In a land far far away…
A probabilistic account of the dative alternation in British, American, and Australian English[1]

DAPHNE THEIJSSEN
Radboud University Nijmegen

JOAN BRESNAN
Stanford University

MARILYN FORD
Griffith University

LOU BOVES
Radboud University Nijmegen

ABSTRACT
This article presents a corpus and judgement study of the dative alternation, i.e. the alternation between the double object construction (e.g. *I'll give him the money*) and the prepositional dative construction (*I'll give the money to him*). The present research is performed in a framework of probabilistic linguistics, in which we assume that syntactic structure is influenced by linguistic factors whose relative importance may vary. With regression models, we compared the dative alternation of British, American, and Australian speakers of English varying in age and gender. We found that both in produced speech and in judgements, the linguistic factors show a consistent pattern (*harmonic alignment*) across different varieties, age groups, and genders: animate objects usually precede inanimate objects, definites precede indefinites, shorter precede longer, and pronouns precede nonpronouns. The two studies also revealed subtle distributional differences between the roles that these linguistic factors play across the different speaker groups.

1 INTRODUCTION

Traditional linguistic theories have attempted to design deterministic rules that would account for all-and-only the sentences of a language that are deemed 'grammatical'. While acknowledging the fact that language use may be variable (graded), conventional theories assume that the underlying human grammar is categorical: a sentence is either grammatical, or it is not. The idea has now gained ground that 'grammaticality' is a graded concept itself, and that human language is essentially probabilistic in nature. A probabilistic theory of language can take various forms (e.g. the studies presented in Bod, Hay & Jannedy 2003),

[1] This study was supported by the Fulbright Center (personal grant to Daphne Theijssen), and by the National Science Foundation (grant numbers IIS-0624345 and BCS-2=1025602 to Stanford University, PI Joan Bresnan).

and is closely related to memory-based and exemplar-based models of language (e.g. Daelemans & van den Bosch 2005, Gahl & Yu 2006).

It is not surprising that many linguists have now moved from studying the dichotomy ('grammatical' and 'ungrammatical'), to studying variation in language. One obvious example of variation is the existence of syntactic alternations, in which there are different grammatical constructions that could be used to express the same core semantics. The alternative grammatical constructions are competing, and language users choose (subconsciously) among these options, depending on the perceived suitability given various syntactic, semantic, and discourse features. One of the best-studied syntactic alternations is the dative alternation in English, in which speakers and writers can choose between structures with a double object (1) or prepositional dative structure (2).[2]

(1) I'll give him the money.
(2) I'll give the money to him.

Previous research has already found sets of predictive syntactic, semantic, and discourse-related features that appear to influence the likelihood of the two dative constructions. For the last decade, researchers have employed multivariate models to combine the features (e.g. Arnold et al. 2000, Bresnan et al. 2007, Theijssen 2010). These models (usually logistic regression models) show that syntactic alternations can largely be explained by the tendency to place animate nouns before inanimate nouns, shorter constituents before longer ones, discourse given before discourse new, pronouns before nonpronouns and definite before indefinite.

The theoretical framework in which we study the dative alternation is that of probabilistic linguistics. We build on existing research on the dative alternation, making use of the predictive features that have been suggested in previous research. In our interpretation of the probabilistic linguistics framework, we assume that these features are suitable for the study of syntactic structure, and we investigate their relative contribution to the likelihood of the two constructions. These assumptions are inspired by the line of research initiated by Bresnan et al. (2007).

The features we employ and the probabilistic approach we take do not necessarily reflect the processes that take place in our brains, but they do form a convenient framework to explain a lot of the variance: Bresnan et al. (2007) were able to predict 94% of the choices in an unseen part of a corpus, with a probabilistic model based on a limited set of features. The prediction accuracy of their model was significantly better than the majority baseline of always selecting the double object construction: 79% (1859 of 2360 instances in the corpus). Also, the relevance of the features employed in Bresnan et al. (2007) has been established independently in psycholinguistic research (e.g. Bock & Irwin 1980, Bock et al. 1992, Prat-Sala & Branigan 2000), in other corpus studies (e.g. Weiner & Labov 1983, Givón 1984, Estival & Myhill 1988, Thompson 1990, 1995, Collins 1995, Snyder 2003, Szmrecsányi 2005, 2006), and in studies that combine experimental and corpus data (e.g. Arnold et al. 2000, Rosenbach 2003, 2005, Bresnan & Ford 2010).

[2] In fact, there are two additional options that are not included in our data: the reversed prepositional dative construction (e.g. *I gave to him a book*) and the reversed double object construction (e.g. *I gave it him*).

Parallels to the English dative alternation also occur in other languages, e.g. in Dutch (Colleman 2006), Korean (Choi 2007), Japanese (Miyagawa & Tsujioka 2004), Greek (Anagnostopoulou 2005) and Spanish (Beavers & Nishida 2010). The study of alternations across languages is very useful for gaining knowledge about syntactic variation (e.g. Levin 2008). However, cross-linguistic studies are not straightforward, because there are many differences between languages that can influence the variation. We therefore choose to study alternation in different variants of the same language: English. Over time, the different varieties of English have evolved in different ways, slowly altering the probabilistic distributions of noun phrases and syntactic constructions in each variety. By studying the dative alternation in different varieties of English, we aim at a (fine-grained) description of the differences in the likelihood of the two syntactic constructions across the varieties, and of the influence of various features on this likelihood.

A number of researchers have already investigated similarities and differences between the dative alternation in different varieties of English. Examples are comparisons of the dative alternation in American and British writing (Grimm & Bresnan 2009), in various written genres in American and British English over time (Wolk et al. 2011), in speech by people from New Zealand and the USA (Bresnan & Hay 2008), in various experiments conducted with American and Australian participants (Bresnan & Ford 2010), in speech and writing in Indian and British English (Mukherjee & Hoffmann 2006), and speech and writing in American and African American English (Kendall et al. 2011). Overall, it seems that the same features play a role in the dative alternation across different varieties of English, but that their relative importance may vary.

Many comparisons between the dative alternation in different varieties of English are still to be made, even in the varieties that have already been included in previous research. For instance, there are several studies that include British or Australian English (e.g. Mukherjee & Hoffmann 2006, Grimm & Bresnan 2009, Bresnan & Ford 2010), but there is no direct comparison between the two varieties. The present article provides multivariate studies of the dative alternation in British English, American English and Australian English. Our first research question is thus: What are the differences and/or similarities in the dative alternation in *British, American and Australian English*?

Given the findings in existing work, we expect to find the same general patterns in the roles that the features play in the three varieties, but with subtle distributional differences. British English was the origin of both American and Australian English, but at different times in history. American English came into existence after the arrival of the British in the early seventeenth century, while Australia was colonised much later, in the late eighteenth century. We therefore could expect to find other differences between American and British English, than between Australian and British English. On the other hand, the ubiquitous cultural influences that are typical of our present society, e.g. the wide-spread influence of American culture, may now be increasing the similarity of the three varieties.

It is now widely known that language is not only influenced by linguistic, but also by *extralinguistic* factors (e.g. Gregory 1967, Biber 1985). Recently, some researchers have extended existing multivariate models, based on linguistic features, by adding extralinguistic factors that may facilitate language variation and change (for an overview see Kristiansen & Dirven 2008, Geeraerts et al. 2010). Language variety, the topic of our

first research question, can be considered an extralinguistic factor. But there are also examples of extralinguistic factors that are nested in a language variety: properties of the text such as genre and modality, and speaker characteristics such as age and gender.

The effect that extralinguistic factors (other than language variety) have on the dative alternation has received only little attention in multivariate analyses. Much more attention has been paid to this in studies on the *genitive* alternation: 'John's book' vs. 'the book of John' (e.g. Hinrichs & Szmrecsányi 2007, Szmrecsányi & Hinrichs 2008, Tagliamonte & Jarmasz 2008, Jankowski 2009, Grafmiller 2010). Szmrecsányi (2010) shows that the linguistic factors play a role across time, varieties, modalities and genres, while the differences in the exact roles of these factors can be explained by extralinguistic factors.

For the *dative* alternation it has been found that in New Zealand English, young and elderly speakers favour the prepositional construction more than middle-aged speakers (Bresnan & Hay 2008). Also, Bresnan & Ford (2010) found a near-significant trend for male Australians to produce more prepositional dative constructions than female Australians. In this article, we study the effect of the sociolinguistic factors *age* and *gender* on the dative alternation in the aforementioned three varieties English. Our second research question is thus an extension of the first: What are the differences and/or similarities in the dative alternation of British, American and Australian language users varying in *age and gender*? Given the finding of Bresnan & Ford (2010), we expect male Australians to be more positive towards the prepositional dative construction than female Australians.

The goal of this article is to answer the two aforementioned research questions, one of which addresses the dative alternation in three varieties of English, and one the effect of age and gender on the dative alternation in these varieties. We perform two studies to reach this goal, placed in a probabilistic framework and making use of multivariate models.

Our first study is a corpus study. In corpus-driven research, theories are based on the *results* of language production that have been collected in natural settings, where speakers and writers are usually unaware of their language behaviour. We include only utterances made in spontaneous (unscripted) speech, since these are the most natural instances of language. In our corpus study, we compare the dative alternation in spontaneous speech in British and American English.[3]

It is rather difficult to investigate the effect of age and gender in corpus data, because (1) corpora often lack sufficient meta-data to provide the information needed, and (2) despite the common use of datives in English, it is difficult to find enough occurrences to investigate the variables, especially since so many other features influence the choice between the two syntactic constructions. For these reasons, we perform a second study: a judgement study in which participants rate the naturalness of the two constructions in context. In fact, the study is an extension of that in Bresnan & Ford (2010), but using a web-based version of the original judgement study on paper, and including participants from the US, Australia and the UK, with varying ages. Bresnan & Ford (2010: 201) found that the participants have 'strong predictive capacities, preferring and anticipating the more probable of two alternative syntactic paraphrases'. Thus, the judgements of the participants closely resembled the probabilities found in their corpus study. This is in line with the

[3] Australian English is not included due to our lack of a comparable Australian corpus (containing spontaneous speech collected in the early 1990s).

recent development of models in which processes used in production are directly linked to the processes used in comprehension. Pickering & Garrod (2005) suggest that people use language production processes to make predictions about the language they hear during comprehension. For this reason, distributional differences found in language production (corpus data) should also be found in perception and comprehension experiments. This motivates our use of a judgement study.

The remainder of this article is structured as follows: Section 2 presents existing work on the dative alternation in English. In Section 3, we present our corpus study; the judgement study is the topic of Section 4. Our general discussion and conclusion is provided in Section 5.

2 RELATED WORK

Many researchers have tried to explain the alternation between the two dative constructions. Some have argued that the change in word order and syntactic structure is likely to cause a change in meaning (Bolinger 1977, Pinker 1989, Levin 1993). Gries & Stefanowitsch (2004: 104) suggested that for the dative alternation: 'the ditransitive should prefer verbs of direct face-to-face transfer, while the *to*-dative should prefer verbs of transfer over distance'. However, empirical studies have shown that in spontaneous speech, speakers sometimes use both alternatives in the same context, repeating the same words, but using the other syntactic construction (Davidse 1996, Bresnan & Nikitina 2009). Research has also been directed at other semantic features that may affect the likelihood of the two constructions. Quirk et al. (1972: 843) for instance mentioned that 'indirect objects are typically animate'. Collins (1995: 47–8) found a discourse effect, seeing a 'strong likelihood' that the recipient will be 'informationally given' and the theme 'informationally new' in the double object construction.

Bresnan et al. (2007) combined the features suggested in individual studies in a multivariate model. Most of these features concern the two objects in the dative alternation: the recipient (*him* in examples 1 and 2) and the theme (*the money* in the examples). Some features are established for only one of the two objects because the other object is too biased to keep in regression models: most themes are inanimate and nonlocal (3rd person), and most recipients are concrete. We adapted this set of features for the current article, as indicated in Table 1. Bresnan et al. (2007) used these fourteen features in multivariate regression models for a data set extracted from the Switchboard corpus of spoken American English (Godfrey et al. 1992). The same fourteen features were employed by Theijssen (2010) in a study of the dative alternation in the ICE-GB corpus (Greenbaum 1996), containing spoken and written British English. In both studies, the same pattern appeared. Everything else being equal:

animate usually precedes inanimate
definite usually precedes indefinite
given usually precedes nongiven
shorter usually precedes longer
pronoun usually precedes nonpronoun

Table 1: *Features adapted from Bresnan et al. (2007). Only the features that are strongest and least correlated with each other (those between the horizontal lines) are included in the present article, using the values provided.*

Feature	**Description**	**Values (shortened)**
AnRec	Animacy of the recipient	animate (a), inanimate (in)
ConTh	Concreteness of theme	concrete (c), inconcrete (in)
DefRec	Definiteness of recipient	definite (d), indefinite (in)
DefTh	Definiteness of theme	definite (d), indefinite (in)
LenDif	Length difference (log of ratio)	ln(# words theme) – ln(#words recipient)
PrnRec	Pronominality of recipient	pronoun (p), nonpronoun (n)
PrnTh	Pronominality of theme	pronoun (p), nonpronoun (n)
—	Discourse givenness of recipient	given, nongiven
—	Discourse givenness of theme	given, nongiven
—	Person of recipient	local (1^{st}/2^{nd}), nonlocal (3^{rd})
—	Number of recipient	singular, plural
—	Number of theme	singular, plural
—	Semantic class	transfer of possession, future transfer of possession, prevention of possession, communication, abstract
—	Structural parallelism in dialogue	1, 0

This consistent pattern is sometimes referred to as *harmonic alignment*, and it has been found in several other varieties of English, e.g. in New Zealand English (Bresnan & Hay 2008), Australian English (Bresnan & Ford 2010) and African American English (Kendall et al. 2011). The length factor (shorter precedes longer) is a proxy for syntactic complexity or *end weight* (Behaghel 1909).

Although the harmonic alignment is consistent across different varieties of English, several studies have found distributional differences in the role that the features play. In a study of American and British writing from the 1960s and 1990s, Grimm & Bresnan (2009) found that in the 1990s British writers were more likely to use a personal pronoun as the second object of a double-object construction (e.g. *give the man it*) than American writers. The British flexibility with respect to pronominality can also be found in the fact that some British dialects allow reversed double object constructions such as *give it him* (Siewierska & Hollmann 2007, Haddican 2010). With respect to changes over time, Grimm & Bresnan (2009) found that both British and American English showed an increasing tendency to use the double object construction. The American data showed that the effect of pronominality was stronger in the 1990s than in the 1960s. For the British data, the effect is the opposite: the effects of pronominality and thematicity (an approximation of discourse givenness) were stronger in the 1960s than in the 1990s.

Bresnan & Ford (2010) used psycholinguistic experiments and found differences in the effects of object length between American and Australian English. In a judgement study, they found that as the recipient increased in length relative to the theme, the Australian

participants showed a greater preference for the prepositional dative than the US participants. In a task measuring reaction times while reading datives, they found that as the theme increased in length in prepositional datives, the US participants showed a more rapid (steeper) slowing down in reaction time than the Australians. Bresnan & Ford suggested that the Australians might have a stronger preference for the prepositional dative compared to US participants.

In the diachronic corpus study by Wolk et al. (2011), based on the ARCHER corpus (A Representative Corpus of Historical English Registers, Biber et al. 1994), the effect of theme length (measured in characters) was stronger in American than in British English writing. There were also some changes over time independent of language variety: the double object construction has become more popular, the effects of the animacy and the length of the recipient have *decreased*, and the effects of the pronominality and definiteness of the recipient have *increased*. Wolk et al. also observe that the more oral registers tend to favour the double object construction, while more literate registers contain relatively more occurrences of prepositional dative constructions. The same has also been observed by Bresnan et al. (2007) in their comparison of the dative alternation in spoken telephone dialogues (Switchboard) and written news paper texts (Wall Street Journal texts in the Penn Treebank, Marcus et al. 1993).

Bresnan & Hay (2008) found a stronger effect of animacy in spoken New Zealand English than in spoken American English. They also found that young and elderly New Zealanders favoured the prepositional construction more than middle-aged speakers. The corpus study by Mukherjee & Hoffmann (2006) showed that the prepositional dative is more frequent in the Indian than in the British English components of the International Corpus of English, ICE. The corpus study by Kendall et al. (2011) revealed no differences between the dative alternation in American and African American English speech and writing.

3 SPEECH CORPUS STUDY: BRITISH AND AMERICAN ENGLISH

Previous studies that compare the dative alternation in the two related varieties British and American English have shown that there are many similarities, but also some subtle distributional differences. This is particularly interesting because American English originated from British English in the early seventeenth century, but has evolved into a clearly different variant of English, embedded in its own nation and culture.

The existing work has focussed on written corpus data. Although written data can be very informative (especially when spoken data is not available, as for the historical study in Wolk et al. 2011), there is no question that spontaneous speech is the most natural language form. Therefore, we compare the dative alternation in British and American English in spontaneous speech corpus data. Our study will shed new light on the role that the linguistic features play in the dative alternation in two fairly different, but closely related, varieties of English. For this reason, it will provide us with new evidence for the universality of the features and their distributional differences in different varieties.

3.1 *Data*

Our American corpus data is a corrected version of the data described in Bresnan et al. (2007). It consists of 2,349 instances taken from the Switchboard corpus of American

telephone dialogues, collected in the early 1990s. All instances were manually checked for relevance and manually annotated with the features in Table 1. For details about the data, refer to Bresnan & Ford (2010). There are 38 different verb types.

The British corpus data is taken from Theijssen (2010) and consists of 930 instances from the British component of ICE. This corpus contains written and spoken English in various genres. It is annotated for the same features as the American corpus data (cf., Theijssen 2010 for details). We limit ourselves to the 491 instances of unscripted speech, containing 41 different verb types. The spoken part of the ICE-GB Corpus was collected in 1990-1992, the same period as Switchboard. It contains all kinds of spontaneous speech, including face-to-face spoken dialogues, but no telephone dialogues.

Given the different verb types included in the two data sets, we decided to remove all instances with verbs present in only one of the two sets. We thus removed from the American data the 264 instances in which the verb was *afford*, *allot*, *allow*, *award*, *bet*, *cost*, *deny*, *flip*, *float*, *loan*, *mail*, *promise*, *serve*, *swap*, *take* and *wish*. From the British data we removed the 35 instances with *bowl*, *circulate*, *deliver*, *explain*, *get*, *guarantee*, *let*, *mouth*, *open*, *pass*, *play*, *pose*, *present*, *report*, *return*, *sign*, *spread*, *square* and *suggest*. The resulting data set consists of 2,541 instances with 22 different verb types (see Table 2).

Table 2: *Verb types and frequencies in the combined corpus data*

Verb	**freq**	**Verb**	**freq**	**Verb**	**freq**	**Verb**	**freq**
give	1,512	*teach*	65	*write*	19	*cause*	12
send	170	*bring*	46	*feed*	16	*read*	11
tell	168	*offer*	46	*leave*	15	*assign*	5
pay	143	*charge*	42	*lend*	15	*make*	5
show	94	*do*	38	*hand*	13	*quote*	3
sell	79	*owe*	24				

The American and British corpus data were created separately, each with their own (single) annotator. To establish the similarity between the annotations in the two sets, the annotator of the British set (the first author of this article) annotated 30 items in the American set. The κ scores for the aforementioned features were all $\geq$ 0.78, showing good overall agreement. We included the feature for the length difference between the theme and the recipient (LenDif), following the definition in Theijssen (2010), as presented in Table 1.

3.2 *Method*

We compare the two varieties by including language Variety as a fixed main factor, as well as its interactions with the other features, in a mixed logistic regression model[4]. The features are potentially correlated; for instance, pronominal objects are generally shorter than full noun phrases because they often consist of a single pronoun only. The point-biserial correlation coefficients confirm this: 0.40 between LenDif and PrnTh, and 0.42 between LenDif and PrnRec. We therefore residualise LenDif: we include PrnRec and

[4] We used lmer() in the lme4 package in R (R Development Core Team, 2008).

PrnTh in a linear regression model that predicts LenDif. The unexplained variance (the residuals) is then included as a fixed factor (rLenDif) in the eventual logistic regression model, replacing LenDif.

The lemma of the verb (Verb) is included as a random factor, as well as the pair of the lemma of the verb and the lemma of the head of the theme (VerbThHead). This is to capture strong biases in certain expressions, for instance 'pay attention to someone' (VerbThHead = *pay_attention*), and 'give someone the creeps' (VerbThHead = *give_creep*). One could decide to exclude such instances, since they do not seem to allow alternation at first sight (cf. Ozón 2009, who excluded instances like 'give it a chance'). However, Google searches indicate that alternation often is possible in these expressions (cf. Bresnan et al. 2007), so we decided to retain them. Since many of these VerbThHeads are infrequent, we group all that have a frequency < 3 in a category 'other'.

The variable selection approach we take is the following: We first include all fixed factors and all interactions with Variety, and remove all *interactions* that are not significant, in a single step. Next, we continue with stepwise backward elimination, removing the most insignificant feature until no insignificant ones remain. If an interaction is significant, we always keep the features it consists of as main features as well. To test the final model for overfitting, we employ 10-fold cross-validation: A regression model is fitted to a subset of 90% of the data. The random effect values (the best linear unbiased predictors, or BLUPs) and coefficients (for the features, or 'fixed factors') are then used to predict the remaining 10% of the data. This is repeated ten times so that the whole data has served as a test set. The division in test sets is random. If a test set contains a Verb or VerbThHead that is not part of the training set, the random effect value is set to that of the category 'other'. The number of items per test set is 254, of which at least 36 and at most 55 are British. We present the average of the concordance *C*, the coefficients and the *p*-values across the ten test sets, together with their standard deviations.

3.3 *Results and discussion*

Table 3 presents information about the regression model we found. The *c*-number[5] for the residualised variables was 10.34 before variable selection, which indicates that there is mild collinearity. The Concordance (*C* index) [6] is above 0.97, also in 10-fold cross-validation, which shows very good overall fit.

Table 3: *Characteristics of the regression model for the combined spoken British and American English corpus data (1 = prepositional dative).*

Description	**Value**
Collinearity (*c*-number) before variable selection	10.34
Concordance (*C* index) of model fitted on all data	0.984
Average Concordance (10-fold cv)	0.972
Standard deviation Concordance (10-fold cv)	0.009
Number of observations	2,541

[5] We used collin.fnc() in the languageR package in R.
[6] We used somers2() in the Hmisc package in R.

In regression, the random factors are treated as normal distributions with mean 0. Table 4 provides the standard deviations of these distributions. Both Verb and VerbThHead are significant, even if both are included at the same time ($p<0.001$, using ANOVAs on models with the same fixed factors and nested random effects). In fact, most of the variance is explained by the random effects, which alone (i.e. without any of the features) yield a Concordance of 0.932.

Table 4: *Number of values and standard deviations of the random effects in the regression model fitted on all corpus data (1 = prepositional dative).*

Random effect	**Nr of values**	**St. Dev.**
VerbThHead	189	2.38
Verb	22	1.88

The coefficients of the features (the main effects), together with their *p*-values, are shown in Table 5. Positive coefficients indicate that the feature value in the first column increases the likelihood that the construction used is the prepositional dative, negative ones increase the likelihood for the double object construction. Features that are not significant, but kept in because they are part of a significant interaction, are written in parentheses. All significant main effects are in line with the *harmonic alignment* explained in Section 2: pronominal themes and inanimate and indefinite recipients favour the prepositional dative construction, longer themes and pronominal recipients the double object construction.

Table 5: *Coefficients and their p-values for the combined spoken British and American English corpus data, fitted on all corpus data (1 = prepositional dative). The average and standard deviation in 10-fold cross-validation are also provided (the models fitted on 90% evaluated on the remaining 10%). Insignificant features are in parentheses.*

Feature	**coef**	**average**	**stdev**	**p-value**	**average**	**stdev**
(Intercept)	-0.35	-0.37	0.23	0.620	0.627	0.213
AnRec=in	1.42	1.43	0.11	0.000	0.000	0.000
DefRec=in	1.42	1.43	0.13	0.000	0.000	0.000
(DefTh=in)	-0.56	-0.55	0.17	0.318	0.371	0.158
PrnRec=p	-4.12	-4.15	0.10	0.000	0.000	0.000
PrnTh=p	2.83	2.86	0.15	0.000	0.000	0.000
rLenDif	-2.58	-2.60	0.09	0.000	0.000	0.000
(Variety=US)	0.67	0.69	0.14	0.150	0.172	0.081
DefTh=in:Variety=US	-1.34	-1.39	0.17	0.026	0.036	0.029
rLenDif:Variety=US	0.93	0.93	0.12	0.027	0.039	0.018

From the interaction term *DefTh=in:Variety=US*, we see that American speakers show a significantly stronger tendency towards the double object construction if the theme is indefinite (e.g. *give him a book*), as compared to British speakers. There is also an

interaction between variety and length difference (*rLenDif:Variety=US*). While there is a main effect of length difference, such that there is a greater preference for the double object dative as the theme increases in length compared to the recipient, this effect is less strong for the American speakers.

In their diachronic corpus study, Wolk et al. (2011) found that in both British and American English writing, the effect of the length of the recipient has decreased, and the effect of the definiteness of the recipient has increased. In our study of spoken data, we found that the definiteness of the theme is more important, and the length difference is less important in American English than in British English. It thus seems that, in comparison with the diachronic study in Wolk et al. (2011), American English has evolved even further than British English, with respect to definiteness and length. However, there are many differences between the study by Wolk et al. (2011) and our study, most notably the medium of data (spoken versus written) and the type of data used (diachronic and modern English). For this reason, the fact that our results seem to be in line with the findings in Wolk et al. should be tested further.

Overall, our corpus study has shown that in contemporary spontaneous spoken English, most of the features suggested in the literature play similar roles in British and American English. These roles are in line with *harmonic alignment*. However, it seems that the definiteness of the theme plays a significant role in American English, but not in British English. With respect to length difference, we saw that it is relevant across British and American English, but that its effect size differs. This provides novel evidence that there are slightly different distributional patterns in the production of two varieties of English.

4 Judgement study: Age and gender in British, American, and Australian English

In this section, we investigate how natural the same dative sentences seem in context, according to speakers of British, American, and Australian English. The participants vary in gender and age, which enables us to answer our second research question: What are the differences/similarities in the judgement of dative sentences in British, American, and Australian English made by participants varying in age and gender?

Effects of the *linguistic* factors used throughout this article have been found in various experimental studies (e.g. Bock & Irwin 1980, Bock et al. 1992, Arnold et al. 2000, Prat-Sala & Branigan 2000, Rosenbach 2005, Bresnan & Hay 2008, Bresnan & Ford 2010). People seem to use language production processes to make predictions in language comprehension (Pickering & Garrod 2005). Therefore, a judgement study will provide insight in the effect of age and gender on the dative alternation in the three varieties.

Until now, the effects of age and gender have been mostly ignored in research on the dative alternation in English. Exceptions are Bresnan & Hay (2008), who found that in New Zealand English, young and elderly speakers favour the prepositional construction more than middle-aged speakers, and Bresnan & Ford (2010), who found a (near-significant) tendency for male Australians to be more likely to produce a prepositional dative construction than female Australians. Since so little is known about the effect of age and gender on the dative alternation in English, we provide such a study for British, American, and Australian English.

4.1 ***Experimental setup***

We extend the judgement study of Bresnan & Ford (2010) by including British English, using a wider age range (20 to 65 years), and conducting it through a website instead of on paper. Participants had to read a short passage followed by two possible continuations: one with a double object construction, one with a prepositional dative. They were asked to rate the naturalness of both options by dividing 100 points between them: the more points, the more natural. A screenshot of the experiment website is provided in Figure 1. All items were presented in random order, and the order in which the two options were presented was alternated.

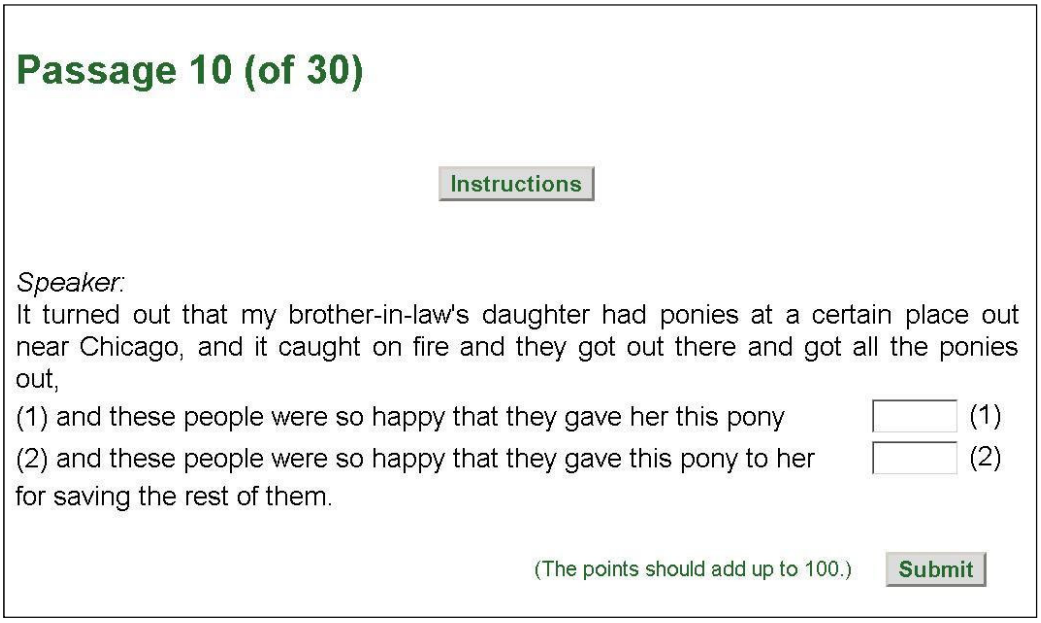

Figure 1
Screenshot of the experiment website

4.2 ***Items and participants***

We use the same 30 items taken from the Switchboard corpus as in Bresnan & Ford (2010), who already localised these for Australian English by replacing US-specific vocabulary and place names; we did the same for British English[7]. The verbs, themes, and recipients in the dative constructions were not altered, thus keeping the items comparable across the three varieties.

The participants were all volunteers who were entered in a prize draw for a gift voucher. Table 6 shows the characteristics of the participants.

[7] We thank Dr. Caroline Piercy for localising the items for British English.

Table 6: *Characteristics of British (UK), American (US) and Australian (Aus) participants in the judgement study*

	Female					***Male***				
	N	**av age**	**sd age**	**min age**	**max age**	**N**	**av age**	**sd age**	**min age**	**max age**
UK	22	32.0	11.7	21	61	18	31.5	10.4	21	63
US	22	37.3	14.3	21	65	13	32.1	11.7	21	61
Aus	23	35.3	12.3	23	63	17	32.1	11.8	20	64

4.3 ***Modelling***

Using linear regression, we modelled the participants' ratings for the prepositional dative variant, being a number between 0 and 100. The features we included as fixed factors are the same as those discussed in Section 3, except that we now residualise length difference on the definiteness and the pronominality of the recipient and the theme. The reason for this is that for the 30 items used, there is not only a high correlation of length difference with the two pronominality features (0.80 for the recipient, and 0.49 for the theme), but also with definiteness (0.29 for the recipient, and 0.32 for the theme).

Besides these main factors, we also look at Age, Gender, and Variety, and their interaction with the other features. We also control for the random order in which the items were presented (ItemOrder, ordinal with values 1-30), the alternating order of the two options (OptionsOrder, binary) and the rating assigned to the previous item (PrevRating, interval scale) by including these as fixed factors.

The verb lemma (Verb) is included as a random factor. Some participants generally award higher scores to the prepositional dative variant than other participants. Also, some participants use the whole range of points (0-100), while others use only a portion of it (e.g. 30-70). With our regression models, we want to establish the relative influence of the explanatory features between participant groups (the three nationalities, age and gender), not within these groups. We therefore follow the approach in Bresnan & Ford (2010) and correct the individual participant differences by including a random intercept for participant (Participant) and a random slope of participant over the centred predicted corpus probabilities (cCorpusProbs). The predicted corpus probabilities are taken from a logistic regression model built on the 2,349 spoken American instances, including the verb theme head as a random intercept. The predicted corpus probabilities are centred by subtracting the mean, as suggested in the literature (e.g. Baayen 2008).

We apply the same variable selection method as in Section 3. We first build three separate regression models for the three varieties of English, including two-way interactions between the features and Age and Gender. Next, we build a single model for all three varieties, only including the interactions that showed up as significant in at least one of the three separate models. The models are evaluated by R^2. We again also do this in a 10-fold cross-validation setting: we randomly divide the 3,450 items in 10 test sets with 345 items each. On each test set, we apply the regression model fitted on the remaining 3,105

items, and calculate the R^2. We also establish the average and standard deviation of the coefficients and *t*-values[8] in the 10 regression models.

4.4 ***Results for the individual models***

Figure 2 presents the significant coefficients in the individual models found for the ratings by participants of the same variety of English. The values for PrevRating and Age have been divided by 10 to increase the coefficients with a factor 10 (to make them more visible in the Figure).

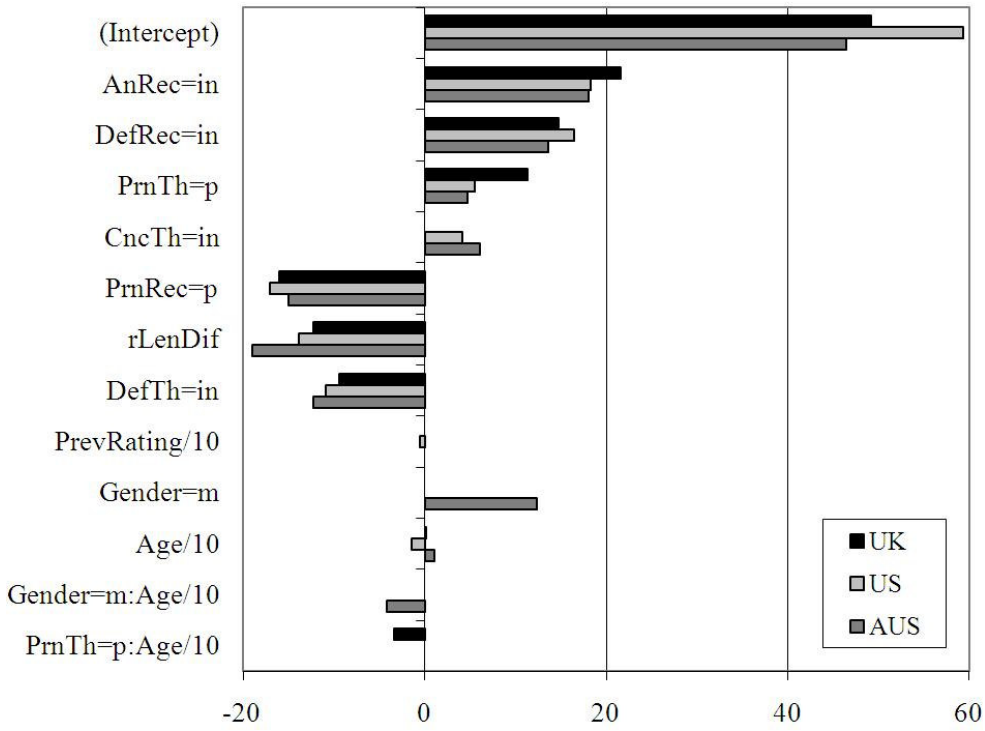

Figure 2: *Coefficients in the final models, made separately for the British, American, and Australian participants. Positive coefficients favour the prepositional dative construction, negative coefficients the double object construction.*

The random effects for Verb, Participant and Participant over cCorpusProbs are significant, also in combination with each other ($p<0.001$, using ANOVAs on models with the same fixed factors and nested random effects). The R^2 is 0.514 for the UK model, 0.499 for the US model and 0.584 for the Aus model, which means that the models are able to explain about half of the variance. This is similar to what was found in the study we extended (R^2 =

[8] *P*-values are not included because there is still uncertainty in the field on how to establish the number of degrees of freedom in mixed-effect linear regression models. We consider significant the features with an absolute *t*-value >2 (Baayen 2008).

0.529, Bresnan & Ford 2010). A large part of the variance is explained by the random effects: the random-effects-only models reach R^2s of 0.475 (UK), 0.457 (US) and 0.505 (Aus).

Overall, the effects of the features are again in accordance with *harmonic alignment*. The bar for *PrnTh=p:Age/10* in Figure 2 shows that older British participants assign lower ratings to the prepositional dative construction when the theme is a pronoun than younger British participants. Younger British people thus seem to have a stronger preference for structures such as *I gave it to the man* (not *gave the man it*) than older people.

Figure 2 also shows that in the judgements made by the American participants, older American participants give lower ratings for the prepositional dative than younger participants (indicated by the bar for Age/10). This seems to indicate that in the US, the prepositional dative construction is becoming slightly more popular.

Male Australian participants give higher ratings to the prepositional dative than female Australian participants (the bar for *Gender=m* in Figure 2), which supports the findings in Bresnan & Ford (2010). The effect is less strong for older than for younger Australian males (because of the significant interaction *Gender=m:Age/10*). We could thus say that younger Australian men seem to be more positive towards using the prepositional dative construction.

4.5 *Results for the combined model*

All features in Figure 2 are next included in the combined model, together with their interaction with language Variety. The characteristics of the final model can be found in Table 7, the standard deviations of the random effects in Table 8.

Table 7: *Characteristics of the regression model for the ratings for the prepositional dative construction by British, American, and Australian participants (combined).*

Description	**Value**
Collinearity (*c*-number) before variable selection	12.62
R^2 of model fitted on all data	0.528
Average R^2 (10-fold cv)	0.485
Standard deviation R^2 (10-fold cv)	0.043
Number of observations	3450

Table 8: *Number of values and standard deviation of the random effects in the regression model, fitted on all (combined) judgement data.*

Random effect	**Nr of values**	**St. Dev.**
Participant	115	4.71
Participant/cCorpusProbs		11.21
Verb	9	14.82
Residual		21.81

The *c*-number of the residualised variables was 12.62 before variable selection, which means a mild correlation. The R^2 shows that the model explains about half of the variance, also in 10-fold cross-validation. A model with only random effects reaches an R^2 of 0.478, which again shows that a lot of the variance is explained by the intercepts for Verb and Participant and the random slope for Participant over cCorpusProbs.

The significant coefficients (together with their *t*-values) are shown in Table 9. The significant effect for *Variety=US* shows that American participants give higher naturalness scores to prepositional dative constructions than British participants. Our data revealed no such difference between British and Australian participants (*Variety=Aus*).

Table 9: *Coefficients and their t-values for the ratings for the prepositional dative construction by the British, American, and Australian participants, fitted on all data. Insignificant features are in parentheses.*

Feature	**coef**	**average**	**stdev**	***t*-value**	**average**	**stdev**
Intercept	48.21	48.24	0.45	9.32	9.27	0.17
AnRec=in	19.06	19.13	0.44	15.05	14.35	0.35
ConTh=in	4.20	4.17	0.40	3.98	3.74	0.37
DefRec=in	14.95	15.00	0.90	7.73	7.36	0.43
DefTh=in	-11.04	-11.06	0.36	-10.48	-9.95	0.33
PrnRec=p	-16.28	-16.38	0.49	-13.56	-13.06	0.48
PrnTh=p	3.72	3.72	0.64	3.02	2.85	0.49
rLenDif	-14.11	-14.13	0.46	-10.03	-9.53	0.30
(Variety=Aus)	1.17	1.17	0.28	0.87	0.84	0.19
Variety=US	3.73	3.72	0.36	2.68	2.60	0.21
rLenDif:Variety=Aus	-3.92	-3.93	0.51	-2.25	-2.14	0.29
(rLenDif:Variety=US)	0.19	0.19	0.83	0.10	0.10	0.44

With respect to Length difference, we see that Australian participants show a significantly stronger effect than British participants (*rLenDif:Variety=Aus*): as the theme increases in length relative to the recipient, the Australians increasingly favour the double object dative more so than the British participants. This difference could not be found in our American and British data.

4.6 *Discussion*

In the individual models per variety (summarised in Figure 2), our data showed that younger British participants tend to be influenced more strongly by the pronominality of the theme than older British participants. In various dialects of British English, speakers are relatively flexible with respect to pronominality: They allow reversed double constructions such as *give it him* (Siewierska & Hollmann 2007, Haddican 2010). Since our participants come from many regions in the UK (see Figure 3), it is likely that many are familiar with

this construction.[9] That the effect of the pronominality is getting stronger could indicate that British English is moving away from the marked dialectal construction. We should note that the two-way interaction was only just significant (t=-2.33), which explains why it was not significant as a three-way interaction (with Variety) in the combined model.

Figure 3: *Google map with the places where the British participants spent most of their youth*

As for the effect of age, the individual model for American English revealed that in the US, the prepositional dative construction is most popular with the younger participants. The individual model for Australian English showed a similar effect for younger Australian men. In the combined model, we also discovered a difference in the preference for the prepositional dative construction across varieties: American participants gave higher naturalness scores to this variant than British participants, and there was no such difference

[9] Because of the many different regions present in the British data, including it in the model would hardly differ from including the individual speaker.

between British and Australian participants. The intercepts in the three individual models in Figure 2 show the same: it is obviously the highest for the US participants. Our corpus study in the previous Section showed no such Variety effect. Recall that to enable comparison between the three varieties, we decided to use the same items for all participants, only slightly adapted to match their nationality (e.g. changing place names). Given the fact that these items were all taken from an American English corpus (Switchboard), our finding for the US participants could be an artefact of the stimuli chosen.

With respect to length difference, our data indicated that Australian participants show a stronger effect than British participants. This difference is not found for American and British participants, making it similar to what was found in the study by Bresnan & Ford (2010) that we extended. In the previous section on corpus data, we found a significant interaction of length difference and Variety in spoken American and British English. This effect has disappeared in the judgement study. The same is true for the significant interaction between Variety and the definiteness of the theme in the corpus study. We will come back to this issue in the next section.

5 General discussion and conclusion

Despite the substantial amount of existing research on the dative alternation in English, various questions still remain to be answered. Many comparisons between the dative alternation in different varieties of English are still to be made. Also, the effect of extralinguistic factors on the dative alternation has so far received only little attention in multivariate analyses. In this article, we therefore aimed for two research objectives: (1) establishing similarities and differences in the dative alternation in British, American, and Australian English, and (2) establishing similarities and differences in the dative alternation of speakers varying in gender and age. The findings were related to previous research on the dative alternation in different varieties of English, all in a framework of probabilistic linguistics. In that framework, we assume that linguistic factors influence syntactic structure, and investigate their relative contribution to the likelihood of the two constructions in different speaker groups.

The two studies have shown that there are certain patterns in the data sets that are in line with each other and with previous work. These patterns show a *harmonic alignment*: speakers, writers, and participants in experiments tend to place or prefer phrases with certain characteristics before phrases with other characteristics:

animate usually precedes inanimate
definite usually precedes indefinite
shorter usually precedes longer
pronoun usually precedes nonpronoun

These patterns have also been found for other syntactic alternations (e.g. also in the genitive alternation, Szmrecsányi 2010), varieties of English (e.g. British, American, Australian, New Zealand, Indian and African-American English), and types of data (speaking versus writing, corpus or experimental). A probabilistic linguistics framework thus seems suitable for modelling effects that certain *linguistic factors* have on syntactic structure.

Besides similarities, the studies have also revealed some distributional differences between varieties and between speakers varying in age and gender, both with respect to the

distribution of the two dative constructions and with respect to the relative contribution of the predictive features.

With respect to variety (the first research objective), we used a corpus and a judgement study to find differences and similarities in dative alternation in British, American, and Australian English. American English and Australian English originated from British English at different times in history. Despite the cultural exchange that dominates our modern society, we expected that we would find other distributional differences between British and American English than between British and Australian English. This is indeed what we found.

In our corpus study, we found that in contemporary spontaneous spoken English, American speakers show a stronger tendency towards the double object construction when the theme is indefinite (e.g. *give him a book*) than British speakers. There is a greater preference for the prepositional dative as the recipient length increases relative to the theme, but the effect is less strong for the American speakers. We suggested that the relative importance of these two features may have evolved further in American English than in British English, following the developments found in the diachronic study by Wolk et al. (2011). But seeing that these effects were not replicated in our judgement study, our conclusion is speculative. Our judgement study indicated that the effect of length difference is stronger for Australian participants than for British participants. This is exactly the same as what was found in the study by Bresnan & Ford (2010) that we extended.

We also investigated the differences and similarities in the dative alternation in British, American and Australian English, made by participants varying in *age and gender* (the second research objective). The British judgements revealed that younger British participants seem to be influenced more strongly by the pronominality of the theme than older British participants. In many dialects in British English, it is common to use double object constructions with pronominal themes, e.g. *give the man it* (Siewierska & Hollmann 2007, Haddican 2010). Our study shows that younger participants are more in favour for the prepositional dative variant (*give it to the man*) than the older participants, which means they are moving away from the dialectal construction. The US judgements showed that the prepositional dative construction is most popular with the younger participants, so it seems that this construction is becoming slightly more popular in American English. A similar effect was found in the Australian judgements: especially younger Australian men seem to be more positive towards using the prepositional dative construction. In general, Australian men are more positive towards using the prepositional dative than Australian women, which supports the findings in Bresnan & Ford (2010).

This study has presented novel evidence that there are universal features playing a role in the dative alternation in English, but that there are subtle distributional differences between the roles that they play across varieties, and across speakers varying in gender and age. We have used a corpus study and a judgement study to reach this goal. Seeing that people seem to use language production processes to predict language input during language comprehension (Pickering & Garrod 2005), we expected to find similar results in the two studies. However, this was not always the case. The results, and comparisons to existing research, could not always be interpreted straightforwardly since the underlying data sets are not fully compatible. For instance, the American corpus data contained spoken telephone dialogues only, while the British corpus data contained all kinds of spontaneous

speech except telephone dialogues. Also, despite the high inter-annotator agreement between the annotations of the British and American corpus data, there are always differences (see also Theijssen et al. 2011, who also used these two data sets). In the experimental data, our choice for using American-based items for all three varieties may have influenced the results.

There is one more observation we should make. The corpus study revealed that most of the variance in the dative alternation in spontaneous speech can be explained by the verb and the combination of the verb and the theme head. In the judgement study, most of the variance could be explained by the verb of the test item and by the individual participant. In both studies, the predictive features under investigation played a significant, but minor, role. This is often the case in (psycho)linguistic studies (cf. Baayen 2008). In order to establish the universality of the features, on top of the effect of frequent lexico-syntactic patterns and participant-specific preferences, future research should be directed at studying the dative alternation and other syntactic alternations in languages other than English.

REFERENCES

Anagnostopoulou, Elena. 2005. Cross-linguistic and cross-categorial variation of datives. In Stavrou & Terzi (eds.), pp 61–126.

Arnold, Jennifer, Thomas Wasow, Anthony Losongco & 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. *Analyzing Linguistic Data. A Practical Introduction to Statistics Using R*. Cambridge: Cambridge University Press.

Beavers, John & Chiyo Nishida. 2010. The Spanish dative alternation revisited. In Colina et al. (eds.), 217–30.

Behaghel, Otto. 1909. Beziehungen zwischen Umfang und Reihenfolge von Satzgliedern. *Indogermanische Forschungen* 25, 110–42.

Belle, William van & Willy van Langendonck. 1996. *The dative, vol. 1: Descriptive studies*. Amsterdam: John Benjamins.

Biber, Douglas. 1985. Investigating macroscopic textual variation through multi-feature/ multi-dimensional analyses. *Linguistics* 23(2), 337–60.

Biber, Douglas, Edward Finegan & Dwight Atkinson. 1994. ARCHER and its challenges: compiling and exploring A Representative Corpus of Historical English Registers. In Fries et al. (eds.), 1–13.

Bock, Kathryn & David Irwin. 1980. Syntactic effects of information availability in sentence production. *Journal of Verbal Learning and Verbal Behavior* 19(4), 467–84.

Bock, Kathryn, Helga Loebell & Randal Morey. 1992. From conceptual roles to structural relations: Bridging the syntactic cleft. *Psychological Review* 99(1), 150–71.

Bod, Rens, Jennifer Hay & Stefanie Jannedy. 2003. *Probabilistic Linguistics*. Cambridge, MA: MIT Press.

Bolinger, Dwight. 1977. *Meaning and form*. London: Longman.

Bouma, Gerlof, Irene Krämer & Joost Zwarts (eds.). 2007. *Cognitive Foundations of Interpretation*. Amsterdam: Royal Netherlands Academy of Science.

Bresnan, Joan & Jennifer Hay. 2008. Gradient Grammar: An Effect of Animacy on the Syntax of give in New Zealand and American English. *Lingua* 118(2), 245–59.

Bresnan, Joan & Marilyn Ford. 2010. Predicting Syntax: Processing dative constructions in American and Australian Varieties of English. *Language* 86(1), 168–213.

Bresnan, Joan, Anna Cueni, Tatiana Nikitina & R. Harald Baayen. 2007. Predicting the Dative Alternation. In Bouma et al. (eds.), 69–94.

Bresnan, Joan & Tatiana Nikitina. 2009. The gradience of the dative alternation. In Uyechi & Wee (eds.), 161–84.

Choi, Hye-Won. 2007. Length and order: A corpus study of Korean dative-accusative construction. *Discourse and Cognition* 14(3), 207–27.

Colina, Sonia, Antxon Olarrea & Ana Maria Carvalho (eds.). 2010. *Romance Linguistics 2009: Selected Papers from the 39th Linguistic Symposium of Romance Languages (LSRL)*. Amsterdam: John Benjamins.

Colleman, Timothy. 2006. *De Nederlandse datiefalternantie: een constructioneel en corpusgebaseerd onderzoek*. Ph.D. dissertation, Ghent University.

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

Daelemans, Walter & Antal van den Bosch. 2005. *Memory-based language processing*. Cambridge, MA: Cambridge University Press.

Davidse, Kristina. 1996. Functional dimensions of the dative in English. In van Belle & van Langendonck (eds.), 289–338.

Edmondson, Jerold A., Crawford Feagin & Peter Mühlhausler. 1990. *Development and diversity: Language variation across time and space*. Dallas, Arlington, TX: Summer Institute of Linguistics and University of Texas at Arlington.

Estival, Dominique & John Myhill. 1988. Formal and functional aspects of the development from passive to ergative systems. In Shibatani (ed.), 441–91.

Fries, Udo, Gunnel Tottie & Peter Schneider (eds.). 1994. *Creating and using English language corpora: papers from the Fourteenth International Conference on English Language Research and Computerized Corpora*. Amsterdam: Rodopi.

Gahl, Susanne & Alan C.L. Yu. 2006. Introduction to the special issue on exemplar-based models in linguistics. *The Linguistic Review* 23(3), 213–216.

Geeraerts, Dirk, Gitte Kristiansen & Yves Peirsman. 2010. *Advances in Cognitive Sociolinguistics*. Berlin: Walter de Gruyter.

Givón, Talmy. 1984. Direct object and dative shifting: Semantic and pragmatic case. In Plank (ed.), 151–82.

Godfrey, John, Edward Holliman & Jane McDaniel. 1992. Switchboard: Telephone speech corpus for research and development. *International Conference on Acoustics, Speech and Signal Processing (ICASSP-92)*, 517–20.

Grafmiller, Jason. 2010. Variation in English genitives across modality and genre. Ms., Stanford University.

Greenbaum, Sidney (ed.). 1996. *Comparing English Worldwide: The International Corpus of English*. Oxford: Clarendon Press.

Gregory, Michael. 1967. Aspects of Varieties Differentiation. *Journal of Linguistics* 3(2), 177–97.

Gries, Stefan Th. & Anatol Stefanowitsch. 2004. Extending Collostructional Analysis: A Corpus-based Perspective on 'Alternations'. *International Journal of Corpus Linguistics* 9(1), 97–129.

Grimm, Scott & Joan Bresnan. 2009. Spatiotemporal variation in the dative alternation: A study of four corpora of British and American English. *Third International Conference Grammar and Corpora*.

Haddican, William. 2010. Theme-Goal Ditransitives and Theme Passivisation in British English Dialects. *Lingua* 120(10), 2424–43.

Hannay, Mike & Gerard J. Steen (eds.). 2007. *Structural-functional studies in English grammar*. Amsterdam: John Benjamins.

Hinrichs, Lars & 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), 437–74.

Icard, Thomas & Reinhard Muskens (eds.). 2010. *Interfaces: Explorations in Logic, Language and Computation*. Berlin: Springer.

Jankowski, Bridget. 2009. Grammatical and register variation and change: A multi-corpora perspective on the English genitive. *American Association for Corpus Linguistics (AACL) 2009*.

Kendall, Tyler, Joan Bresnan & Gerard van Herk. 2011. The Dative Alternation in African American English: Researching Syntactic Variation and Change across Sociolinguistic Datasets. *Corpus Linguistics and Linguistic Theory* 7(2), 229–44.

Kristiansen, Gitte & René Dirven (eds.). 2008. *Cognitive Sociolinguistics: Language Variation, Cultural Models, Social Systems*. Berlin: Mouton de Gruyter.

Landsberg, Marge E. (ed.). 1995. *Syntactic iconicity and linguistic freezes*. Berlin: Mouton de Gruyter.

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

Levin, Beth. 2008. Dative verbs: A crosslinguistic perspective. *Lingvisticæ Investigationes* 31(2), 285–312.

Marcus, Mitchell P., Mary Ann Marcinkiewicz & Beatrice Santorini. 1993. Building a large annotated corpus of English: the Penn Treebank. *Computational Linguistics* 19(2), 313–30.

Miyagawa, Shigeru & Takae Tsujioka. 2004. Argument structure and ditransitive verbs in Japanese. *Journal of East Asian Linguistics* 13(1), 1–38.

Mondorf, Britta & Günther Rohdenburg (eds.). 2003. *Determinants of Grammatical Variation in English.* Berlin: Mouton de Gruyter.

Mukherjee, Joybrato & Sebastian Hoffman. 2006. Describing verb-complementational profiles of New Englishes. *English World-Wide* 27(2), 147–73.

Nevalainen, Terttu, Irma Taavitsainen, Päivi Pahta & Minna Korhonen (eds.). 2008. *The Dynamics of Linguistic Variation: Corpus evidence on English past and present*. Amsterdam: John Benjamins.

Ozón, Gabriel Alejandro. 2009. *Alternating Ditransitives in English: A Corpus-Based Study*. Ph.D. dissertation, University College London.

Pickering, Martin J. & Simon Garrod. 2005. Do people use language production to make predictions during comprehension? *Trends in Cognitive Sciences* 11(3), 105–10.

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

Plank, Frans (ed.). 1984. *Objects: Towards a theory of grammatical relations*. New York: Academic Press.

Prat-Sala, Mercè & 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(2), 168–82.

Quirk, Randolph, Sidney Greenbaum, Geoffrey Leech & Jan Svartvik. 1972. *A Grammar of Contemporary English*. London: Longman.

R Development Core Team. 2008. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna. http://www.R-project.org.

Rosenbach, Anette. 2003. Aspects of iconicity and economy in the choice between the *s*-genitive and the *of*-genitive in English. In Mondorf & Rohdenburg (eds.), 379–411.

Rosenbach, Anette. 2005. Animacy versus weight as determinants of grammatical variation in English. *Language* 81(3), 613–44.

Shibatani, Masayoshi (ed.). 1988. *Passive and Voice*. Amsterdam: Benjamins.

Siewierska, Anna & Willem Hollmann. 2007. Ditransitive clauses in English with special reference to Lancashire dialect. In Hannay & Steen (eds.), 83–102.

Snyder, Kieran. 2003. The relationship between form and function in ditransitive constructions. Ph.D. dissertation, University of Pennsylvania, Philadelphia, PA.

Stavrou, Melita & Arhonto Terzi (eds.). 2005. *Advances in Greek generative grammar*. Amsterdam: John Benjamins.

Szmrecsányi, Benedikt. 2005. Language users as creatures of habit: A corpus-based analysis of persistence in spoken English. *Corpus Linguistics and Linguistic Theory* 1(1), 113–49.

Szmrecsányi, Benedikt. 2006. *Morphosyntactic persistence in Spoken English: A corpus study at the intersection of variationist sociolinguistics, psycholinguistics, and discourse analysis*. Berlin: Mouton de Gruyter.

Szmrecsányi, Benedikt. 2010. The English genitive alternation in a cognitive sociolinguistics perspective. In Geeraerts et al. (eds.), 141–66.

Szmrecsányi, Benedikt & Lars Hinrichs. 2008. Probabilistic determinants of genitive variation in spoken and written English: A multivariate comparison across time, space and genres. In Nevalainen et al. (eds.), 291–309.

Tagliamonte, Sali & Lidia Jarmasz. 2008. Variation and change in the English genitive: A sociolinguistic perspective. *Annual Meeting of the Linguistic Society of America*.

Theijssen, Daphne. 2010. Variable selection in Logistic Regression: The British English dative alternation. In Icard & Muskens (eds.), 87–101.

Theijssen, Daphne, Lou Boves, Hans van Halteren & Nelleke Oostdijk. 2011. Evaluating automatic annotation: Automatically detecting and enriching instances of the dative alternation. *Language Resources and Evaluation*.

Thompson, Sandra A. 1990. Information flow and dative shift in English discourse. In Edmondson et al. (eds.), 239–53.

Thompson, Sandra A. 1995. The iconicity of 'dative shift' in English: Considerations from information flow in discourse. In Landsberg (ed.), 155–75.

Uyechi, Linda & Lian-Hee Wee (eds.). 2009. *Reality exploration and discovery: Pattern interaction in language and life*. Stanford, CA: CSLI Publications.

Weiner, E. Judith & William Labov. 1983. Constraints on the agentless passive. *Journal of Linguistics* 19(1), 29–58.

Wolk, Christoph, Joan Bresnan, Anette Rosenbach & Benedikt Szmrecsányi. 2011. Dative and genitive variability in Late Modern English: Exploring cross-constructional variation and change. Ms., University of Freiburg and Stanford University.