

A regression model for the English benefactive alternation

An efficient, practical, actually usable approach

Daphne Theijssen, Hans van Halteren, Karin Fikkers, Frederike Groothoff, Lian van Hoof, Eva van de Sande, Jorieke Tiems, Véronique Verhagen and Patrick van der Zande

Department of Linguistics, Radboud University Nijmegen

Abstract

In this paper, we use logistic regression modelling to predict the English benefactive alternation (*He baked me a cake* vs. *He baked a cake for me*). We developed a data set consisting of 107 instances in adult writing and 36 in the writing of 8-to-12-year-olds, and annotated them with 13 syntactic, semantic and discourse features. We show that a model trained and tested on the adult data reaches a prediction accuracy of 86.9%. Due to the small number of data instances, our model includes only 4 significant effects and shows considerable overfit (reaching 79.6% accuracy in a ten-fold cross-validation setting). The regression coefficients found are similar to those found in the model for the *to*-dative alternation (Bresnan et al. 2007). When applying the adult model to the instances in child writing, 80.6% is predicted correctly. We conclude that there are no indications of major differences either between the *to*-dative and benefactive alternation in adult language, nor between the benefactive alternation in adult language and that in child language.

1 Introduction

There are many situations where speakers can choose between several syntactic options that are equally grammatical, but that may differ in their acceptability in the given context. An example is the English benefactive alternation, where speakers choose between the double object variant (1) and the prepositional dative variant with the preposition *for* (2).

- (1) He baked me a cake.
- (2) He baked a cake for me.

The benefactive alternation is very similar to the *to*-dative alternation, where the prepositional dative variant is formed with the preposition *to*. In both the benefactive and the dative alternation, the constructions contain two objects: a ‘theme’ (*a cake* in the examples above) and a ‘recipient’ (*me*)¹.

¹In the benefactive alternation, the latter is also referred to as the ‘beneficiary’. In this paper, however, we stick to the term ‘recipient’ to keep the comparison to the *to*-dative alternation as transparent as possible.

The question how speakers and writers decide which construction to use has been addressed by many researchers in various fields. Approaches that have been taken include syntactic (Quirk et al. 1972), semantic (Gries and Stefanowitsch 2004) and discourse (Collins 1995) approaches. Probabilistic modelling techniques can help in combining these approaches and testing the combination on corpus data. Bresnan et al. (2007) have applied such an approach to explain the *to*-dative alternation in the language produced by adult speakers of American English. With the help of 14 features taken from the literature, they constructed logistic regression models which could predict the choice made with an accuracy of 94%.

In this paper, we apply Bresnan et al.'s probabilistic approach to explain the alternation with the preposition *for* (i.e. the benefactive alternation). More specifically, we want to answer the following question: Are the features and the logistic regression methodology that were successfully applied to the *to*-dative alternation also suitable for explaining the benefactive alternation? Once we have answered this question, we can establish whether the model can also be applied successfully to the benefactive alternation in child language. This is our second, minor, research question.

In the research reported here², we focus on written text produced by adult native speakers of American and British English. For the child language, we limit ourselves to writing by children aged 8 to 12 years.

The structure of the paper is as follows: An overview of the related work is presented in Section 2. Section 3 focuses on the design and collection of our data set. The description of the features and the annotation process follows in Section 4. In Section 5, we present and discuss the logistic regression models. Final conclusions are provided in Section 6.

2 Related work

Bresnan et al. (2007) extracted 2360 instances from the three-million word Switchboard Corpus of transcribed American English telephone dialogues, and annotated them for features suggested in the literature. With the help of a logistic mixed-effect regression model with *verb* as a random effect, they were able to explain 95% of the variation. On previously unseen data, 94% of the test instances was predicted correctly, showing only a slight overfit. Next, they added 905 instances from the Wall Street Journal texts in the Penn Treebank and concluded that the model generalized well to the instances taken from financial newspaper articles

²The research was all carried out as part of the Master course 'Corpus-based Methods' in the research master Language and Communication, being a collaboration between Radboud University Nijmegen and Tilburg University. The students were involved in all steps of the research process: deciding on a topic, searching and reading related literature, finding relevant corpora, extracting and annotating the benefactive instances, applying logistic regression, evaluating the models, presenting at the CLIN conference and writing this paper. We believe the students have greatly benefited from this 'hands-on' experience and therefore recommend that such student projects are designed more frequently, especially in research masters.

Students appear in alphabetical order in the author list.

(accuracy of 93% for the combined data set).

Many of the features in the model concern the theme and the recipient. In prepositional dative constructions, the first-mentioned object is the theme, and the second the recipient. In double object constructions, the recipient precedes the theme. Bresnan et al. found that the first object is typically (headed by) a pronoun, mentioned previously in the discourse (*discourse given*), animate, definite and shorter than the second object. The characteristics of the second object are generally the opposite: discourse new, nonpronominal, inanimate, indefinite and relatively long.

Researchers have adopted Bresnan et al.'s approach (2007) to model the *to*-dative alternation in other varieties of English. Bresnan and Hay (2008) compared the *to*-dative alternation with the verb *give* in American and New Zealand English. The model built on the 1667 American and 1127 New Zealand English instances showed that the *to*-dative alternation is very similar in both varieties. The only clear difference they found was that speakers of New Zealand English are more sensitive to the animacy of the recipient. Theijssen (2009) used the features and logistic regression methodology in Bresnan et al. (2007) to study the *to*-dative alternation in spoken and written British English. The models found appear to be very similar to those found by Bresnan et al., though with fewer significant features and with less success in generalizing to previously unseen data. This was probably a result of the relatively small number of data instances (915).

Some research has also been directed at explaining the *to*-dative alternation in child language. De Marneffe et al. (2007) investigated the *to*-dative alternation as it occurs in the production of seven children, aged between 2 and 5. They concluded that the results of the adult and child data were "neither identical nor radically different" (2007, 13). The differences that did occur mainly concerned the strength of the influence. An important observation was that each child produced both constructions from early on in their language.

3 Constructing a data set

In order to compile a data set, we employed several existing corpora containing British and American English texts written by adults and children³. Initially, we only selected corpora that contained syntactic annotations and/or Part-of-Speech (POS) tagging to enable automatic extraction of the instances. For the adult data, we employed the written component of the ICE-GB Corpus (appr. 424,000 words, Greenbaum 1996) and the SUSANNE Corpus (appr. 130,000 words, Sampson 1995). With respect to child language, we decided to include texts written by 8-to-12-year-olds only; the resemblance between teenager and adult writing could be too large to detect differences, whereas the writing of children younger than 8 years old is probably too limited to find enough data. These criteria led to the selection of part of the LUCY corpus (appr. 30,000 words,

³Given the low frequency of benefactive constructions (as described later on in this section) and the consequent small size of our data set, we include both types of English. We are aware of the fact that this is a suboptimal solution.

Sampson 2003). Because this amount of child data was insufficient, the LCCPW corpus (appr. 78,500 words, Smith et al. 1998) was selected as well, although this corpus lacks syntactic and POS annotation. In total, our final corpus set consisted of approximately 662,500 words.

Next, we extracted the benefactive constructions from the corpora in several steps (the numbers of instances found at each step are presented in Table 1). First, we used Perl scripts to find sentences containing the construction with the preposition *for*. For the ICE-GB, SUSANNE and LUCY corpora, the available syntactic annotation or POS tags were used. For LCCPW, where no such information was available, we simply selected all sentences containing the word *for*.

Each automatically extracted *for*-instance was then randomly assigned to at least two annotators, who independently judged its relevance. Many of the instances were irrelevant, even when found with the help of syntactic information in the corpus. For instance, in example 3, the prepositional phrase *for the classification* has the meaning of *for the purpose of classification*.

- (3) The four shape categories associated with weld defects have been highlighted and two methods ... are now proposed for the classification of these shape types. -ICE-GB W2A-036 #88:1

The relevance judgements were compared pairwise. After that, doubtful instances were discussed by the annotators and a final decision about relevance was made.

In order to minimize the effect of other syntactic phenomena, we excluded from the resulting instances those with a phrasal verb, with a split object, and with fronting of one of the objects. This procedure resulted in a final list of 99 *for*-prepositional instances with 58 different main verbs.

Next, potentially relevant double object instances were extracted from the corpora using the 58 verbs. In ICE-GB, SUSANNE and LUCY, ditransitive verbs are marked as such, which facilitated an automatic selection process. For LCCPW, we extracted all sentences with any form of one of these 58 main verbs.

Irrelevant instances and instances with the syntactic characteristics mentioned above were removed manually using the same procedure as with the prepositional instances. For the LCCPW corpus, numerous instances (812 out of 819) had to be deleted, since the broad selection on the basis of main verbs yielded many clauses that did not contain a benefactive construction. In total, 44 instances with a double object construction remained.

As Table 1 shows, our final data set consists of 143 instances, 107 of which concern adults' writing and 36 concern children's writing. This is a rather small number, especially compared to the 3265 instances in Bresnan et al.'s (2007) data set. There are two reasons why our data set is smaller. First of all, Bresnan et al. use the three-million word Switchboard corpus, while our set of corpora contains 'only' 662,500 words. Secondly, the benefactive alternation appears to be five times less frequent than the *to*-dative alternation: Bresnan et al. found 1.09 instances per 1,000 words on average, while we obtain only 0.22 instances per 1,000 words. The small number of data instances obviously affects our method and the conclusions to be drawn from the results (see Section 5).

Table 1: Number of *for*-prepositional and double object instances found after automatic and manual selection

Type	Corpus	Automatic selection		Manual selection		<i>Total</i>
		<i>for</i> -prep.	double obj.	<i>for</i> -prep.	double obj.	
Adult	ICE-GB	767	30	56	23	79
	SUSANNE	165	10	19	9	28
	Total	932	40	75	32	107
Child	LUCY	20	5	6	5	11
	LCCPW	461	819	18	7	25
	Total	481	824	24	12	36

4 Feature annotation

4.1 Description of the features

For the annotation of the instances, we adopted the features suggested by Bresnan et al. (2007). An exception was ‘structure parallelism’. Research has shown that the writers’ choices are influenced by the construction that has been used previously in the text (e.g. Szmrecsanyi 2005). However, as concluded in the previous section, the benefactive alternation is very infrequent. It is therefore unlikely that any preceding instance of the benefactive alternation influences the current, especially since many other types of constructions can intervene. As a (suboptimal) solution to this problem, we decided to remove the feature altogether.

For the remaining thirteen features, the annotators employed a manual⁴ that follows the descriptions in Bresnan et al. (2007) as strictly as possible. Below, we briefly explain the different features.

Animacy of recipient (1)

Following Bresnan et al. (2007), the animacy of the recipient was annotated as a binary feature: it was labelled either *animate* (human and animal) or *inanimate* (not human or animal). Companies and organizations were considered animate when it was evident from the context that the writer meant the people working in these institutions. This is for example the case for *Dallas* in example 4, which refers to the players in the Dallas football team.

- (4) The Texans made themselves a comforting break on the opening kickoff when Denver’s Al Carmichael was jarred loose from the ball when Dave Grayson, the speedy halfback, hit him and Guard Al Reynolds claimed it for Dallas. –SUSANNE A12

⁴The annotation manual can be found at <http://lands.let.ru.nl/~daphne/phdproject/AnnotationManual/Benefactive2008.pdf>.

Concreteness of theme (2)

For the annotation of the concreteness of the theme, the instructions in Bresnan et al. (2007) were not very clear, except that the feature again allowed only two values: either *concrete* or *inconcrete*. We decided to follow Garretson (2003), in which a noun phrase is deemed concrete if it is prototypically concrete (e.g. *this new table* in 5). We assumed that prototypically concrete objects have a known physical size. The themes that did not fit this description were labelled *inconcrete* (e.g. *quite a bit of fun* in 6).

- (5) She 'd buy all the women, who were her mother, this new table and they'd still use the old one. –ICE-GB W2F-020 #122:1
- (6) There is a small brook – here too which provides quite a bit of fun for us and the dogs as we both can jump across it. –LUCY F10

Definiteness of recipient and theme (3, 4)

For both the recipient and the theme we annotated the definiteness. All (syntactic) object heads that were preceded by a definite article or a definite pronoun (e.g. demonstrative and possessive pronouns), and all objects that were proper nouns or definite pronouns themselves, were annotated *definite*. The remaining objects were given the value *indefinite*.

Discourse givenness of recipient and theme (5, 6)

A recipient or theme was labelled *given* when it was mentioned in the preceding context (maximally 20 clauses before). In example 7, the recipient “*themselves*” is given because it refers to “*The Texans*” at the beginning of the sentence. We also considered an object given when it was stereotypical of something mentioned in the preceding context, or when it was part of the writing context (e.g. the news paper article itself, or the reader). *You*, *one* and *us* as impersonal pronouns (e.g. *you* in example 8) were annotated as given as well. All remaining objects received the value *new*.

- (7) The Texans made themselves a comforting break on the opening kickoff when Denver's Al Carmichael was jarred loose from the ball ... –SUSANNE A12
- (8) Sewing is a very useful thing when you want to make yourself an apron or something like that –LUCY M09

Number of recipient and theme (7, 8)

Recipients and themes were annotated for number: *singular* or *plural*. In case a recipient or theme could refer to something singular or plural (which is especially the case with the pronoun *you*), the antecedent was checked.

Person of recipient (9)

Person of recipient was annotated by giving it the value *local* or *nonlocal*. Local recipients are in first or second person (e.g. *I, me, yourself*), non-local ones in third person.

Pronominality of recipient and theme (10, 11)

We also annotated whether or not the recipient and the theme were (syntactically) headed by a pronoun and thus *pronominal*, or not (*nonpronominal*). We treated all types of pronouns as such, including for instance indefinite and relative pronouns like *all* and *that*.

Semantic verb class (12)

In Bresnan et al. (2007), five semantic classes were distinguished for the dative alternation with 'to', mostly being different forms of *transfer of possession*. For the annotation of the verbs with the benefactive 'for', we encountered problems with their subdivision since there is generally no transferring. Therefore we formed new semantic classes for the benefactive verbs. From Pinker (1989, 84), we copied *creation of possession* (e.g. *produce* in example 9) and *obtaining of possession* (*get* in 10), which we supplemented with *keeping of possession* (*keep* in 11). In these classes, the theme is typically created, obtained or kept to cause the recipient to possess it. Following Bresnan et al. (2007), we add a fourth class, *abstract*, for instances with a figurative meaning (e.g. *do someone a favor* in 12). Also, instances where a situation is created (or arranged) for the recipient were considered abstract (e.g. *arrange something for someone* in 13).

- (9) The rotting bark also produces food for the stinkbug. –LCCPW DS62HRZ
- (10) Thanks again for getting all the stuff for me! –ICE-GB W1B-002 #58:1
- (11) Please could I have the following reprints (keep one of each for yourself as I've kept the originals). –ICE-GB W1B-004 #168:5
- (12) I'm sure you won't mind doing me a small favor. –SUSANNE N12
- (13) If you would like to have your book displayed at the conference again this year, please let me know and I will be happy to arrange this for you. –ICE-GB W1B-030 #96:5

Length difference (13)

An important factor in clause word order is the so-called *principle of end weight* (e.g. Quirk et al. 1972), which states that language users tend to place the more complex constituents at the end of an utterance. Bresnan et al. (2007) therefore included a feature indicating the length difference between the recipient and the theme. Following their approach, we counted the number of words in the recipient and the theme, and took the natural log of these lengths to smoothen outliers. The

length difference was then calculated by subtracting the recipient length from the theme length.

4.2 Annotation procedure

During the data extraction phase, the theme, recipient and verb were manually checked (for ICE-GB, SUSANNE, and LUCY) or established (LCCPW) for each instance. The lengths of the theme and recipient (in number of words) could be extracted automatically by splitting them on white space. The remaining twelve features had to be annotated manually for all corpora. The annotators were each assigned 50 instances from ICE-GB, SUSANNE, and LUCY and the 25 relevant instances from LCCPW. All instances were annotated by at least two annotators.

When there was doubt about the exact meaning of a feature, this was discussed in a meeting. For example, there was discussion about what should be done with recipients or themes that consisted of a coordination of two elements with opposite properties (e.g. a definite and an indefinite element). To keep the annotators' decisions as comparable as possible, we decided that these objects should be annotated according to the first element in the coordination⁵. After the discussions, all features that were experienced as difficult were checked again by the annotators. They were asked to make one decision for each case, with the option to add a question mark if they were unsure about it.

After the annotation, the inter-annotator agreement was measured for the twelve features that had been annotated manually. Each pair of annotators was given a score if they agreed on the feature value for a particular instance. This score was 1 if both annotators were confident about their decision, 0.5 if one of them had expressed doubt, and 0.25 if both doubted their choice. If the annotators disagreed, the score was 0. The relative observed agreement for each feature was then established by adding all agreement scores for this feature, and dividing it by the number of comparisons (the total number of annotator pairs in all instances). Using this value, we calculated Cohen's Kappa score (K):

$$K = \frac{P(a) - P(e)}{1 - P(e)}$$

where $P(a)$ is the relative observed agreement among the annotators as explained above, and $P(e)$ the hypothetical probability of chance agreement. The latter is calculated by adding for each feature value, the proportion of this value in the annotations of annotator A times the proportion of this value in the annotations of B. The Kappa scores for the twelve features are presented in Table 2.

Given the low Kappa scores for some features, we checked all cases where the annotators disagreed on the feature value. Moreover, for the features that had a Kappa below 0.5 (discourse givenness of the theme, pronominality of the theme, and semantic verb class), we checked the feature values for *all* instances. Given

⁵Of the 9 coordinations in recipients (5) and themes (4), only 4 contained elements with opposite characteristics (2 for both object types). Given this small number, we believe our rather crude approach is justified.

Table 2: Kappa scores (K) for inter-annotator agreement for each of the features (ordered by Kappa score)

Feature	K	Feature	K
Number of recipient	0.92	Semantic verb class	0.48
Animacy of recipient	0.87	Pronominality of theme	0.39
Person of recipient	0.86	Givenness of theme	0.32
Pronominality of recipient	0.80		
Definiteness of theme	0.78		
Definiteness of recipient	0.74		
Number of theme	0.71		
Concreteness of theme	0.55		
Givenness of recipient	0.55		

our thorough checks we believe the annotations are sufficiently accurate to proceed with them.

5 Regression models

5.1 Modelling technique and procedure

In order to gain insight into the way the features may simultaneously influence the choice for one of two dative constructions, Bresnan et al. (2007) successfully used a probabilistic modelling technique called logistic regression modelling. Since we want to compare our findings for the benefactive alternation to their findings for the dative alternation, we use the same technique⁶. This means we refrain from the use of feature selection procedures such as stepwise backward or forward selection⁷. Instead, we simply model the benefactive alternation as depending on the values assigned to the 13 features of the benefactive instances as described in the previous section.

Because logistic regression models can only handle binary or continuous variables, the values of all nominal variables were recoded so that they reflected the presence of one of the feature values (e.g. recipient=animate is 1), or the absence of that same feature value (recipient=animate is 0)⁸. The same was done for the construction used: if the construction was prepositional dative, it received value 1, if it was a double object construction, the value was 0. Apart from the nominal features, there was also one feature with a continuous scale: the length difference between the theme and the recipient. The feature is treated as such in regression.

⁶We use the function `lrm()` in the `Design` package in R (R Development Core Team 2008).

⁷Theijssen (2009) describes the effect of different feature selection procedures on logistic regression models for the British English dative alternation. She concludes that choosing the best procedure depends on ones goal. Since our goal is to compare it to Bresnan et al.'s work, we adopt their approach.

⁸Unlike the other nominal features, the feature 'semantic verb class' has four, not two, annotation values. Here, three values of this feature (all but *abstract*) are treated as a binary variable, which could again adopt the values 0 and 1. When all values were 0, the verb class was thus *abstract*.

The feature values and the construction used for all instances form the input matrix that is fed to the modelling technique.

In logistic regression modelling, a regression function is established that fits the data matrix best. It outputs the log of the odds that the construction used (c) is 1 (a prepositional dative construction) rather than 0 (a double object construction):

$$\text{logit}[P(c = 1)] = \alpha + \vec{\beta}X$$

where α is the model intercept, X are the feature values and $\vec{\beta}$ are the coefficients. The coefficients β can be understood as the weights assigned to the features by the model, where positive values favour the prepositional dative construction, and negative values the double object construction. The optimal values for α and $\vec{\beta}$ are estimated using Maximum Likelihood Estimation.

The regression models found can be employed for two main purposes. First, it enables us to establish which of the feature values have significantly affected the benefactive alternation, and which of the two benefactive constructions they favour. These observations can be compared to the effects found for the *to*-dative alternation (Bresnan et al. 2007). Second, the log odds outputted by the regression function can be used to predict which construction has been used by the writer. The percentage of correctly predicted instances can be used to evaluate the performance of the model.

The remainder of this section is structured as follows: The regression model built and evaluated on the adult data is presented and discussed in Section 5.2. Section 5.3 contains the results and discussion for the child data.

5.2 A model for the benefactive alternation in adult writing

Prediction

In order to evaluate how suitable the features used to explain the *to*-dative alternation are for explaining the benefactive alternation, we use them to establish a logistic regression model for the adult data. We do this by building the optimal regression model for the complete adult data set and applying it to exactly the same data. The percentage of instances that this model predicted correctly during the testing phase expresses its ‘model fit’ accuracy.

Our optimal logistic regression model, trained and tested on the 107 instances in adult writing, reaches a model fit accuracy of 86.9%. The rank coefficient between the predicted log odds and the construction used (Somers’ D_{xy}) is 0.88. These results indicate that even though the model fit accuracy is lower than the scores of 95% reported by Bresnan et al. (2007), the selected features can explain the benefactive variation in our data well enough.

In order to test how well the model can predict previously unseen data, we use 10-fold cross-validation. We build a model on 90% percent of the adult data, and then test it on the remaining 10% of the data. This is repeated ten times, in such a way that each data instance is tested once.

For the adult data, this procedure results in a prediction accuracy of 79.6%. This score is again remarkably lower than the prediction accuracy of 94% obtained by Bresnan et al. (2007). Still, the prediction accuracy we reach is significantly higher than the majority baseline of 70.1% ($p < 0.001$, McNemar test). The baseline reflects the proportion of instances that would be predicted correctly if always choosing the most frequent construction (being the prepositional dative). Given the significant improvement over the baseline, we believe it is justified to interpret the significance and coefficients of the features in order to compare them to those found for the *to*-dative alternation.

Feature relevance

As mentioned in 5.1, the coefficients that were estimated for the features in the optimal model express the strength and direction of the features' influences in the benefactive alternation. Positive coefficients favour the prepositional dative construction, negative ones the double object construction.

In our model the intercept has a significant effect: $\alpha=8.1$ (standard error 2.83, 95% confidence interval 2.41 – 13.75, $p < 0.005$). This implies that when all feature values are 0, there is a strong bias towards using the prepositional dative variant. In Bresnan et al. (2007), the intercept is only 0.95, but no information is given about its significance. The large difference between the intercepts for the benefactive and the *to*-dative alternation can be easily explained. In the benefactive alternation, the majority of the instances is prepositional dative (70.1% in our data set), while in the *to*-dative alternation, the majority has a double object construction (79%)⁹.

The significant β -coefficients in our model and those found by Bresnan et al. (2007) are given in Table 3. Only four features significantly affect the choice for the two benefactive constructions: semantic verb class, length difference, discourse givenness of theme and number of theme.

Clearly, there is a large difference between the number of significant effects found by Bresnan et al. (2007) for the *to*-dative alternation and the number we find for the benefactive alternation. This does not necessarily mean that the benefactive alternation is less complicated than the *to*-dative alternation. If we compare the number of instances in our data set (107) to that in Bresnan et al.'s data set (2360), it is not surprising that many features do not reach significance in our model. Apparently, the four effects listed in Table 3 are strong enough to be significant even in a data set as small as ours. The directions of these effects are the same as found in the model for the *to*-dative alternation (Bresnan et al. 2007)¹⁰, but their strengths differ.

The strongest effect towards the double object construction is found for the

⁹Apparently, the bias towards the double object construction in the *to*-dative alternation is completely accounted for by the β -coefficients, since the intercept favours the prepositional dative construction.

¹⁰One of the significant features, *semantic class = obtaining of possession*, is not present in the models for the *to*-dative alternation. Its direction and strength can therefore not be compared to effects in the *to*-dative alternation model.

Table 3: Coefficients (β) of significant effects in regression models ($p < 0.05$). The coefficients for the *to*-dative are copied from Bresnan et al. (2007), page 17. For our benefactive model, the standard errors (SE) and 95% confidence intervals (CI) are provided as well.

Effect	<i>dative</i>	<i>benefactive</i>			
	β	β	SE	CI	
theme = indefinite	-1.4				
semantic class = communication	-1.3				NP-NP
semantic class = obtaining of possession		-2.8	1.4	-5.7 – 0.0	↑
length difference (th-rec)	-1.2	-2.3	0.8	-3.9 – -0.7	
theme = nonpronominal	-1.2				
theme = discourse new	-1.1	-3.0	1.5	-6.1 – 0.0	
structure parallellism = yes	-1.1				
theme = inconcrete	-0.5				
recipient = nonlocal	0.5				
theme = plural	0.5	2.6	1.0	4.6 – 0.6	
recipient = indefinite	0.9				
semantic class = transfer of possession	1.0				
recipient = discourse new	1.0				
recipient = nonpronominal	1.2				NP-PP
recipient = inanimate	2.5				

feature *theme = discourse new*. This means that when the theme has not been mentioned previously in the text, the writer is expected to put it after the recipient, resulting in a double object construction. In the literature (e.g. Collins 1995), it is generally agreed that language users tend to place information that has already been introduced before information that is new to the discourse. The model we found, as well as the model found by Bresnan et al. (2007) for the *to*-dative alternation, shows the same effect.

Another strong effect is found for length difference. If the theme is longer (in words) than the recipient the probability that the construction with *for* is used decreases, while it increases if the recipient is longer than the theme. In other words, the longer object is generally placed at the end. This finding corresponds with the *principle of end weight* (Quirk et al. 1972).

A significant effect showing a preference for the prepositional dative construction is found when the theme is plural. The effect was also found for the *to*-dative alternation (Bresnan et al. 2007), though less strong ($\beta = 0.5$ compared to $\beta = 2.6$). In cross-linguistic studies, Haspelmath (2004) and Aissen (2003) predict that number plays a role in the syntactic realization of objects (the use of bound

object pronouns¹¹ and the case marking of objects), though in combination with pronominality and/or definiteness. Since our data set is too small to check these interactions, we can only conclude that the number of the theme is also relevant in the benefactive alternation in English.

The fourth significant effect concerns the semantic verb class. A comparison to the *to*-dative alternation is not possible here since it entails different semantic classes. For the benefactive alternation, we see that the obtaining of a possession is more often expressed in the double object construction than the other verb classes. Pinker (1989) argues that the double object construction will be selected especially if the subject (or agent) can cause the recipient to possess the theme. He states this is typically possible for verbs denoting creation or obtaining of the theme. In our data, the effect is only apparent for ‘obtaining of possession’.

5.3 Applying the adult model to the benefactive alternation in child writing

Now we have found a model for the benefactive alternation in adult writing, we want to compare it with the benefactive alternation in child writing.

However, during the manual annotations of the child data we noticed that many sentences (for instance example 14) seemed to have an adult writing level. This led us to suspect that the children had included benefactive sentences in their writing that were copied from, or strongly inspired by, texts produced by adults. We therefore checked the instances in our child data set, and found that almost one third of all identified data instances can be characterized as strongly ‘adult-like’ and might not be spontaneously produced by children.

- (14) They have a large keel-shaped Sternum to provide an anchorage for the large flight muscles –LCCPW AK53HRZ

Due to the small number of benefactive constructions in our data set and the large proportion of adult-like sentences, it is undesirable to build a separate model for the child data. We therefore decide to let the optimal adult model (built on all 107 adult instances) predict the child data instances.

When using the data instances produced by children as test material for the optimal model for the adult material, the model correctly predicts 80.6% of the instances. This is similar to that reached for adults in the 10-fold cross-validation setting (79.6%). The prediction accuracy is significantly higher than the majority baseline of 66.7% ($p < 0.001$, McNemar test).

Given the small data set, we cannot draw strong conclusions about the way children acquire the use of both syntactic constructions. The only thing we can say is that we have found no indications that there are differences between the use of the benefactive alternation in adult and child writing.

¹¹ A bound pronoun is a pronoun that is added to a word (e.g. a verb) as a suffix or prefix.

6 Conclusion

In this paper, we have studied the benefactive alternation in English writing. We followed the successful approach of Bresnan et al. (2007), which combines various existing theories in a single regression model to explain the *to*-dative alternation. We have researched: (1) whether the features used in Bresnan et al. (2007) are also suitable for predicting the benefactive alternation, and (2) whether the model found for adult language can be successfully applied to child data.

In order to answer these research questions, we extracted instances of the benefactive alternation from language written by American and British English native speakers. The instances were extracted automatically and were then manually checked. The final data set included only 107 instances produced by adults and 36 by children. Apparently, establishing a considerably sized data set for the benefactive alternation in a limited time frame is not possible. Also, we learned that the benefactive alternation itself is relatively infrequent (compared to the *to*-dative alternation).

The 143 instances were annotated with the desired semantic, discourse and syntactic information. Due to the complexity of the features, we had to check many of the annotations, making the annotation process very labour-intensive. With the resulting, qualitatively acceptable, data set, we could proceed with achieving our research goals.

First of all, we built a logistic regression model on the adult data. When applying it to the same set of adult instances, we reached a model fit accuracy of 86.9%, which is lower than the 95% reached by Bresnan et al. (2007), but significantly higher than the majority baseline (70.1%). We saw that the coefficients in the regression function for the benefactive alternation are similar to those for the *to*-dative alternation, but the model does not generalize as well (the prediction accuracy decreases to 79.6% in ten-fold cross-validation) and has fewer significant effects (only 4). With the current data set size we can only conclude that the use of the features for predicting the *to*-dative alternation for the prediction of the benefactive alternation seems promising.

For the child data, the set was so small that it was undesirable to build a regression model on it. Moreover, many of the instances in the texts written by children seemed to be copied from (or at least strongly inspired by) adult writing. Therefore, we decided to simply apply the adult model to the child data and compare the prediction accuracy to that found for the adult data. The accuracy reached (80.6%) significantly beat the majority baseline (66.7%), and was comparable to the 79.6% reached for adults in a ten-fold cross-validation setting. We thus have no reason to assume that the benefactive alternation in the writing of 8-to-12-year-olds differs from that in adult writing.

Apparently, in order to thoroughly study the benefactive alternation, more data is needed. We are currently developing an approach for automatically extending a data set for the *to*-dative alternation in written British English. If the approach proves successful, it can also be applied to the benefactive alternation.

References

- Aissen, Judith (2003), Differential object marking: Iconicity vs. economy, *Natural Language & Linguistic Theory* **21**, pp. 435–483.
- 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), pp. 245–259.
- Bresnan, Joan, Anna Cueni, Tatiana Nikitina, and R. Harald Baayen (2007), Predicting the Dative Alternation, in Bouma, Gerlof, Irene Kraemer, and Joost Zwarts, editors, *Cognitive Foundations of Interpretation*, Royal Netherlands Academy of Science, Amsterdam, pp. 69–94.
- Collins, Peter (1995), The indirect object construction in English: an informational approach, *Linguistics* **33**, pp. 35–49.
- De Marneffe, Marie-Catherine, Uriel Cohen Priva, Sander Lestrade, Scott Grimm, Gorkem Ozbek, Tyler Schnoebelen, Susannah Kirby, Misha Becker, Vivienne Fong, and Joan Bresnan (2007), A statistical model of grammatical choices in children's productions of dative sentences, Manuscript, Stanford University and Radboud University Nijmegen.
- Garretson, Gregory (2003), Coding manual for the project "optimal typology of determiner phrases", Unpublished manuscript, Boston University.
- Greenbaum, Sidney, editor (1996), *Comparing English Worldwide: The International Corpus of English*, Clarendon Press, Oxford.
- Gries, Stefan Th. and Anatol Stefanowitsch (2004), Extending Collostructional Analysis: A Corpus-based Perspective on 'Alternations', *International Journal of Corpus Linguistics* (9), pp. 97–129.
- Haspelmath, Martin (2004), Explaining the Ditransitive Person-Role Constraint: a usage-based account, *Constructions* (2/2004), pp. 1–71. Free online journal, University of Düsseldorf.
- Pinker, Steven (1989), *Learnability and Cognition: The Acquisition of Argument Structure*, The MIT Press, Cambridge.
- Quirk, Randolph, Sidney Greenbaum, Geoffrey Leech, and Jan Svartvik (1972), *A Grammar of Contemporary English*, Longman, London.
- R Development Core Team (2008), *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna. <http://www.R-project.org>.
- Sampson, Geoffrey (1995), *English for the Computer: The SUSANNE Corpus and Analytic Scheme*, Clarendon Press, Oxford.
- Sampson, Geoffrey (2003), The LUCY Corpus: Documentation, University of Sussex.
- Smith, Nicholas, Tony McEnery, and Roz Ivanic (1998), Issues in Transcribing a Corpus of Children's Handwritten Projects, *Literary and Linguistic Computing* **13** (4), pp. 217–225.
- Szmrecsanyi, Benedikt (2005), Language users as creatures of habit: A corpus-based analysis of persistence in spoken English, *Corpus Linguistics and Linguistic Theory* **1** (1), pp. 113–150.

- Theijssen, Daphne (2009), Comparing Logistic Regression models for the British English dative alternation, *Proceedings of the Student Session of the 21st European Summer School in Logic, Language and Information (ESSLLI)*, Bordeaux, France, pp. 85–95.