Towards a corpus-based identification of prototypical instances of constructions*

Stefan Th. Gries
Southern Denmark University

The notion of prototype is relevant to many analyses within cognitive linguistics. On the basis of a syntactic example (the so-called dative alternation), this analysis introduces a corpus-based technique for (i) the identification of prototypical instances of categories and (ii) the quantification of the degree of similarity between different category members. First, several desiderata for such a statistical prototype identification are motivated. Second, I propose a multifactorial technique to sort instances of constructions according to their degree of prototypicality. Third, I demonstrate that the resulting sorting of constructions is supported both by investigating the resulting typicality continuum and an in-depth experimental study of judgement data. Finally, I briefly discuss some implications and point to further applications of the proposed technique.

Keywords: prototype, grammatical variation, ditransitive

1. Introduction

1.1 Models of categorisation: The 'classical' theory vs. prototype theory

The ability to categorise, to classify elements into classes, is one of the most basic abilities of human beings. It does, therefore, not come as a surprise that this ability was investigated in a multitude of studies. Until about 30 years ago, the so-called classical theory was the central theory of categorisation. According to this theory (ultimately based on Aristotle's ideas), a category C is defined by a set of n necessary and sufficient criteria (or conditions $c_1, \ldots, c_n$), which often come in the form of binary attributes. Thus, an entity e is a member of C if and only if it exhibits all n criteria. It also follows that all members of C are equally good members of C because they all exhibit the required criteria c_1 to c_n; also, the entity

Annual Review of Cognitive Linguistics 1 (2003), **1–27.**
ISSN 1572–0268 / E-ISSN 1572–0276

either is or is not a member of C (depending on whether it exhibits the criteria c_1 to c_n or not) — a quasi interim status of entity e is not allowed for. Finally, note that the criteria c_1 to c_n are considered objective and context-independent.

Although the classical theory has some intuitive appeal, researchers in the 1970s started to mount evidence against, or at least problems of, this approach. In a by now classic series of experiments, Labov (1973) showed that subjects experienced difficulties to categorise particular containers as either cups or mugs. These difficulties were especially obvious when the visually-presented containers resembled both cups and mugs. This finding received further empirical support from psychological studies (especially from research by Rosch), yielding a series of so-called prototype effects contradicting predictions from / conclusions of the classical theory. For instance, given appropriate contextual motivation, subjects were able to extend the boundaries of categories quite flexibly (cf. Labov 1973). In addition, categories could overlap since there are objects which can belong to two categories that do not participate in a taxonomic relation. Finally, subjects judge some members of a category to be better examples of that category (a robin is a more typical bird that an ostrich or a penguin); these elements were often referred to as the prototypes of that category (i.e. its best, most typical and most representative members);[1] the prototypical members of a category exhibit the following characteristics: The prototypical members of a category can be identified faster than, say, peripheral members of the very same category (McCloskey and Glucksberg 1978); they can be identified faster than non-prototypical members (Rosch 1973; Rips, Shoben and Smith 1973; Armstrong, Gleitman and Gleitman 1983); they are more frequently named as category members than non-prototypical members (cf. conjoint frequency by Wilkins 1971; Rosch, Simpson and Miller 1976); finally, they are more often used as cognitive reference points and as a basis for inferences than non-prototypical members (Rips 1975).[2]

In what way, however, are prototypes defined — put differently, how can prototypical members be identified non-experimentally? There are several, ultimately related, answers to that question: a category member is typical insofar as

– it corresponds to the central tendency of the category as a whole (cf. Barsalou 1987: 104) — on the other hand, findings by Hampton (1981) and Barsalou (1985) lead us to suspect that the importance of the central tendency of a category is sometimes overestimated;
– it corresponds to the ideal of the category as a whole (Barsalou 1985, 1987: 105);

– it is similar to a cluster of attributes each of which has a high cue validity for this category (Rosch and Mervis 1975: 515f.); cue validity is defined as follows:

> [T]he validity of a cue is defined in terms of its total frequency within a category and its proportional frequency in that category relative to contrasting categories. Mathematically, cue validity has been defined as a conditional probability — specifically, the frequency of a cue being associated with the category in question divided by the total frequency of that cue over all relevant categories (Rosch and Mervis 1975: 515; cf. also Barsalou 1985, 1987: 105; Taylor 1995: 59f.; for the first mention, cf. Beach 1964)

– it exhibits the highest similarity to members of the same category and the lowest degree of similarity to members of other categories (Rosch and Mervis 1975: 598).[3]

Still though, according to Barsalou (1985: 105), there does not seem to be a single universally valid criterion for the prototypicality of a particular category member: "It is safe to say that there are many reasons why exemplars are typical and that no single factor or invariant set of factors is solely responsible", a statement to which we will return below.

1.2 Linguistic categories

While the psychological research mentioned above has mainly, though not exclusively, concentrated on concrete objects, colours etc., there was also converging evidence that linguistic categories display similar prototype effects. Nevertheless, there is one crucial difference with respect to the evidence pointing to the prototypicality of category members. Given the high inter-correlation of experimental results, one can identify prototypes fairly straightforwardly and directly, e.g. by asking subjects to simply name typical exemplars, to directly assess similarity between entities etc. — more problematic, since obviously less direct, is the analysis of linguistic categories where ratings of typicality or similarity are much more difficult to obtain. Naïve subjects will probably face difficulties to identify typical characteristics of English transitive constructions (a question asking for properties) or which of the following two questions is more typical of English preposition stranding as in (1) (a question asking for a similarity judgment).

(1) a. Who did you give the book to?
    b. Who did you see a picture of?

Therefore, (cognitive) linguistic research has pursued other strategies: In some early studies, Ross (1972, 1973a, b) has shown that the category NOUN exhibits prototype effects that manifest themselves in the applicability of syntactic distributional tests. The prototypical noun, e.g., can occur in many syntactic environments, but the environments of other, less typical nouns are more restricted. Further evidence for prototype effects within linguistics include Rice's (1987) study on transitive clauses; Bybee and Moder's analysis of strong verbs (1983) in English; cf. Lakoff (1987: chap. 3) and Taylor (1995) for more examples. Contrary to psychological research on natural kinds or artefacts, prototypicality is therefore determined more on the basis of qualitative criteria such as markedness, sequence of acquisition, cognitive simplicity etc.

While this kind of research has yielded interesting and far-reaching results, it is not without problems that need to be considered: First, the intersubjectivity of the results is more threatened than that of results obtained from dozens of subjects participating in standardised experimental settings (although this of course introduces the question of purely experimental effects).

Second, previous research has mainly focussed on the degree of typicality of exemplars or types, i.e. of subcategories of categories — for some research purposes it could, however, also be necessary to determine the degree of prototypicality of particular instances (i.e. category members, that is tokens); cf. below Section 4.

Third, the question may be posed whether the prototypes that are being determined on the basis of linguistic constructs represent the conceptual system of the speaker: what is reasonable within a given linguistic frame of reference should, prior to acceptance as fact, be validated in terms of natural (as opposed to elicited) behaviour of native speakers.

Finally, it can become very difficult (or even impossible) to measure the similarity between category members, which is problematic since the notion of similarity plays a vital role for determining prototypicality. Imagine an analysis of 15 transitive clauses with respect to the criteria c_1, …, c_{10}. Imagine further that there are two sentences of which each fulfils nine criteria, but the criteria the sentences fulfil are not identical, e.g. sentence A does not fulfil c_1 and sentence B does not fulfil c_{10}. Without a decision that can be motivated a priori, the decision as to (i) which sentence is (closer to) the prototype and (ii) how similar the sentences are to each other cannot be made.

2. Scope and methods of the study

2.1 Introduction

This study intends to show how one can (i) determine the degree of prototypicality of a category member and (ii) objectively determine the degree of similarity between category members on the basis of corpus data and multifactorial statistical techniques (cf. Tversky 1977: 344 on the interplay between similarity and categorisation). The scope of the study is, therefore, mainly methodological in nature, and the method to be introduced can be applied to various similar phenomena; I will introduce the technique on the basis of a syntactic example, namely the so-called dative alternation in English; examples include (2) and (3).

(2) a. John gave [$_{NP\_Goal}$ him] [$_{NP\_Pat}$ the book]. = ditransitive constr. (DK_{to})
b. John gave [$_{NP\_Pat}$ the book] [$_{PP}$ to [$_{NP\_Goal}$ him]]. = prepositional constr. (PK_{to})

(3) a. John made [$_{NP\_Ben}$ him] [$_{NP\_Pat}$ a cake]. = ditransitive constr. (DK_{for})
b. John made [$_{NP\_Pat}$ a cake] [$_{PP}$ for [$_{NP\_Ben}$ him]]. = prepositional constr. (PK_{for})

Given Ross's results mentioned above, the question arises how prototypical exemplars of the three constructions look like;[4] for the present paper, I will restrict my attention to the alternation between DK_{to} and PK_{to}. Since direct similarity or prototypicality judgments by native speakers of English are less likely to yield promising results (cf. above), one might pursue a (top-down) approach and define a prototypical construction on the level of exemplars and, from that, derive a prototypical instance. In this spirit, Goldberg argues that the prototypical ditransitive construction denotes the intended transfer of the referent of the direct object (NP_{Pat}) from the referent of the subject to the willingly receiving referent of the indirect object (NP_{Goal}) (cf. Goldberg 1992: 51, 56; 1995: 34ff., chap. 6).

This approach has provided insightful results on the more abstract level of grammatical theory, but it still suffers from the above-mentioned drawbacks: it is, e.g., not possible to apply the same technique to a set of constructions from a corpus and identify the prototypical instance of, say, the ditransitive construction. Also, once extensions of the prototypes are to be discussed (for instance in

Goldberg's form a metaphorical network), the proximity of both exemplars and instances to the postulated prototype cannot be determined objectively, although such considerations are obviously relevant to the characterisation of such networks that aim at capturing conceptually relevant distinctions, cf. further below. I would thus contend that an additional methodology is required. Let me start by pointing out some desiderata of the new methodology that can be derived from previous empirical results concerning categorization, prototypicality and similarity.

2.2 Desiderata

Whatever method is to be used for the identification of prototypical category members, it needs to tie in with previous empirical (psychological) results on categorization and similarity. To my mind, the following criteria are essential.

First, the category members that the method characterises as prototypical should exhibit the following characteristics:

- the prototypical members of each category (DK_{to} and PK_{to}) should correspond to the ideal of their respective category;
- the prototypical members of each category (DK_{to} and PK_{to}) should possess those attributes that have a high cue validity for their respective category; more precisely: the attributes that DK_{to} possesses should ideally occur frequently with co-members of DK_{to} and rather infrequently with PK_{to} and vice versa.

Second, a further desideratum concerning the quantification of similarity between members of a category and its prototype results from work by Rips, Shoben and Smith (1973), Smith, Shoben and Rips (1974), and Rosch and Mervis (1975). The latter argued (1975: 601f.) that the notion of family resemblance (that was explicitly equated with cue validity) serves as a foundation for the analysis of prototypicality as a central tendency within an attribute/similarity space. In experiment two of Rips, Shoben and Smith (1973), it is demonstrated that subjects provide so-called semantic distance judgements for the distances between attributes between members of the categories Bird and Mammal.

For our purposes, two important conclusions can be drawn from this experiment: on the one hand, the rated semantic distances obtained correlate highly and significantly with judgments of prototypicality provided by Heider ($r=.89$; $p<.01$; Rips, Shoben and Smith 1973: 9); on the other hand, they show how one can, with multidimensional scaling techniques, transform the semantic distance judgements into Euclidean distances (so-called derived semantic

distances) between category members. In other words, the result of the scaling procedure is a geometric similarity model where the similarity between e.g. two category members follows from their distance in the n-dimensional space resulting from the analysis. Thus, a third desideratum is that the method by means of which prototypical instances are identified should enable us to determine the similarity between members (and, thus, also the similarity of members to their category's prototype) in terms of their distance in an n-dimensional space.[5]

Finally, the probably most essential desideratum, that of validity: The categories and prototypes determined by the method must not contradict elicited native speaker intuition or, even more importantly, natural native speaker usage. What does that mean? Well, it means that, given a particular parameter setting, most multidimensional statistical techniques can generate a statistically reasonable model and/or space — what is not guaranteed, however, is that the model/space is compatible with linguistic/psychological structures of native speakers. Nothing is gained by developing a mathematically reasonable model the relations of which to what speakers think and do are opaque, if existent at all. But how do we meet all these desiderata in the analysis of the dative alternation? This question is addressed in the following section.

2.3 Variables and data

One central problem was only briefly dealt with so far: while the psychological results mentioned above are all based on experimental settings, it is difficult, though not impossible, to apply similar techniques to the analysis of syntactic constructions. Assuming that native speakers have acquired/formed usage-based categories for each of the two constructions, I advocate the use of corpus data. In other words, while subjects in the psychological experiments went through a decision process yielding a similarity judgement or a (more or less) important feature of a category, in the present approach I assume that speakers go through an unconscious decision process resulting in the production of one out of two possible constructions. This procedure has as a positive side-effect that the data to be analysed are from natural rather than experimental settings, virtually ruling out the possibility of experimentally introduced interference or biases.

But which attributes can be utilised for the process of categorization and the identification of prototypical constructions in the first place? The answer to this question is obvious, once the question is rephrased: which attributes support the choice of the ditransitive construction and which support the choice of the

prepositional construction? For this, one can refer to a fairly large number of linguistic studies resulting in a variety of variables that determine (i) whether the dative alternation is possible at all or not and/or (ii), if both constructions are possible, which construction is actually chosen. The latter variables include attributes of the NPs involved in the two constructions, of which all those amenable to straightforward corpus analysis will be included into the analysis; cf. Table 1 and also Section 3.2 below.

In order to base one's findings on a relatively representative corpus, one should first determine the most frequent verbs allowing the dative alternation in the first place. To my knowledge, however, there is no empirical/statistical study providing such figures, which is why these figures had to be determined approximatively. I extracted from both previous literature and a variety of dictionaries all verbs that were claimed to allow the dative alternation; the total number of verbs listed was 586. However, since the vast majority of previous studies were based on introspective evidence alone and since dialectal variation is likely to occur, it comes as no surprise that for many verbs opinions as to dativisation differed. Therefore, in a second step, I included only those verbs which were claimed to allow both constructions unanimously by more than one source; 128 verbs survived this step. Finally, of this 128 verbs I determined the 10 most frequent ones, which were then extracted from files of the British National Corpus (BNC version 1.0).[6]

This study forms part of a larger project on grammatical variation in English and, as yet, the data base is limited. In this study, I analyse 60 cases of DK_{to} (51.28%) and 57 cases of PK_{to} (48.72%), and all variables mentioned in Table 1 were hand-coded as follows:

- the process described: intended transfer of NP_{Pat} from NP_{Subj} to NP_{Goal} or not;
- the animacy of NP_{Goal} and NP_{Pat} and their lengths in syllables and words;
- the distance to last mention of the referents of NP_{Goal} and NP_{Pat} in the 10 preceding clauses;
- the times of preceding mention of NP_{Goal} and NP_{Pat} in the 10 preceding clauses;
- the determiner of NP_{Goal} and NP_{Pat}: none vs. indefinite (*a, an, some* etc.) vs. definite (*the, their, my, all* etc.);
- the kind of NP: pronominal (*I, him*, etc.) vs. lexical (*the car, flowers*, etc.) vs. proper name (Mr Heseltine etc.).[7]

Table 1. Variables and their effect on the choice of construction

Value for DK_{to}	Variable (source/reference)	Value for PK_{to}
NP_{Subj} transfers NP_{Pat} to NP_{Goal}	Process described by the utterance (Goldberg 1992: 51, 56; 1995: 34ff.)	
animate/ human	animacy of the referent of NP_{Goal} (Green 1974: 103; Ransom 1979: 215f., 225; Welte 1979: 297; Thompson 1990: 247)	not/less animate/ human
inanimate/ non-human	animacy of the referent of NP_{Pat} (Ransom 1979: 225; Thompson 1990: 243)	
short	Length of NP_{Goal} (Thompson 1990: 246; Hawkins 1994: 213ff.; Arnold et al. 2000: 36ff.)	long
long	Length of NP_{Pat} (Thompson 1990: 246; Hawkins 1994: 213ff.; Arnold et al. 2000: 36ff.)	short
definite	Determiner of NP_{Goal} (Ransom 1979: 221)	indefinite
pronominal	Kind of NP of NP_{Goal} (Mazurkewhich and White 1984: 267 n. 8; Thompson 1990: 244)	lexical
indefinite	Determiner of NP_{Pat} (Ransom 1979: 219; Allerton 1978: 24, 28)	definite
lexical	Kind of NP of NP_{Pat} (Thomson and Martinet 1988: 77; Thompson 1990: 244)	pronominal
given	Newness of the referent of NP_{Goal} (Givón 1984: 154; Thompson 1990: 245f.; Panther 1997: 110)	new
new	Newness of the referent of NP_{Pat} (Givón 1984: 154; given Thompson 1990: 245f.; Panther 1997: 110)	

Then, a linear discriminant analysis (LDA) was computed; the choice of construction was the dependent variable and as prior probabilities of the two constructions I entered the constructional probabilities in the corpus.[8] As output, we obtain an equation with a factor loading for each independent variable and a discriminant score for each sentence:

– if a sentence exhibits mainly those variable values that have a high cue validity for DK_{to} / PK_{to}, a high (positive) / low (negative) discriminant score will result;
– if a sentence mainly exhibits variable values supporting different constructional preferences, a discriminant score around zero will result.

The computation of discriminant scores satisfies our second desideratum. If one sorts the sentences in order of their discriminant scores, we obtain a continuum between two extreme points. The sentences represented by the two most extreme points are the ideals of the two categories made up by the sets of positive values and the set of negative values and, thus, represent the prototypes of the two constructions: they exhibit exactly those characteristics that have a high cue validity for the construction they instantiate. Thus, the first desideratum is also met. Note, however, that the researcher's ability to mathematically model the constructional choice must not be taken to imply that native speakers perform such analyses online, but still the correspondence between the mathematical results and the cognitive processes is too high to be simply ignored (cf. also Gries 2001, 2003a): the precision of the LDA can be checked on the basis of the constructional decisions of the native speakers, we can at the same time determine which of the independent variables in Table 1 are most important for the constructional choice of the native speaker, and we can generalise from the sample to hitherto unanalysed instances. These points and additional ones will be the topic of the following section.

3. Results and discussion

3.1 Identification of prototypes

The LDA shows that the independent variables do indeed discriminate well between the two constructions, i.e. they predict the constructional choice well (canonical R=.821; $\chi^2(30)=112.12$; $p<.001$ ***). From the values of each variable of each sentence, the LDA computes how important each variable is for the discrimination between the two constructions (expressing that in a factor loading) and computes the discriminant score. When the discriminant score is smaller than an empirically a posteriori determined threshold value, then the LDA predicts that the speaker will produce PK_{to} — if, on the other hand, the discriminant score is higher than the threshold value, the LDA predicts that the speaker will utter DK_{to}. On the whole, an exceptional result is obtained: the LDA predicts 88.9% of all cases (=104 out of 117) correctly; consider Table 2.

Table 2. Classification matrix: Absolute frequencies and row percentages

		Predicted: DK_{to}		Predicted: PK_{to}	Row totals
Observed: DK_{to}	①	50 (83.3%)	②	10 (16.7%)	60 (100%)
Observed: PK_{to}	③	3 (5.3%)	④	54 (94.7%)	57 (100%)

While the number of miscategorised instances of DK_{to} (cell ②) clearly outnumbers the number of miscategorised instances of PK_{to} (cell ③), this difference is, according to a two-tailed binomial test, not significant (p=.094 ns).[9] Cells ① and ④, on the other hand, contain the frequencies/percentages of the constructional choices predicted correctly: 83.3% of DK_{to} and 94.7% of PK_{to} were identified correctly, given the variables' weight in the analysis. This is especially important since it provides evidence that our fourth desideratum is also met: the LDA did not just result in some mathematically optimal model, but the model makes it possible to predict native speakers' behaviour with a precision seldom found in multifactorial analyses in the behavioural sciences. Finally, note that the success rate is similarly high for a priori predictions: Following Werner (1997: Section 3.8), 117 LDAs were computed where in each LDA the constructional choice of one sentence is predicted on the basis of all other sentences; in this way, no sentence figures in its own prediction (a cross-validation technique called 'leave-one-out method'), and the success rate achieved is still very high, namely 82.1% (=96 out of 117).[10]

Once we know that the LDA ties in with native speaker behaviour, it is for our main purpose equally important to look at the discriminant scores of individual sentences. Figure 1 shows the sorted discriminant scores relative to whether the prediction based on them is correct or false; each dot represents the sentence with the corresponding score.

How do these results bear on a prototype-based definition of categories? First, we see that the absolute size of the discriminant scores correlates negatively with the number of misclassifications: misclassifications occur only in the area (-1.62≤discriminant score≤.92); most of the misclassifications (10 out of 13) are even located in the much smaller area of (-.53≤discriminant score≤.32). Thus, PK_{to} and DK_{to} differ so strongly that they do instantiate two different constructions (each possessing highly characteristic attributes) that can be distinguished well along the lines of previous research. Moreover, the results conform to what was to be expected on a prototype-based approach. According to prototype theory, one would expect that the more a category member

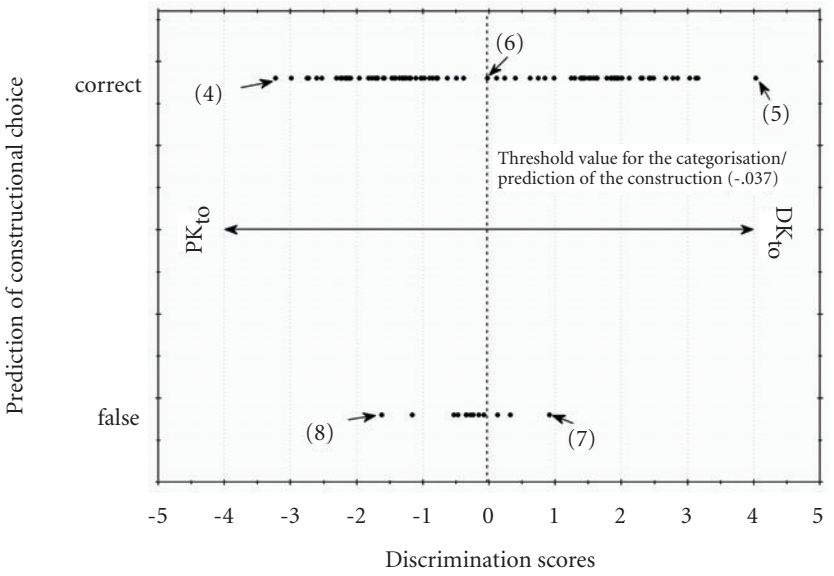

Figure 1. Discriminant scores relative to prediction accuracy.

possesses the high-cue-validity attributes of its category, the more prototypical it is for its category and, thus, the more likely it is that this instance is recognised as a member of the category it instantiates. Accordingly, once again, (cf. Section 2.1) the two sentences with the most extreme discriminant scores ((4) and (5)) are the most typical (namely ideal) representatives of their category.

(4) Jean Floud, A. H. Halsey, and F. M. Martin gave a new impetus both to the study of these themes and to action upon them (discr. score=-3.22)

(5) going round beer festivals gave me the idea of doing it for a living (discr. score=4.03)

From what we know about native speaker usage, example (4) is indeed a very representative member of PK_{to}: the sentence does not denote the prototypical transfer expected for DK_{to} (although the verb *give* is used), NP_{Pat} is short, NP_{Goal} is very long and complex and not a typical (i.e. willing and conscious) recipient — only the nature of the determiners does not fit previous findings: the determiner of NP_{Pat} is indefinite and that of NP_{Goal} is definite, which does not tie in to the information distribution (given>new, where ">" means precedes) that has been associated with the two constructions. Nevertheless, I believe that the status of (4) as a very typical instance of PK_{to} is beyond dispute,

an assessment which is corroborated once one reformulates (4) as DK_{to} and obtains a somewhat awkward sentence: *Jean Floud, A. H. Halsey, and F. M. Martin gave both the study of these themes and action upon them a new impetus.*

With (5), the situation is similar: most variables exhibit exactly those values that, according to previous studies, are to be expected from the prototype of DK_{to}: NP_{Goal} is short, pronominal, human (and, thus, a typical recipient) and given information whereas NP_{Pat} is comparatively long and complex. The sentence again (as with (4)) denotes only a metaphorical kind of transfer, but still its meaning is close to that of prototypical transfer since, e.g., after having heard (5) you can say *so then I had the idea* (a similar test does not work with (4), supporting the above analysis). The determiner of NP_{Pat} is definite, which superficially contradicts our expectation, but the NP's referent is still new — the definite determiner is possible here because the referent is further elaborated by the following PP; that is to say, the expected structure of given>new is confirmed. Finally, the reformulation of (5) as PK_{to} results in a less natural utterance: *going round beer festivals gave the idea of doing it for a living to me.*

Thus, the analysis has indeed picked out instances that approximate our intuitive (qualitative, non-empirical) understanding of the prototypical instance quite well, thereby providing further strong evidence that the analysis meets the standard imposed by desideratum four and, at the same time, integrates previous findings well. But let us still look at some other salient constructions marked in Figure 1.

3.2 Further salient examples

(6) represents a sentence where the discriminant score is extremely close to the threshold value of -.037 determining the choice of construction.

(6) The Iran/Contra scandal gave the French much cause for amusement (discr. score=-.02)

More precisely, for (6) the discriminant score is slightly higher than the threshold value, which is why the analysis (correctly) predicted the choice of DK_{to} — however, the proximity to the threshold value suggests that the speaker might nearly equally well also have chosen PK_{to}, an assessment that turns out to be supported by a closer look at the data. For instance, the utterance does not denote a case of prototypical transfer, but rather a somewhat intermediate case of metaphorical transfer: while NP_{Goal} corresponds to the expected value for DK_{to} by being a human recipient (though not a single individual), NP_{Subj},

which should be animate for a prototypical case, is inanimate (more precisely, an event); also, NP_{Pat} is abstract. Thus, as expected from the analysis, the semantic variables prefer PK_{to} since the sentence does not refer to prototypical transfer in the sense of Goldberg. The discourse-functional variables (and, under their influence, the morphosyntactic variables) tend toward DK_{to}: NP_{Goal} is short (two words and syllables) and was mentioned three times in the preceding ten clauses (for expository reasons not shown above). NP_{Pat}, by contrast, is much longer (four words and six syllables) and was not mentioned before at all. In this case, different preferences clash so the decision for one construction over the other is not a clear case as in the prototypical cases discussed above. The important point is, thus, the other construction is nearly equally possible, especially in the presence of certain communicative intentions: *The Iran/Contra scandal gave much cause for amusement to the French.*

Let us finally turn to (7) and (8). These sentences are noteworthy since they were misclassified such that, although they possess many attributes of one construction, the other construction was chosen by the speaker. For (7), the analysis predicted DK_{to} with a very high probability (posterior p for DK_{to}=.941), but still the speaker decided to produce PK_{to}. For (8), on the other hand, the analysis predicted PK_{to} (posterior p for PK_{to}=.989), but the speaker preferred DK_{to}.

(7) Fans wrote letters to the band (discr. score=.92)

(8) Marco van Basten, the European Footballer of the Year, gave European champions Milan a timely boost (discr. score=-1.62)

While the analysis has failed to predict these constructions correctly, these misclassifications nevertheless support the approach. How can that be? More to the point, when even wrong predictions are counted as somehow supporting the analysis, how can the approach be falsified at all? The answer is straightforward: the analysis could be falsified by

– an insignificant discrimination between the constructions given the above variables;
– a large number of misclassifications;
– a strong interpretive mismatch between the categories the analysis creates and the category usage we actually find or a completely unmotivated category structure.

The first two possible ways of falsification do not play a role here, given the above results. As to the third point, (7) and (8) are still not incompatible with

the proposed analysis since, even though the predictions are false, they are not completely unmotivated. As for (7), the prediction of DK_{to} by the analysis results from the fact that, while the morphosyntactic variables do not prefer either construction (both NP_{Goal} and NP_{Pat} are equally long (two syllables) and lexical), the semantic characteristics of the utterance included in the analysis support DK_{to} because (7) denotes nearly prototypical transfer (and, a variable not included in the present analysis, the direct object is effected, not affected; cf. Hawkins 1981: 1f.). Add to this that that discourse-functional variables rather prefer DK_{to}: the referent of NP_{Goal} was mentioned four times in the ten preceding clauses (the last mention was in the immediately preceding clause) — the referent of NP_{Pat}, by contrast, is discourse-new so, given a default preference of given > new we would expect $NP_{Goal} > NP_{Pat}$, i.e. DK_{to}. But how do we explain the native speaker's choice of the other construction? A possible explanation is based on Thompson and Koide's (1987) account of the dative alternation in terms of (conceptual) distance, a variable hat was not included here because of the difficulty to operationalise it objectively. According to Thompson and Koide, the conceptual distance between NP_{Pat} and NP_{Goal} influences the choice of construction such that smaller conceptual distance correlates positively with smaller linguistic distance,. i.e. DK_{to}. Conceptual distance in turn derives from a variety of factors, two of which may be at play here, namely geographical distance and the degree to which the action denoted by the verb impinges on the referent of NP_{Goal}. In this case, the geographical distance is probably quite large (at any rate, it is larger than the distance in sentences such as *John gave Mary a book*) and the degree of impingement on NP_{Goal} is left open since reception of the letters is not entailed or presupposed (again, contrary to what we find for *John gave Mary a book*). Thus, it seems as if (part of) the speaker's motivation to not choose the predicted construction can be explained with reference to an additional cognitively salient variable that has, for methodological reasons alone, not been included so far.

In the case of (8), the morphosyntactic variables alone lead one to expect PK_{to}: NP_{Pat} and NP_{Goal} are four syllables and seven syllables long respectively, so the short > long tendency that is so strong in English (cf., e.g., Hawkins 1994) should have resulted in PK_{to}. A strong semantic preference does not exist since we have seen above that metaphorical transfer can occur in both constructions. Also, no preference from the discourse-functional variables can be inferred from the context since both NP_{Goal} and NP_{Pat} are discourse-new. That is to say, although there are no semantic and discourse-functional preferences and a morphosyntactic preference for PK_{to}, the speaker unexpectedly chose DK_{to}. But

again we see that, although the prediction is wrong in this particular case, it is not arbitrarily wrong: First, it conforms to what native speakers or linguists would expect on the basis of what they know about other dative alternation cases; in other words, the choice of this particular speaker is, at least given the variables included in the present analysis, somewhat unexpected, given how other native speakers decided in other cases. Second, the construction predicted by the analyses (cf. (9)) is perfectly acceptable.

(9) Marco van Basten, the European Footballer of the Year, gave a timely boost to European champions Milan

While I could propose a plausible motivation for the misclassification in the case of (7), the erroneous prediction of (8) cannot be resolved so easily. We need to bear in mind, however, that the length of the discussion of these examples does not reflect their frequency: there is only a very small number of cases displaying these explanatory difficulties since 88.9% cases were predicted correctly. Also, it is highly unlikely that we will ever be able to predict native speakers' behaviour completely flawlessly irrespective of the number of variables we might still want to include in the analysis. For instance, possible candidates for variables other than affectum-vs.-effectum direct objects and conceptual distance that might prove valuable for the prediction of the dative alternation are structural priming (cf., e.g., Bock 1986; Gries 2003a) and rhythmic alternation preferences (Gries 2003b).

3.3 Experimental validation of the results

Finally, I would like to draw attention to a very important (methodological) point. In the last section, I repeatedly argued what (kind of) sentence native speakers would find more likely or would accept more than others. Since I am not even a native speaker of English, the reader might ask what evidence other than my argumentative desires supports the claims made above. Moreover, even if I were a native speaker of English, an empirically responsible audience would probably hope for more 'evidence' than my own intuition even if this kind if evidence is unfortunately considered sufficient in several contemporary theories. Since it would be highly unfortunate to, on the one hand, advocate more objective and reliable techniques and, on the other hand, fall back on techniques whose drawbacks we are all too familiar with, I conducted a questionnaire experiment in order to back up my claims as to which sentences are supposedly more acceptable in which constructions. This was done as follows.

Twelve sentences (four with high, low and intermediate discriminant scores respectively) were chosen from the corpus data. In order to be able to test for constructional preferences, each sentence was converted to the construction it did not instantiate in the corpus data, yielding a total of 24 experimental sentences. For each of these experimental sentences, I formulated a context sentence to simulate the degree of givenness of the referents from the context preceding the experimental sentence (or its converted counterpart) in the corpus data. For instance, NP_{Goal} and NP_{Pat} of sentence (6) (repeated here for ease of reference as (10)) were highly given and discourse-new respectively (cf. above). Since this information contributed to the constructional choice of the native speaker and the corresponding discriminant score, this information needs to be provided in the elicitation of acceptability judgements, too. Thus, the experimental sentence (6) was preceded by the context sentence given in (11), on the basis of which, again, NP_{Goal} is given and NP_{Pat} is not.

(10) The Iran/Contra scandal gave [NP_{Goal} the French] [NP_{Pat} much cause for amusement].

(11) The French government claimed it would never negotiate with terrorists.

Various controls discussed in the relevant literature (e.g. Schütze 1996 and Cowart 1997) were implemented in order to guarantee valid and representative results. The experimental stimulus sentences were added to a list of several hundred filler items (experimental items with varying though balanced degrees of acceptability from other ongoing projects). The items were also ordered pseudo-randomly such that no subject judged more than one sentence from each token set (using Cowart's 1997: 48f. terminology). 36 native speakers of British English received questionnaires with 32 sentences each, where the judgement process was explained and exemplified; the examples anchored only the endpoints of the rating scale. The subjects were then asked to provide acceptability judgements ranging from -3 (strange/unnatural) to +3 (natural English/perfect). '0' was considered the middle of -3 and +3 and the subjects were additionally offered the opportunity to answer 'I don't know.' On the results, a two-way ANOVA (construction × discriminant score) was computed. According to my claims from above we would expect a significant interaction of these two independent variables as represented in Table 3.

The statistical results are clear: Not only is the two-way interaction very significant ($F_{2, 173}=17.56$; $p<.001$) — every prediction of Table 3 is supported as can be seen in Figure 2.

Table 3. Factorial design of my questionnaire study and the predictions

statistical prediction of preferred syntactic structure	high discriminant score: DK_{to}		intermediate discriminant score: DK_{to} or PK_{to}		low discriminant score: PK_{to}	
structure of the stimulus sentence to be judged	DK_{to}	PK_{to}	DK_{to}	PK_{to}	DK_{to}	PK_{to}
predicted degree of acceptability	positive	negative	positive	positive	negative	positive

When the discriminant score based on the corpus data is high, then the analysis considers the sentence to be highly typical of DK_{to}, which is reflected exactly by speakers finding DK_{to} acceptable and PK_{to} unacceptable. Conversely, when the discriminant score is low, then the analysis considers the sentence to be highly typical of PK_{to}, which is again reflected by the speakers finding PK_{to} acceptable and DK_{to} unacceptable. Lastly, when the discriminant score is intermediate (i.e. about 0), then the analysis considers the sentence to be a borderline case where both constructions are possible; this result is borne out by the acceptability judgements as well.

Given these results, there is little doubt that the LDA is an appropriate way to identify prototypical instances of linguistic categories (in this case, syntactic

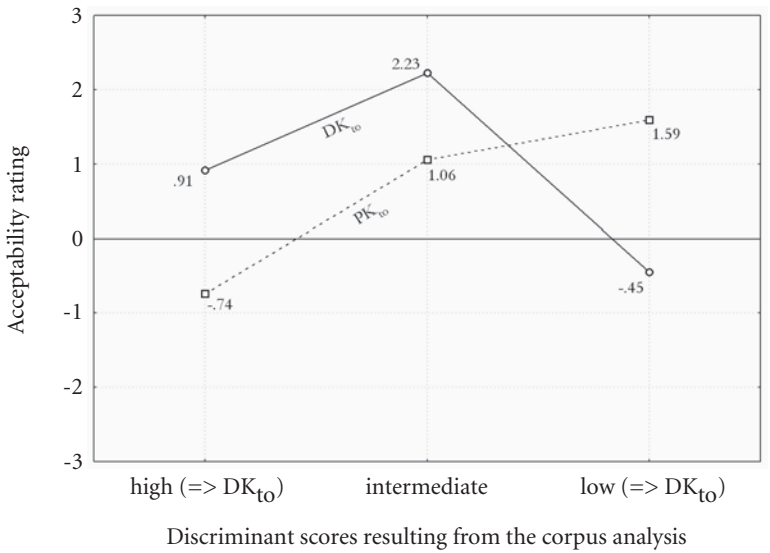

Figure 2. Mean acceptability relative to corpus prediction and constructional choice.

constructions): the overall results of the statistical analysis and the arguments based thereupon tie in with (i) previous psychological and linguistic accounts (actually, making the latter much more precise), (ii) actual usage by native speakers and (iii) elicited acceptability judgements from native speakers.

3.4 The importance of the variables for prediction and category cohesion

In the psychological studies hinted at above, multidimensional scaling techniques were used to define an n-dimensional attribute space whose dimensions could be interpreted meaningfully; for instance, Rips, Shoben and Smith's (1973:9f.) analysis of similarities between different mammals yields two dimensions which can be reasonably interpreted as the animals' size and predacity. The present analysis does not allow for exactly that sort of dimensional interpretation (for that, a factor analysis would be necessary), but it enables us to determine exactly each variable's importance for the discrimination between the variables. Put differently, from each variable's weight, we can (i) infer exactly how important it is for speakers' subconscious online decisions for constructions and (ii) evaluate previous findings: is a variable really as important as a previous analysis had made us believe? While this is not at the heart of the present question, I nevertheless believe that this opens up interesting possibilities for the empirical and objective confirmation (or falsification, for that matter) of previous work on the basis of authentic data. Given lack of space, let me just mention that the variables having the highest discriminatory power are (i) all properties of NP_{Goal} and (ii) morphosyntactic variables in general — other variables, e.g. properties of NP_{Subj}, are less decisive.

4. Summary and conclusions/implications

The previous sections have demonstrated how we can identify the prototypical instances of two syntactic categories from corpus data. This quantitative approach is based on results from experimental psychology and aims to bridge the gap between cognitive linguistics, where constructions and prototypes are postulated and investigated on a more theoretical basis, and psychological approaches where precise quantitative measures of the (degree of) prototypicality of category members and similarity have already been developed a long time ago. The results of the application of such concepts to the instance/token level has provided results that are indeed highly compatible with previous

linguistic and psychological findings; concepts such as cue validity, family resemblance and prototypes as ideal representatives of a category are supported.

At first sight, the results might seem to be just a replication of existing claims — at a second glance, however, I hope to have shown the potential of this approach as well as some interesting and more general (methodological) implications, which I would like to summarise briefly below.

First, the identification of prototypes in corpus data with an LDA constitutes an additional technique which improves upon previous ways of prototype identification: on the one hand, it is based on authentic data only; there is neither the danger of an unconscious bias on the part of the investigating linguist nor can experimental artefacts distort the results. All that is needed is a linguist who can identify variables from previous research and a (sufficiently representative) corpus. On the other hand, given the quantitative make-up of the present approach, intuitive and vague assessments of what is somehow more prototypical or more similar to some standard of comparison can be replaced by precise statements. Finally, note that the multifactorial analysis of corpus data yields results comparable (but much more precise) to the traditional linguistic methodology of acceptability judgements (once the latter are obtained in scientifically/experimentally reasonable ways). Thus, while some researchers have frequently argued against corpus analyses (for reasons concerned with, among other things, the finiteness of corpora and the difficulties of obtaining representative corpora), this analysis shows that corpus data and judgement data can complement each other naturally (cf. for a similar result Gries 2002).

Second, a previous point of critique concerning such quantitative analyses involving similarity and prototypicality does not apply here. The approach by Tversky (1977) was criticised for not being able to include cognitively complex factors (cf. Taylor 1995: 61). For instance, Taylor's prototype-based approach of possessive constructions involves a prototype of possession defined by eight criteria (1995: 202ff.). In the present approach, there are several possibilities to also integrate such complex concepts: the eight variables are all individually coded and/or one constructs a single variable encoding whether all eight criteria are satisfied or not and/or one constructs an additive index out of the individual variables, representing the degree of prototypicality.

Third, the present approach enables us to test whether the patterns one predicts to find on linguistic grounds alone do in fact exist, once actual usage is taken into consideration; put differently, do instances really fall in the predicted exemplar categories? Consider, for instance, network analyses of syntactic, semantic/lexicographic or pragmatic phenomena such as Goldberg's analysis of

the ditransitive construction, network analyses of polysemous lexemes and Thornburg and Panther's (1997) analysis of speech-act constructions. When extensions from prototypes (however to be determined) are involved, it is possible to (i) reliably test to what extent the extensions postulated do indeed manifest themselves predictably in the data (in the way DK_{to} and PK_{to} can be predicted here) and (ii) determine which extensions are more remote from the prototype and which are not, in order to

– provide a more adequate account of the internal structure of a category;
– motivate why some extensions are possible whereas others are not, given their (objectively determined) distance to the prototype;
– explain the diachronic development of (constructional) categories on the basis of corpus data by showing how prototypes change over time and, thus, license different extensions and developments;
– explain speed/ease of acquisition of category members or extensions etc.[11]

Note, that, e.g., the acquisition of DK_{to} has been the focus of a variety of studies, given that it is so difficult to account for how children learn to use DK_{to} in the right circumstances only. Since it is well known that prototypes are very much relevant to the formation of categories (cf. Rosch and Mervis 1975), a possibly interesting strategy for further research on this as yet debated issue would be to investigate different levels of acquisition as to (i) which constructions can be identified as prototypical (and, thus, as constitutive of the syntactic category being acquired) and (ii) which variables are most influential for prototype formation and category extension.

A fourth and final advantage of the way of analysis proposed is that, apart from the hypothesis-testing approach described so far, it can also be used for exploratory data analysis. One could start an analysis of some alternation phenomenon by identifying the ideal cases of the alternatives where a lot of different variables, so to say, converge and investigate the attributes forming these categories. Also, one can immediately determine which variables are important for the proposed classification and which are not. Finally, the misclassified constructions deserve special attention since they may point to either exceptions from the general rule or aspects of the phenomenon that have hitherto escaped one's notice.

Let me finally address a possible comment. The conceptual core of the present study is based on a variety of experimental results on studies of similarity (cf. especially Tversky 1977, Tversky and Gati 1978, Gati and Tversky 1982) and prototype theory; a major role played, of course, results by Rosch and

further studies on, among other things, similarity based on her work. Frequently, Rosch's results were (at first by herself, too; Lakoff (1987: 42–45) and Kleiber (1993: 29 and passim)) interpreted as if they were statements on the structure of mental representations as such; cf. the *effects=structure fallacy* and the *prototype=representation fallacy*. I do not wish to support such interpretations. The discriminant scores used for the prediction of the constructional choice are not meant to be interpreted as reflecting a graded structure of the two syntactic categories. Put differently, the fact that one instance of DK_{to} has a higher discriminant score than another instance of DK_{to} does not entail that the former is more of a member of DK_{to}. Still, even if the form of analysis does not translate into statements on mental representations, the high predictive power (measured in terms of the obtained success rates) shows that the cognitive factors underlying the choice of construction have been identified properly and weighted in accordance with their importance for actual usage. Thus, given the four advantages enumerated above, I believe that the combination of qualitative and corpus-based quantitative approaches in the present study makes it possible to gain further insights into processes of categorisation and production.

Notes

* I would like to express my thanks to Constanze Bühner of Southern Denmark University for her assistance in hand-coding a part of the data. Also, I would like to thank (in alphabetical order) Hans Boas (University of Texas at Austin), Barbara Lohse (University of Southern California) and Debra Ziegeler (University of Manchester) for their help in obtaining the judgement data (by distributing questionnaires). Without the kind assistance of all of these people, the huge amount of data necessary for this study could not have been obtained in time. Finally, I would like to thank some participants of CSDL 2000 at UCSB and the anonymous referees for their comments. Naturally, all remaining inadequacies are my own.

1. By this I do not mean to imply that prototypes are the mental representations of the best exemplar; i.e. the prototype of birds of a particular conceptualiser is not the mental image of a particular robin (cf. below).

2. It is interesting to note in passing that this kind of prototype effects also occurred with classical categories, i.e. categories conforming to the axioms of the classical theory: subjects were able to identify prototypical prime numbers (Wanner 1979) and odd numbers (Armstrong, Gleitman and Gleitman 1983).

3. Rosch and Mervis (1975: 575) define the degree of family resemblance between two category members in terms of the number of shared attributes; thus, this criterion is intimately related to cue validity as well.

4. The three constructions I am referring to are the ditransitive construction, the prepositional construction with *to* and the prepositional construction with *for*.

5. The line of reasoning behind this desideratum was criticised in a well-known study by Tversky. In Tversky's (1977) model, (i) the decision of whether to classify an entity e as a member of the category C or not and (ii) the similarity of e to other members of the same category is based on two aspects: first, one counts both the attributes of e that are also attributes of C and the attributes of e that are not attributes of C. Second, the properties involved are weighted according to their salience and their diagnostic value. That is to say, an entity e is identified as prototypical for the category C when it exhibits as many highly weighted attributes of C and as few highly weighted attributes of other categories as possible (cf. Tversky 1977: 347f.). This definition is obviously intimately related with the notion of cue validity. Rosch and Mervis (1975) demonstrated that measures of similarity and prototypicality correlated highly positively with subjects' similarity judgements on an ordinal scale. For the present approach, however, Tversky's criticism does not necessitate a fundamental revision of the present approach since the integration of cue validity into the set of desiderata considers his well-taken point — what should also be integrated, however, is the idea of weighting attributes according to their diagnostic value, something that is difficult to do in traditional linguistic treatments of categorization and similarity and easy to do, once a more quantitative perspective is adopted.

6. This procedure is far more complex than might be expected from the superficial description above and will, thus, be illustrated briefly. A concordance software identified all occurrences of finite forms of the 128 verbs in a frequency list of 95m words in the Guardian (cf. http: //www.liv.ac.uk/~ms2928.). However, many of the items in this frequency list (e.g. *make* and *makes*) did not only occur as verb forms, but also as nouns; add to this a substantial number of instances where the tagger used for the BNC assigned a so-called portmanteau tag to a word form. Since I was only interested in verbal frequencies, the frequencies had to be adjusted.

First, I determined which of the verb forms of the 128 verbs could principally also occur as nouns (using the Collins Cobuild Dictionary on CD-ROM (1995)), in the case of, say, *make* this is *make* and *makes*. I then extracted all occurrences of these verbs from a random sample of BNC files and noted (i) how often they were unambiguously classified as verb or noun (in the case of *make* 15,514 and 71 cases respectively, amounting to 15,585 cases) and (ii) how often they were assigned a portmanteau tag (e.g. <w NN2-VVZ>), namely 163 cases. From the number of portmanteau tags, I classified as many percent as verbs as there were unambiguous verb tags and as many percent as nouns as there were unambiguous noun tags. In the example of *make*, 99.544% (15,514 von 15,585) of the 163 portmanteau tags were classified as verbal usages as verbal (i.e. 162), and this frequency was added to the number of unambiguous tags, yielding a total of 15,514+162=15,676 verbal usages of *make*/*makes*. This process was repeated for all 128 verbs in order to obtain an approximate frequency of each verb in the Guardian frequency list. The ten verbs with the highest numbers of verbal usages were then used as the search words for the search in the BNC files; including verbs occurring in DK_{for}, these are (in descending order) *make, take, get, give, leave, find, win, tell, write.*

One weakness of this strategy is probably immediately obvious: the frequencies of the verbs in the corpus data are not only based on their frequencies in DK_{to} and PK_{to}. More

concisely, some part of the usages of *make* that entered into the approximation process is based on the frequency of *make* in, e.g., transitive constructions. However, given limited corpus resources, a computation of the frequencies in DK_{to} and PK_{to} alone was not possible, which is probably also the reason why no such statistics seem to be available at present. Be that as it may, the results are probably not too bad after all since the verb *give* was identified as the most frequent verb entering into DK_{to} and PK_{to} (which also represents the consensus of the field) and Stefanowitsch and Gries (under review) showed on the basis of the ICE-GB corpus that *give* is the verb most strongly associated with the ditransitive construction.

7. The first five variables could be used without further refinements; the latter two were recoded using the dummy coding of the General Linear Model (GLM) cf. e.g. Bortz 1999: chap. 14), where "1" and "0" refer to the presence or absence of an attribute.

8. At this point, I would like to anticipate the methodological objection that the data do not conform to a multivariate normal distribution. Gries (2003a), however, addresses this problem in some detail; suffice it here to say that previous research allows for the application of LDAs in such cases, and the resulting possibilities of interpretation outweigh the small risk to obtain slightly skewed results.

9. The binomial test was used to calculate the possibility that the 13 misclassified sentences could be distributed 3: 10 or more extreme when the prior probabilities are those used in the LDA (i.e. 51.28% vs. 48.72%).

10. Note in passing that the present results are identical to an analysis with .5: .5 as prior probabilities.

11. While space does not permit an exhaustive discussion of these aspects of the present work, let me briefly mention an example. In an analysis of the alternation of English verb-particle constructions (*He picked up the book* vs. *He picked the book up*), I found that the usage and predictability of the two constructions renders implausible the assumption that the two constructions are derived from a single underlying structure. More importantly, it was also found that the construction with VP-final position of the particle considered prototypical exhibits exactly those properties that characterise the verb-particle construction that is primary from the perspective of both language acquisition and diachrony. That is, in this analysis synchronic, diachronic and acquisition data mutually supported each other.

References

Allerton, K. (1978). Generating Indirect Objects in English. *Journal of Linguistics* 14, 21–33.

Armstrong, S. L., L. R. Gleitman & H. Gleitman. (1983). What Some Concepts Might Not Be. *Cognition* 13, 263–308.

Arnold, J. E., T. Wasow, A. Losongco & R. Ginstrom. (2000). Heaviness vs. Newness: The Effects of Structural Complexity and Discourse Status on Constituent Ordering. *Language* 76, 28–55.

Barsalou, L. D. (1985). Ideals, Central Tendency, and Frequency Instantiation. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 11, 629–54.

Barsalou, L. D. (1987). The Instability of Graded Structure: Implications for the Nature of Concepts. In U. Neisser (Ed.), *Concepts and Conceptual Development* (pp. 101–40). Cambridge: Cambridge University Press.

Beach, L. R. (1964). Cue Probabilism and Inference Behavior. *Psychological Monographs* 78, 1–20.

Bock, J. K. (1986). Syntactic Persistence in Language Production. *Cognitive Psychology* 18, 355–87.

Bortz, J. (1999). *Statistik für Sozialwissenschaftler*. Berlin/Heidelberg/New York: Springer.

Bybee, J. & C. Moder. (1983). Morphological Classes as Natural Categories. *Language* 59, 251–70.

Cobuild on CD-ROM. (1995). HarperCollins Publishers.

Cowart, W. (1997). *Experimental Syntax: Applying Objective Methods to Sentence Judgments*. Thousand Oaks: Sage Publications.

Gati, I. & A. Tversky. (1982). Representations of Qualitative and Quantitative Dimensions. *Journal of Experimental Psychology: Human Perception and Performance* 8, 325–40.

Givón, T. (1984). Direct Object and Dative Shifting: Semantic and Pragmatic Case. In F. Plank (Ed.), *Objects* (pp. 151–82). New York: Academic Press.

Goldberg, A. (1992). The Inherent Semantics of Argument Structure: The Case of the English Ditransitive Construction. *Cognitive Linguistics* 3, 37–74.

Goldberg, A. (1995). *Constructions: A Construction Grammar Approach to Argument Structure*. Chicago, IL: The University of Chicago Press.

Green, G. M. (1974). *Semantics and Syntactic Regularity*. Bloomington and London: Indiana University Press.

Gries, S. T. (2001). A Multifactorial Analysis of Syntactic Variation: Particle Movement Revisited. *Journal of Quantitative Linguistics* 8, 33–50.

Gries, S. T. (2002). Evidence in Linguistics: Three Approaches to Genitives in English. In R. M. Brend, W. J. Sullivan & A. R. Lommel (Eds.), *LACUS Forum XXVIII: What Constitutes Evidence in Linguistics?* (pp. 17–31). Fullerton, CA: Linguistic Association of Canada and the United States.

Gries, S. T. (2003a). *Multifactorial Analysis In Corpus Linguistics: A Study of Particle Placement*. London, New York: Continuum Press.

Gries, S. T. (2003b). New Perspectives on Old Alternations. Paper presented at the Annual Meeting of the Chicago Linguistics Society.

Hampton, J. (1981). An Investigation of the Nature of Abstract Concepts. *Memory and Cognition* 9, 149–56.

Hawkins, J. A. (1981). On Generating Indirect Objects in English: A Reply to Allerton. *Journal of Linguistics* 17, 1–9.

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

Kleiber, G. (1993). *Prototypensemantik: Eine Einführung*. Tübingen: Narr.

Labov, W. (1973). The Boundaries of Words and Their Meanings. In J. Fishman (ed.), *New Ways of Analyzing Variation in English* (pp. 340–73). Washington, D. C.: Georgetown University Press.

Lakoff, G. (1987). *Women, Fire, And Dangerous Things*. Chicago: University of Chicago Press.

Mazurkewich, I. & L. White. (1984). The Acquisition of the Dative Alternation: Unlearning Generalizations. *Cognition* 16, 261–83.

McCloskey, M. & S. Glucksberg. (1978). Natural Categories: Well-defined or Fuzzy Sets?. *Memory and Cognition* 6, 462–72.

Panther, K.-U. (1997). Dative Alternation from a Cognitive Perspective. In B. Smieja & M. Tasch (Eds.), *Human Contact Through Language and Linguistics* (pp. 107–26). Frankfurt a. M.: Peter Lang.

Ransom, E. (1979). Definiteness and Animacy Constraints on Passive and Double-Object Constructions in English. *Glossa* 13, 215–40.

Rice, S. (1987). Towards a Transitive Prototype: Evidence from Some Atypical English Passives. In L. Michaelis et al. (Eds.), *Proceedings of the Thirteenth Regional Meeting of the Berkeley Linguistics Society* (pp. 422–34). Berkeley, CA: Berkeley Linguistics Society.

Rips, L. J. (1975). Inductive Judgments about Natural Categories. *Journal of Verbal Learning and Verbal Behavior* 14, 665–81.

Rips, L. J., E. J. Shoben & E. E. Smith. (1973). Semantic Distance and the Verification of Semantic Relations. *Journal of Verbal Learning and Verbal Behavior* 12, 1–20.

Rosch, E. (1973). On the Internal Structure of Perceptual and Semantic Categories. In T. E. Moore (Ed.), *Cognitive Development and the Acquisition of Language* (pp. 111–44). New York: Academic Press.

Rosch, E. & C. B. Mervis. (1975). Family Resemblances: Studies in the Internal Structure of Categories. *Cognitive Psychology* 7, 573–605.

Rosch, E., C. Simpson & R. S. Miller. (1976). Structural Bases of Typicality Effects. *Journal of Experimental Psychology: Human Perception and Performance* 2, 491–502.

Ross, J. R. (1972). The Category Squish: Endstation Hauptwort. In Paul Peranteau et al. (Eds.), *Papers from the Eighth Regional Meeting of the Chicago Linguistics Society* (pp. 316–28). Chicago, IL: Chicago Linguistics Society.

Ross, J. R. (1973a). A Fake NP Squish. In C.-J. Bailey & R. Shuy (Eds.), *New Ways of Analyzing Variation in English* (pp. 96–140). Washington, D. C.: Georgetown University Press.

Ross, J. R. (1973b). Nouniness. In O. Fujimura (Ed.), *Three Dimensions of Linguistic Theory* (pp. 137–257). Tokyo: TEC Corporation.

Schütze, C. T. (1996). *The Empirical Base of Linguistics: Grammaticality Judgements and Linguistic Methodology*. Chicago: University of Chicago Press.

Smith, E. E., E. J. Shoben & L. J. Rips. (1974). Structure and Process in Semantic Memory: A Featural Model for Semantic Decision. *Psychological Review* 81, 214–41.

Stefanowitsch, A. & S. T. Gries. (under review). Collostructions: On the Interaction between Verbs and Constructions.

Taylor, J. (1995). *Linguistic Categorization: Prototypes in Linguistic Theory*. 2nd ed. Oxford: Clarendon Press.

Thompson, S. A. (1990). Information Flow and Dative Shift in English Discourse. In J. A. Edmondson, C. Feagin & P. Mühlhäusler (Eds.), *Development and Diversity: Language Variation across Time and Space* (pp. 239–53). Dallas: SIL and University of Arlington, TX.

Thompson, S. A. & Y. Koide. (1987). Iconicity and 'Indirect Objects' in English. *Journal of Pragmatics* 11, 399–406.

Thomson, A. J. & A. V. Martinet. (1988). *A Practical English Grammar for Foreign Students*. Oxford: Oxford University Press.

Thornburg, L. & K.-U. Panther. (1997). Speech Act Metonymies. In W.A. Liebert, G. Redeker & L. Waugh (Eds.), *Discourse and Perspectives in Cognitive Linguistics* (pp. 205–19). Amsterdam, Philadelphia: John Benjamins.

Tversky, A. (1977). Features of Similarity. *Psychological Review* 84, 327–52.

Tversky, A. & I. Gati. (1978). Studies of Similarity. In E. Rosch & B.B. Lloyd (Eds.), *Cognition and Categorization* (pp. 81–98) Hillsdale, NJ: Erlbaum.

Wanner, E. (1979). False Identification of Prime Numbers. Paper presented at the Society for Philosophy and Psychology, New York, N.Y.

Welte, W. (1979). Dative Movement: Ein kritischer Überblick über Motivation and Formulierungsprobleme einer Transformation in der Syntaxtheorie des Englischen. *Anglia* 97, 290–303.

Werner, J. (1997). *Lineare Statistik: Das Allgemeine Lineare Modell.* Weinheim: Psychologie Verlags Union.

Wilkins, A.J. (1971). Conjoint Frequency, Category Size, and Categorization Time. *Journal of Verbal Learning and Behavior* 10, 382–5.

Author's address

Stefan Th. Gries
STGries@sitkom.sdu.dk
Southern Denmark University / IFKI
Grundtvigs Allé 150
6400 Sønderborg
Denmark

About the author

Stefan Th. Gries did his MA and his PhD in English linguistics at the University of Hamburg, Germany; he is now Associate Professor of English at Southern Denmark University. His current research interests include (i) corpus-based investigations of language within a broadly cognitive and constructionist framework and (ii) methodological issues.