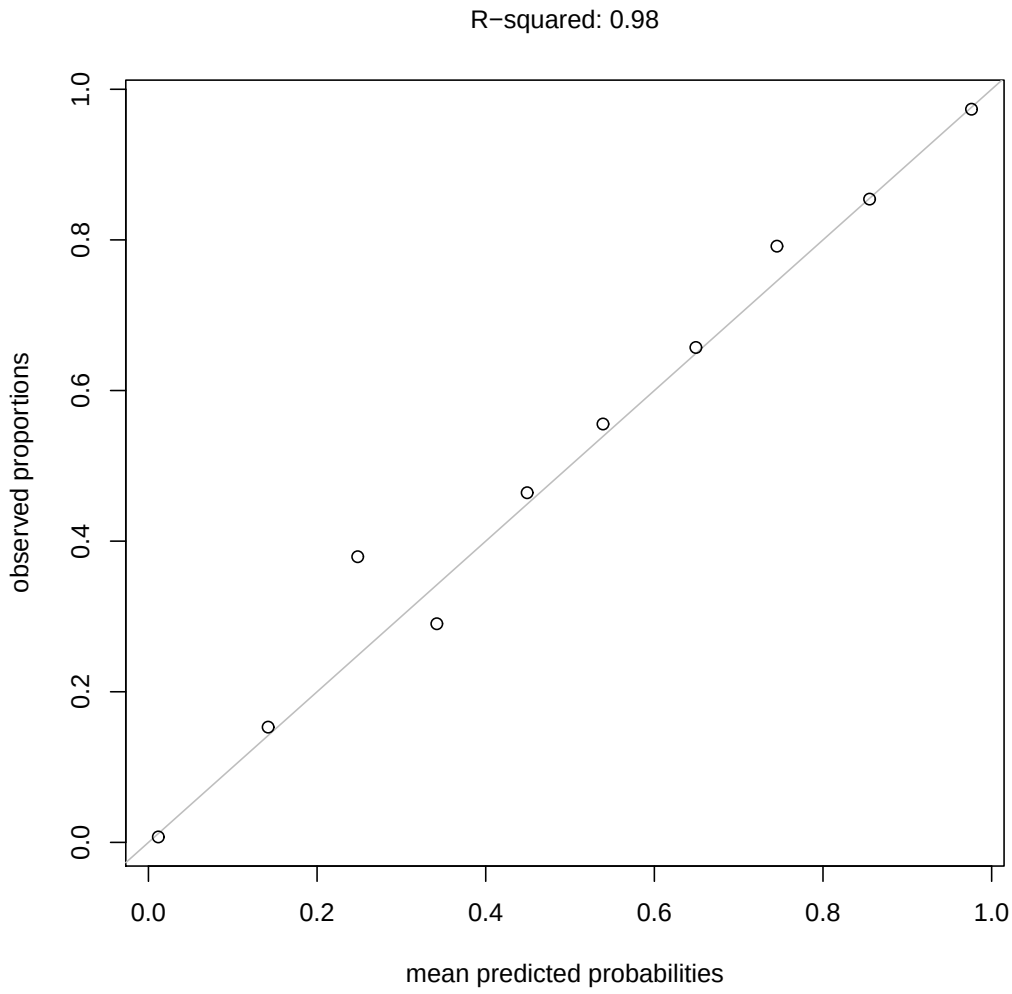


Figure 3: Corpus model fit between grouped observations and mean predicted probabilities

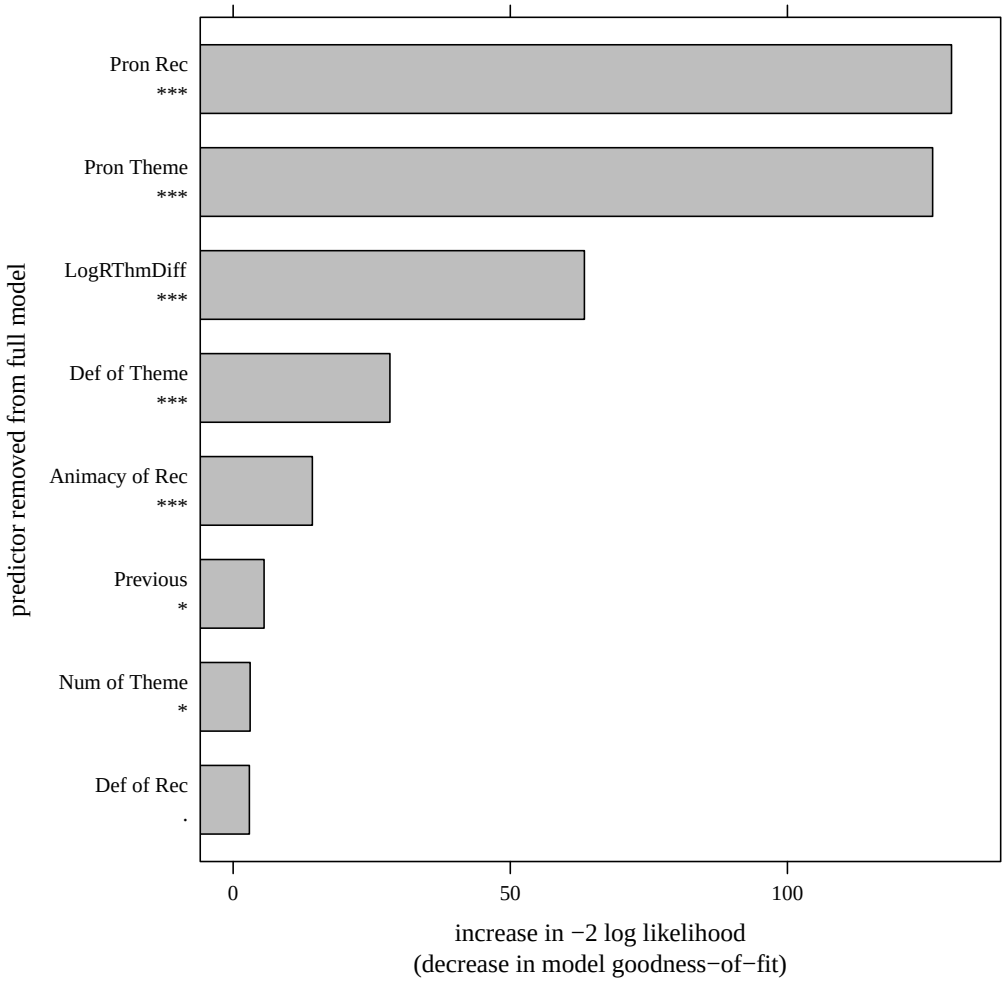


Figure 4: Effect sizes of the corpus model predictors
[‘.’ $p < 0.1$, ‘*’ $p < 0.05$, ‘**’ $p < 0.001$, ‘***’ $p < 0.0001$]

if the recipient argument is a lexical noun phrase, inanimate, indefinite, or longer, it will tend to appear in the prepositional dative construction; see the bolded recipient in (11a,b). Conversely, if the theme argument is a non-pronoun, inanimate, indefinite, or longer, it will tend to appear in the double object construction; see the bolded theme in (12a,b).

(11) a. *give those to **a man*** (more probable)

b. *give **a man** those* (less probable)

(12) a. *give **a backpack** to me* (less probable)

b. *gave me **a backpack*** (more probable)

In general, the choice of construction tends to be made in such a way as to place the inanimate, indefinite, nominal, or longer argument in the final complement position, and conversely to place the animate, definite, pronominal, or shorter argument in the position next to the verb where it precedes the other complement.

The magnitudes and directions of the estimated parameters of the model (the numbers and their signs in Figure 1) quantitatively reveal the same kind of pattern illustrated in (11)–(12). To see this, notice that the coefficients of the predictors for pronominality and definiteness in the model formula in Figure 1 have opposite signs when they apply to recipients and themes. For example, the coefficient for an indefinite recipient is positive ($+0.5412$), while that for an indefinite theme is negative (-1.5075). Because positive coefficients favor the prepositional dative and negative, the double object, this pattern implies that, all else being equal, an indefinite recipient favors the prepositional dative construction while an indefinite theme favors the double object construction—in other words, the model favors whichever construction places an indefinite complement in final position. A similar finding holds for pronominality: the model favors whichever construction places a pronominal complement adjacent to the verb. An inanimate recipient is favored in the prepositional dative, which places it in final position. As for the relative complexity predictor, the positive coefficient is multiplied by the difference between the log length of the recipient and the log length of the theme. This value is positive when the recipient is longer than the theme (hence favoring the prepositional dative and placing the recipient in final position), and negative when the theme is longer than the recipient (favoring the double object construction which places the theme in final position). When the two complements are equal in length, the value is zero and the other parameters influence the outcome.

Similar effects have been found in psycholinguistic studies of sentence production using English dative constructions (Bock 1982, Bock and Irwin 1980, Bock 1986, Bock, Loebell, and Morey 1992, McDonald, Bock, and Kelly 1993, Prat-Sala and Branigan 2000, Pickering, Branigan, and McLean 2002, Branigan, Pickering, and Tanaka 2008). Various production models have been proposed to explain them (cf. Ferreira 1996, Bock 1982; Bock and Levelt 1994; Chang et al. 2006; Branigan, Pickering, and Tanaka 2008).

To assist in interpreting the data patterns, we provide a qualitative view of the quantitative interpretation of the model in Figure 5.[21] The '$\succ$' symbol refers to relative prominence on a linguistic scale or hierarchy (see below). The arrows connecting the complements show the alternative positions of theme and recipient in the two constructions. When the theme or recipient has bolded properties, it is preferred in its bolded structural position; when it has unbolded properties, it is preferred in its unbolded structural position. The corpus model shows that though aligned, the multiple effects cannot be reduced to any one of them, whether it be syntactic complexity (cf. Hawkins 1994, Arnold et al. 2000) or any other single property (Bresnan et al. 2007)a.[22]

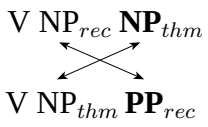
Figure 5: Qualitative view of Quantitative Harmonic Alignment

animate $\succ$ **inanimate**

definite $\succ$ **indefinite**

pronoun $\succ$ **non-pronoun**

less complex $\succ$ **more complex**



This statistical pattern is a kind of *harmonic alignment*. The term *harmonic alignment* is used here phenomenologically to refer to the tendency for linguistic el-

[21]—omitting model properties that do not apply to both recipients and themes, like plurality of the theme and previous occurrence of a prepositional dative construction, as well as those properties like discourse accessibility and person that are not retained in the present model.

[22]See also Rosenbach 2002, 2005; O'Connor, Anttila, Fong, and Maling 2004; and Strunk 2005 for parallel conclusions on determinants of possessive construction choice.

ements which are more or less prominent on a scale (such as the animacy or nominal expression type scales) to be disproportionately distributed in respectively more or less prominent syntactic positions (such as preceding in word order or occupying a superordinate syntactic position).[23] Thus, example (11a) is a harmonically aligned prepositional dative, and (12b) is a harmonically aligned double object dative. The bolded phrases are more harmonic in the final position because they are indefinite, lexical noun phrases; they also happen to be longer than the non-bolded definite pronominal phrases (the recipient *a man* consists of two words and the theme *those*, of one), but the alignment of (in-)definiteness and (non-)pronominality with final complement position is significant even when relative length is held constant.

Importantly, Australian English datives show a similar pattern of *end weight* and quantitative harmonic alignment, for *givenness, definiteness, pronounhood*. This fact can be inferred from Collins' (1995: p. 47) discovery of a frequency pattern of 'Receiver/Entity Differentiation' in the Australian corpus datives, by considering the proportional distribution of these properties across the alternative constructions in his data (Bresnan et al. 2007: pp. 74–75).[24]

It should be borne in mind that what we are calling quantitative harmonic alignment here is a probabilistic pattern arising from the frequencies and distributions of linguistic properties in the production of dative constructions. We use this terminology because the hierarchical and qualitative relations among these properties are far more familiar than the probabilities themselves. A more harmonically aligned construction is in our dataset a more probable construction.

3 Experiment 1: Sentence ratings

We have already highlighted the link between the corpus study of production of the genitive alternation in British and American English (Hinrichs and Szmrecsányi 2007) and a psycholinguistic study of preferred genitives as continuations after reading a passage by British and American experimental participants (Rosenbach 2003). As in Bresnan (2007b) we formulated the hypothesis that English speakers implicitly

[23] In Optimality Theoretic (OT) syntax the term refers to a formal operation of constraint conjunction that is designed to preserve hierarchical structure between different prominence hierarchies of constraints (see Aissen 1999, Prince and Smolensky 1993). We use the term purely phenomenologically as described above.

[24] Across languages, relative syntactic complexity or *end weight* is not always aligned with definiteness, animacy, or pronominality (cf. Hawkins 1994, 2004, 2007, Gibson 2000, Yamashita and Chang 2001, Yamashita 2002, Rosenbach 2005, 2008, Vasishth and Lewis 2006, Temperley 2007, Choi 2007).

know the quantitative usage patterns of production in their own variety and can use them to predict syntactic choices just as the corpus model does. Where the model predicts high probabilities, the experimental participants will, too. Where the model predicts more even probabilities, participants will, too.

3.1 Method

3.1.1 Participants

The participants were 19 volunteers from the Stanford University community and 20 volunteers from the Griffith University and Queensland University of Technology communities. They were paid for their participation. There was a balance of males and females in both groups. All participants were native speakers of English, did not speak another language as fluently as English, had not taken a syntax course, and had grown up in the US (the Stanford participants) or Australia (the Griffith participants).

3.1.2 Materials

There were 30 items, each consisting of a context followed by the two alternative dative continuations. The items were edited transcriptions obtained from actual speakers in dialogues and this was explained to the participants.

A sample item is given in (13).

(13) *Speaker:*
I'm in college, and I'm only twenty-one but I had a speech class last semester, and there was a girl in my class who did a speech on home care of the elderly. And I was so surprised to hear how many people, you know, the older people, are like, fastened to their beds so they can't get out just because, you know, they wander the halls. And they get the wrong medicine, just because, you know,

(1) the aides or whoever just give the wrong medicine to them.
(2) the aides or whoever just give them the wrong medicine.

One continuation was the observed continuation in the corpus and one was the constructed alternative.

The items were randomly sampled from throughout the range of corpus model probabilities in the 2349-observation corrected dative dataset and checked for obvi-

ous ambiguities in either alternative.

The fitted corpus model values and confidence intervals for the 30 items are shown on the $logit$ scale in Figure 6 (n. 13). From this plot we see that the items' probabilities are well differentiated by the corpus model, except for those in the middle ranage. We will use this scale in our linguistic comprehension and judgment tasks.

The items were presented in pseudo-random order, manually adjusted to avoid apparent patterns. Also, the order of the alternative dative constructions was alternated. The Australian participants received the same 30 items as the US participants, though with the context altered slightly to Australian conditions. Where necessary, place names, spelling, and atypical lexical items were changed; for example, for (13), *in college* was changed to *at university*.

3.1.3 Procedure

Participants were tested individually in their own country. They were given a booklet containing the instructions and the 30 items. They were told that we were interested in how people choose between different ways of saying the same thing in informal conversations. They were told that in the passages given in the booklet, one or two speakers were talking informally about different topics and that each passage included a choice of two ways of saying the same thing. The participants were required to read each passage and to rate the relative naturalness of the given alternatives in their context. They had 100 points to express their rating, so that the ratings for any pair of alternatives added up to 100.

The task differs from a forced choice task, such as Rosenbach's (2003), in that participants are free to assign equal value to both alternatives as well as a graded preference to one of them. The only choice that is not provided is the rejection of both alternative constructions. Because one of the alternatives was always from the actual spoken discourse in the corpus, we were not concerned by this limitation. The same procedure had been used in an explicit prediction task, asking participants to guess which construction was used by the speaker in the original dialogue and give a numerical rating of the likelihood, and similar results were obtained (Bresnan 2007b).

3.2 Results and discussion

The results of Experiment 1 have interesting properties which motivate the design of the linear mixed-effect regression model used for the analysis of the data. We first

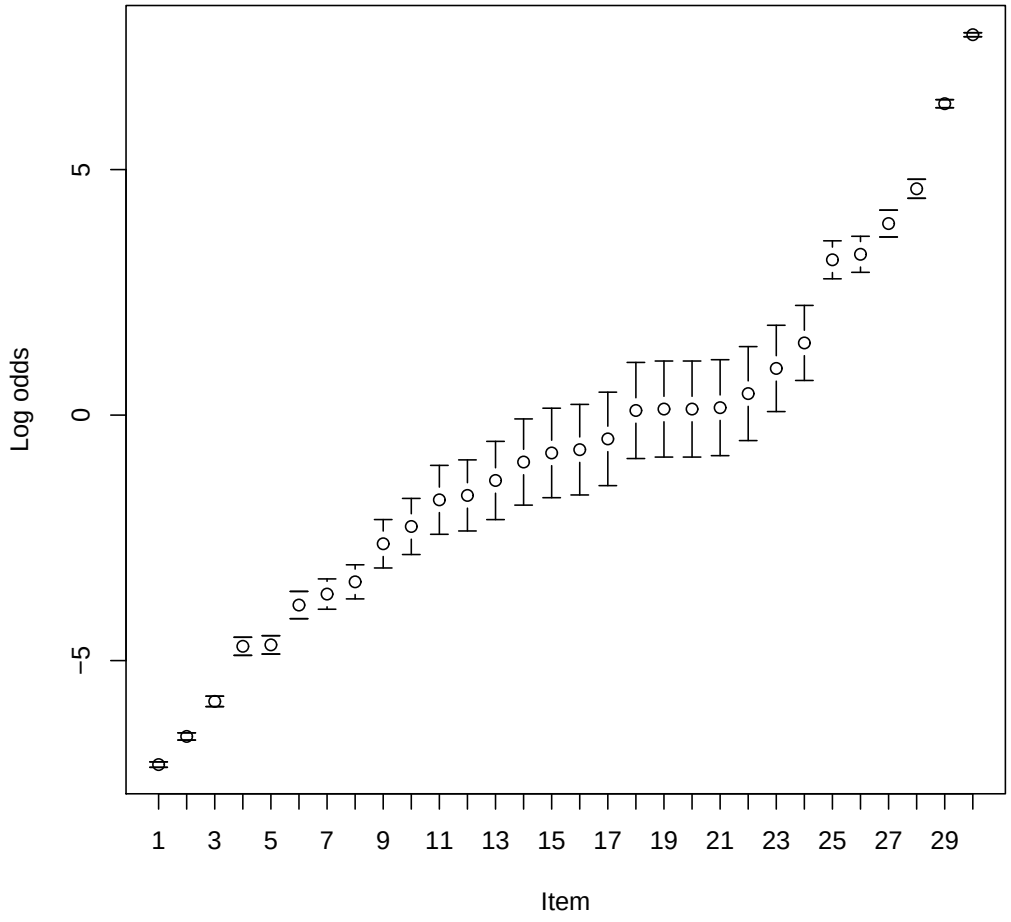


Figure 6: Experiment 1 items: fitted values (corpus model log odds) and 95% confidence intervals. Except for the middle range, the items are well differentiated by the model values.

discuss the random effect structure of the model and then turn to the fixed effects.

Figure 7 displays the mean ratings for the US and Australian participants for the V NP PP version of each of the 30 items, plotted against the corpus model log odds for V NP PP.[25] The regression lines show a linear correspondence between the mean ratings of each group and the corpus model log odds. But Figure 7 gives no idea of individual performance. Does the linear correspondence it depicts between corpus model probability (on the $logit$ scale) and ratings by the group also hold for the individuals in the group?

The trellis plots in Figure 8 show the ratings of the items by probability for each of the 19 US participants (top trellis plot) and 20 Australians (bottom trelllis plot). Each panel within the two trellis plots displays the relation between ratings and corpus log odds of one of the participants. The plots thus display every data point of every experimental participant in both groups. Even though the unaveraged raw data of the individuals shows much more variance, as expected, a roughly linear relationship between ratings and corpus probability remarkably appears in almost every panel, as shown by the individual regression lines.

From these plots we can also see that participants varied in how much of the rating scale they used. For example, in the top trellis plot of Figure 8, S4.us's ratings cluster closely around the middle band of the ratings scale from 40 to 60, while S5.us's ratings extend from near 0 to 100. This difference in rating range or amplitude is reflected by *slopes* of the regression lines in each plot: a steeper slope corresponds to a wider range of ratings given. Participants also varied somewhat in the baseline they appeared to be using. For example, S5.us and S10.us have approximately similar slopes, but S10.us's regression line intercepts the vertical axis higher up, suggesting a higher baseline, or average rating.

The individual regression lines in Figure 8 are all from a single linear mixed effect regression model with random slopes and intercepts together with a fixed effect of corpus log odds and an interaction between variety and log odds.[26] The main ef-

[25]Individual ratings are often standardized in order to reduce inter-individual variability as much as possible (e.g. Bard et al. 1996). The models we fit to the ratings data adjust for individual variation in the range of the ratings scale, in a way explained below.

[26]Recall that the random effect values are not parameters of the model but best linear unbiased predictors (BLUPS). See n. 16. Hence only five degrees of freedom in all are used to fit the random inter-subject variation shown in Figure 8: three for the random intercepts and slopes—one for the standard deviation (s.d.) of the random intercepts of subject, one for the s.d. of the by-subject slopes, and one for the s.d. of the random verb intercepts—and two for the fixed effects of the intercept and the corpus model probability predictor. Likelihood ratio tests show that each of the random effects contributes significantly to the model fit. (Because of the fitting algorithm, likelihood ratio tests are used for the random effect structures of linear mixed effect regression models, and not for comparing

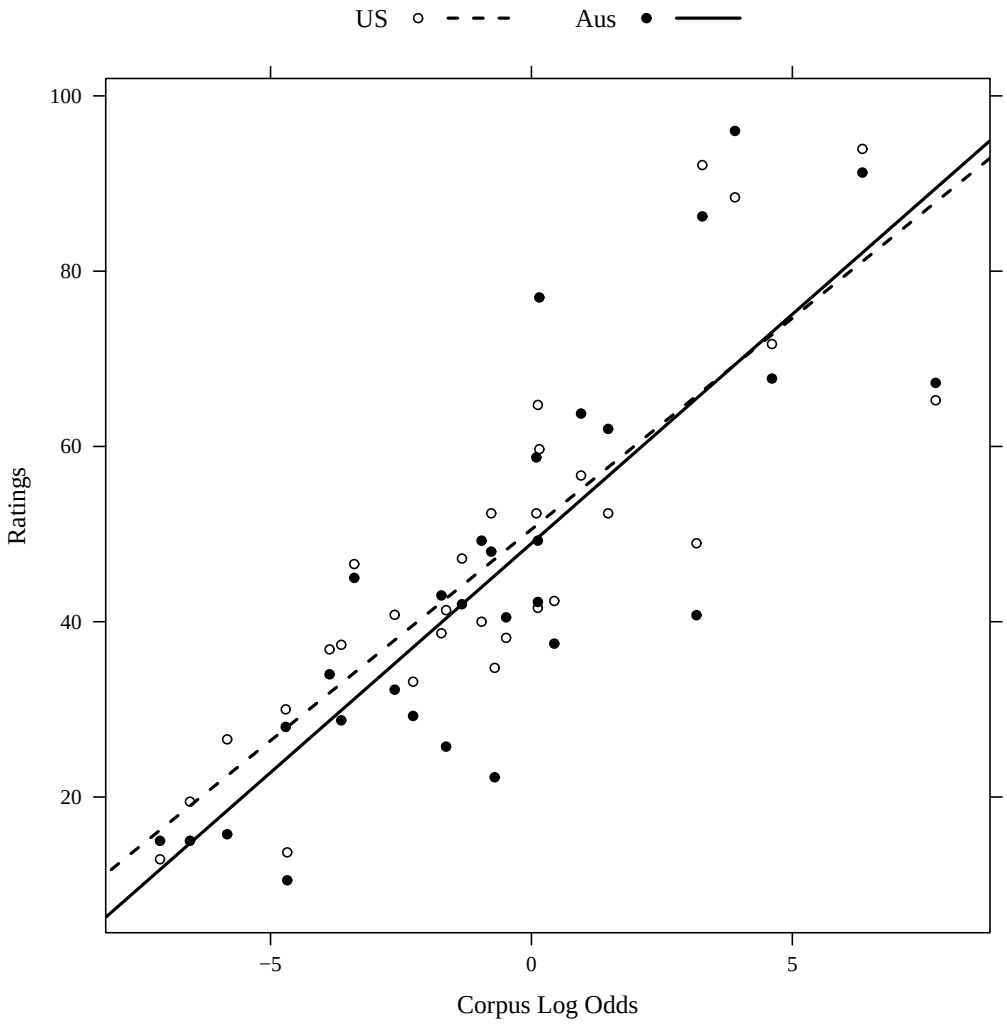


Figure 7: Mean ratings of US and Australian participants for each item by corpus log odds. Regression lines show ratings increasing with probability for both groups.

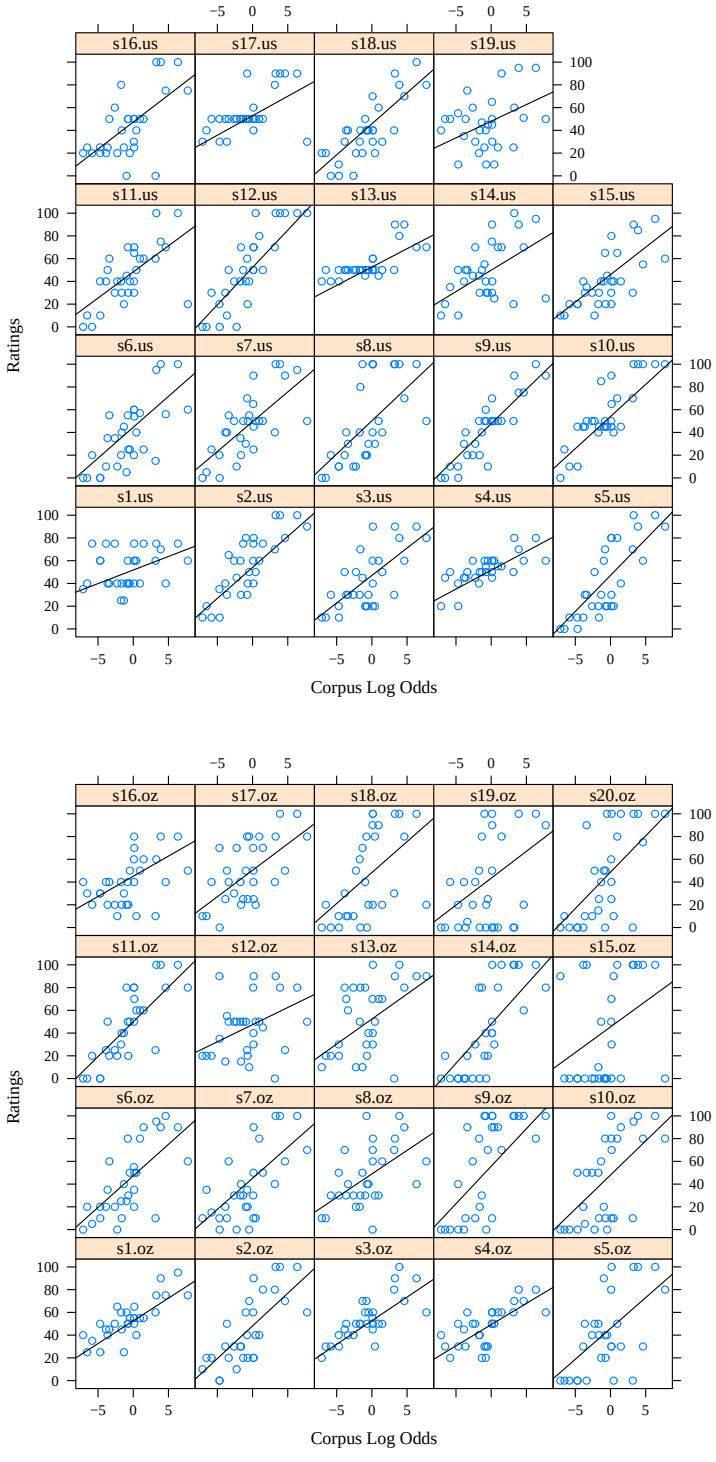


Figure 8: Fit of a single linear mixed effect regression model with random intercepts and slopes to data of US and Australian participants. Regression lines show variation between subjects.

fect of corpus log odds is highly significant, $t = 15.65$, $\Pr(> |t|) = 0$. The random effect structure of the model provides an impressive fit to the individual variation in rating trends. We adopt this random effect structure for our final model of the data: random intercepts for participant and verb, random by-subject slopes for the corpus model log-odds.

As seen in Figure 8, the random effects structure of a mixed effect model of the ratings data allows direct modeling of inter-subject variation in both means and slopes. The fixed effects are shared across the groupings of the random effects. The general structure of our model is shown in (14).

(14) **A Linear Model with Random Intercepts and Slopes**

$$\boldsymbol{y_{ij}} = \boldsymbol{X_{ij}\beta} + \boldsymbol{S_i s_i} + \boldsymbol{V_j v_j} + \boldsymbol{\epsilon_{ij}}$$

In this model the ratings are described in terms of the cross-group explanatory variables $\boldsymbol{X_{ij}}$, two randomly varying group effects, one for experimental subjects $\boldsymbol{S_i}$ and one for verbs $\boldsymbol{V_j}$, and a random error $\boldsymbol{\epsilon_{ij}}$. The random effect of verb is modeled as a random intercept. The random effect of subjects has two components: a random intercept representing a subject's rating baseline, and a random slope representing a subject's rating range modeled as a by-subject slope adjustment to the corpus log odds. In other words, the random effect of participant adjusts for differences in mean ratings between participants. The by-subject adjustment in slope adjusts for differences in the range of the rating scale used by participants. The random effect of verb adjusts for item variation attributable to dative verb bias toward one of the alternative dative constructions.

Let us now turn to the fixed effect structure of our final model $\boldsymbol{X_{ij}}$. In Section 1 we showed how multiple information sources characterizing context, meaning, and form of the construction contribute systematically and quantitatively to the binary choice between alternative dative constructions in production. We hypothesize that in comprehension-based tasks, including ratings of sentences read in context, speakers are able to recognize probabilistic differences in alternative grammatical constructions from their implicit knowledge of these information patterns—relative complexity, definiteness, and pronominality of theme and recipient, animacy of recipient, number of theme, and previous occurrence of a prepositional dative. If so, behavioral sensitivity to probabilistic differences in alternations may vary in speakers of different varieties of a language where probabilistic syntactic variation exists.

fixed effects; see n. 15).

We fit a model to the combined data from the Australian and US participants. The random effect structure remained as just described: random verb and participant intercepts and random by-subject adjustments to slope with respect to corpus model log-odds. In the initial model the fixed effects consisted of all of the fixed effect predictors in the corpus model, to each of which we added an interaction with variety. This model was then simplified by removing all interaction predictors whose estimate was less than twice the standard error. This left all the predictors in the model as main effects, together with one interaction, between variety and relative syntactic complexity.

Inspection of the residuals and the density plots of the posterior distributions of the estimates showed that the model assumptions are reasonably satisfied (Baayen, Davidson, and Bates 2008; Baayen 2008). Model collinearity was calculated; the 'condition number' $c = 7.65$ shows very slight collinearity, which is not surprising given the log odds included in the random effects (see n. 14). The model also shows a reasonable fit to the data. Overall, it accounts for over half of the variance in the data ($R^2 = 0.529$). Not surprisingly, a very large component of the variance is explained by the random effects: for a model consisting of the random effects only, $R^2 = 0.49$.

The coefficients for the final resulting model are shown in Table 5. The p-values and the upper and lower 95% confidence limits are derived from posterior distributions using the pvals.fnc function in the languageR package for linear mixed effects regression modeling (Baayen, Davidson, and Bates 2008). The independent main effects and the interaction term are reliable.

The model shows that (after adjusting for the random effects of subject, verb, and by-subject sensitivity to corpus probability) there were significant main effects of previous occurrence of a *to*-dative in the previous context, pronominality of recipient, pronominality of theme, definiteness of recipient, definiteness of theme, animacy of the recipient, and number of theme. All of these effects are consistent with harmonic alignment (Figure 5), in the sense that the more harmonic patterns are rated more highly than the less harmonic patterns. This is to be expected if participants are sensitive to the corpus model probabilities, because the more harmonic patterns are the more probable, based on the frequencies and distributions in the data. The effect sizes are plotted in Figure 9 (see n. 20). Each predictor contributes significantly to the model fit; the separate removal of variety and relative syntactic complexity of recipient and theme also removes the interaction term, and significantly degrades the model fit. Given the very powerful method of modeling individual differences in the use of the rating scale, it is notable that variety still contributes to the explanatory value of the model.

Table 5: Model parameters for Experiment 1

Fixed Effects:

	Estimate	95% Confidence Limits lower	upper	p-values
(Intercept)	50.251	39.8172	61.175	0.0001
variety = Aus	−1.802	−5.2705	1.956	0.3176
log rec-theme diff	3.406	−0.9040	7.891	0.1114
previous *to*-dative	11.032	4.6250	15.767	0.0004
recipient = pronoun	−16.791	−22.0959	−10.237	0.0001
theme = pronoun	14.445	5.9744	23.351	0.0010
recipient = indefinite	16.304	10.2424	22.033	0.0001
theme = indefinite	−25.800	−29.7793	−20.730	0.0001
recipient = inanimate	21.609	15.4746	28.489	0.0001
number of theme = singular	11.889	2.189	5.430	0.0001
variety : log rec-theme diff	3.224	0.3832	6.168	0.0378

Random effects:

Group	Std.Dev.	HPD95lower	HPD95upper
subject slope (corp model log odds)	1.5431	1.0263	2.1670
subject (Intercept)	4.3139	2.4073	6.0369
verb (Intercept)	17.8646	8.8611	20.4336
Residual	20.833	20.0458	21.7873

number of observations: 1170, groups: subject, 39; verb, 9

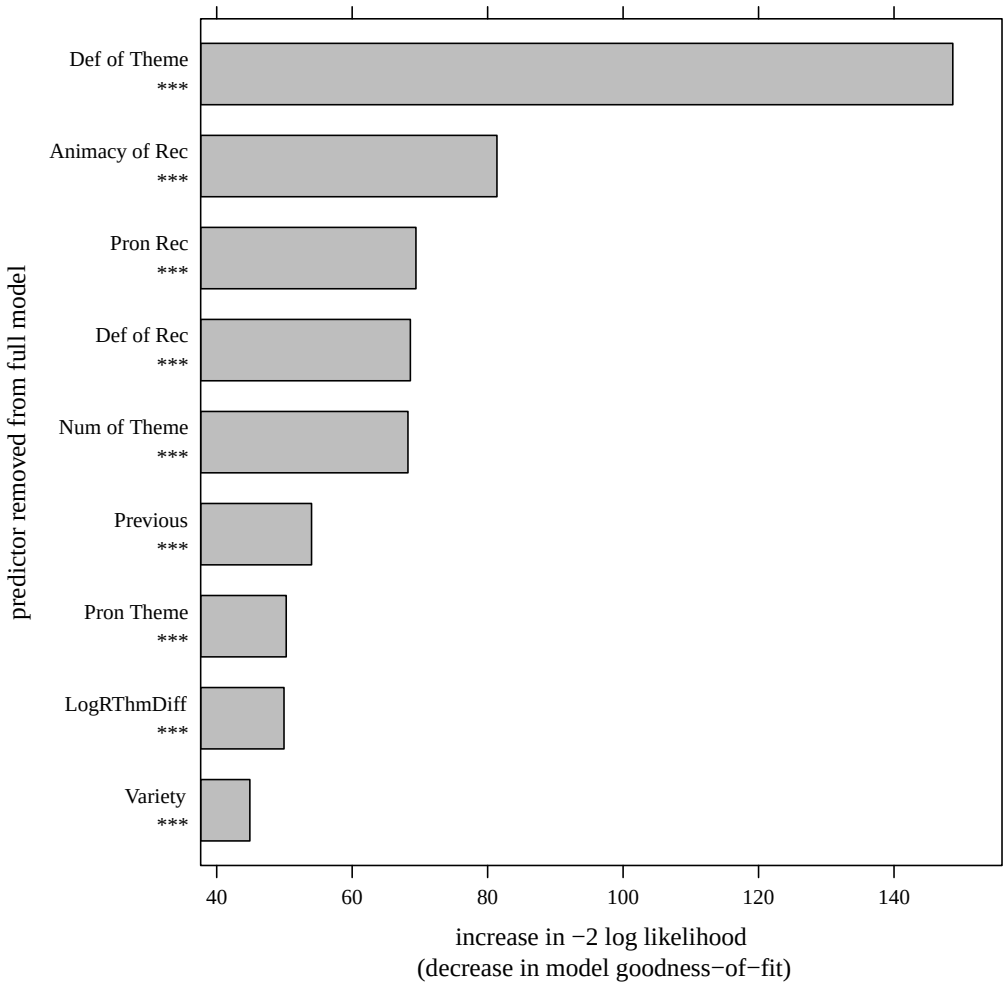


Figure 9: Effect Sizes of Predictors in Experiment 1
[‘.’ $p < 0.1$, ‘*’ $p < 0.05$, ‘**’ $p < 0.001$, ‘***’ $p < 0.0001$]

Notice that variety of English reliably interacts with the relative syntactic complexity of recipient and theme: for longer recipients the Australians favor the NP PP construction more compared to the US participants. The model predictions for the interaction are shown in Figure 10, together with the 95% confidence intervals.

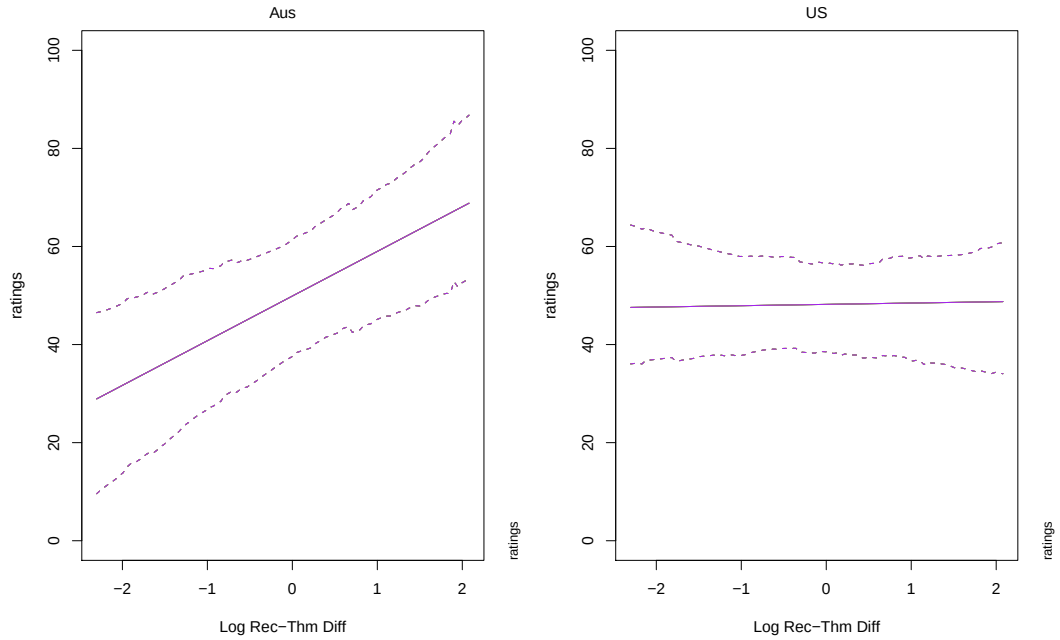


Figure 10: Model predictions of interaction between variety and relative complexity, with 95% confidence intervals from posterior simulations

For an illustration, consider the item in (15). The recipient and theme are both two words in length; the probability of the prepositional dative is 0.3309 and the double object dative 0.6691. If the recipient length is increased (for example, from *my kids* to *my kids and their cousin who is staying with us*, the model predicts that the ratings of the prepositional dative would increase more for the Australians than it would for the US participants. That is, the Australian participants would tolerate the change to *Instead of giving my kids and their cousin who is staying with us an allowance* less than the US participants.

(15) Speaker A:
I wish they had just one central place, you know, where you can just dump

all the recycling. Because really I am not really looking for the money portion of it, you know.

Speaker B:
Well I used to. It used to be a good days work.

(i) Instead of giving my kids an allowance,

(ii) Instead of giving an allowance to my kids,

I just told them they could go around the neighborhood and collect things to be recycled and then I would drive them over and they would get some money.

Experiment 1 shows clearly that both groups of participants are sensitive to the corpus probabilities and that the Australians show a greater effect of relative syntactic complexity or end weight than the American subjects; as the recipient argument of a dative gets longer relative to the theme, the Australians have a greater liking of the V NP PP dative than the Americans.

4 Experiment 2: Continuous lexical decision

The ratings data obtained in Experiment 1 possibly reflect processes that come into play only after reading a sentence. But we have good reason to believe that higher-level corpus probabilities are involved much more immediately in *producing* dative constructions during spontaneous speech. In a corpus study of the acoustically time-aligned and corrected dataset of spoken dative constructions (Recchia 2007), Tily et al. (2009) found that the corpus model probabilities of dative production in Bresnan et al. (2007a) predicted the acoustic duration of the word *to* in prepositional datives even after holding the low-level transitional probabilities of the adjacent words constant.

Experiment 2 is a comprehension experiment structured very much like the observational production study of Tily et al. (2009). It was designed to obtain data during sentence processing. More specifically, we conducted an experiment with American and Australian English speakers to investigate whether lexical decision latencies during a reading task would reflect the corpus probabilities and whether there were interactions between variety of English and the linguistic predictors of the corpus model. The task used was the Continuous Lexical Decision Task (Ford 1983) in which subjects read a sentence (or part of a sentence) word by word at their

own pace, but making a lexical decision as they read each word. The purpose of requiring a lexical decision, and not just a press of a button to get the next word, is to prevent any rhythmic responding, which is a known cause of spillover processing (see Ford 1983: 204).[27] The lexical decision task is made, though, in the context of fitting each word into the current syntactic construction. Ford showed that this method is sensitive to subject- and object-relative differences, which have been very well established and replicated in subsequent work (see Gennari and MacDonald 2008: 162). In Experiment 2 we were interested in responses to the word *to* in the dative NP PP as a function of linguistic predictors of the corpus model and also of variety. Given that the recipient does not occur before the word *to*, new probabilities were calculated by omitting any predictors related to the recipient and using log(length of theme) for the length measure. We call these new probabilities "partial-construction probabilities."

4.1 Method

4.1.1 Participants

The participants were 20 volunteers from the Stanford University community and 20 from the Griffith University community. They were paid for their participation. There were 10 males and 10 females in both groups. All participants were native speakers of English, did not speak another language as fluently as English, had not taken a syntax course, and had grown up in the US (the Stanford participants) or Australia (the Griffith participants). None had taken part in Experiment 1.

4.1.2 Materials

The experimental items for Experiment 2 consisted of 24 of the 30 items from Experiment 1. Those omitted were from the middle segment of corpus model probabilities for the prepositional dative construction. As seen in Figure 6 items in the middle range of probabilities were overrepresented: having the widest confidence intervals, these were poorly differentiated by the fitted values of the model from their flanking items.

Each experimental item consisted of a context passage, which was to be read normally, and a continuation of the passage in the prepositional dative form, which

[27] In spillover processing, button presses advance rapidly beyond comprehension, requiring catch-up to process material that came earlier in the sequence. Compare for example the discussion of self-paced reading and EEG by Van Berkum et al. 2005.

was to be read while performing the Continuous Lexical Decision Task. The continuation was either the same as the original from the corpus or it was the constructed prepositional alternative. The continuation always began with the word before the dative verb and all lexical items in the experimental items, up to and including the word after *to*, were real words. Some experimental items included nonwords after that point, simply to give more opportunities for responding no to the lexical decision. An example of an item is given in (16).

(16) *Speaker:*
I'm in college, and I'm only twenty-one but I had a speech class last semester, and there was a girl in my class who did a speech on home care of the elderly. And I was so surprised to hear how many people, you know, the older people, are like, fastened to their beds so they can't get out just because, you know, they wander the halls. And they get the wrong medicine, just because, you know, the aides or whoever

just give the wrong medicine to them just sornly

The 6 omitted items served as fillers, with the continuation being given in the NP NP structure. A sample item is given in (17).

(17) *Speaker A:*
The technology is really, you know, going crazy with PCs.

Speaker B:
It's clearly a productivity enhancement device and allows you to do –

Speaker A:
Originally I didn't think it was. I thought that what, you know, we ended up doing was doing all of the secretarial work and the secretaries had nothing to do. And I guess part of that is true. I do all my own typing. I

don't give the secretary paper to lorm vlob any more

As can be seen, the continuation of these fillers sometimes also contained nonwords. Apart from these 6 fillers, another 10 were constructed. These consisted of a passage and a continuation that did not have a dative construction. The continuations of these fillers always contained one or more nonwords.

Each item was followed by a yes/no question that appeared on a new screen after a response had been made to the last lexical item in a continuation. This was to encourage participants to read each passage and continuation. Thus, for example,

after the response to *sornly* in (16), the question in (18) appeared on a new screen.

(18) Was the speech about the good care elderly get?

For the 24 experimental items, the partial-construction probability, that is, the corpus model probability based on the context, verb, and theme, but not the recipient, was calculated. The range of these partial-construction probabilities (in log odds) was from -4.53 to 3.08, with a mean of -0.87. The partial-construction corpus probabilities for the prepositional dative construction for the 24 experimental items are shown in Figure 11.

4.1.3 Procedure

The participants were tested individually in their own countries. Participants were given written instructions outlining the procedure (see Appendix 1). They were told that they would see the beginning of a conversation on the computer screen, followed by the next word of the continuation of the conversation and a line of dashes. They were given (19) as the example.

(19) *Speaker A:*
I just spoke to Peter on the phone. He didn't sound very well.

Speaker B:
Has he got this cold that is going around?

Speaker A: No. He

says _

For each item, the subject read the conversation, then the first word of the continuation. They then decided whether the first word of the continuation was a word or not and pressed the appropriate button (yes or no). Once a decision was made, the next word appeared to its right and the preceding word became dashes. A lexical decision was then made about the second word. This procedure continued until the last lexical item in the continuation. At the end of the continuation, the context and continuation disappeared and a yes/no question appeared relating to what had just been read. Participants were told that there were no tricks and that it would be obvious if something was a word or not. They were asked to read the conversations as naturally as possible, making sure they understand what they read. E-Prime software for Windows (Schneider et al. 2002a,b) was used to run the Continuous Lexical Decision Task. E-Prime gives software checks for whether the computer being used

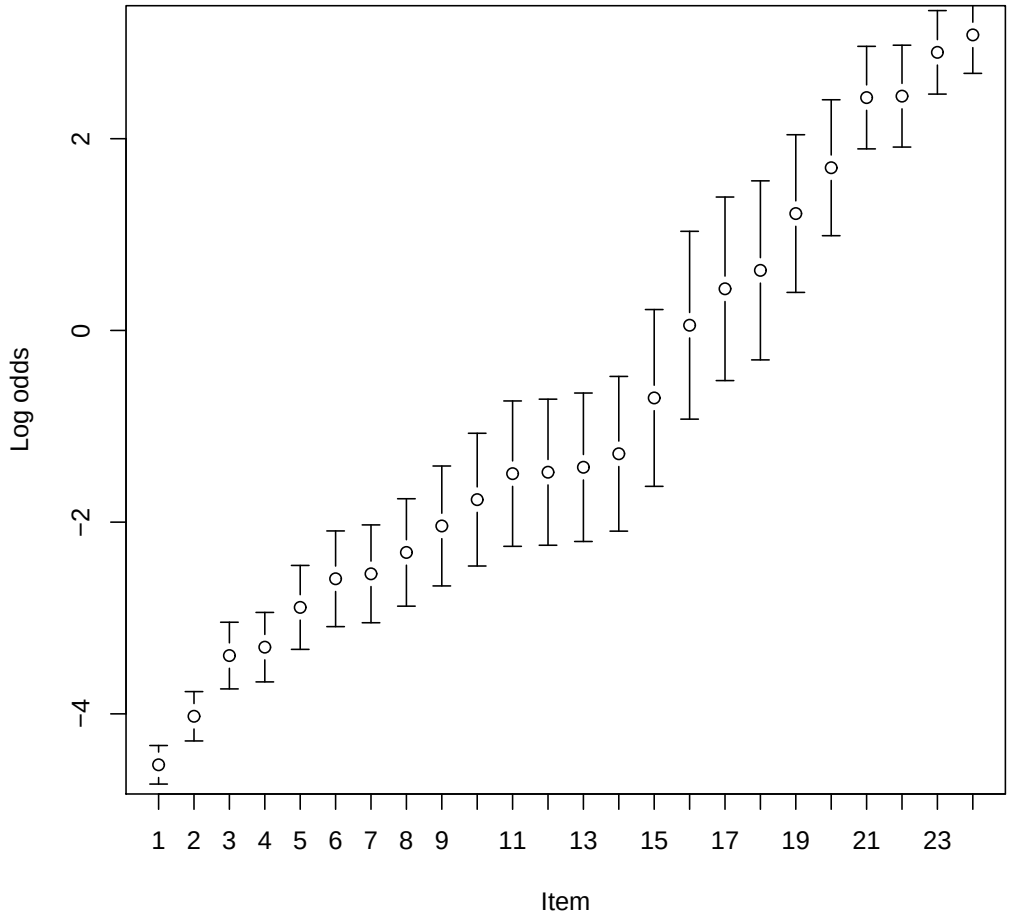


Figure 11: Partial-construction corpus log odds of Experiment 2 items with 95% confidence intervals

is suitable for millisecond timing and both computers used were shown by these procedures to be good.

4.2 Results and discussion

As an indication of whether participants had comprehended the passages and their continuations, an analysis of responses to the comprehension questions following the 24 experimental items was carried out. Results showed that comprehension was high and did not differ significantly for the Australians and Americans; the average number of correct responses was 20.5 for Australian males, 20.5 for Australian females, 20.9 for American males, and 21.4 for American females.

To reduce the effect of extreme reaction times, the raw RTs were first investigated for outliers. It was clear that there were three outliers. Two RTs of 10156 and 5584 milliseconds were well above the next highest RT (1496 milliseconds). One of 99 milliseconds was well under the next lowest RTs (239 milliseconds). The two extremely high reaction times were probably due to distraction and not any linguistic feature. The reaction time of 99 milliseconds was probably a mistaken press; the response time being unrealistically low as a true reaction time. Thus, a decision was made that all reaction times greater than 1500 milliseconds or less than 100 milliseconds should be eliminated. To further reduce the effect of extreme reaction times, all reaction times were logged. By logging both the dependent and predictor variables, the relation between the variables now describes how the proportional change in the reaction times on *to* varies with the proportional change in the corpus odds of the prepositional dative, given the partial information available to the reader.

Again we visualize the data to motivate the model for analysis. Figure 12 gives the mean reaction times at the word *to* for each item for the Australian (Aus) and American (US) subjects plotted against the partial-construction log odds of the corpus data, together with a nonparametric smoother for both varieties, which shows the trend by averaging local values in the data. From this plot we see that the Australian reaction times are somewhat slower overall. There is an apparent general trend in both groups for reaction times to decrease as the corpus log odds increase.

There are a number of incidental variables that must be controlled for in this task. Specifically, controls are needed for random effects of the dative verb in the item and for subject and inter-subject differences in response to item order. We used a random intercept for the former and a random slope for the latter: during the task some participants may improve their speed with practice, while others may tire and slow. A random effect of the word preceding *to* was tested but did not significantly improve the model fit, and so was eliminated from modeling. We also added several

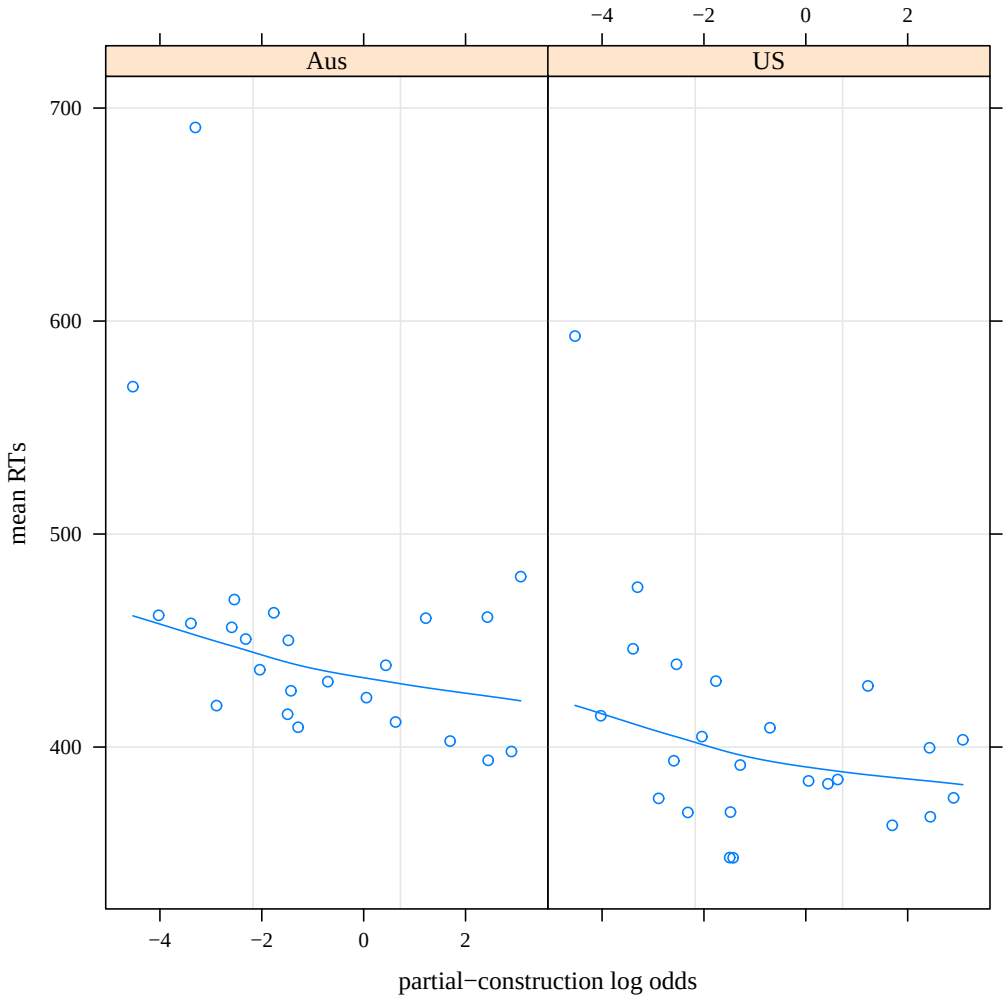


Figure 12: Mean reaction times by partial-construction log odds for both varieties

controls to the main effects: it is well known that the best predictor of reaction time to a word is reaction time to the preceding word, so that was included in the fixed effects as a shared effect across all groupings. Item order was also added as a possible cross-group shared factor in the fixed effects, to model any general effect of the task trial order such as fatigue (causing slowing) or practice (causing speed ups). With all of these controls in place, the partial construction log odds remained a significant predictor of the reaction times to word recognition, $t = -2.14$, Pr$(> |t|) = 0.0324$.

Having established the random effect structure and controls, as well as a reliable relation between linguistic probabilities and our dependent variable, we turned to our main question: whether the two groups varied in the importance of the linguistic predictors that are components of the corpus model probabilities. We began with a full model containing all of the coded predictors present in the data (excluding recipient predictors) in addition to the random effects described above, and we selected models by eliminating predictors where the magnitude of the estimated coefficient was less than twice the standard error. In the final model two linguistic predictors remained: definiteness of the theme and length of the theme.

We measured the multi-collinearity of the model using the method in n. 14. The condition number of 3.31 shows that collinearity is of no concern. Inspection of the residuals and the density plots of the posterior distributions of the estimates showed that the model assumptions were reasonably satisfied (Baayen et al. 2008; Baayen 2008b). For this model $R^2 = 0.5668$. For a model consisting of the random effects only, $R^2 = 0.4643$. Thus the fixed effects account for a substantial amount of the variance.

The model parameters are shown in Table 6. These results show that there was a significant main effect of length of the theme and that variety significantly interacted with length of theme. The main effect of variety was also significant, with Australian subjects responding more slowly than the Americans. Definiteness of theme was significant. Reaction time to the word preceding *to* was significant. Item order was also significant. The effect sizes are shown in Figure 13 (see n. 20). Given the low signal-to-noise ratios that are common in reaction time experiments, it is remarkable that variety contributes substantially to the explanatory value of the model.

To explore the group difference in response to the length or syntactic complexity of the theme, we found that nesting the subject variable within a random effect of variety added nothing to the model (by a likelihood ratio test, $\chi^2 = 0, Df = 1, Pr(> \chi^2) = 1$) and did not alter the interaction with length of theme.

Could the interaction with length of theme somehow be traced to specific subjects who were faster or slower than average in their lexical decision times? The random intercept for subject already represents random individual differences in re-

Table 6: Model parameters for Experiment 2

Fixed Effects:

	Estimate	95% Confidence Limits lower	upper	p-values
(Intercept)	5.9998	5.9064	6.0913	0.0001
variety = Aus	0.1098	0.0455	0.1708	0.0008
log theme length	0.1164	0.0820	0.1542	0.0001
theme = indefinite	0.0337	0.0046	0.0632	0.0190
log RT to preceding word	0.3378	0.2905	0.3874	0.0001
item order	−0.0021	−0.0034	−0.0008	0.0026
log theme length : variety = Aus	−0.0741	−0.1133	−0.0364	0.0002

Random effects:

Group	Std.Dev.	HPD95lower	HPD95upper
subject slope (item order)	0.0027	0.0009	0.0045
subject (Intercept)	0.1159	0.0762	0.1148
verb (Intercept)	0.1161	0.0599	0.1674
Residual	0.1679	0.1620	0.1782

number of observations: 953, groups: subject, 40; verb, 8

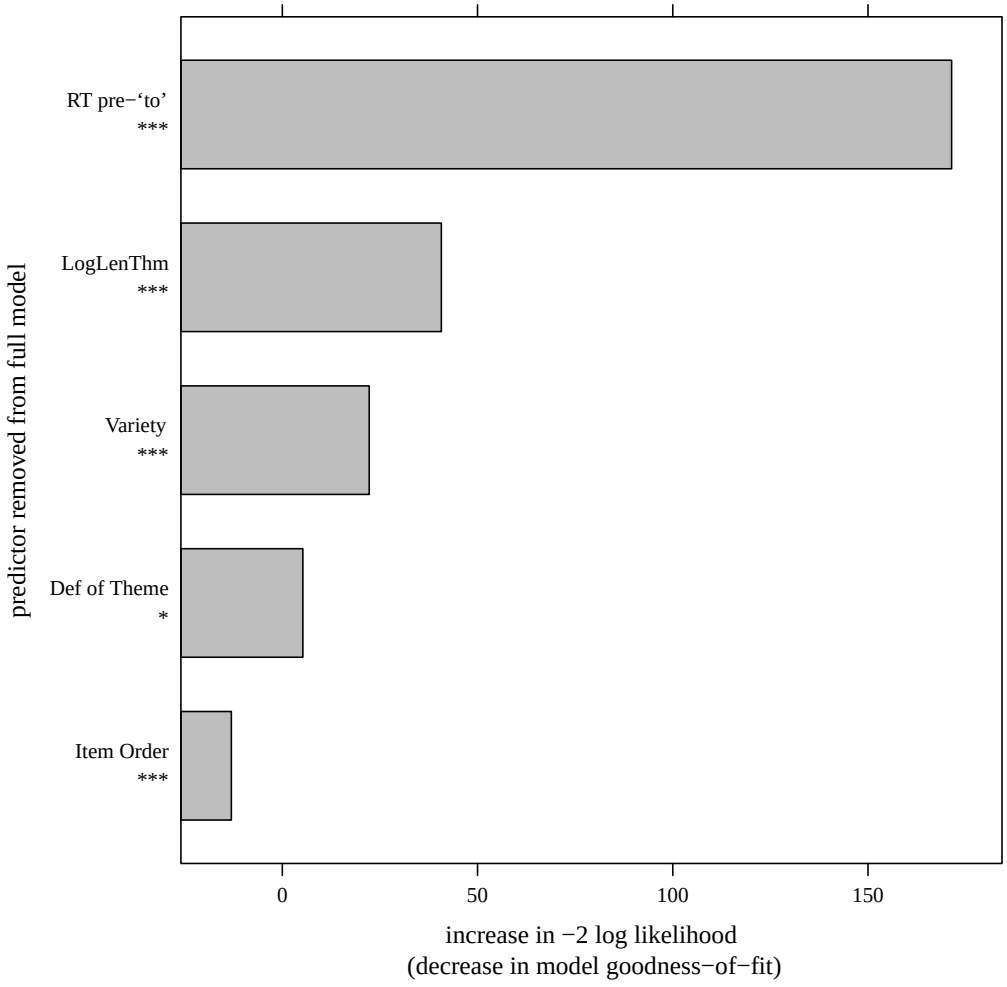


Figure 13: Effect Sizes of Predictors in Experiment 2
[‘.’ $p < 0.1$, ‘*’ $p < 0.05$, ‘**’ $p < 0.001$, ‘***’ $p < 0.0001$]

action time, and the interaction with variety holds after adjusting for these individual differences. Nevertheless, we further checked this hypothesis by classifying subjects as "fast" or "slow", depending on whether their mean reaction time to *to* was above or below the mean for all subjects. Speed class shared across all random groups was then added as a control in the fixed effects of the regression analysis. Results showed that all other effects including the interaction of variety with length of theme were robust after adjusting for this hypothetical speed effect.

Interestingly, the direction of the main effect of length of theme is consistent with the harmonic alignment pattern of Figure 5, in the following sense. More complex themes (approximated by length in words) favor the double object construction over the prepositional dative and thus reaction times to *to* increase with length of theme. The interaction with variety indicates that the Americans show this effect much more sharply than the Australians; see Figure 14. As an example, consider the item shown in (16), containing *give the wrong medicine*. While the Australian speakers would apparently show little increase in reaction time at *to* if the theme were lengthened to *the wrong and often dangerous medicine*, the US participants would show a marked slowing of reaction time. The range of theme lengths was from one word to ten, with a mean of 2.35 and a median of 2. Adjusting for all other variables, the models showed that American reaction times started exceeding the Australians' as theme length increased beyond three words.

Analyses showed that this interaction between variety and length of theme is very robust. It cannot be attributed to group differences in speed caused, for example, by a hypothetical ceiling effect in the slower Australian subjects' decision latencies. There is no significant group effect of speed to explain the interaction. Adding speed as a controlling shared factor across groups does not eliminate the interaction. The Americans were slowing much faster than the Australians in their lexical decisions on the word *to* and were exceeding the Australians' RTs when themes grew longer than three words.

It might be thought that the Australians could, in fact, show a sharper increase in reaction time as length of theme increases, but perhaps as a delayed effect. Thus, a linear mixed effects regression model was fit to the data using log RTs on the word after *to* as the dependent variable and adding the log RT to the word *to* as a possible predictor. The regression also used the word after *to* as a control random effect. This is a standard technique for modeling possible spillover effects in word-by-word reading tasks with mixed-effect regression; see Kuperman et al. (2009), for example. Results showed that there was no interaction between variety and length of theme at this post-*to* position. Moreover, at this point in the sentence, there was a significant main effect of length of theme such that reaction times *decreased* after

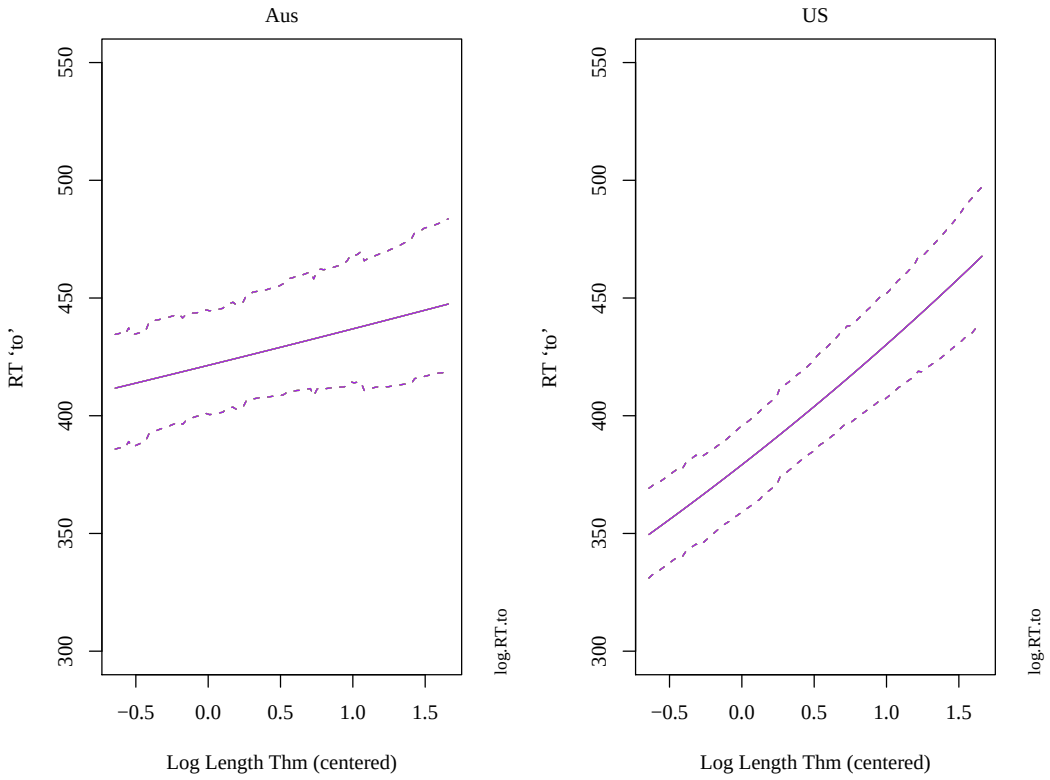


Figure 14: Predicted reaction times (RTs) on *to* showing the the partial effects of the interaction between length of theme and variety, with 95% confidence intervals from posterior simulations

longer themes.

Regarding the main effect of definiteness of theme, it was also consistent with harmonic alignment (Figure 5), with reaction times at the word *to* increasing after an indefinite theme where the probability of a *to*-dative decreases. The speakers of the two varieties do not differ in this effect.

5 Experiment 3: Sentence completion

In both Experiments 1 and 2 there were shared effects showing sensitivity to harmonically aligned linguistic properties. In Experiment 2, as we have just seen, indefiniteness of the theme (which is harmonically aligned in the double object construction: see Figure 5) was associated with slowing reaction times to recognize the preposition *to* as a word for both the American and the Australian participants. In Experiment 1 inanimacy of the recipient, indefiniteness of the recipient and of the theme, and pronominality of the recipient and of the theme were all associated with subjects' ratings in exactly the way that harmonic alignment would lead us to expect: for example, an indefinite theme is harmonically aligned in the double object construction, so it is associated with subjects' lower ratings of the prepositional dative; conversely, an indefinite recipient is harmonically aligned in the prepositional dative construction, so it is associated with subjects' higher ratings of the prepositional dative. These shared main effects reveal sensitivity to corpus model probabilities in both comprehension-based tasks, sentence ratings and continuous lexical decision while reading.

Yet in both experiments, interactions with variety existed. At first glance, the results of Experiment 2 might seem to contradict those of Experiment 1. In Experiment 1, Australians showed an end-weight effect of the recipient and the Americans did not, while in Experiment 2 the Americans showed a much stronger end-weight effect of the theme than the Australians. If one thinks of the results only in terms of end-weight then it is difficult to reconcile the results of the two experiments. However, when one reflects on the results in terms of whether the linguistic predictors favor or disfavor the prepositional dative construction V NP PP (recalling Figure 5), then a consistent pattern emerges.

Consider Table 7, which summarises how variety interacts with the linguistic predictors favoring or disfavoring V NP PP in Experiments 1 and 2. The top half of Table 7 concerns decision latencies from Experiment 2 and thus an increase suggests greater difficulty and a decrease suggests greater ease in processing, while the bottom half of Table 7 concerns ratings from Experiment 1 and thus an increase sug-

gests greater acceptance. Compared to the Americans, the Australians show more effect of a property that favors prepositional datives and less effect of a property disfavoring them. One possibility is that the Australian group has a higher expectation of prepositional datives than the US group. Increases in theme length disfavor V NP PP, but, unlike the Americans, the Australians' reaction times at the word *to* are increasing much more slowly as theme length increases, as though they are more tolerant of V NP(LongTheme) PP than the Americans. Turning to the ratings experiment, increases in recipient length favor NP PP, and while the Australians show a large effect of favoring NP PP in ratings as recipient length grows, the Americans show no effect (both the variety contrast and the main effect of log recipient-theme difference lacking significance), as though the Americans are more tolerant of V NP(LongRecipient) NP than the Australians.

Table 7: Summary of variety differences in Experiments 1 and 2

Decision latency experiment:

property	expectation	RT on *to*
theme length grows	disfavors V NP PP	US RTs increase faster (slowing down)

Rating experiment:

property	expectation	rating of V NP PP
relative recipient length grows	favors V NP PP	only Aus increases

Reflecting on the results in terms of whether the linguistic predictors favor or disfavor an V NP PP suggests that the two groups may be more or less tolerant of different stuctures. One possibility is that the Australians have a higher expectation of V NP PP than the US group. If so, it might be expected that they would produce more prepositional datives than the Americans do, *in the same contexts*. To obtain evidence about differences in production, we used a sentence completion task in Experiment 3.

5.1 Method

5.1.1 Participants

The participants were 20 volunteers from the Stanford University community and 20 from the Griffith University community. They were paid for their participation.

There were 10 males and 10 females in both groups. All participants were native speakers of English, did not speak another language as fluently as English, had not taken a syntax course, and had grown up in the US (the Stanford participants) or Australia (the Griffith participants). None had taken part in Experiments 1 or 2.

5.1.2 Materials

The items for Experiment 3 consisted of all 30 items from Experiment 1. As with Experiments 1 and 2, the context was given for each item, though each item ended after the dative verb and was followed by lines where a completion could be entered. The items were given in a random order for each subject.

5.1.3 Procedure

Each participant was tested in her or his own country. Participants were given a booklet with instructions and the 30 items. The instructions stated that in each of the given passages one or two speakers were talking informally about different topics. They were also told that the final sentence in each item was left unfinished. They were instructed to read each passage and then complete the unfinished sentence in the way that felt most natural to them. They were instructed that they need not spend a lot of time deciding how to complete it, but to just write down what seemed natural.

5.2 Results

The transcripts of each subject were checked separately by each author for NP NP and NP PP *to*-dative completions. Examples of a double object and a prepositional dative completion, as well as a completion that was neither are given in (20).

(20) *Speaker A:*
The technology is really, you know, going crazy with PCs.

Speaker B:
It's clearly a productivity enhancement device and allows you to do –

Speaker A:
Originally I didn't think it was. I thought that what, you know, we ended up doing was doing all of the secretarial work and the secretaries had nothing to do. And I guess part of that is true. I do all my own typing.

I dont give

—myself time to relax when I could be working. (US: NP NP)

—menial jobs to my secretary anymore because the computer does it all for me. (Australian: NP PP)

—dictation when I can do it myself just as quickly and neatly. (US: neither NP NP nor NP PP)

The average level of production of datives for the 30 items was 0.55 for the Australians and 0.56 for the Americans. For the Australians, 0.42 of their datives were NP PP *to*-datives, while for the US, the corresponding figure was 0.33. The data were tabulated as numbers of *to*-dative, double-object, and other constructions by subject, gender, and variety. For analysis, a logistic generalized linear model was fit to the tabular responses of *to*-datives vs. double-object datives per subject, as a function of variety interacting with gender. The model statistics and odds ratios are displayed in Tables 8 and 9. From Table 8 we see that the Australians produce significantly more *to*-datives than the Americans, after controlling for gender. From Table 9 we can infer that the Australian males were more than three times as likely to produce *to*-datives as the US males, in the same contexts. (The intercept gives the odds of producing prepositional datives over double-object constructions for the reference levels of the variety and gender predictors, which are US and male, respectively. The intercept odds of producing a prepositional dative over a double object construction are about one in two—0.4867. In contrast, the Australian participants who were not female had odds of producing prepositional datives over double-object constructions of not quite two to one—1.8677.) The interaction term shows that the female Australians have a reduced odds of producing prepositional datives, bringing them closer to the US participants than the Australian males.

Table 8: Wald Statistics for Experiment 3 Model

	Df	Deviance	Resid. Df	Resid. Dev	P($>|\chi^2|$)
NULL			39	63.153	
variety	1	4.186	38	58.967	0.04076
gender	1	1.557	37	57.410	0.21213
variety:gender	1	3.635	36	53.775	0.05657

Table 9: Odds Ratios from Experiment 3 Model

	Odds Ratio	Estimate	P($>\chi^2$)
intercept	0.4867	−0.7201	0.0000
variety = Aus	1.8677	0.6247	0.0056
gender = f	1.121	0.1139	0.6200
variety = Aus : gender = f	−0.5422	−0.6121	0.057

6 Theoretical implications

In the experimental tasks of sentence rating and continuous lexical decision while reading, both the American and Australian subjects showed sensitivity to the spoken English corpus model probabilities of the dative construction (or partial construction). In Experiment 1 subjects gave higher or lower ratings to prepositional datives according to their higher or lower probabilities of occurrence in the given contexts. In Experiment 2, subjects while reading prepositional datives had faster or slower lexical decision latencies at the word *to* according to the higher or lower probability of occurrence of the partial prepositional dative in its context. The experiments show that subjects have strong predictive capacities, preferring and anticipating the more probable of two alternative syntactic paraphrases.

How could the subjects accomplish these predictive tasks? In both experiments, subjects' responses showed significant relations to the component linguistic variables of the corpus model. In Experiment 1 preference for type of dative construction was overwhelmingly in accordance with quantitative harmonic alignment (Section 2), with the main effects of relative complexity of recipient and theme, definiteness of recipient and theme, pronominality of recipient and theme, and animacy of recipient all going in the directions consistent with harmonic alignment. In Experiment 2

the partial-construction properties of length and definitenes of theme argument were among the main effect predictors of reaction time, in the directions expected from the harmonic alignment pattern shown in Figure 5: a definite theme favors a prepositional dative, and leads to faster decision latencies on *to* after controlling for all of the other variables; a longer theme favors a double object construction, leading to slower decision latencies on *to*.

According to a class of 'parsing-based' or 'memory-storage' theories (see Hawkins 1994, 2004, 2007; Gibson 1998; Grodner and Gibson 2005; Temperley 2007 for representative work), language users' sensitivity to what is more or less unlikely in linguistic contexts could simply derive from their sensitivity to what is more or less difficult to parse. On these theories limited cognitive resources for memory storage and semantic integration cause difficulty in parsing complex syntactic dependencies. Because speakers tend to accommodate hearers by avoiding sources of linguistic difficulty, the easiest linguistic variants to process in comprehension may become the most likely to be used.

Figure 15 illustrates one specific parsing theory analysis of data like ours. (See Temperley 2007 for discussion of similar predictions of others.) On this theory the difficulty of integrating a second argument with a ditransitive verb increases with the dependency length of the intervening first object (Chen et al. 2005: 284). The dependency length is calculated as the number of words that introduce new discourse entities between the start and end of the syntactic dependency—hence, as the number of lexical words.[28] The difference in length is illustrated for the head-argument dependencies between the verb *brought* and the beginning of its second complement in the Figure: there are zero lexical words spanned by the dependency arrow in the first and third examples and there are two lexical words (*children*, *van* and *pony*, *van*) spanned by the arrow in the second and fourth examples. Such differences in dependency length are predicted to yield inverse effects on reaction times.

There are two problems for this type of resource-limited explanation of our findings. First, our models show that the harmonic alignment effects mentioned above cannot be reduced to the relative complexity of the recipient and theme (Experiment 1), nor to variation in the length of the theme (Experiment 2), because the models all control for these complexity effects. Thus the shared harmonic alignment patterns in production that recur across varieties and modalities of English and are reflected in judgment and comprehension behavior remain unexplained. Secondly,

[28]Dependency length measured by length in lexical words is highly correlated with the simple length-in-words measure used here. On the set of 2349 theme NPs in the dative database of Bresnan et al. (2007a), the two measures have a Spearman's $\rho > 0.91, p < 2.2 \times 10^{-16}$. See also Temperley (2007).

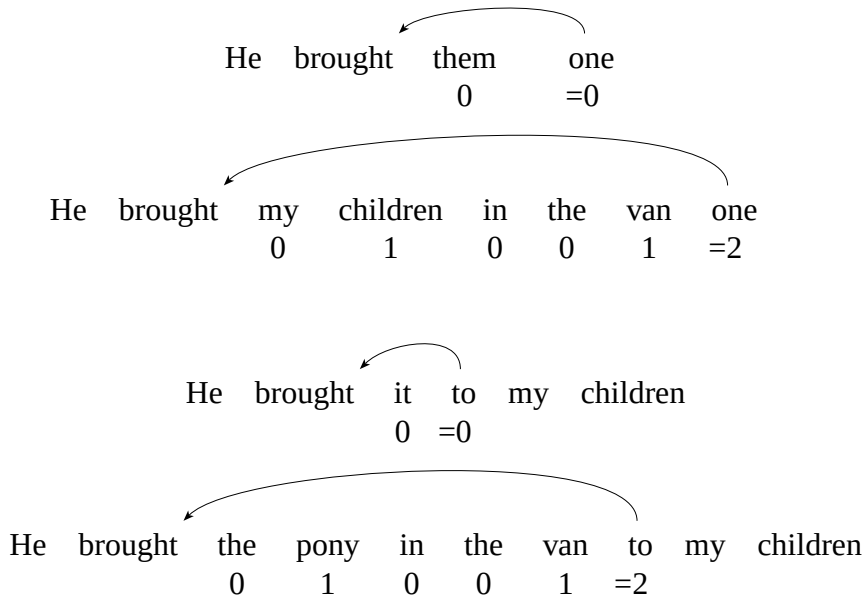


Figure 15: Dependency Length Theory

these theories offer no explanation for the covariation between the Australians and the Americans in the complexity effects themselves. In the ratings experiment, the Australian subjects showed a strong preference for V NP PP as relative length of recipient increased, while the US subjects showed none (Table 5 and Figure 10). And, in contrast, in the Continuous Lexical Decision Task, the US showed a much sharper slowing of reaction times as the length of the theme increased than the Australians. The Australians' reaction times increased only mildly at the word *to* and showed no lagging effect on the following word (Table 6 and Figure 14). The Australian subjects had slower decision latencies on average, which might have reflected a possible ceiling effect on reaction times, but this possibility was eliminated because the experimental analysis controlled for the mean speed of each subject as a random effect in the task. Moreover, when we added speed as a control in the fixed effects of the regression analysis, we found that our effects remained robust. The sentence completion task showed that the covariation in the judgment and comprehension tasks in Experiments 1 and 2 matched the production task in Experiment 3, with the Australians favoring prepositional datives more than the US participants.

Do other theories offer a better explanation of our findings? In their study of speakers' pronunciation of the preposition *to* and other words in spontaneously produced dative constructions in spoken English, Tily et al. (2009) found a significant

effect of the Bresnan et al. (2007a) corpus model probabilities. They proposed the principle of *uniform information density* (also known as *constant entropy rate*) as a possible theoretical explanation of these effects. In accordance with information theory (Shannon 1948), information is measured in such a way that the more probable an item in a sequence, the less informative it is, and conversely the less probable, the more informative. If the rate at which information is conveyed in the speech stream is roughly constant, then more predictable words, which carry little information, should take less time to pronounce during production than less predictable words. The efficiency of this strategy for communication over a speech channel lies in the fact that it allows utterances to be shorter and easier to produce without reducing the less predictable words that the hearer would have the most difficulty reconstructing (Aylett and Turk 2004). Above the word level, the theory further implies that information density may increase toward the ends of sentences as continuations become more predictable (Genzel and Charniak 2002). Under this theory, generally speaking, less predictable arguments should follow more predictable ones in word order. The harmonic alignment effects could be explained in these terms, provided that animate, definite, pronominal, and relatively less complex arguments are indeed more predictable in information-theoretic terms, as measured in texts.

Although uniform information density applies to production, there is empirical work connecting it to perception and to expectation-based theories of comprehension (Hale 2001, Keller 2004, Levy 2008, Kuperman et al. 2009). On these theories processing difficulties are a function of the probabilities of continuations as measured by informativeness: during incremental processing, resources are allocated to more expected alternatives among the parallel set of possibilities that unfold at each point. Violations of expectation cause difficulties as resources are re-allocated. Recently it has been shown that this theory can explain opposite weight effects found in the visual comprehension of sentences containing verb-particle constructions in Dutch and English, in terms of the very different conditional probabilities of a particle following a verb at various distances in production corpora of the two languages (Kuperman et al. 2009).

Because uniform information density and expectation-based theories of processing are grounded in usage probabilities, they offer a principled explanation at the computational level for the shared harmonic alignment effects found in our corpus, rating, and reading with lexical decision studies. Thus an explanation for our findings of covariation is that the Australians may have had a greater anticipation of prepositional datives than the Americans because of differences in the probabilities of the dative alternation in productions of speakers of the two varieties of English. This hypothesis is consistent with the greater frequency of prepositional datives in an

Australian dative database (Collins 1995) compared to an American dative database (Bresnan et al. 2007a) (but see n. 3) and with our findings in Experiment 3.

If the Australians had a greater expectation of the prepositional dative than the Americans because of greater production frequencies of the prepositional dative in their variety of English, we would predict that in the same contexts, Australians would produce more prepositional datives than Americans. Experiment 3 tested this prediction with a sentence completion task using the materials of Experiment 1, and the prediction was borne out.

7 Limitations of the present study

Why should there be a greater preference for prepositional datives among the Australians than the Americans, and why should it manifest itself in end-weight rather than some other predictor? These causal questions cannot be answered with our present data. But we can observe that elsewhere weight has been found to have a role in syntactic divergence: increasing weight of the possessum has affected the English genitive alternation in the past 150 years by favoring the Saxon (*'s*) over the Norman (*of*) genitive (Szmrecsányi 2009), and in the present day, possessum length is an important predictor of the genitive alternation in American and not British English (Hinrichs and Szmrecsányi 2007). As we have observed at the outset of the present study, historical variationist studies have shown that the dative alternation has been diverging in varieties of English for several centuries (Rohdenburg 2008, Mukherjee and Hoffman 2006, Bresnan and Hay 2008). American English is leading British English in greater use of the double object dative (Grimm and Bresnan 2009). Various causes of these kinds of changes have been proposed, such as increasing drift toward oral language styles or colloquialization (Biber and Finegan 1989, Leech and Smith 1994) and greater pressure toward economy in writing (Biber 2003), but it is unlikely that a single cause accounts for all of the ongoing changes in the dative alternation (cf. Hinrichs and Szmrecsányi 2007, Szmrecsányi and Hinrichs 2008, Szmrecsányi 2009 on the genitive alternation). Social perceptions of formality or ideologies of standard norms (Kroch and Small 1978) might be playing a role. Notably only the recipient argument can be marked by a preposition; the theme remains unmarked in both constructions. If the prepositional construction were perceived as more formal, then in some situations and possibly with less common recipient types, the prepositional construction might be preferred. Finally, prosodic differences might also affect construction choice in different varieties (Anttila 2008, Shih et al. 2009). These speculative remarks are only intended to suggest directions for further investigation.

Another important limitation of this study is that we cannot overgeneralize from small samples of speakers of different varieties, because of many other differences between the groups. Most of the Australian subjects were from a Queensland state university which admits many students of lesser means than the elite and expensive private university of the American subjects located in a wealthy Californian suburb. However, since quantitative trends in the dative alternation are entirely beneath the awareness of speakers, there may be few social influences on the variable. Some might wonder if there were differences in working memory span or reading span between the two groups that could have caused our variety differences (cf. Just and Carpenter 1992, Wells et al. 2009). This seems unlikely. One would expect that if there were a group difference in memory or reading span, it would be one that would favor the US students, from the elite university. However, it was the US participants and not the Australian participants who had the longer reaction times to the word *to* as theme length increased beyond three words in Experiment 2. Further, it was shown that the result could not be due to a difference in speed between the two groups.

Another limitation is the absence of a corpus of spoken Australian usage comparable to the corpus of spoken US English. Nevertheless, no matter how closely matched, corpus data across varieties are full of potential confounds and unmeasured contingent differences. Our use of identical experimental materials drawn from the same corpus and then localized provides one solution to this problem. In future work, we hope to develop and more fully analyze larger-scale contextualized sentence-completion datasets across the varieties.

While our corpus model captures effects of spontaneous spoken production, the behaviors tapped by our three experiments—sentence rating, reading with continuous lexical decision, and sentence completion—are disparate, spanning the range of linguistic judgment, comprehension, and production. This range extends the scope of predictivity beyond a skill for a single task to a more general base of implicit probabilistic knowledge that supports, or perhaps constitutes, the language faculty. But much future research will be required to replicate these findings and deepen our understanding of predictive syntactic behavior across varieties and its relation to production probabilities.

If we can reliably detect the psycholinguistic effects of differing syntactic probabilities between varieties, as the present study suggests, it will open many new questions in the intersection of language variation and cognition, in language development, and in the historical development of language. There are also potential applications in reading, second language education (Frishkoff et al. 2008), and language impairment. We hope the present study stimulates more collaborative research

within the boundaries of our field and across them.

References

Aissen, Judith. 1999. Markedness and subject choice in Optimality Theory. *Natural Language and Linguistic Theory* 17(4):673–711.

Altmann, Gerry T. M., and Yuki Kamide. 1999. Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition* 73(3):247–264.

Anttila, Arto. 2008. Phonological constraints on constituent ordering. In C. B. Chang and H. J. Haynie (Eds.), *Proceedings of the 26th West Coast Conference on Formal Linguistics*, 51–59.

Aoun, Josef, and Y.H Li. 1989. Scope and constituency. *Linguistic Inquiry* 20:141–172.

Arnold, Jennifer, Thomas Wasow, Anthony Losongco, and Ryan Ginstrom. 2000. Heaviness vs. newness: the effects of complexity and information structure on constituent ordering. *Language* 76(1):28–55.

Aylett, Matthew, and Alice Turk. 2004. The smooth signal redundancy hypothesis: A functional explanation for relationships between redundancy, prosodic prominence, and duration in spontaneous speech. *Language and Speech* 47(1):31–56.

Baayen, R. Harald. 2008a. *Analyzing Linguistic Data: A Practical Introduction to Statistics Using R*. Cambridge: Cambridge University Press.

Baayen, R. Harald. 2008b. *languageR: Data sets and functions with "Analyzing Linguistic Data: A practical introduction to statistics"*. R package version 0.953.

Baayen, R. Harald, Douglas J. Davidson, and Douglas M. Bates. 2008. Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language* 59(4):390–412. (Special Issue: Emerging Data Analysis).

Bard, Ellen G., Dan Robertson, and Antonella Sorace. 1996. Magnitude estimation of linguistic acceptability. *Language* 72(1):32–68.

Bates, Douglas, Martin Maechler, and Bin Dai. 2009. *lme4: Linear mixed-effects models using S4 classes*. R package version 0.999375-31.

Belsley, David A., Edwin Kuh, and Roy E. Welsch. 1980. *Regression Diagnostics: Identifying Influential Data and Sources of Collinearity*. New York: Wiley.

Bender, Emily. 2005. On the boundaries of linguistic competence: Matched-guise experiments as evidence of knowledge of grammar. *Lingua* 115:1579–1598.

Biber, Douglas. 2003. Compressed noun-phrase structure in newspaper discourse: The competing demands of popularization vs. economy. In J. Aitchison and D. Lewis (Eds.), *New Media Language*, 169–181. London: Routledge.

Biber, Douglas, and Edward Finegan. 1989. Drift and the evolution of English style: A history of three genres. *Language* 65(3):487–517.

Bock, J. Kathryn, , Helge Loebell, and R. Morey. 1992. From conceptual roles to structural relations: bridging the syntactic cleft. *Psychological Review* 99:150–171.

Bock, J. Kathryn. 1982. Toward a cognitive psychology of syntax: Information processing contributions to sentence formulation. *Psychological Review* 89(1):1–47.

Bock, J. Kathryn. 1986. Syntactic persistence in language production. *Cognitive Psychology* 18(33):355–387.

Bock, J. Kathryn, Anne Cutler, Kathleen M. Eberhard, Sally Butterfield, J. Cooper Cutting, and Karin R. Humphreys. 2006. Number agreement in British and American English: Disagreeing to agree collectively. *Language* 82(1):64–113.

Bock, J. Kathryn, and David E. Irwin. 1980. Syntactic effects of information availability in sentence production. *Journal of Verbal Learning and Verbal Behavior* 19(4):467–484.

Bock, J. Kathryn, and Willem J. Levelt. 1994. Language production: grammatical encoding. In M. A. Gernsbacher (Ed.), *Handbook of Psycholinguistics*, 945–984. San Diego: Academic Press.

Bolinger, Dwight. 1977. *Meaning and Form*. London and New York: Longman.

Branigan, Holly P., Martin J. Pickering, and Mikihiro Tanaka. 2008. Contributions of animacy to grammatical function assignment and word order during production. *Lingua* 118(2):172–189.

Bresnan, Joan. 2002. The lexicon in Optimality Theory. In S. Stevenson and P. Merlo (Eds.), *The Lexical Basis of Syntactic Processing: Formal, Computational, and Experimental Issues*, 39–58. Amsterdam: John Benjamins.

Bresnan, Joan. 2007a. A few lessons from typology. *Linguistic Typology* 11(1):297–306.

Bresnan, Joan. 2007b. Is knowledge of syntax probabilistic? Experiments with the English dative alternation. In S. Featherston and W. Sternefeld (Eds.), *Roots: Linguistics in Search of Its Evidential Base, Series: Studies in Generative Grammar*, 75–96. Berlin and New York: Mouton de Gruyter.

Bresnan, Joan, Anna Cueni, Tatiana Nikitina, and R. Harald Baayen. 2007a. Predicting the dative alternation. In G. Boume, I. Krämer, and J. Zwarts (Eds.), *Cognitive Foundations of Interpretation*, 69–94. Amsterdam: Royal Netherlands Academy of Science.

Bresnan, Joan, Ashwini Deo, and Devyani Sharma. 2007b. Typology in variation: A probabilistic approach to *be* and *-n't* in the Survey of English Dialects. *English Language and Linguistics* 11(2):301–346.

Bresnan, Joan, Shipra Dingare, and Christopher D. Manning. 2001. Soft constraints mirror hard constraints: Voice and person in English and Lummi. In *Proceedings of the LFG '01 Conference*.

Bresnan, Joan, and Jennifer Hay. 2008. Gradient grammar: An effect of animacy on the syntax of *give* in New Zealand and American English. *Lingua* 118(2):245–259. (Special Issue *Animacy, Argument Structure, and Argument Encoding* edited by M. Lamers, S. Lestrade, and P. de Swart).

Bresnan, Joan, and Tatiana Nikitina. 2009. The gradience of the dative alternation. In L. Uyechi and L. H. Wee (Eds.), *Reality Exploration and Discovery. Pattern Interaction in Language and Life*, 161–184. CSLI.

Chang, Franklin, Gary S. Dell, and J. Kathryn Bock. 2006. Becoming syntactic. *Psychological Review* 113(2):234–272.

Chen, Evan, Edward Gibson, and Florian Wolf. 2005. Online syntactic storage costs in sentence comprehension. *Journal of Memory and Language* 52(1):144–169.

Cheshire, Jenny. 1987. Syntactic variation, the linguistic variable and sociolinguistic theory. *Linguistics* 25(2):257–282.

Choi, Hye-Won. 2007. Length and order: A corpus study of korean dative-accusative construction. *Discourse and Congnition* 14(3):207–227.

Clark, Eve. 1987. The principle of contrast: A constraint on language acquisition. In B. MacWhinney (Ed.), *Mechanisms of language acquisition*, 1–33. LEA.

Collins, Peter. 1995. The indirect object construction in English: an informational approach. *Linguistics* 33(1):35–49.

Cueni, Anna. 2004. Coding notes. Unpublished ms., Manuscript, Stanford University.

Davidse, Kristina. 1996. Functional dimensions of the dative in English. In W. V. Belle and W. V. Langendonck (Eds.), *The Dative, Volume 1: Descriptive Studies*, 289–338. Amsterdam: John Benjamins.

DeLong, Katherine A., Thomas P. Urbach, and Marta Kutas. 2005. Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. *Nature Neuroscience* 8(8):1117–1121.

den Dikken, Marcel. 1995. *Particles*. Oxford: Oxford University Press.

Deshmukh, Neeraj, Aravind Ganapathiraju, Andi Gleeson, Jonathan Hamaker, and Joseph Picone. 1998. Resegmentation of Switchboard. In *ICSLP-1998. Proceedings of the* 5^{th} *International Conference on Spoken Language Processing, Sydney, Australia November 30–December 4, 1998*. On-line ICSA archive: http://www.isca-speech.org/archive/icslp_1998/i98_0685.html.

Diessel, Holger. 2007. Frequency effects in language acquisition, language use, and diachronic change. *New Ideas in Psychology* 25(2):108–127.

Erteschik-Shir, Nomi. 1979. Discourse constraints on dative movement. In T. Givón (Ed.), *Discourse and Syntax. Syntax and Semantics*, 441–467. New York: Academic Press.

Fellbaum, Christiane. 2005. Examining the constraints on the benefactive alternation by using the world wide web as a corpus. In M. Reis and S. Kepser (Eds.), *Evidence in Linguistics: Empirical, Theoretical, and Computational Perspectives*, 207–236. Berlin and New York: Mouton de Gruyter.

Ferreira, Victor S. 1996. Is it better to give than to donate? Syntactic flexibility in language production. *Journal of Memory and Language* 35(5):724–755.

Ford, Marilyn. 1983. A method for obtaining measures of local parsing complexity throughout sentences. *Journal of Verbal Learning and Verbal Behavior* 22(2):203–218.

Frishkoff, Gwen, Lori Levin, Phil Pavlik, Kaori Idemaru, and Nel de Jong. 2008. A model-based approach to second-language learning of grammatical constructions. In *Proceedings of the 30th Conference of the Cognitive Science Society*, 916–921, Washington, D.C.

Garretson, Gregory. 2003. *Optimal Typology of Determiner Phrases Coding Manual Excerpt Section 5: "Applying Tags to the Examples"*. Boston University.

Garretson, Gregory, M.C. O'Connor, B. Skarabela, and M. Hogan. 2004. Coding practices used in the project optimal typology of determiner phrases. Unpublished ms., On-line, Boston University: http://npcorpus.bu.edu/documentation/index.html.

Gennari, Silvia P., and Maryellen C. MacDonald. 2008. Semantic indeterminacy in object relative clauses. *Journal of Memory and Language* 58(2):161–187.

Genzel, Dimitry, and Eugene Charniak. 2002. Entropy rate constancy in text. In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL), Philadelphia, July 2002*, 199–206, Philadelphia, PA.

Gibson, Edward. 1998. Linguistic complexity: Locality of syntactic dependencies. *Cognition* 68(1):1–76.

Givón, Talmy. 1984. Direct object and dative shifting: Semantic and pragmatic case. In F. Plank (Ed.), *Objects. Towards a Theory of Grammatical Relations*, 151–182. New York: Academic Press.

Godfrey, J., E. Holliman, and J. McDaniel. 1992. Telephone speech corpus for research and development. In *Proceedings of ICASSP-92*, 517–520.

Goldberg, Adele E. 2002. Surface generalizations: An alternative to alternations. *Cognitive Linguistics* 13–14:327–356.

Green, Georgia. 1971. Some implications of an interaction among constraints. *CLS* 7:85–100.

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

Greenberg, Joseph. 1966. *Language Universals, with Special Reference to Feature Hierarchies*. Vol. 59 of *Janua Linguarum, Series Minor*. The Hague: Mouton de Gruyter.

Gries, Stefan Th. 2005a. *Multifactorial Analysis in Corpus Linguistics*. London: Continuum International Publishing Group.

Gries, Stefan Th. 2005b. Syntactic priming: A corpus-based approach. *Journal of psycholinguistic research* 34:365–399.

Gries, Stephan Th., and A. Stefanowitsch. 2004. Extending collostructional analysis. *International Journal of Corpus Linguistics* 9:97–129.

Grimm, Scott, and Joan Bresnan. 2009. Spatiotemporal variation in the dative alternation: A study of four corpora of British and American English. Paper presented at the Third International Conference *Grammar and Corpora*, Mannheim, 22–24 September 2009.

Grodner, Daniel, and Edward Gibson. 2005. Consequences of the serial nature of linguistic input for sentenial complexity. *Cognitive Science: A Multidisciplinary Journal* 29(2):261–290.

Gropen, J., S. Pinker, M. Hollander, R. Goldberg, and R. Wilson. 1989. The learnability and acquisition of the dative alternation in English. *Language* 65(2):203–257.

Gundel, J. K., N. Hedberg, and R. Zacharsky. 1993. Referring expressions in discourse. *Language* 69:274–307.

Hale, John. 2001. A probabilistic Earley parser as a psycholinguistic model. In *Proceedings of NAACL, vol. 2*, 159–166, Pittsburg, PA.

Halliday, M. A. K. 1970. Language structure and language function. In J. Lyons (Ed.), *New Horizons in Linguistics*, 140–165. Harmondsworth, Middlesex: Penguin.

Harley, Heidi. 2002. Possession and the double object construction. In P. Pica and J. Rooryck (Eds.), *The Yearbook of Linguistic Variation 2*, 29–68. Amsterdam: John Benjamins.

Harrell, Jr., Frank E. 2001. *Regression Modeling Strategies. With Applications to Linear Models, Logistic Regression, and Survival Analysis*. New York: Springer.

Harrell, Jr., Frank E. 2009. *Design: Design Package*. R package version 2.2-0.

Hawkins, John A. 1994. *A Performance Theory of Order and Constituency*. Cambridge: Cambridge University Press.

Hawkins, John A. 2004. *Efficiency and Complexity in Grammars*. Oxford: Oxford University Press.

Hawkins, John A. 2007. Processing typology and why psychologists need to know about it. *New Ideas in Psychology* 25(2):87–107.

Hinrichs, Lars, and Benedikt Szmrecsányi. 2007. Recent changes in the function and frequency of Standard English genitive constructions: A multivariate analysis of tagged corpora. *English Language and Linguistics* 11(3):437–474.

Ito, Takayuki, Mark Tiedea, and David J. Ostry. 2009. Somatosensory function in speech perception. *PNAS* 6(4):1245–1248.

Jacobson, Sven. 1980. Issues in the study of syntactic variation. In S. Jacobson (Ed.), *Papers from the Scandinavian Symposium on Syntactic Variation, Stockholm, May 18-19, 1979*, 23–36. Stockholm: Almvqvist and Wiksell.

Jaeger, T. Florian. 2008. Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language* 59(4):434–446.

Johnson, Keith. 2008. *Quantitative Methods in Linguistics*. Oxford: Blackwell.

Just, M. A., and P. A. Carpenter. 1992. A capacity theory of comprehension: Individual differences in working memory. *Psychological Review* 99:122–149.

Kamide, Yuki, Gerry T. M. Altmann, and Sarah L. Haywood. 2003a. The time-course of prediction in incremental sentence processing: Evidence from anticipatory eye movements. *Journal of Memory and Language* 49(1):133–156.

Kamide, Yuki, Christoph Scheepers, and Gerry T. M. Altmann. 2003b. Integration of syntactic and semantic information in predictive processing: Cross-linguistic evidence from German and English. *Journal of Psycholinguistic Research* 32(1):37–55.

Keller, Frank. 2004. The entropy rate principle as a predictor of processing effort: An evaluation against eye-tracking data. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing, Barcelona*, 317–324, Barcelona.

Krifka, Manfred. 2003. Semantic and pragmatic conditions for the dative alternation. *Korean Journal of English Language and Linguistics* 4:1–32.

Kroch, Anthony. 1989. Reflexes of grammar in patterns of language change. *Language Variation and Change* 1(3):199–244.

Kroch, Anthony. 1994. Morphosyntactic variation. In K. Beals, J. Denton, R. Knippen, L. Melnar, H. Suzuki, and E. Zeinfeld (Eds.), *Papers from the 30th regional meeting of the Chicago Linguistics Society: Volume 2: The Parasession on Variation and Linguistic Theory*, 180–201. Chicago Linguistics Society.

Kroch, Anthony, and Cathy Small. 1978. Grammatical ideology and its effect on speech. In D. Sankoff (Ed.), *Linguistic Variation. Models and Methods*, 44–55. New York: Academic Press.

Kroch, Anthony, and Ann Taylor. 1997. Verb movement in Old and Middle English: Dialect variation and language contact. In A. van Kemenade and N. Vincant (Eds.), *Parameters of Morphosyntactic Change*, 297–325. Cambridge: Cambridge University Press.

Kuno, Susumo, and Ken ichi Takami. 1993. *Grammar and Discourse Principles: Functional Syntax and GB Theory*. Chicago: The University of Chicago Press.

Kuperman, Victor, Vitória Magalhães Piai, Robert Schreuder, and Thomas Wasow. 2009. Processing of discontinuous syntactic dependencies: Verb-particle constructions in dutch and english. Unpublished ms., submitted MS., Stanford University and Radboud University Nijmegen.

Lamers, Monique, Sander Lestrade, and Peter de Swart (Eds.). 2008. *Lingua. Special Issue on Animacy, Argument Structure, and Argument Encoding*. Amsterdam: Elsevier.

Lapata, Maria. 1999. Acquiring lexical generalizations from corpora: a case study for diathesis alternations. In *Proceedings of the* 37^{th} *Meeting of the North American Chapter of the Association for Computational Linguistics*, 397–404, College Park, Maryland.

Larson, Richard K. 1988a. Double objects revisited: Reply to jackendoff. *Linguistic Inquiry* 21:589–632.

Larson, Richard K. 1988b. On the double object construction. *Linguistic Inquiry* 19:335–391.

Lavendera, Beatriz R. 1978. Where does the sociolinguistic variable stop? *Language in Society* 7:171–182.

Leech, Geoffrey, and Nicholas Smith. 1994. Recent grammatical change in written English 1961–1992. some preliminary findings of a comparison of American with British English. In A. Renouf and A. Kehoe (Eds.), *The Changing Face of Corpus Linguistics*, 185–204. Amsterdam: Rodopi.

Levin, Beth. 1993. *English verb classes and alternations: A preliminary investigation*. Chicago and London: University of Chicago Press.

Levy, Roger. 2008. Expectation-based syntactic comprehension. *Cognition* 106(3):1126–1177.

MacDonald, Maryellen C. 1999. Distributional information in language and acquisition: Three puzzles and a moral. In B. MacWhinney (Ed.), *The Emergence of Language*, 177–196. Mahwah, NJ: Erlbaum.

Manning, Christopher. 2003. Probabilistic syntax. In R. Bod, J. Hay, and S. Jannedy (Eds.), *Probabilistic Linguistics*, 289–341. MA: MIT Press.

Marcus, M., B. Santorini, and M. A. Marcinkiewicz. 1993. Building a large annotated corpus of English: The Penn Treebank. *Computational Linguistics* 19:313–330.

McDonald, Janet L., J. Kathryn Bock, and Michael H. Kelly. 1993. Word and world order: Semantic, phonological, and metrical determinants of serial position. *Cognitive Psychology* 25(2):188–230.

Michaelis, Laura A., and S. Francis Hartwell. 2007. Lexical subjects and the conflation strategy. In N. Hedberg and R. Zacharski (Eds.), *Topics in the Grammar-Pragmatics Interface: Papers in Honor of Jeanette K. Gundel*, 19–48. Amsterdam: Benjamins.

Mukherjee, Joybrato, and Sebastian Hoffman. 2006. Describing verb-complementational profiles of New Englishes. *English World-Wide* 27(2):147–173.

Nasir, Sazzad M., and David J. Ostry. 2009. Auditory plasticity and speech motor learning. *PNAS* 6(4):1245–1248.

O'Connor, Mary Catherine, Arto Anttila, Vivienne Fong, and Joan Maling. 2004. Differential possessor expression in English: Re-evaluating animacy and topicality effects. Paper presented at the Annual Meeting of the Linguistics Society of America, January 9–11, 2004, Boston.

Oehrle, Richard T. 1976. *The grammar of the English dative alternation*. PhD thesis, MIT.

Pickering, Martin J., Holly P. Branigan, and Janet F. McLean. 2002. Constituent structure is formulated in one stage. *Journal of Memory and Language* 46(3):586–605.

Pickering, Martin J., and Simon Garrod. 2005. Do people use language production to make predictions during comprehension? *TRENDS in Cognitive Sciences* 11(3):105–110.

Pinheiro, José C., and Douglas M. Bates. 2000. *Mixed-Effects Models in S and S-PLUS*. New York: Springer.

Pinker, Steven. 1989. *Learnability and Cognition: The acquisition of argument structure*. Cambridge, Mass.: MIT Press.

Pintzuk, Susan. 1993. Verb seconding in Old English: Verb movement to Infl. *The Linguistic Review* 10:5–35.

Pintzuk, Susan. 1996. Variation and change in Old English clause structure. *Language Variation and Change* 7(2):29–260.

Pintzuk, Susan, and Anthony Kroch. 1989. The rightward movement of complements and adjuncts in the Old English of Beowulf. *Language Variation and Change* 1:115–143.

Pito, Richard. 1994. Tgrepdoc. Manual page for tgrep, Linguistics Data Consortium, University of Pennsylania.

Polinsky, Maria. 1998. A non-syntactic account of some asymmetries in the double object construction. In J.-P. Koenig (Ed.), *Conceptual Structure and Language: Bridging the Gap*, 403–423. Stanford: CSLI.

Prat Sala, Mercè, and Holly P. Branigan. 2000. Discourse constraints on syntactic processing in language production: a cross-linguistic study in English and Spanish. *Journal of Memory and Language* 42:168–182.

Prince, Alan, and Paul Smolensky. 1993. Optimality theory: Constraint interaction in generative grammar. Technical report, Rutgers University Center for Cognitive Science Technical Report 2.

Prince, Ellen F. 1981. Toward a taxonomy of given-new information. In P. Cole (Ed.), *Radical Pragmatics*, 223–256. New York: Academic Press.

Quené, Hugo, and Huub van den Bergh. 2008. Examples of mixed-effects modeling with crossed random effects and with binomial data. *Journal of Memory and Language* 59(4):413–425. (Special Issue: Emerging Data Analysis).

R Development Core Team. 2009. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.

Ransom, Evelyn N. 1977. Definiteness, animacy, and np ordering. *BLS* 3:418–429.

Ransom, Evelyn N. 1979. Definiteness and animacy constraints on passive and double-object constructions in english. *Glossa* 13:215–240.

Recchia, Gabriel. 2007. STRATA: Search tools for richly annotated and time-aligned linguistic data. Stanford University Symbolic Systems Program Honors Thesis.

Rohdenburg, Günter, and Julia Schlüter (Eds.). 2009. *One Language, Two Grammars? Differences between British and American English*. Cambridge: Cambridge University Press.

Rohdenburg, Günther. 2007. Grammatical divergence between British and American English in the 19^{th} and Early 20^{th} centuries. Paper presented at the Third Late Modern English Conference, the University of Leiden, September 1, 2007.

Romaine, Suzanne. 1984. On the problem of syntactic variation and pragmatic meaning in sociolinguistics theory. *Folia Linguistica* 18(3–4):409–437.

Rosenbach, Anette. 2002. *Genitive Variation in English. Conceptual Factors in Synchronic and Diachronic Studies (Topics in English Linguistics, 42)*. Berlin/New York: Mouton de Gruyter.

Rosenbach, Anette. 2003. Aspects of iconicity and economy in the choice between the *s*-genitive and the *of* -genitive in English. In G. Rohdenburg and B. Mondorf (Eds.), *Determinants of Grammatical Variation in English*, 379–411. Berlin and New York: Mouton de Gruyter.

Rosenbach, Anette. 2005. Animacy versus weight as determinants of grammatical variation in English. *Language* 81(3):613–644.

Rosenbach, Anette. 2008. Animacy and grammatical variation–Findings from English genitive variation. *Lingua* 118(2):151–171.

Sankoff., David. 1982. Sociolinguistic method and linguistic theory. In L. J. Cohen, J. Los, H. Pfeiffer, and K. Peter (Eds.), *Logic, Methodology, Philosophy of Science 6*, Vol. 1, 677–689. Amsterdam and Warsaw: North Holland and Polish Scientific.

Sankoff, David. 1988. Sociolinguistic and syntactic variation. In F. J. Newmeyer (Ed.), *Linguistics: The Cambridge Survey*, 140–161. Cambridge: Cambridge University Press.

Sankoff, David, and Pierrette Thibault. 1981. Weak complementarity: Tense and aspect in Montreal French. In B. B. Johns and D. R. Strong (Eds.), *Syntactic Change*, 205–216. Ann Arbor: University of Michigan Press.

Santorini, Beatrice. 2000. *Part-of-Speech Tagging Guidelines for the Penn Treebank Project (3rd revision, 2nd printing)*. University of Pennsylvania, Pittsburgh.

Schneider, Edgar W. 2007. *Postcolonial English: Varieties Around the World*. Cambridge: Cambridge University Press.

Schneider, Walter, Amy Eschman, and Anthony Zuccolotto. 2002a. *E-Prime Reference Guide*. Pittsburgh: Psychology Software Tools Inc.

Schneider, Walter, Amy Eschman, and Anthony Zuccolotto. 2002b. *E-Prime Users Guide*. Pittsburgh: Psychology Software Tools Inc.

Shannon, Claude E. 1948. The mathematical theory of communication. *Bell System Technical Journal* 27:379–423, 623–656.

Shih, Stephanie, Jason Grafmiller, Richard Futrell, and Joan Bresnan. 2009. Rhythm's role in genitive and dative construction choice in spoken English. Presented at the 31^{st} Annual Meeting of the Linguistics Association of Germany (DGfS), the University of Osnabrück, Germany, March 4, 2009.

Silva-Corvalán, Carmen. 1986. On the problem of meaning in sociolinguistic studies of syntactic variation. In D. Kastovsky and A. Swedek (Eds.), *Linguistics and Historical and Geographical Boundaries: In Honour of Jacek Fisiak on the Occasion of his 50th Birthday, I: Linguistic Theory and Historical Linguistics*, 111–123. Berlin and New York: Mouton de Gruyter.

Silverstein, Michael. 1976. Hierarchy of features and ergativity. In R. M. W. Dixon (Ed.), *Grammatical Categories in Australian Languages*, 112–171. Canberra and New Jersey: Australian Institute of Aboriginal Studies and Humanities Press.

Snider, Neal. 2009. Similarity and structural priming. In N. A. Taatgen and H. van Rijn (Eds.), *Proceedings of the 31th Annual Conference of the Cognitive Science Society*, 816–822, Austin, Texas. Cognitive Science Society.

Snyder, Kieran. 2003. *The relationship between form and function in ditransitive constructions*. PhD thesis, University of Pennsylvania.

Stallings, Lynne M., Maryellen C. MacDonald, and Padraig G. O'Seaghdha. 1998. Phrasal ordering constraints in sentence production: Phrase length and verb disposition in Heavy-NP Shift. *Journal of Memory and Language* 39(3):392–417.

Staum Casasanto, Laura. 2008. Does social information influence sentence processing? In *30^{th} Annual Meeting of the Cognitive Science Society, July 23–26, 2008, Washington D.C.*

Strunk, Jan. 2005. The role of animacy in the nominal possessive constructions of Modern Low Saxon. Paper presented at the Pionier workshop on 'Animacy', Radboud University Nijmegen, May 19–20, 2005.

Szmrecsányi, Benedikt. 2004. On operationalizing syntactic complexity. In G. Purnelle, C. Fairon, and A. Dister (Eds.), *Proceedings of the 7th International Conference on Textual Data Statistical Analysis. Louvain-la-Neuve, March 10–12, 2004*, 1032–1039, Louvain-la-Neuve. Presses universitaires de Louvain.

Szmrecsányi, Benedikt. 2005. Language users as creatures of habit: a corpus-based analysis of persistence in spoken English. *Corpus Linguistics and Linguistics Theory* 1(1):113–149.

Szmrecsányi, Benedikt. 2006. *Morphosyntactic Persistence in Spoken English. A Corpus Study at the Intersection of Variationist Sociolinguistics, Psycholinguistics, and Discourse Analysis*. Berlin and New York: Mouton de Gruyter.

Szmrecsányi, Benedikt. 2009. The great regression: genitive variability in Late Modern English news texts. Unpublished ms., MS., Freiburg Institute for Advanced Studies, Albert-Ludwigs Universität Freiburg.

Szmrecsányi, Benedikt, and Lars Hinrichs. 2008. Probabilistic determinants of genitive variation in spoken and written English: A multivariate comparison across time, space, and genres. In T. Nevalainen, I. Taavitsainen, P. Pahta, and M. Korhonen (Eds.), *The Dynamics of Linguistic Variation: Corpus Evidence on English Past and Present*, 291–309. Amsterdam and Philadelphia: John Benjamins.

Tagliamonte, Sali A. 2006. *Analysing Sociolinguistic Variation*. Cambridge: Cambridge University Press.

Temperley, David. 2007. Minimization of dependency length in written English. *Cognition* 105(2):300–333.

Theijssen, Daphne. 2008. Modelling the English dative alternation in varied written and spoken text. Paper presented at the Third Workshop on Quantitative Investigations in Theoretical Linguistics (QITL-3), June 2–4, 2008, Helsinki, Finland.

Thompson, Sandra A. 1990. Information flow and dative shift in English discourse. In J. A. Edmondson, C. Feagin, and P. Mühlhausler (Eds.), *Development and*

Diversity: Language Variation across Time and Space, 239–253. Dallas: Summer Institute of Linguistics and University of Texas at Arlington.

Thompson, Sandra A. 1995. The iconicity of "dative shift" in English: considerations from information flow in discourse. In M. E. Landsberg (Ed.), *Syntactic Iconicity and Linguistic Freezes*, 155–175. Berlin: Mouton de Gruyter.

Tily, Harry, Susanne Gahl, Inbal Arnon, Neal Snider, Anubha Kothari, and Joan Bresnan. 2009. Syntactic probabilities affect pronunciation variation in spontaneous speech. *Language and Cognition* 1(2):147–165.

Tily, Harry, Barbara Hemforth, Inbal Arnon, Noa Shuval, Neal Snider, and Thomas Wasow. 2008. Eye movements reflect comprehenders' knowledge of syntactic structure probability. Paper presented at the 14^{th} Annual Conference on Architectures and Mechanisms for Language Processing, Cambridge, UK.

Trousdale, Graeme, and Lynn Clark. 2008. Phonological variation in a Scottish community: Method and theory in cognitive sociolinguistics. Paper presented at the Symposium on Approaches to Variation and Change in English, Bamberg, Germany, July 21–23, 2008.

van Bergen, Geertje, and Peter de Swart. 2009. Scrambling in spoken Dutch. Unpublished ms., submitted, Radboud University Nijmegen and University of Groningen.

Van Berkum, Jos J. A., Colin M. Brown, Pienie Zwitserlood, Valesca Kooijman, and Peter Hagoort. 2005. Anticipating upcoming words in discourse: Evidence from ERPs and reading times. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 31(3):443–467.

Vasishth, Shravan, and Richard L. Lewis. 2006. Argument-head distance and processing complexity: explaining both locality and antilocality effects. *Language* 82(4):767–794.

Wasow, Thomas. 2002. *Postverbal Behavior*. Stanford: CSLI.

Weiner, E. Judith, and William Labov. 1983. Constraints on the agentless passive. *Journal of Linguistics* 19:29–58.

Wells, Justine B., Morten H. Christiansen, David S. Race, Daniel J. Acheson, and Maryellen C. MacDonald. 2009. Experience and sentence processing: Statistical learning and relative clause comprehension. *Cognitive Psychology* 58:250–271.

Yamashita, Hiroko. 2002. Scrambled sentences in Japanese: Linguistic properties and motivations for production. *Text—Interdisciplinary Journal for the Study of Discourse* 22(4):597–633.

Yamashita, Hiroko, and Franklin Chang. 2001. "Long before short" preference in the production of a head-final language. *Cognition* 81(2):45–55.

Zaenen, Annie, Jean Carletta, Gregory Garretson, Joan Bresnan, Andrew Koontz-Garboden, Tatiana Nikitina, Catherine O'Connor, and Tom Wasow. 2004. Animacy encoding in English: Why and how. In D. Byron and B. Webber (Eds.), *Proceedings of the 2004 ACL Workshop on Discourse Annotation, Barcelona, July 2004*, 118–125.

Appendix 1

Instructions for the continuous lexical decision task of Experiment 2

Instructions

Welcome. In this experiment you will be reading some paragraphs on the computer screen.

For each item, you will first see the beginning of a conversation, followed by the next word of the conversation and dashes. An example would be:

(21) ***Speaker A:***
I just spoke to Peter on the phone. He didn't sound very well.

Speaker B:
Has he got this cold that is going around?

Speaker A: **No. He**

says _

The dashes are covering the words that continue the conversation.

Once you have read the conversation that is presented, you must read the first string of letters in the continuation (says in this example) and decide whether it is a word or not. If it is a word, press the key marked Y (for Yes) and if it is not, press N (for No). Once you have pressed Y or N, a new string of letters will appear and the last one will become dashes again. There are no tricks. It will be obvious if something is a word or not.

You should try to read the conversations as naturally as possible, making sure that you understand what you read. Please do not rush the task, but be as quick as you can, while still reading naturally.

When you have finished a conversation, you will see a question about what you have just read. To answer the question press the Y (for Yes) or N (for No) key. Sometimes you will be instructed to press the Space Bar one or more times before you get the question.

You should keep your thumbs resting on the Space Bar and your fingers on the keys marked Y and N. Use the fingers next to your thumbs. Use your thumb to press the Space bar and your fingers for the keys marked Y and N.

You can take breaks as you need them, but please try to do so before youve started reading a paragraph.

Thats all there is to it. Just to review:

1. Once you have read the conversation, read the next string of letters and press Y if it is a word and N if it isnt.

2. Once you have pressed Y or N, the next string of letters will appear. Again press Y or N.

3. Read as naturally as possible, comprehending what you read.

4. After each conversation you will see a Yes/No question. Press Y for Yes and N for No.

When the experiment is over, a screen will appear telling you to stop. At that point, you should let the experimenter know that you have finished.