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

Joan Bresnan & Marilyn Ford*

March 18, 2009

ABSTRACT

Probabilistic models of corpus data can be used to predict higher-level grammatical choices and to quantify changes in such choices across different speaker groups in geographic or social space and in historical time. The present study uses probabilistic models in a novel way, to measure and compare the syntactic predictive capacities of speakers of different varieties of the same language. The present study shows that speakers knowledge of probabilistic grammatical choices can vary across different varieties of the same language and can be detected psycholinguistically in the individual. Given evidence of probabilistic changes in the English dative alternation across varieties of English, we examined responses to the verb-argument dependencies in the English dative alternation by six different groups of American and Australian subjects in three parallel psycholinguistic experiments

*We thank Gabriel Recchia, Nick Romero, and Richard Futrell for assistance with data collection. 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. Thanks also to Victor Kuperman and Anette Rosenbach for helpful comments at several stages. 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.

involving sentence ratings, decision latencies during reading, and sentence completion. The experimental items were all sampled from a database of 2349 spoken corpus datives stratified by corpus model probabilities. The findings show that the Australian and the American subjects can make reliable probabilistic predictions of the syntactic choices of others, that in both groups lexical decision latencies during reading vary inversely with syntactic probabilities, and that there is subtle covariation in these psycholinguistic tasks, which can be explained by differences in patterns of usage in language production between the Australian and American subjects.

Probabilistic models of corpus data have been used both to predict higher-level grammatical choices (as discussed, for example, in Bresnan, Cueni, Nikitina, and Baayen 2007) and to quantify changes in such choices across different speaker groups in geographic or social space and in historical time (as variationists have shown in decades of studies). The present study uses probabilistic models in a novel way, to measure and compare the syntactic predictive capacities of speakers of different varieties of the same language in parallel psycholinguistic tasks. We provide a new kind of evidence that speakers of English have detailed probabilistic knowledge of higher-level grammatical structures in their language, which can be tapped in multiple tasks.

The possibility of probabilistic prediction and the existence of probabilistic variation are already well established with some English constructions. For example, the choice between the *'s* and *of* genitive (*the woman's shadow vs. the shadow of the woman*) can be largely predicted by a generalized linear model based on the animacy, phonology, and complexity of the possessor, together with other variables (Leech, Francis, and Xu 1994, Hinrichs and Szmrecsányi 2007, Tagliamonte and Jarasz 2008, Shih, Grafmiller, Futrell, and Bresnan 2009). Historical changes in the English genitive alternation are also well studied and widely known (see Altenberg 1982, Rosenbach 2002, Allen 2008). Probabilistic models of corpus data show that the choice of the *'s* genitive over the *of* genitive has been increasing within a timespan of thirty years in both British and American journalistic texts from the 1960's and 1990's, with Americans leading British writers in this increase (Hinrichs and Szmrecsányi 2007). In spoken Toronto English, animacy is by far the most important predictor of the genitive alternation, but speaker gender and level of education significantly influence construction choice where there is variation (Tagliamonte and Jarasz 2008).

The predictability of the English dative alternation, like the genitive alternation, is also well established. To fix terminology, the two constructions illustrated in (1a,b) are paraphrases describing the same transfer of an entity *that wonderful watch* (the 'theme') to the goal of the transfer *you* (the 'recipient').

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

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

The choice of prepositional dative or double object construction depends on multiple, often conflicting syntactic, informational, and semantic properties. The probability of a construction, all else being equal, is increased when the first phrase following the verb is a pronoun, is definite, refers to a highly accessible referent, refers to a human, or is short (Bock and Irwin 1980, Bock, Loebell, and Morey 1992, Hawkins 1994, Thompson 1995, Collins 1995, Prat-Sala and Branigan 2000, Arnold et al. 2000, Snyder 2003, Wasow 2002, Gries 2003, a.o.). From these and other variables such as the previous occurrence of a parallel structure (Bock 1986; Pickering, Branigan, and McLean 2002; Szmrecsányi 2005) and the lexical bias of the verb (Lapata 1999), it is possible to predict the choice of construction for dative verbs in spoken English by means of a generalized linear mixed model with 94% accuracy (Bresnan, Cueni, Nikitina, and Baayen 2007).

Like the genitive alternation, the English dative alternation also shows historical and inter-variety changes. 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 parallel to the rise in the *'s* genitive, according to a preliminary corpus study which controlled for verb lemma as well as length, pronominality, and text frequency of recipient and theme (Grimm and Bresnan 2007); 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. (2007): 34.5% vs. 25%.[1]

[1]However, the selection criteria of the two datasets differ (for example, Collins included both

Such differing distributions of grammatical constructions as we see with the English genitive and dative alternations may become, at some historical stage in each of the varieties of postcolonial English, a component of group identity, in a process referred to as 'structural nativization' by Schneider (2007: 87ff). Is structural nativization internalized in the cognitive processes of individuals during speaking and reading? Can it be detected and measured psycholinguistically? In the case of the dative alternation in American English, recent studies have found effects of syntactic probabilities on sentence ratings (Bresnan 2007), phonetic production (Tily, Gahl, Arnon, Snider, Kothari, and Bresnan to appear), and effects of verb bias on eye movements (Tily, Hemforth, Arnon, Shuval, Snider, and Wasow 2008), and earlier work has shown that there are important parallels between the comprehension and production of such constructions in the use of distributional information (MacDonald 1999: 189; Stallings, MacDonald, and O'Seaghdha 1998). But it has not yet been shown 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 fact, Grodner and Gibson (2005) have argued against psycholinguistic processing theories based on different distributions in usage or probabilities—"experience-based" theories—in the domain of comprehending syntactically unambiguous sentences. They argue in favor of classical parsing models, according to which difficulty in comprehension is a function of the serial, resource-limited processing of syntactic dependencies (e.g. Hawkins 1994, 2004; Gibson 1998, 2000). The important explanatory hypotheses of these models are that nonlocal syntactic dependencies impose greater memory or integration burdens on "the human parser," and these processing difficulties can influence alternative word orders in construction choice.[2] These theories can also be extended from comprehension to production in various ways (Clark 1994, Wasow 1997, Yamashita 2002, Temperley 2007, Hawkins 2007).

However, the evidence that would favor classical models over experience-based theories of language processing is mixed (see Levy 2008 for a recent review). There are findings supporting a statistical basis for some of the processing complexity in filler-gap syntactic dependencies in English, though not all (Riali and Christiansen

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. 2007).

[2]Hawkins (2007) criticizes a memory-based explanation of syntactic processing difficulty, but Gibson's (2000) processing theory, based on integration cost rather than memory cost, escapes these criticisms. See Temperley (2007) for a review of differences between the (closely related) resource-limited theories of Hawkins and Gibson.

2007; Roland, Dick, and Elman 2007), and there are mixed findings and alternative explanations with verb-argument dependencies and varying word orders in Japanese (Yamashita and Chang 2001), German (Konieczny 2000), Hindi (Vasishth and Lewis 2006), and Russian (Levy, Fedorenko, and Gibson 2007).

While crosslinguistic approaches to studying theories of language processing are important and fruitful (as emphasized by Hawkins 2007), it remains true that varying the dependency length in typologically different languages brings with it many co-varying and interacting linguistic properties such as morphology, agreement, word order, alternative construction types, and information structure, for which model predictions may be unclear or undefined. In contrast, different varieties of the same language—Australian and American English, for example—are usually typologically identical in syntactic structure, while showing the subtle distributional differences demonstrated by variationist research. They are therefore ideal test cases for probabilistic, experience-based theories. Conversely, psycholinguistic methods and accurate probabilistic models provide a magnifying window into syntactic variation at the micro-level, allowing us to probe for processing effects of structural nativization phenomena.

Combining the independent lines of variationist and psycholinguistic research within a probabilistic approach leads us to look for linkages between syntactic variation at very different time scales. That is, subtle variations in the experiences of the English dative alternation in historically and spatially divergent speaker groups could create differences in internalized expectations and preferences in individuals, measurable in predictive psycholinguistic tasks, down to the millisecond level during the rapid time-course of word-recognition latencies in reading.

The present study instantiates this approach by examining responses to the verb-argument dependencies in the English dative alternation (1a,b) by six different groups of American and Australian subjects in three parallel psycholinguistic experiments involving sentence ratings (Bresnan 2007), decision latencies during reading (Ford 1983), and sentence completion. The experimental items, together with their contexts, were all sampled from the database of corpus datives of Bresnan et al. (2007), stratified by corpus model probabilities.

1 The Corpus Model

To measure predictive capacities of both Australian and US subjects, we used the Bresnan et al. corpus model of American dative choices during spontaneous conversations. Bresnan et al. collected a database of 2360 instances of dative constructions from the three-million word Switchboard corpus of telephone conversations in English, manually annotated the data for multiple variables, fit a mixed effect multivariable model to the data and evaluated the model on randomly selected subsets of training and testing data. For the present project we used a corrected version of the database, which has 2349 observations of dative constructions.[3]

The annotated variables of the original dataset are verb lemmas and broad classes of verb senses, concreteness of the theme argument, the presence of structural parallelism in the dialogue, and for both theme and recipient arguments the syntactic complexity (approximated by length in words), the discourse accessibility, pronominality, definiteness, animacy, number, and person. Details about the data sampling and annotation can be found in Bresnan et al. (2007) and Bresnan and Hay (2008). The annotated variables are incorporated by Bresnan et al. as predictors in a series of generalized linear models and generalized linear mixed models of the data.

For the present project we re-fit the Bresnan et al. model to the corrected dataset to re-derive the corpus probabilities of the binary choice of a *to*-dative construction conditioned on all of the model parameters. The final predictors in our final model of the spoken data are *semantic class, givenness of the recipient, givenness of the theme, pronominality of the recipient, pronominality of the theme, definiteness of the recipient, definiteness of the theme, animacy of the recipient, person of the recipient, number of the recipient, number of the theme, concreteness of the theme, presence of parallel dative construction in the dialogue, and the length difference of recipient and theme,* together with verb sense as a random effect.

Tables 1–3 provide simplified illustrations of the kind of data contained in the database, simplified in that they refer only to verb, pronominality, and givenness. For each probability level, both alternative constructions occur naturally; they differ not in grammaticality but in frequency. The italicized expressions illustrated are the ones actually observed. For example, the Table 3 expression *so he gave me a backpack* was observed as a double object construction and sentences of that type (in terms of verb, pronominality, and givenness) were very infrequently found in the *to*-dative construction (which would be *so he gave a backpack to me* in this case). Because the model is estimating the probability of a *to*-dative occuring, this is example had low probability of being a *to*-dative (and in fact was not realized as a *to*-dative). Corpus frequencies illustrated in the tables are used by the model fitting algorithm to weight the predictors so as to maximize the likelihood of the observed data.

[3]The corrected version was 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, Ganapathiraju, Gleeson, Hamaker, and Picone 2998). This was a case study for Recchia (2007).

Table 1: Example of a high-probability *to*-dative

My Dad's given it to me

verb:	*give*	$=$	transfer
theme:	*it*	$=$	pronoun, given
recipient:	*me*	$=$	pronoun, given

frequency of realization: NP NP $= 4$, NP PP $= 58$

Table 2: Example of an even-probability *to*-dative

whenever we give arms to people

verb:	*give*	$=$	transfer
theme:	*arms*	$=$	non-pronoun, non-given
recipient:	*people*	$=$	non-pronoun, non-given

frequency of realization: NP NP $= 16$, NP PP $= 17$

Table 3: Example with low probability of being a *to*-dative

so he gave me a backpack

verb:	*gave*	$=$	transfer
theme:	*a backpack*	$=$	non-pronoun, non-given
recipient:	*me*	$=$	pronoun, given

frequency of realization: NP NP $= 198$, NP PP $= 3$

As already mentioned, the Bresnan et al. model predicts the choice of dative construction (for *give* and thirty-seven other dative verbs in spoken English) with 94% accuracy on (unseen) test data (against a baseline of 79%).

2 Quantative Harmonic Alignment

One of the main findings of Bresnan et al. (2007), building on previous corpus work by Thompson (1995), Collins (1995), and others, is the existence of a statistical pattern in which animate, given, definite, pronominal, and shorter arguments tend to precede inanimate, non-given, indefinite, non-pronominal, and longer arguments in both dative constructions (1a,b), after adjusting for verb sense biases. For example, if the recipient argument is a non-pronoun (that is, a lexical noun phrase), inanimate, not given, indefinite, or longer, it will tend to appear in the prepositional dative construction, which places the recipient in the final position where it follows the theme; see the bolded recipient in (2a,b). Conversely, if the theme argument is a non-pronoun, inanimate, not given, indefinite, or longer, it will tend to appear in the double object construction, which positions it in the final position, following the recipient; see the bolded theme in (3a,b).

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

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

(3) 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, non-given, indefinite, nominal, or longer argument in the final complement position, and conversely to place the animate, given, definite, pronominal, or shorter argument in the position preceding the other complement.

A qualitative view of the quantitative findings of the Bresnan et al. model is given in Table 4. 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 models of Bresnan et al. show that the predictors contribute independently to this effect, so that it cannot be reduced to any one of them, whether it be syntactic complexity (cf. Hawkins 1994, Arnold et al. 2000), givenness (Snyder 2003), or any other single property.[4]

Table 4: Qualitative view of Quantitative Harmonic Alignment

discourse given ≻ **not given**

animate ≻ **inanimate**

definite ≻ **indefinite**

pronoun ≻ **non-pronoun**

less complex ≻ **more complex**

V NP_{rec} $\textbf{NP}_{thm}$

V NP_{thm} $\textbf{PP}_{rec}$

(Adjusted for verb biases)

This statistical pattern is a kind of *harmonic alignment*. The term ‘harmonic alignment’ is used here phenomenologically to refer to the tendency for linguistic elements which are more or less prominent on a scale (such animacy or discourse accessibility) to be disproportionately distributed in respectively more or less prominent syntactic positions.[5] Thus, example (2a) is a harmonically aligned prepositional dative, and (3b) is a harmonically aligned double object dative. The bolded phrases are more harmonic in the final position because they are indefinite, lexical noun phrases, longer than the non-bolded definite pronominal phrases.

Importantly, Australian English datives show a similar pattern of quantitative harmonic alignment, for *givenness, definiteness, pronounhood, end weight*. 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).

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

[5] 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).

One explanation for the observed harmonic alignment phenomena comes from incremental models of syntactic production with variable lexical activation (e.g. Bock 1982; Bock and Levelt 1994; Chang, Dell, and Bock 2006). The more activated units in the abstract cognitive representation of the message being formulated are expressed earlier in the incremental process of linearizing the sentence structure. Activation is increased by lexical frequency, discourse accessibility, animacy, and effects of prior processing (Bock 1982, Bock and Irwin 1980, Prat-Sala and Branigan 2000, Bock, Loebell, and Morey 1992, Bock 1986; Pickering, Branigan, and McLean 2002). Ferreira (1996) provides an implementation of these ideas within a very simple interactive activation model of production of dative constructions, while symbolic computational models of incremental production have been developed that have wide syntactic scope (Kempen and Hoenkamp 1987, De Smedt and Kempen 1991). Nevertheless, the precise mechanisms of harmonic alignment remain to be worked out (cf. McDonald, Bock, and Kelly 1993; Rosenbach 2005, 2008; Branigan, Pickering, and Tanaka 2008). Syntactic complexity or "end weight" effects show typological variation which suggests independence from givenness and animacy effects (Hawkins 1994, 2004, 2007; Gibson 2000; Yamashita and Chang 2001; Yamashita 2002; Rosenbach 2005, 2008; Temperley 2007; Choi 2007).

How can an experience-based model capture both the variability and the generality—perhaps universality—of the harmonic alignment phenomena? The answer is that the range of information sources, such as animacy, prior reference, and rhythmic pattern, to which we are attuned while speaking and understanding may well be universal—but the specific degrees of cognitive/perceptual activation associated with each source may vary subtly as a function of learning from experience, affecting their combination and the resulting outputs. In the present study the aim is not to test a specific model of these hypothesized processing mechanisms, but to examine the basic question of whether speakers' knowledge of probabilistic grammatical choices can vary across typologically identical varieties of the same language, which experience-based models would predict.

3 Experiment 1: Sentence Ratings

As in Bresnan (2007) we formulated the hypothesis that English speakers implicitly know the quantitative usage patterns of harmonic alignment in their own variety and can use them to predict syntactic choices just as the corpus model does. Where the model predicts most decisively, subjects will, too. Where the model is indecisive, subjects will be, too. We conducted an experiment inspired by Rosenbach's (2003) work on the English genitive alternation.

3.1 Method

3.1.1 Participants

The participants were 19 volunteers from the Stanford University community and 20 volunteers from the Griffith University community. 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 U.S. (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 subjects.

A sample item is given in (4).

(4) *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.

The items were randomly sampled from the 2349 observation corrected dataset of Bresnan et al. (2007) (see fn. 3) and checked for obvious ambiguities in either alternative. One continuation was the observed continuation in the corpus and one was the constructed alternative. The items were presented in pseudo-random order, manually adjusted to avoid obvious patterns. Also, the order of the alternative dative constructions was alternated. The items were sampled from throughout the range of corpus model probabilities for the NP PP construction, and were selected primarily from the centers of five probability bins. The probabilities for the prepositional construction ranged from 0.000939 to 0.999004, with a mean of 0.476990. The corpus model probabilities for the prepositional dative construction for the 30 items are shown in Figure 1.

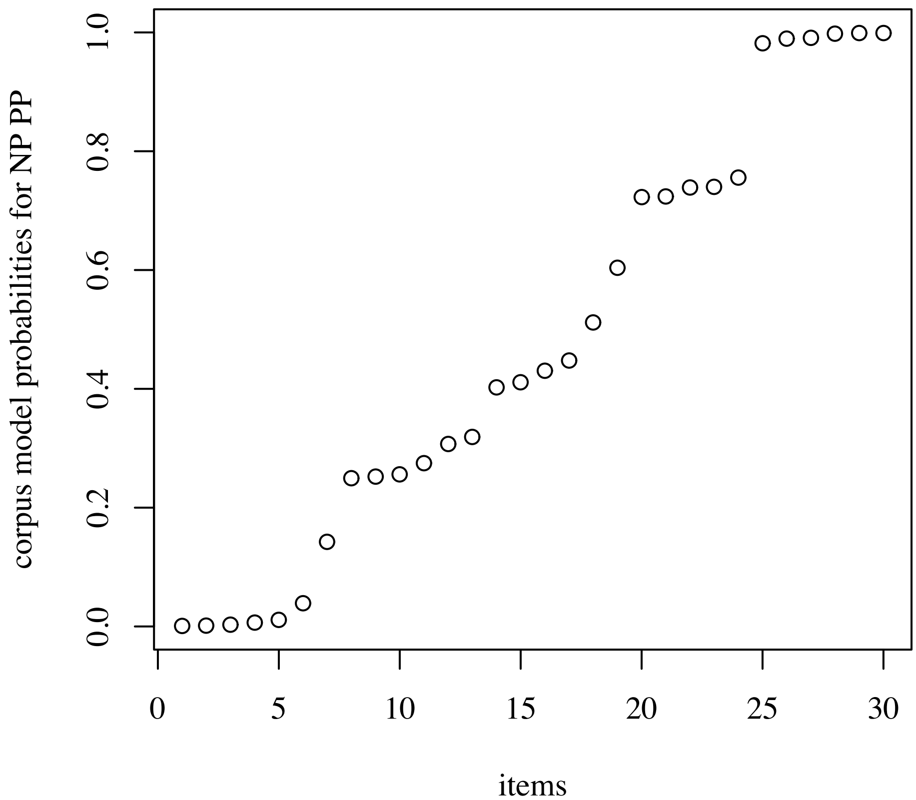

Figure 1: Corpus Model Probabilities of Experiment 1 Items

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 (4), *in college* was changed to *at university*.

3.1.3 Procedure

Each participant was 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.

3.2 Results and discussion

The mean ratings for the US subjects for the NP PP version of each of the 30 items are plotted against the corpus model probabilities for NP PP in Figure 2.[6] The line is a nonparametric smoother indicating the trend of the data obtained by averaging local values. There is a roughly linear correspondence between the mean ratings and the corpus model probabilities. It can be seen that for items at the extreme range of probabilities according to the corpus model, subjects are, in general, giving ratings that are correspondingly low or high. For the middle ranges, where the model is giving less decisive probabilities for NP PP, subjects are also giving ratings showing less certainty about the probability of the NP PP version.

To gain a clearer picture of the relationship between mean ratings, corpus probabilities, and individual subject performance, items were classified into five bins of six items each, classified as very low, low, medium, high, or very high probability for NP PP according to the corpus model. Figure 3 shows the mean rating of each bin for each of the 19 US subjects. All subjects had a lower mean rating for the lowest probability bin than for the highest probability bin. There is more variability in middle bins, as would be expected by the fact that the model gives less certain probabilities for these items.

From Figure 3 we can see that subjects varied in how much of the rating scale they used. For example, Subject S4's average ratings per probability bin cluster closely around the middle band of the ratings scale from 40 to 60, while the adjacent

[6]It is customary to standardize individual subject ratings in order to reduce subject variability as much as possible (e.g. Bard, Robertson, and Sorace 1996). The models we fit to the ratings data automatically adjusted for individual variation in both the baseline and the range of the ratings scale, in a way explained below with respect to the plots of the raw ratings data.

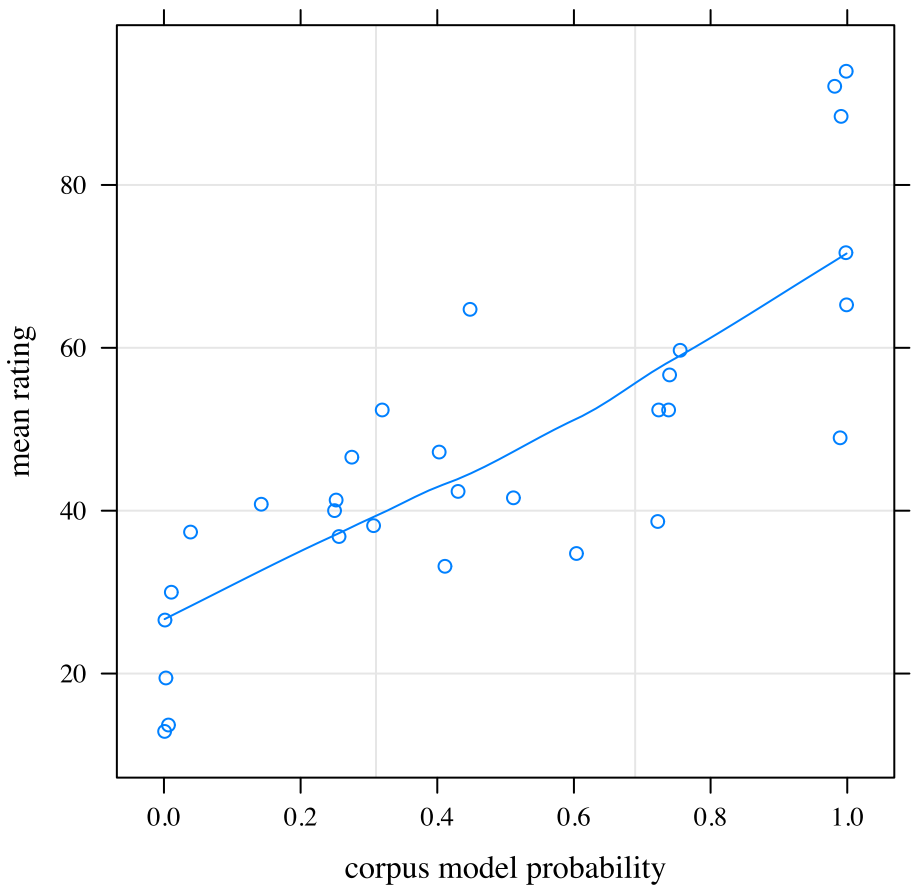

Figure 2: Mean ratings by probability for US subjects

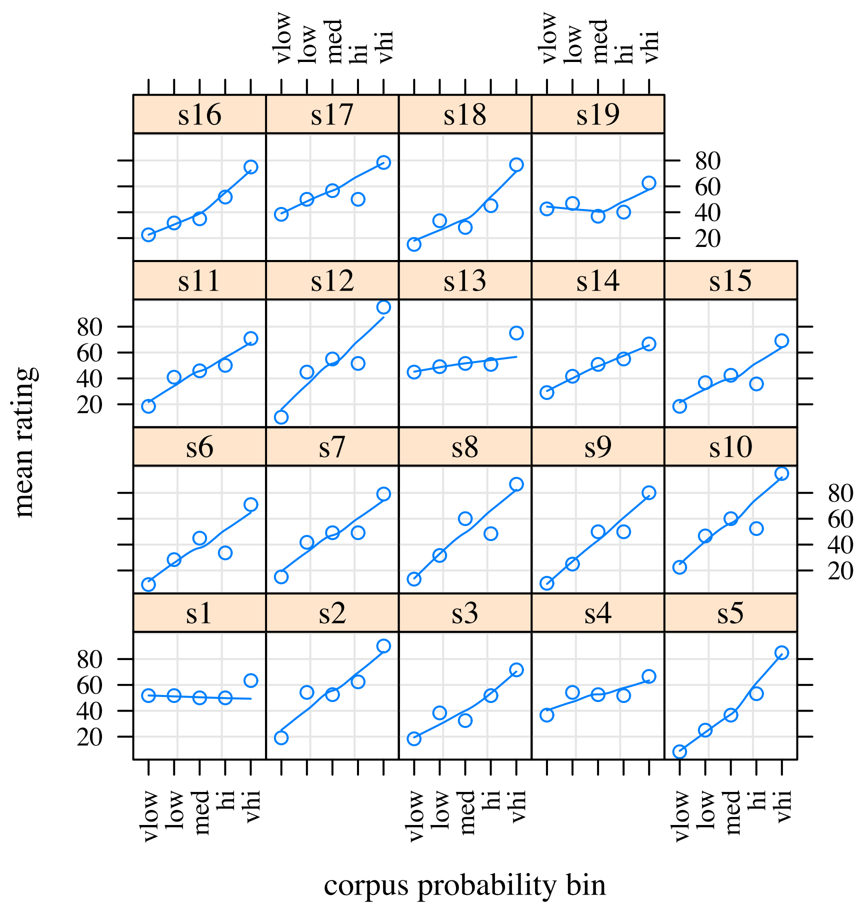

Figure 3: Mean ratings for each probability bin for each US subject

subject S5's ratings extend from far below 20 to 100. This difference in rating range or amplitude is modeled by the *slopes* of the lines in each plot: a steeper slope corresponds to a wider range of ratings given. Subjects also varied somewhat in the baseline they appeared to be using. For example, subjects S6 and S10 have approximately similar slopes and ranges, but subject S10's average ratings in each bin are higher, suggesting a higher baseline. This difference can be modeled by the rating *means* over the entire probability spectrum of items. The structure of a mixed effect model of the ratings data allows direct modeling of inter-subject variation in both means and slopes, in the random effects. The fixed effects, including any interactions, are conditioned on these random effects.

To determine the significance of the corpus probabilities in subjects' ratings, the data were analysed using a linear mixed effects regression model (Pinheiro and Bates 2000; Bates, Maechler, and Dai 2008; Baayen, Davidson, and Bates 2008). The model used corpus probability as a fixed effect and verb, subject, and an interaction between subject and corpus probability as random effects. The random effect of subject modeled inter-subject variation in the mean, or baseline, rating. The interaction between subject and corpus probability modeled inter-subject variation in the slope, or range, of ratings. The random effect of verb modeled verb bias toward the *to*-dative. Thus after controlling for the random effects of the nineteen individual subjects, their varying interactions with the corpus model probabilities, and the nine verbs used, the model shows that corpus probability was a highly significant main effect, with $p = 0.0001$.

The Australian subjects also gave ratings that were in line with the corpus model probabilities. Figure 4 gives the mean ratings for the Australian subjects for the NP PP versions of each item plotted against the corpus model probabilities. Figure 5 gives the mean rating of each of the five probability bins for each of the 20 Australian subjects.

As with the US subjects, there is a roughly linear correspondence between the mean ratings and the corpus model probabilities. Also, all Australian subjects had a lower mean rating for the lowest probability bin than for the highest probability bin, with more variability in middle bins. The linear mixed effects regression model, controlling again for verb bias and variation in subject means and slopes, showed that corpus probability was a highly significant effect, with $p = 0.0001$.

The pattern of responding for both groups of subjects suggests that people have a knowledge, at some level, of the quantitative patterns of usage found in spontaneous production; all sentences are grammatical, but people's ratings of naturalness are aligned to the corpus model probabilities. If people do have knowledge of the patterns of usage found in spontaneous speech, we might expect subtle differences

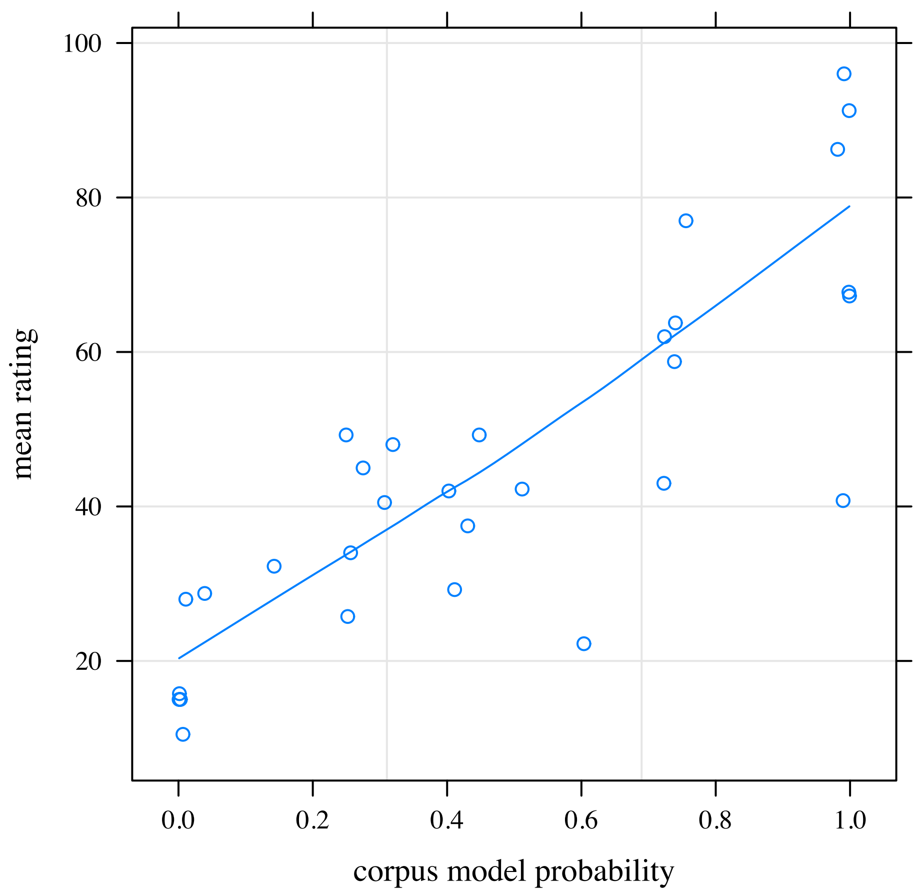

Figure 4: Mean ratings by probability for Australian subjects

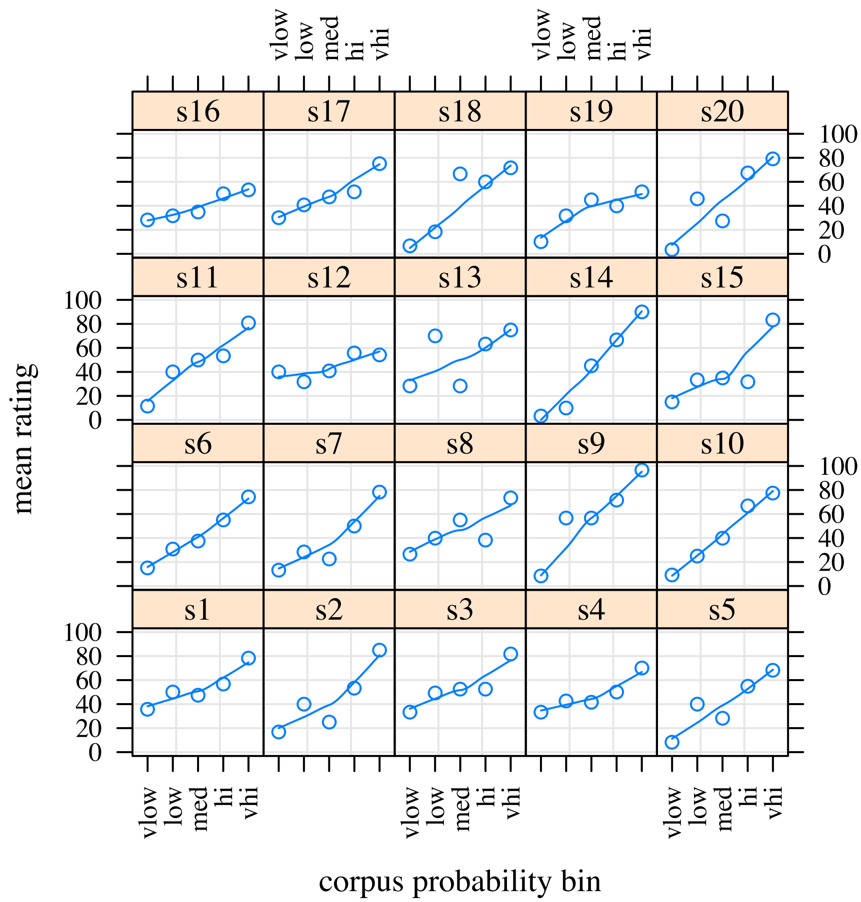

Figure 5: Mean ratings for each probability bin for each Australian subject

even across varieties of the same language. To determine whether there are such differences between the US and Australian subjects, the data were analysed using a linear mixed effects regression model that incorporated variety of English interacting with the linguistic predictors of the corpus model probabilities.

Given that there were only 30 items, the regression model for the experiment could not include all of the original corpus model predictors described in Section 1. The fixed effects included in the initial experiment model were these: variety of English, givenness of the theme, givenness of the recipient, pronominality of the recipient, pronominality of the theme, definiteness of the theme, length of the recipient (logged and centered), and length of the theme (logged and centered). The numerical covariates length of recipient and length of theme were first logged to reduce any effect of extreme values and then centered so that 0 would represent the mean values. The parallelism variable in the original corpus study indicated the occurrence of a parallel construction in the entire dialogue. This variable was replaced for modeling the experiment data by manually re-annotating to indicate the presence or absence of a *to*-dative construction in the short context passage of each item. Definiteness of the recipient was not included because only 2 out of 30 recipients were indefinite. The random effects of verb, subject, and subject interacting with corpus probability were included. In the initial model, variety was given as possibly interacting with all other fixed effects. For all interactions with variety, except for length of recipient, pronominality of the recipient, and givenness of the theme, the estimated coefficient was less than the standard error. These interactions were thus eliminated. In the next regression, the estimated coefficient for givenness of the theme interacting with variety was also found to be less than the standard error and thus was also eliminated.

The fixed-effect coefficients for the final resulting model are shown in Table 5.[7] The model shows that (after adjusting for the random effects of subject, verb, and corpus probability interacting with subject) there were significant main effects of length of theme, givenness of the theme, definiteness of the theme, and pronominality of the theme. There was also a tendency for the occurrence of a parallel *to*-dative in the dialogue to influence ratings; the p-value is $< .05$, but the upper and lower confidence intervals cross zero, perhaps because there were too few observations, with there being only 5 items with a prior *to*-dative in the short context passage. *All of the main effects with* $p < .05$ *in Table 5, except for pronominality of the theme,*

[7]The p-values in Table 5 are derived from the t-values using the pvals.fnc function in the languageR package for linear mixed effects regression model (Baayen, Davidson, and Bates 2008). These are normally appropriate if there are hundreds of observations over items and subjects. The upper and lower 95% confidence limits, derived by computational simulations from the posterior distributions, are also given. The upper and lower limits should not cross zero if the result is significant.

were in the direction consistent with harmonic alignment (Table 4).[8]

On further investigation it was realised that pronominality of theme was interacting with definiteness of the theme, with the indefinite pronominal themes favoring NP NP, while the definite pronominal themes favored NP PP. Because there were only 7 examples of pronominal themes in the items, it was not feasible to add this interaction to the model. Length of the recipient interacted significantly with variety; as the recipient increases in length, the Australians favor the NP PP construction while the US subjects do not seem to have a length of recipient effect. The The estimates of the model for the intercept (favoring NP PP), length of recipient, and variety differences were used to plot the model interaction. This interaction is shown in Figure 6.

Table 5: Model coefficients for the linguistic predictors in Experiment 1

Fixed Effects	Estimate	95% Confidence Limits		p-values
		lower	upper	
(Intercept)	50.1930	53.158	88.6025	0.0000
variety = Aus	−5.2814	−11.291	2.8382	0.2683
recipient length	−0.0787	−1.236	11.5337	0.9805
theme length	−20.5661	−35.051	−21.7377	0.0000
recipient = non-given	−6.0691	−15.965	0.8844	0.1331
theme = non-given	−6.5040	−12.412	−4.4502	0.0008
theme = indefinite	−15.2051	−24.117	−14.6491	0.0000
recipient = pronoun	−2.1236	−19.426	−0.1559	0.6629
theme = pronoun	−9.6107	−15.691	−6.5004	0.0000
parallel *to*-dative = yes	7.3419	−2.195	11.7947	0.0260
recipient length : variety = Aus	8.6696	1.245	16.6656	0.0247
recip = pronoun : variety = Aus	5.4426	−4.059	13.9874	0.2352

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

Inspection of the residuals and the density plots of the posterior distributions of the estimates showed that the model assumptions were reasonably satisfied (Baayen, Davidson, and Bates 2008; Baayen 2008).[9]

[8] Negative estimates of the fixed effects favor the double object construction; positive estimates favor the prepositional dative. See Bresnan et al. (2007).

[9] The model including fixed effects accounts for 53.27% of the variance in the data, compared to 47.09% accounted for by a baseline model consisting of an intercept and the random effects only.

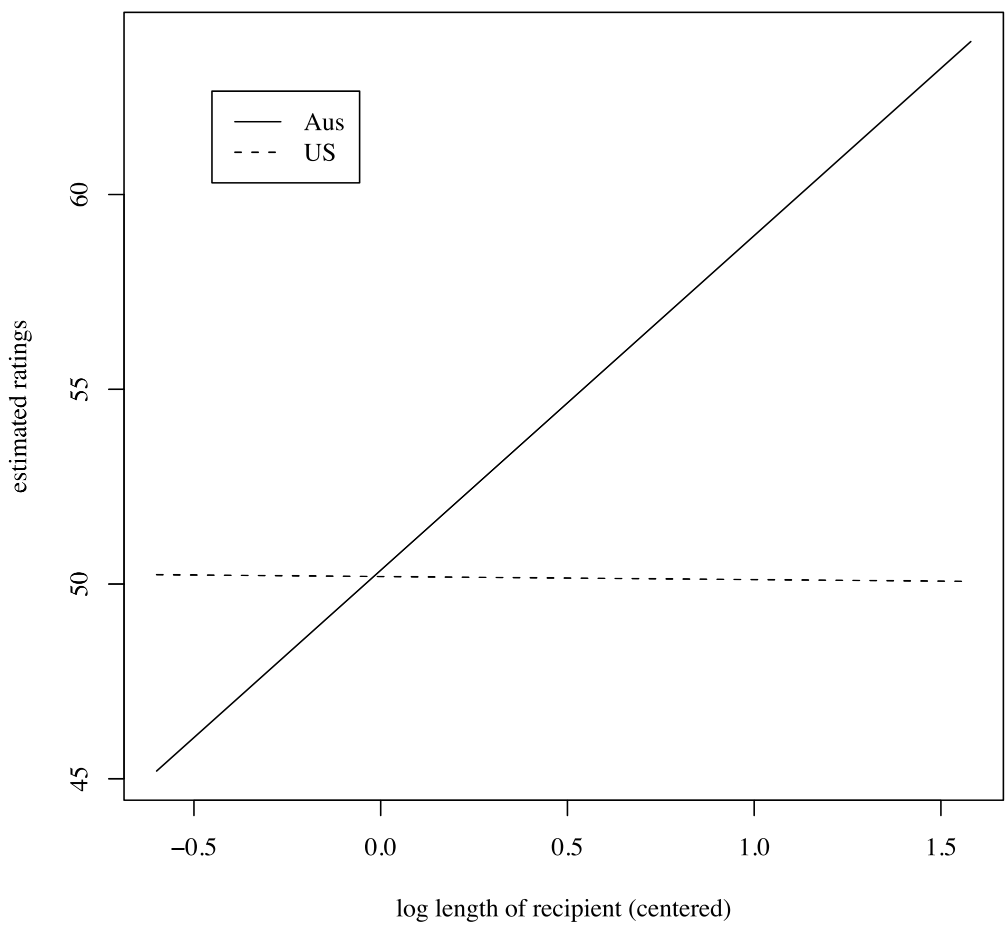

Figure 6: Estimated ratings showing the length of recipient : variety interaction

Experiment 1 shows clearly that both populations of subjects are sensitive to the corpus probabilities. However, there is evidence that the Australians show a greater end-weight effect of the recipient than the American subjects; as the recipient argument of a dative gets longer, the Australians have a greater liking of the V NP PP dative. Equivalently, given the binary choice of construction type, it could be said that the Australians have less tolerance for V NP(LongRecipient) NP than the American subjects.

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. Experiment 2 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 self-paced 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 (see Ford 1983: 204). 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 the present study, 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. We call these new probabilities "partial-construction probabilities." The design of this experiment was inspired in part by Tily et al.'s (to appear) production study, which showed that durations in the pronunciation of *to* varied as a function of corpus model probabilities for the prepositional dative construction.

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 U.S. (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 bin of corpus model probabilities for the prepositional dative construction. 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 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 (5).

(5) *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 (6).

(6) *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 (5), the question in (7) appeared on a new screen.

(7) 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 was from 0.006317 to 0.87506, with a mean of 0.355492. The partial-construction corpus probabilities for the prepositional dative construction for the 24 experimental items are shown in Figure 7.

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 the example in (8).

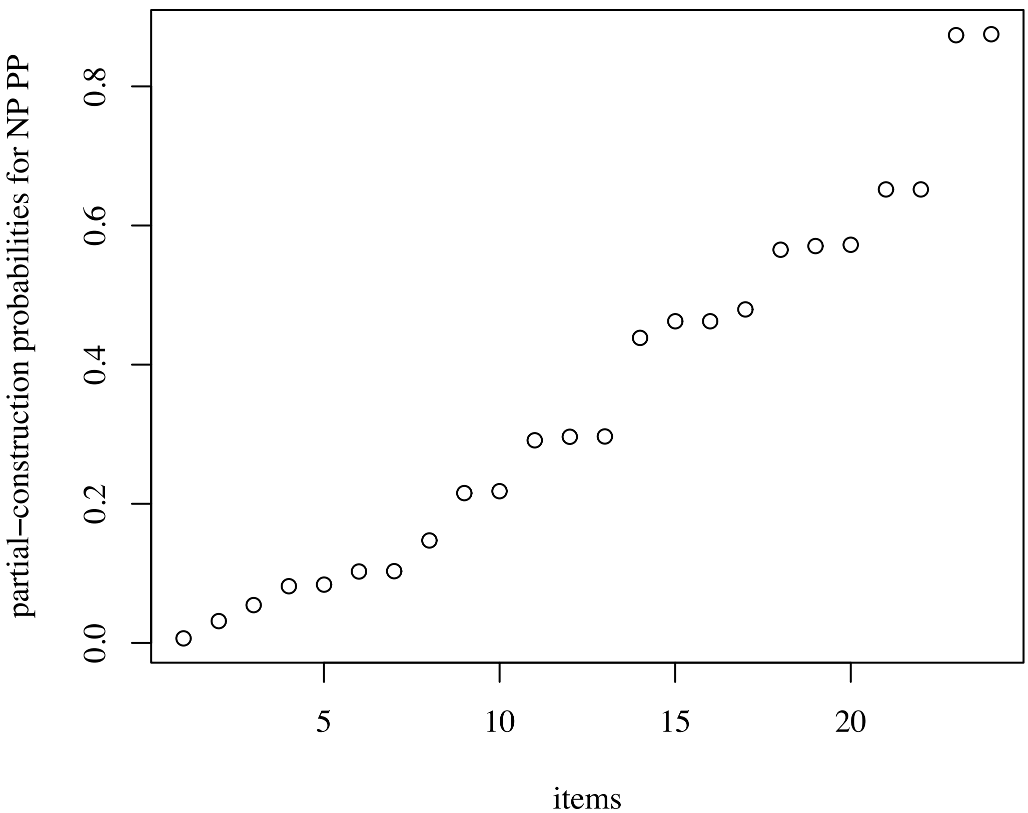

Figure 7: Partial-construction corpus probabilities of Experiment 2 items

(8) *Speaker A:*
I just spoke to Peter on the phone. He didnt 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 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 (Schneider, Eschman, and Zuccolotto 2002a,b) was used to run the Continuous Lexical Decision Task.

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. We also transformed the predictor partial-construction probabilities into log odds to obtain a better fitting relation to the response. By logging both the dependent and predictor variables, the model 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. Figure 8 gives the mean log 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 the nonparametric smoother for both varieties.

As would be predicted, there is a general trend for reaction times to decrease as the corpus log odds increase. To gain a picture of the performance of individual participants, items were classified into four bins of six items each according to the corpus model partial-construction log odds, that is, "very low", "low", "high", and "very high". Figures 9 and 10 show the mean log reaction times for each bin for each of the US subjects and the Australian subjects, respectively. As these figures show, the trend for most subjects is downwards; that is, the bin with very low log odds tends to have the highest mean log reaction times, while the bin with the very high log odds tends to have the lowest mean log reaction times.

To determine the significance of the corpus log odds in determining subjects' reaction times to the word *to*, the data were analysed using a linear mixed effects regression model. The regression model used partial-construction log odds and variety, together with their interaction, as fixed effects and subject and verb as random effects. The corpus log odds were highly significant, $p = 0.0000$. Variety was also significant, $p = 0.0077$, with the Americans having faster reaction times than the Australians. There was no significant interaction.

Given that other variables apart from partial-construction log odds could influence reaction times, in the second regression analysis adjustments were made to control for several other variables. Specifically, controls were added for length of the theme, and any interaction of theme length with variety, the word preceding *to*, the reaction time to that preceding word, item order, and any interaction between item order and subject. It was found that models with and without the random effect of word preceding *to* did not differ significantly and so that control was eliminated. The partial-construction log odds were still significant, with $p < 0.0022$.

To answer the question of whether the two groups varied in the importance of the linguistic predictors that are components of the corpus model probabilities, fixed effects for each component were added to the experiment model in place of the partial-construction log odds. After eliminating predictors where the estimated coefficient

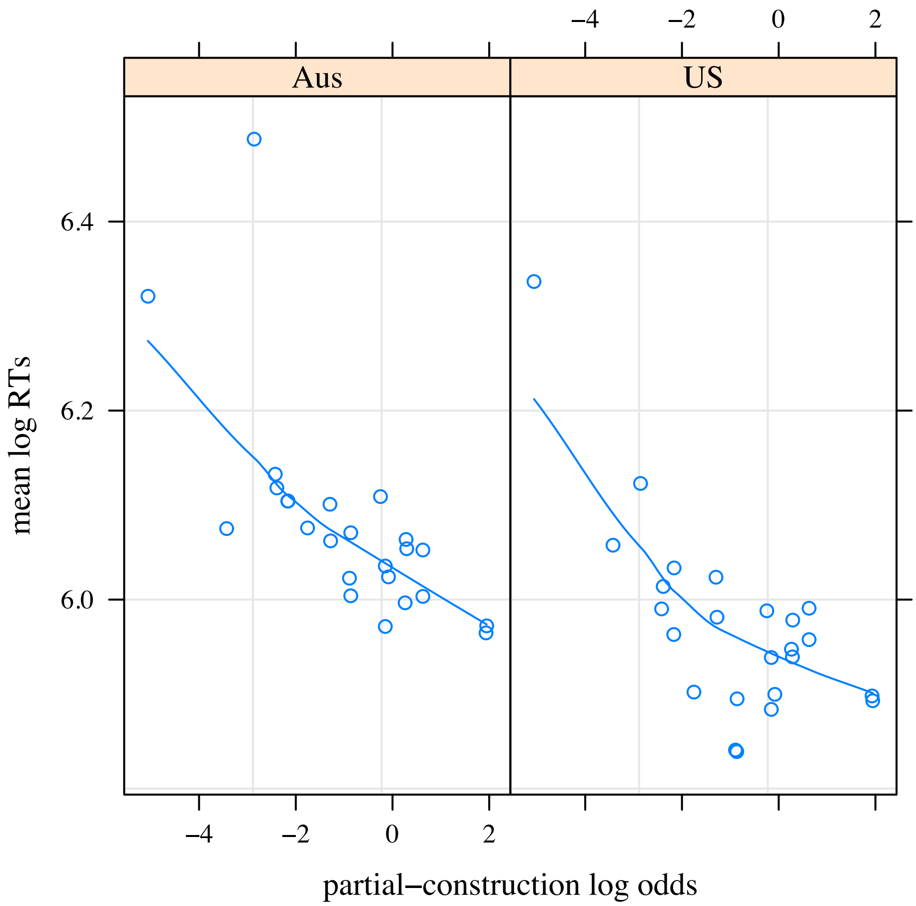

Figure 8: Mean log reaction times by partial-construction log odds for both varieties

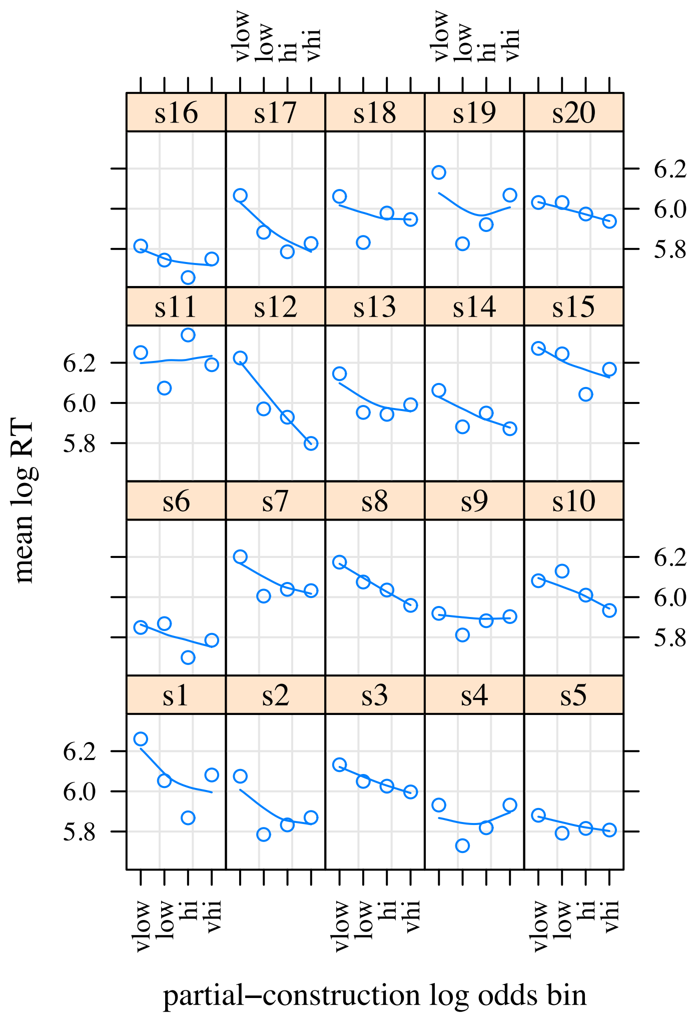

Figure 9: Mean log reaction times for each corpus partial-construction log odds bin for each US subject

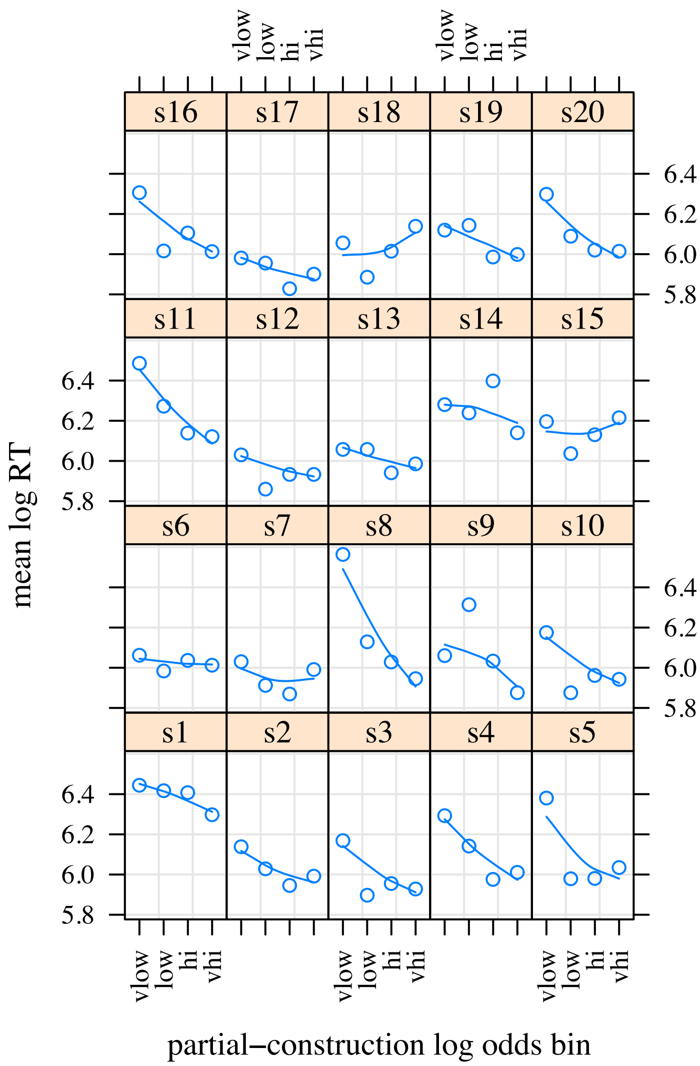

Figure 10: Mean log reaction times for each corpus partial-construction log odds bin for each Australian subject

was less than the standard error, two linguistic predictors remained: pronominality of the theme and length of the theme. Results showed that there was a significant main effect of length of the theme ($p = 0.0007$) and that variety significantly interacted with pronominality of theme ($p = 0.0347$) and length of theme ($p = 0.0000$). The main effect of variety was also significant ($p = 0.0027$), with Australian subjects responding more slowly than the Americans. Reaction time to the word preceding *to* was significant ($p = 0.0000$). Item order was also significant ($p = 0.0079$).

Given that the two groups differ in speed, it is important to see whether interactions with variety hold when speed is controlled for. Subjects were thus classified as "fast" or "slow", depending on whether their mean reaction time to *to* was above or below the mean for all subjects. Speed was then added as a control in the regression analysis. Results showed that the effects were robust. The model coefficients for the regression are given in Table 6.

Table 6: Model coefficients for the linguistic predictors of reaction times in Experiment 2

Fixed Effects	Estimate	95% Confidence Limits		p-values
		lower	upper	
(Intercept)	5.9648	5.8607	6.0554	0.0000
variety = Aus	0.0805	0.0540	0.1074	0.0030
theme length	0.0749	0.0272	0.1229	0.0007
theme = pronoun	−0.0193	−0.0720	0.0354	0.4376
log RT to preceding word	0.4873	0.4106	0.5375	0.0000
item order	−0.0017	−0.0031	−0.0004	0.0076
speed of subject	0.1792	0.1565	0.2027	0.0000
theme = pronoun : variety = Aus	−0.0670	−0.1374	−0.0024	0.0346
theme length : variety = Aus	−0.0984	−0.1471	−0.0503	0.0000

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

For all p-values $< .05$ in Table 6, the upper and lower limits did not cross zero. However, the upper limit for the interaction of pronominality of theme and variety is very close to zero and a p-value based on these limits hovers around $.05$. Inspection of the residuals and the density plots of the posterior distributions of the estimates showed that the model assumptions were reasonably satisfied (Baayen, Davidson, and Bates 2008; Baayen 2008).[10]

[10]The model including fixed effects accounts for 60.42% of the variance in the data, compared

Interestingly, the direction of the main effect of length of theme is consistent with the harmonic alignment pattern of Table 4. More complex themes favor the double object construction over the prepositional dative and thus reaction times to *to* increase with length of theme. Yet the interaction with variety indicates that it is only the Americans who show this effect, as seen in Figure 11.

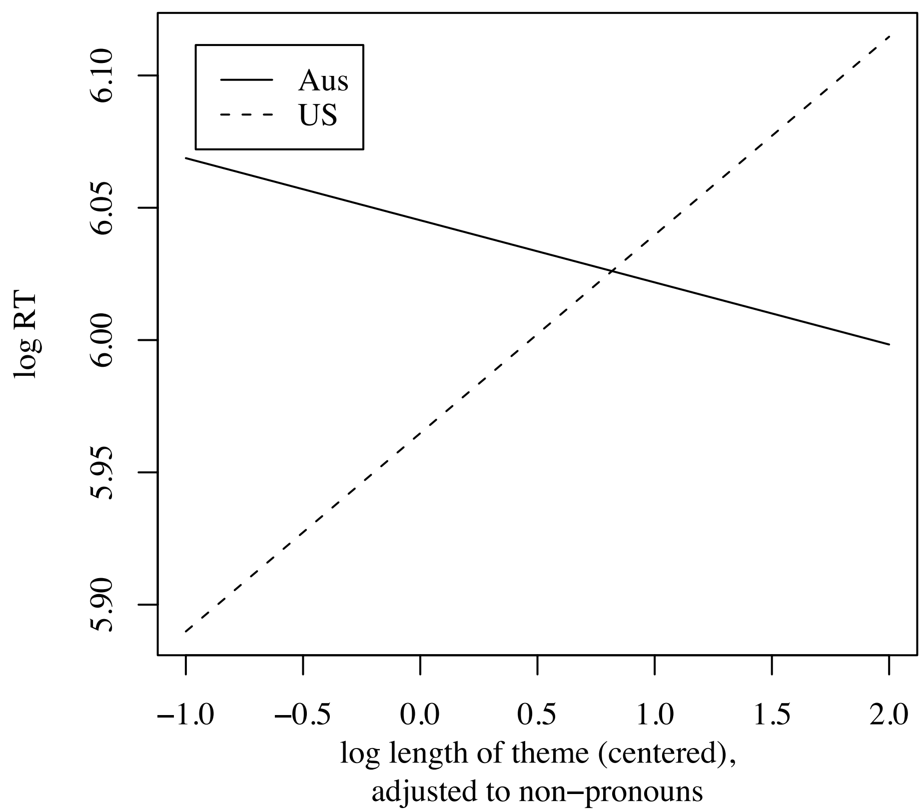

Figure 11: Predicted log RTs showing the length of theme : variety interaction

to 47.46% by a baseline model consisting of an intercept with random effects only. Given the low signal-to-noise ratio typical of reaction time experiments, this value is quite good.

Analyses showed that this interaction between variety and length of theme is very robust. It cannot be attributed to differences in speed—such as a ceiling effect in slower subjects' decision latencies—as the model included speed as a control. Moreover, a second regression analysis where speed was substituted for variety showed that no potential interaction with speed approached significance. It might be thought that the Australians could, in fact, show an 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. Results showed that there was no interaction between variety and either length of theme or pronominality 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 longer themes.

Regarding the main effect of pronominality, it was consistent with harmonic alignment (Table 4), with reaction times at the word *to* decreasing after a pronoun where the probability of a *to*-dative increases. Both varieties show the effect, though the Australians show a greater effect than the Americans.

5 Experiment 3: Sentence Completion

Experiments 1 and 2 generated results where linguistic predictors showed harmonic alignment. However, 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 a greater end-weight effect of the recipient than the Americans, while in Experiment 2 the Americans showed a very strong end-weight effect of the theme and the Australians showed no such effect. 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 an NP PP, then a consistent pattern emerges.

Consider Table 7, which summarises how variety interacts with certain linguistic predictors favoring or disfavoring NP PP.

Compared to the Americans, the Australians show more effect of properties that favor 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 NP PP, but, unlike the Americans, the Australians do not have increased reaction times at the word *to* as theme

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

Decision latency experiment:

property	expectation	RT on *to*
theme length grows	disfavors NP PP	only US increases
theme is pronoun	favors NP PP	Aus decreases more

Rating experiment:

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

length increases, as though they are more tolerant of V NP(LongTheme) PP than the Americans. 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 less effect, as though they are more tolerant of V NP(LongRecipient) NP than the Australians. Pronominality of the theme favors NP PP and it is the Australians who show decreased reaction times to *to* after a pronominal theme. The Americans have decreased reaction times to *to* after a pronominal theme, but the effect is less, seeming more tolerant of V NP(NonPronoun) NP than the Australians.

Reflecting on the results in terms of whether the linguistic predictors favor or disfavor an 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 NP PP than the US group. If so, it might be expected that they would produce more prepositional datives than the Americans do, all else being equal, as when the preceding discourse contexts are identical. 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 U.S. (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 their 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.

5.1.4 Results

The transcripts of each subject were checked separately by each author for NP NP and NP PP *to*-dative completions. 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 analysed using a generalized linear model, controlling for gender. The greater preference for NP PP by the Australians was significant ($p < 0.05$).

6 Concluding Discussion

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 length of theme, givenness of the theme, and definiteness of the theme going in the direction predicted by harmonic alignment. In Experiment 2 the partial-construction properties of length and pronominality of theme argument were among the main effect predictors of reaction time, in the directions expected from the harmonic alignment pattern shown in Table 4: a pronoun 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*. Surprisingly, though, in the ratings experiment, the US subjects, unlike the Australian subjects, did not show a greater preference for NP PP as length of recipient increased (Table 6 and Figure11). And, in contrast, in the Continuous Lexical Decision Task, the Australian subjects did not show increased processing time as a function of increasing the theme length–neither at the word *to* nor as a lagging effect on the following word (Table 6 and Figure 11).

Previous work has argued that the difficulty of integrating a second argument with a ditransitive verb increases with the length of the intervening first object (Chen, Gibson, and Wolf 2005: 284), as shown in the Dependency Length Theory analysis in Figure 12. 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.[11] The difference in length is illustrated for the head-argument dependencies between the verb *brought* and the preposition *to* in the Figure: there are zero lexical words spanned by the dependency arrow in the top example and there are two lexical words (*pony, van*) spanned by the arrow in the bottom example. Such differences in dependency length are predicted to yield inverse effects on reaction times by several of the 'classical' parsing theories discussed at the outset.

How can the differences between the Australian and American subjects in Experiment 2 be explained? The Australian subjects had slower decision latencies on

[11]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. (2007), the two measures have a Spearman's $\rho > 0.91, p < 2.2 \times 10^{-16}$. See also Temperley (2007).

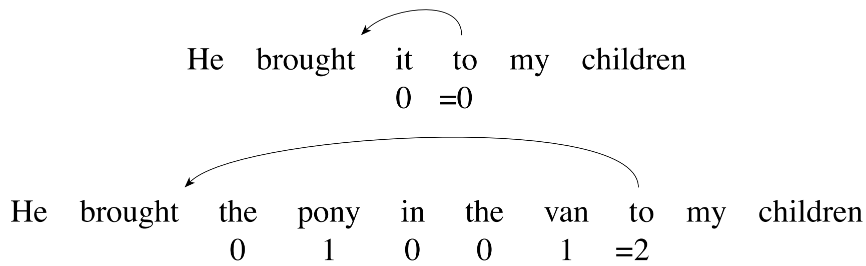

Figure 12: Dependency Length Theory

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 in the task. The predicted increased processing effects of increasing theme length might have shown up as a lagging effect on the next word, but an analysis of reaction times on the word following *to* eliminated this possibility. A third hypothesis is that the Australians may have had a greater anticipation of prepositional datives in the longer-theme contexts than the Americans because of differences in the usage distribution of the dative alternation in 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. 2007) (though not much weight can be placed on comparisons of summary statistics of different corpora). It is also consistent with the Australians' increased end-weight effect of the recipient in Experiment 1, which reveals their stronger bias toward prepositional datives, at least with longer recipients.

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.

An 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 students of lower socioeconomic status than the elite and extremely expensive private university of the American subjects located in a wealthy Californian suburb. But any difference between the groups lends support to our hypothesis of an important effect of differences in language experience on language processing, which cannot be explained in terms of universal parsing architecture alone.

Our general conclusions are thus that language users can make reliable probabilistic predictions of the syntactic choices of others, that lexical decision latencies during reading vary inversely with syntactic probabilities, and that Australian and American subjects showed subtle covariation in these psycholinguistic tasks, which can be explained by different patterns of usage in language production.

The present study also provides several interesting methodological conclusions. First, accurate corpus models can be used to measure language-users' predictive capacities, even across different varieties of English. Secondly, simple psycholinguistic tasks such as sentence rating and sentence completion with natural linguistic materials can be used to confirm and supplement sparse or unavailable corpus data. And thirdly, combining methods from different disciplines can shed light on the dynamics of probabilistic grammar over different timescales.

References

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

Allen, Cynthia L. 2008. *Genitives in Early English Typology and Evidence*. Oxford: Oxford University Press.

Altenberg, Bengt. 1982. *The Genitive v. the Of-Construction. A Study of Syntactic Variation in* 17^{th} *Century English*. Malmö: CWK Gleerup.

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.

Baayen, R. Harald. 2008. *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. 2008. *lme4: Linear mixed-effects models using S4 classes*. R package version 0.999375-23.

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, and David E. Irwin. 1977. 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.

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. 2007. 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. 2007. 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, 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).

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.

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

Clark, Herbert H. 1994. Managing problems in speaking. *Speech Communication* 15(3–4):243–250.

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

De Smedt, Koenraad, and Gerard Kempen. 1991. Segment grammar: A formalism for incremental sentence generation. In C. L. Paris, W. R. Swartout, and W. C. Mann (Eds.), *Natural Language Generation in Artificial Intelligence and Computational linguistics*, 329–349. Boston/Dordrecht/London: Kluwer Academic Publishers.

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.

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.

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

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

Gibson, Edward. 2000. The dependency locality theory: a distance-based theory of linguistic complexity. In A. Marantz, Y. Miyashita, and W. O'Neil (Eds.), *Image, Language, Brain*, 95–126. Cambridge, MA: MIT Press.

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

Grimm, Scott, and Joan Bresnan. 2007. Spatiotemporal variation in the dative alternation: a study of four parallel corpora of British and American English. Unpublished ms., Stanford University Department of Linguistics.

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.

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.

Kempen, Gerard, and Edward Hoenkamp. 1987. An incremental procedural grammar for sentence formulation. *Cognitive Science* 11(2):201–258.

Konieczny, Lars. 2000. Locality and parsing complexity. *Journal of Psycholinguistic Research* 29(6):627–645.

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.

Leech, Geoffrey, Brian Francis, and Xunfeng Xu. 1994. The use of computer corpora in the textual demonstrability of gradience in linguistic categories. In C. Fuchs and B. Victorri (Eds.), *Continuity in Linguistic Semantics*, 57–76. Amsterdam and Philadelphia: John Benjamins.

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

Levy, Roger, Evelina Fedorenko, and Edward Gibson. 2007. The syntactic complexity of Russian relative clauses. Paper presented at the CUNY-2007 sentence processing conference on Sentence-Processing, Lajolla, CA, March 2007.

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.

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.

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

O'Connor, Mary Catherine, Arto Anttila, Vivienne Fong, and Joan Maling. 2005. 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.

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.

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

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.

Reali, Florencia, and Morten H. Christiansen. 2007. Processing of relative clauses is made easier by frequency of occurrence. *Journal of Memory and Language* 57(1):1–23.

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

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.

Roland, Douglas, Frederic Dick, and Jeffrey L. Elman. 2007. Frequency of basic English grammatical structures: A corpus analysis. *Journal of Memory and Language* 57(3):348–379.

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.

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.

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.

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.

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. 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.

Tagliamonte, Sali A., and Lidia Jarmasz. 2008. Variation and change in the English genitive: a sociolinguistic perspective. Paper presented at the 82^{nd} Annual Meeting of the Linguistic Society of America, Chicago, Illinois, January 4, 2008.

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

Thompson, Sandra. 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. To appear. Syntactic probabilities affect pronunciation variation in spontaneous speech. *Language and Cognition* 1(2).

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.

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. 1997. End-weight from the speaker's perspective. *Journal of Psycholinguistic Research* 26(3):347–361.

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

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.

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:

(9) ***Speaker A:***
I just spoke to Peter on the phone. He didnt 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.