# Acquiring Lexical Generalizations from Corpora:
## A Case Study for Diathesis Alternations

**Maria Lapata**
School of Cognitive Science
Division of Informatics, University of Edinburgh
2 Buccleuch Place, Edinburgh EH8 9LW, UK
mlap@cogsci.ed.ac.uk

## Abstract

This paper examines the extent to which verb diathesis alternations are empirically attested in corpus data. We automatically acquire alternating verbs from large balanced corpora by using partial-parsing methods and taxonomic information, and discuss how corpus data can be used to quantify linguistic generalizations. We estimate the productivity of an alternation and the typicality of its members using type and token frequencies.

## 1 Introduction

Diathesis alternations are changes in the realization of the argument structure of a verb that are sometimes accompanied by changes in meaning (Levin, 1993). The phenomenon in English is illustrated in (1)–(2) below.

(1) a. John offers shares to his employees.
    b. John offers his employees shares.
(2) a. Leave a note for her.
    b. Leave her a note.

Example (1) illustrates the dative alternation, which is characterized by an alternation between the prepositional frame 'V NP1 to NP2' and the double object frame 'V NP1 NP2'. The benefactive alternation (cf. (2)) is structurally similar to the dative, the difference being that it involves the preposition *for* rather than *to*.

Levin (1993) assumes that the syntactic realization of a verb's arguments is directly correlated with its meaning (cf. also Pinker (1989) for a similar proposal). Thus one would expect verbs that undergo the same alternations to form a semantically coherent class. Levin's study on diathesis alternations has influenced recent work on word sense disambiguation (Dorr and Jones, 1996), machine translation (Dang et al., 1998), and automatic lexical acquisition (McCarthy and Korhonen, 1998; Schulte im Walde, 1998).

The objective of this paper is to investigate the extent to which diathesis alternations are empirically attested in corpus data. Using the dative and benefactive alternations as a test case we attempt to determine: (a) if some alternations are more frequent than others, (b) if alternating verbs have frame preferences and (c) what the representative members of an alternation are.

In section 2 we describe and evaluate the set of automatic methods we used to acquire verbs undergoing the dative and benefactive alternations. We assess the acquired frames using a filtering method presented in section 3. The results are detailed in section 4. Sections 5 and 6 discuss how the derived type and token frequencies can be used to estimate how productive an alternation is for a given verb semantic class and how typical its members are. Finally, section 7 offers some discussion on future work and section 8 conclusive remarks.

## 2 Method

### 2.1 The parser

The part-of-speech tagged version of the British National Corpus (BNC), a 100 million word collection of written and spoken British English (Burnard, 1995), was used to acquire the frames characteristic of the dative and benefactive alternations. Surface syntactic structure was identified using Gsearch (Keller et al., 1999), a tool which allows the search of arbitrary POS-tagged corpora for shallow syntactic patterns based on a user-specified context-free grammar and a syntactic query. It achieves this by combining a left-corner parser with a regular expression matcher.

Depending on the grammar specification (i.e., recursive or not) Gsearch can be used as a full context-free parser or a chunk parser. Depending on the syntactic query, Gsearch can parse full sentences, identify syntactic relations (e.g., verb-object, adjective-noun) or even single words (e.g., all indefinite pro-

nouns in the corpus).

Gsearch outputs all corpus sentences containing substrings that match a given syntactic query. Given two possible parses that begin at the same point in the sentence, the parser chooses the longest match. If there are two possible parses that can be produced for the same substring, only one parse is returned. This means that if the number of ambiguous rules in the grammar is large, the correctness of the parsed output is not guaranteed.

## 2.2 Acquisition

We used Gsearch to extract tokens matching the patterns 'V NP1 NP2', 'VP NP1 *to* NP2', and 'V NP1 *for* NP2' by specifying a chunk grammar for recognizing the verbal complex and NPs. POS-tags were retained in the parser's output which was post-processed to remove adverbials and interjections.

Examples of the parser's output are given in (3). Although there are cases where Gsearch produces the right parse (cf. (3a)), the parser wrongly identifies as instances of the double object frame tokens containing compounds (cf. (3b)), bare relative clauses (cf. (3c)) and NPs in apposition (cf. (3d)). Sometimes the parser attaches prepositional phrases to the wrong site (cf. (3e)) and cannot distinguish between arguments and adjuncts (cf. (3f)) or between different types of adjuncts (e.g., temporal (cf. (3f)) versus benefactive (cf. (3g))). Erroneous output also arises from tagging mistakes.

(3) a. The police driver [v shot] [NP Jamie] [NP a look of enquiry] which he missed.

  b. Some also [v offer] [NPa free bus] [NP service], to encourage customers who do not have their own transport.

  c. A Jaffna schoolboy [v shows] [NP a drawing] [NP he] made of helicopters strafing his home town.

  d. For the latter catalogue Barr [v chose] [NP the Surrealist writer] [NP Georges Hugnet] to write a historical essay.

  e. It [v controlled] [NP access] [PP to [NP the vault]].

  f. Yesterday he [v rang] [NP the bell] [PP for [NP a long time]].

  g. Don't [v save] [NP the bread] [PP for [NP the birds]].

We identified erroneous subcategorization frames (cf. (3b)-(3d)) by using linguistic heuristics and a process for compound noun detection (cf. section 2.3). We disambiguated the attachment site of

PPs (cf. (3e)) using Hindle and Rooth's (1993) lexical association score (cf. section 2.4). Finally, we recognized benefactive PPs (cf. (3g)) by exploiting the WordNet taxonomy (cf. section 2.5).

## 2.3 Guessing the double object frame

We developed a process which assesses whether the syntactic patterns (called cues below) derived from the corpus are instances of the double object frame.

**Linguistic Heuristics.** We applied several heuristics to the parser's output which determined whether corpus tokens were instances of the double object frame. The 'Reject' heuristics below identified erroneous matches (cf. (3b-d)), whereas the 'Accept' heuristics identified true instances of the double object frame (cf. (3a)).

1. **Reject** if cue contains at least two proper names adjacent to each other (e.g., *killed Henry Phipps*).

2. **Reject** if cue contains possessive noun phrases (e.g., *give a showman's award*).

3. **Reject** if cue's last word is a pronoun or an anaphor (e.g., *ask the subjects themselves*).

4. **Accept** if verb is followed by a personal or indefinite pronoun (e.g., *found him a home*).

5. **Accept** if verb is followed by an anaphor (e.g., *made herself a snack*).

6. **Accept** if cue's surface structure is either 'V MOD[1] NP MOD NP' or 'V NP MOD NP' (e.g., *send Bailey a postcard*).

7. **Cannot decide** if cue's surface structure is 'V MOD* N N+' (e.g., *offer a free bus service*).

**Compound Noun Detection.** Tokens identified by heuristic (7) were dealt with separately by a procedure which guesses whether the nouns following the verb are two distinct arguments or parts of a compound. This procedure was applied only to noun sequences of length 2 and 3 which were extracted from the parser's output[2] and compared against a compound noun dictionary (48,661 entries) compiled from WordNet. 13.9% of the noun sequences were identified as compounds in the dictionary.

---

[1]Here MOD represents any prenominal modifier (e.g., articles, pronouns, adjectives, quantifiers, ordinals).

[2]Tokens containing noun sequences with length larger than 3 (450 in total) were considered negative instances of the double object frame.

| G-score | 2-word compound |
|---------|-----------------|
| 1967.68 | *bank manager* |
| 775.21 | *tax liability* |
| 87.02 | *income tax* |
| 45.40 | *book reviewer* |
| 30.58 | *designer gear* |
| 29.94 | *safety plan* |
| 24.04 | *drama school* |

Table 1: Random sample of two word compounds

| G-score | 3-word compound |
|---------|-----------------|
| 574.48 | [[*energy efficiency*] *office*] |
| 382.92 | [[*council tax*] *bills*] |
| 77.78 | [*alcohol* [*education course*]] |
| 48.84 | [*hospital* [*out-patient department*]] |
| 36.44 | [[*tumour suppressor*] *function*] |
| 32.35 | [[*nature conservation*] *resources*] |
| 23.98 | [[*quality amplifier*] *circuits*] |

Table 2: Random sample of three word compounds

For sequences of length 2 not found in WordNet, we used the log-likelihood ratio (G-score) to estimate the lexical association between the nouns, in order to determine if they formed a compound noun. We preferred the log-likelihood ratio to other statistical scores, such as the association ratio (Church and Hanks, 1990) or $\chi^2$, since it adequately takes into account the frequency of the co-occurring words and is less sensitive to rare events and corpus-size (Dunning, 1993; Daille, 1996). We assumed that two nouns cannot be disjoint arguments of the verb if they are lexically associated. On this basis, tokens were rejected as instances of the double object frame if they contained two nouns whose G-score had a p-value less than 0.05.

A two-step process was applied to noun sequences of length 3: first their bracketing was determined and second the G-score was computed between the single noun and the 2-noun sequence. We inferred the bracketing by modifying an algorithm initially proposed by Pustejovsky et al. (1993). Given three nouns $n_1, n_2, n_3$, if either $[n_1 \ n_2]$ or $[n_2 \ n_3]$ are in the compound noun dictionary, we built structures $[[n_1 \ n_2] \ n_3]$ or $[n_1 \ [n_2 \ n_3]]$ accordingly; if both $[n_1 \ n_2]$ and $[n_2 \ n_3]$ appear in the dictionary, we chose the most frequent pair; if neither $[n_1 \ n_2]$ nor $[n_2 \ n_3]$ appear in WordNet, we computed the G-score for $[n_1 \ n_2]$ and $[n_2 \ n_3]$ and chose the pair with highest value ($p \leq 0.05$). Tables 1 and 2 display a random sample of the compounds the method found ($p \leq 0.05$).

### 2.3.1 Evaluation

The performance of the linguistic heuristics and the compound detection procedure were evaluated by randomly selecting approximately 3,000 corpus tokens which were previously accepted or rejected as instances of the double object frame. Two judges decided whether the tokens were classified correctly. The judges' agreement on the classification task was calculated using the Kappa coefficient (Siegel and

| Method | Prec | Kappa |
|--------|------|-------|
| Reject heuristics | 96.9% | $K = 0.76, N = 1000$ |
| Accept heuristics | 73.6% | $K = 0.82, N = 1000$ |
| 2-word compounds | 98.9% | $K = 0.83, N = 553$ |
| 3-word compounds | 99.1% | $K = 0.70, N = 447$ |
| Verb attach-to | 74.4% | $K = 0.78, N = 494$ |
| Noun attach-to | 80.0% | $K = 0.80, N = 500$ |
| Verb attach-for | 73.6% | $K = 0.85, N = 630$ |
| Noun attach-for | 36.0% | $K = 0.88, N = 500$ |

Table 3: Precision of heuristics, compound noun detection and lexical association

Castellan, 1988) which measures inter-rater agreement among a set of coders making category judgments.

The Kappa coefficient of agreement ($K$) is the ratio of the proportion of times, $P(A)$, that $k$ raters agree to the proportion of times, $P(E)$, that we would expect the raters to agree by chance (cf. (4)). If there is a complete agreement among the raters, then $K = 1$.

$$(4) \quad K = \frac{P(A) - P(E)}{1 - P(E)}$$

Precision figures[3] (Prec) and inter-judge agreement (Kappa) are summarized in table 3. In sum, the heuristics achieved a high accuracy in classifying cues for the double object frame. Agreement on the classification was good given that the judges were given minimal instructions and no prior training.

### 2.4 Guessing the prepositional frames

In order to consider verbs with prepositional frames as candidates for the dative and benefactive alternations the following requirements needed to be met:

1. the PP must be attached to the verb;

---

[3]Throught the paper the reported percentages are the average of the judges' individual classifications.

2. in the case of the 'V NP1 *to* NP2' structure, the *to*-PP must be an argument of the verb;

3. in the case of the 'V NP1 *for* NP2' structure, the *for*-PP must be benefactive.[4]

In order to meet requirements (1)–(3), we first determined the attachment site (e.g., verb or noun) of the PP and secondly developed a procedure for distinguishing benefactive from non-benefactive PPs.

Several approaches have statistically addressed the problem of prepositional phrase ambiguity, with comparable results (Hindle and Rooth, 1993; Collins and Brooks, 1995; Ratnaparkhi, 1998). Hindle and Rooth (1993) used a partial parser to extract $(v, n, p)$ tuples from a corpus, where $p$ is the preposition whose attachment is ambiguous between the verb $v$ and the noun $n$. We used a variant of the method described in Hindle and Rooth (1993), the main difference being that we applied their lexical association score (a log-likelihood ratio which compares the probability of noun versus verb attachment) in an unsupervised non-iterative manner. Furthermore, the procedure was applied to the special case of tuples containing the prepositions *to* and *for* only.

### 2.4.1 Evaluation

We evaluated the procedure by randomly selecting 2,124 tokens containing *to*-PPs and *for*-PPs for which the procedure guessed verb or noun attachment. The tokens were disambiguated by two judges. Precision figures are reported in table 3.

The lexical association score was highly accurate on guessing both verb and noun attachment for *to*-PPs. Further evaluation revealed that for 98.6% ($K = 0.9$, $N = 494$, $k = 2$) of the tokens classified as instances of verb attachment, the *to*-PP was an argument of the verb, which meant that the log-likelihood ratio satisfied both requirements (1) and (2) for *to*-PPs.

A low precision of 36% was achieved in detecting instances of noun attachment for *for*-PPs. One reason for this is the polysemy of the preposition *for*: *for*-PPs can be temporal, purposive, benefactive or causal adjuncts and consequently can attach to various sites. Another difficulty is that benefactive *for*-PPs semantically license both attachment sites.

To further analyze the poor performance of the log-likelihood ratio on this task, 500 tokens con-

taining *for*-PPs were randomly selected from the parser's output and disambiguated. Of these 73.9% ($K = 0.9$, $N = 500$, $k = 2$) were instances of verb attachment, which indicates that verb attachments outnumber noun attachments for *for*-PPs, and therefore a higher precision for verb attachment (cf. requirement (1)) can be achieved without applying the log-likelihood ratio, but instead classifying all instances as verb attachment.

### 2.5 Benefactive PPs

Although surface syntactic cues can be important for determining the attachment site of prepositional phrases, they provide no indication of the semantic role of the preposition in question. This is particularly the case for the preposition *for* which can have several roles, besides the benefactive.

Two judges discriminated benefactive from non-benefactive PPs for 500 tokens, randomly selected from the parser's output. Only 18.5% ($K = 0.73$, $N = 500$, $k = 2$) of the sample contained benefactive PPs. An analysis of the nouns headed by the preposition *for* revealed that 59.6% were animate, 17% were collective, 4.9% denoted locations, and the remaining 18.5% denoted events, artifacts, body parts, or actions. Animate, collective and location nouns account for 81.5% of the benefactive data.

We used the WordNet taxonomy (Miller et al., 1990) to recognize benefactive PPs (cf. requirement (3)). Nouns in WordNet are organized into an inheritance system defined by hypernymic relations. Instead of being contained in a single hierarchy, nouns are partitioned into a set of semantic primitives (e.g., *act*, *animal*, *time*) which are treated as the unique beginners of separate hierarchies. We compiled a "concept dictionary" from WordNet (87,642 entries), where each entry consisted of the noun and the semantic primitive distinguishing each noun sense (cf. table 4).

We considered a *for*-PP to be benefactive if the noun headed by *for* was listed in the concept dictionary and the semantic primitive of its prime sense (Sense 1) was *person*, *animal*, *group* or *location*. PPs with head nouns not listed in the dictionary were considered benefactive only if their head nouns were proper names. Tokens containing personal, indefinite and anaphoric pronouns were also considered benefactive (e.g., *build a home for him*).

Two judges evaluated the procedure by judging 1,000 randomly selected tokens, which were accepted or rejected as benefactive. The procedure achieved a precision of 48.8% ($K = 0.89$, $N =$

---

[4]Syntactically speaking, benefactive *for*-PPs are not arguments but adjuncts (Jackendoff, 1990) and can appear on any verb with which they are semantically compatible.

| | Sense 1 | Sense 2 | Sense 3 |
|---|---|---|---|
| gift | *possession* | *cognition* | *act* |
| cooking | *food* | *act* | ———— |
| teacher | *person* | *cognition* | ———— |
| university | *group* | *artifact* | *group* |
| city | *location* | *location* | *group* |
| pencil | *artifact* | ———— | ———— |

Table 4: Sample entries from WordNet concept dictionary

500, $k = 2$) in detecting benefactive tokens and 90.9% ($K = .94$, $N = 499$, $k = 2$) in detecting non-benefactive ones.

## 3 Filtering

Filtering assesses how probable it is for a verb to be associated with a wrong frame. Erroneous frames can be the result of tagging errors, parsing mistakes, or errors introduced by the heuristics and procedures we used to guess syntactic structure.

We discarded verbs for which we had very little evidence (frame frequency = 1) and applied a relative frequency cutoff: the verb's acquired frame frequency was compared against its overall frequency in the BNC. Verbs whose relative frame frequency was lower than an empirically established threshold were discarded. The threshold values varied from frame to frame but not from verb to verb and were determined by taking into account for each frame its overall frame frequency which was estimated from the COMLEX subcategorization dictionary (6,000 verbs) (Grishman et al., 1994). This meant that the threshold was higher for less frequent frames (e.g., the double object frame for which only 79 verbs are listed in COMLEX).

We also experimented with a method suggested by Brent (1993) which applies the binomial test on frame frequency data. Both methods yielded comparable results. However, the relative frequency threshold worked slightly better and the results reported in the following section are based on this method.

## 4 Results

We acquired 162 verbs for the double object frame, 426 verbs for the 'V NP1 *to* NP2' frame and 962 for the 'V NP1 *for* NP2' frame. Membership in alternations was judged as follows: (a) a verb participates in the dative alternation if it has the double object and 'V NP1 *to* NP2' frames and (b) a verb

| Dative Alternation | |
|---|---|
| Alternating | allot, assign, bring, fax, feed, flick, give, grant, guarantee, leave, lend offer, owe, take pass, pay, render, repay, sell, show, teach, tell, throw, toss, write, serve, send, award |
| V NP1 NP2 | allocate, bequeath, carry, catapult, cede, concede, drag, drive, extend, ferry, fly, haul, hoist, issue, lease, peddle, pose, preach, push, relay, ship, tug, yield |
| V NP1 *to* NP2 | ask, chuck, promise, quote, read, shoot, slip |
| **Benefactive Alternation** | |
| Alternating | bake, build, buy, cast, cook, earn, fetch, find, fix, forge, gain, get, keep, knit, leave, make, pour, save procure, secure, set, toss, win, write |
| V NP1 NP2 | arrange, assemble, carve, choose, compile, design, develop, dig, gather, grind, hire, play, prepare, reserve, run, sew |
| V NP1 *for* NP2 | boil, call, shoot |

Table 5: Verbs common in corpus and Levin

participates in the benefactive alternation if it has the double object and 'V NP1 *for* NP2' frames. Table 5 shows a comparison of the verbs found in the corpus against Levin's list of verbs;[5] rows 'V NP1 *to* NP2' and 'V NP1 *for* NP2' contain verbs listed as alternating in Levin but for which we acquired only one frame. In Levin 115 verbs license the dative and 103 license the benefactive alternation. Of these we acquired 68 for the dative and 43 for the benefactive alternation (in both cases including verbs for which only one frame was acquired).

The dative and benefactive alternations were also acquired for 52 verbs not listed in Levin. Of these, 10 correctly alternate (*cause, deliver, hand, refuse, report* and *set* for the dative alternation and *cause, spoil, afford* and *prescribe* for the benefactive), and 12 can appear in either frame but do not alternate (e.g., *appoint, fix, proclaim*). For 18 verbs two frames were acquired but only one was correct (e.g., *swap* and *forgive* which take only the double object frame), and finally 12 verbs neither alternated nor had the acquired frames. A random sample of the acquired verb frames and their (log-transformed) frequencies is shown in figure 1.

---

[5]The comparisons reported henceforth exclude verbs listed in Levin with overall corpus frequency less than 1 per million.
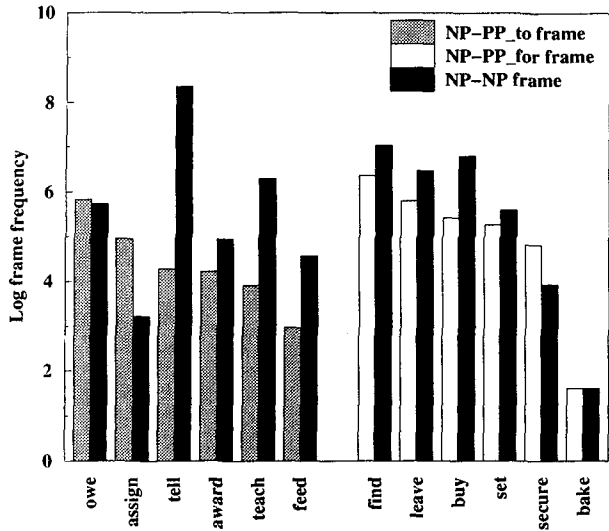
401

Figure 1: Random sample of acquired frequencies for the dative and benefactive alternations
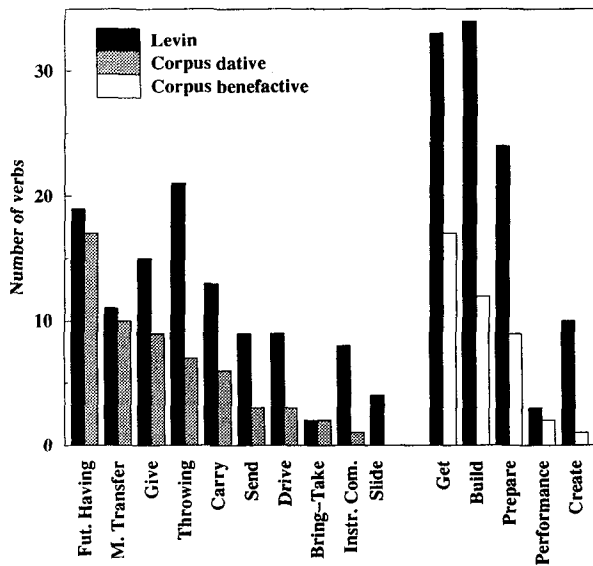


Figure 2: Semantic classes for the dative and benefactive alternations

Levin defines 10 semantic classes of verbs for which the dative alternation applies (e.g., GIVE verbs, verbs of FUTURE HAVING, SEND verbs), and 5 classes for which the benefactive alternation applies (e.g., BUILD, CREATE, PREPARE verbs), assuming that verbs participating in the same class share certain meaning components.

We partitioned our data according to Levin's predefined classes. Figure 2 shows for each semantic class the number of verbs acquired from the corpus against the number of verbs listed in Levin. As can be seen in figure 2, Levin and the corpus approximate each other for verbs of FUTURE HAVING (e.g., guarantee), verbs of MESSAGE TRANSFER (e.g., tell) and BRING-TAKE verbs (e.g., bring). The semantic classes of GIVE (e.g., sell), CARRY (e.g., drag), SEND (e.g., ship), GET (e.g., buy) and PREPARE (e.g., bake) verbs are also fairly well represented in the corpus, in contrast to SLIDE verbs (e.g., bounce) for which no instances were found.

Note that the corpus and Levin did not agree with respect to the most popular classes licensing the dative and benefactive alternations: THROWING (e.g., toss) and BUILD verbs (e.g., carve) are the biggest classes in Levin allowing the dative and benefactive alternations respectively, in contrast to FUTURE HAVING and GET verbs in the corpus. This can be explained by looking at the average corpus frequency of the verbs belonging to the semantic classes in question: FUTURE HAVING and GET verbs outnumber THROWING and BUILD verbs by a factor of two to one.

## 5 Productivity

The relative productivity of an alternation for a semantic class can be estimated by calculating the ratio of acquired to possible verbs undergoing the alternation (Aronoff, 1976; Briscoe and Copestake, 1996):

$$(5) \quad P(acquired|class) = \frac{f(acquired, class)}{f(class)}$$

We express the productivity of an alternation for a given class as $f(acquired, class)$, the number of verbs which were found in the corpus and are members of the class, over $f(class)$, the total number of verbs which are listed in Levin as members of the class (Total). The productivity values (Prod) for both the dative and the benefactive alternation (Alt) are summarized in table 6.

Note that productivity is sensitive to class size. The productivity of BRING-TAKE verbs is estimated to be 1 since it contains only 2 members which were also found in the corpus. This is intuitively correct, as we would expect the alternation to be more productive for specialized classes.

The productivity estimates discussed here can be potentially useful for treating lexical rules probabilistically, and for quantifying the degree to which language users are willing to apply a rule in order

| Dative alternation | | | | |
|---|---|---|---|---|
| Class | Total | Alt | Prod | Typ |
| BRING-TAKE | 2 | 2 | 1 | 0.327 |
| FUTURE HAVING | 19 | 17 | 0.89 | 0.313 |
| GIVE | 15 | 9 | 0.6 | 0.55 |
| M.TRANSFER | 17 | 10 | 0.58 | 0.66 |
| CARRY | 15 | 6 | 0.4 | 0.056 |
| DRIVE | 11 | 3 | 0.27 | 0.03 |
| THROWING | 30 | 7 | 0.23 | 0.658 |
| SEND | 23 | 3 | 0.13 | 0.181 |
| INSTR. COM. | 18 | 1 | 0.05 | 0.648 |
| SLIDE | 5 | 0 | 0 | 0 |
| Benefactive alternation | | | | |
| Class | Total | Alt | Prod | Typ |
| GET | 33 | 17 | 0.51 | 0.54 |
| PREPARE | 26 | 9 | 0.346 | 0.55 |
| BUILD | 35 | 12 | 0.342 | 0.34 |
| PERFORMANCE | 19 | 1 | 0.05 | 0.56 |
| CREATE | 20 | 2 | 0.1 | 0.05 |

Table 6: Productivity estimates and typicality values for the dative and benefactive alternation

to produce a novel form (Briscoe and Copestake, 1996).

## 6 Typicality

Estimating the productivity of an alternation for a given class does not incorporate information about the frequency of the verbs undergoing the alternation. We propose to use frequency data to quantify the typicality of a verb or verb class for a given alternation. The underlying assumption is that a verb is typical for an alternation if it is equally frequent for both frames which are characteristic for the alternation. Thus the typicality of a verb can be defined as the conditional probability of the frame given the verb:

$$(6) \quad P(frame_i|verb) = \frac{f(frame_i, verb)}{\sum_n f(frame_n, verb)}$$

We calculate $P(frame_i|verb)$ by dividing $f(frame_i, verb)$, the number of times the verb was attested in the corpus with frame $i$, by $\sum_n f(frame_n, verb)$, the overall number of times the verb was attested. In our case a verb has two frames, hence $P(frame_i|verb)$ is close to 0.5 for typical verbs (i.e., verbs with balanced frequencies) and close to either 0 or 1 for peripheral verbs, depending on their preferred frame. Consider the verb owe as an example (cf. figure 1). 648 instances

of owe were found, of which 309 were instances of the double object frame. By dividing the latter by the former we can see that owe is highly typical of the dative alternation: its typicality score for the double object frame is 0.48.

By taking the average of $P(frame_i, verb)$ for all verbs which undergo the alternation and belong to the same semantic class, we can estimate how typical this class is for the alternation. Table 6 illustrates the typicality (Typ) of the semantic classes for the two alternations. (The typicality values were computed for the double object frame). For the dative alternation, the most typical class is GIVE, and the most peripheral is DRIVE (e.g., ferry). For the benefactive alternation, PERFORMANCE (e.g., sing), PREPARE (e.g., bake) and GET (e.g., buy) verbs are the most typical, whereas CREATE verbs (e.g., compose) are peripheral, which seems intuitively correct.

## 7 Future Work

The work reported in this paper relies on frame frequencies acquired from corpora using partial-parsing methods. For instance, frame frequency data was used to estimate whether alternating verbs exhibit different preferences for a given frame (typicality).

However, it has been shown that corpus idiosyncrasies can affect subcategorization frequencies (cf. Roland and Jurafsky (1998) for an extensive discussion). This suggests that different corpora may give different results with respect to verb alternations. For instance, the to-PP frame is poorly represented in the syntactically annotated version of the Penn Treebank (Marcus et al., 1993). There are only 26 verbs taking the to-PP frame, of which 20 have frame frequency of 1. This indicates that a very small number of verbs undergoing the dative alternation can be potentially acquired from this corpus. In future work we plan to investigate the degree to which corpus differences affect the productivity and typicality estimates for verb alternations.

## 8 Conclusions

This paper explored the degree to which diathesis alternations can be identified in corpus data via shallow syntactic processing. Alternating verbs were acquired from the BNC by using Gsearch as a chunk parser. Erroneous frames were discarded by applying linguistic heuristics, statistical scores (the log-likelihood ratio) and large-scale lexical resources

(e.g., WordNet).

We have shown that corpus frequencies can be used to quantify linguistic intuitions and lexical generalizations such as Levin's (1993) semantic classification. Furthermore, corpus frequencies can make explicit predictions about word use. This was demonstrated by using the frequencies to estimate the productivity of an alternation for a given semantic class and the typicality of its members.

## Acknowledgments

## References

Mark Aronoff. 1976. *Word Formation in Generative Grammar*. Linguistic Inquiry Monograph 1. MIT Press, Cambridge, MA.

Michael Brent. 1993. From grammar to lexicon: Unsupervised learning of lexical syntax. *Computational Linguistics*, 19(3):243–262.

Ted Briscoe and Ann Copestake. 1996. Contolling the application of lexical rules. In *Proceedings of ACL SIGLEX Workshop on Breadth and Depth of Semantic Lexicons*, pages 7–19, Santa Cruz, CA.

Lou Burnard, 1995. *Users Guide for the British National Corpus*. British National Corpus Consortium, Oxford University Computing Service.

Kenneth Ward Church and Patrick Hanks. 1990. Word association norms, mutual information, and lexicography. *Computational Linguistics*, 16(1):22–29.

COLING/ACL 1998. *Proceedings of the 17th International Conference on Computational Linguistics and 36th Annual Meeting of the Association for Computational Linguistics*, Montréal.

Michael Collins and James Brooks. 1995. Prepositional phrase attachment through a backed-off model. In *Proceedings of the 3rdWorkshop on Very Large Corpora*, pages 27–38.

Béatrice Daille. 1996. Study and implementation of combined techniques for automatic extraction of terminology. In Judith Klavans and Philip Resnik, editors, *The Balancing Act: Combining Symbolic and Statistical Approaches to Language*, pages 49–66. MIT Press, Cambridge, MA.

Hoa Trang Dang, Karin Kipper, Martha Palmer, and Joseph Rosenzweig. 1998. Investigating regular sense extensions based on intersective Levin classes. In COLING/ACL 1998, pages 293–299.

Bonnie J. Dorr and Doug Jones. 1996. Role of word sense disambiguation in lexical acquisition: Predicting semantics from syntactic cues. In *Proceedings of the 16th International Conference on Computational Linguistics*, pages 322–327, Copenhagen.

Ted Dunning. 1993. Accurate methods for the statistics of surprise and coincidence. *Computational Linguistics*, 19(1):61–74.

Ralph Grishman, Catherine Macleod, and Adam Meyers. 1994. Comlex syntax: Building a computational lexicon. In *Proceedings of the 15th International Conference on Computational Linguistics*, pages 268–272, Kyoto.

Donald Hindle and Mats Rooth. 1993. Structural ambiguity and lexical relations. *Computational Linguistics*, 19(1):103–120.

Ray Jackendoff. 1990. *Semantic Structures*. MIT Press, Cambridge, MA.

Frank Keller, Martin Corley, Steffan Corley, Matthew W. Crocker, and Shari Trewin. 1999. Gsearch: A tool for syntactic investigation of unparsed corpora. In *Proceedings of the EACL Workshop on Linguistically Interpreted Corpora*, Bergen.

Beth Levin. 1993. *English Verb Classes and Alternations: A Preliminary Investigation*. University of Chicago Press, Chicago.

Mitchell P. Marcus, Beatrice Santorini, and Mary Ann Marcinkiewicz. 1993. Building a large annotated corpus of english: The penn treebank. *Computational Linguistics*, 19(2):313–330.

Diana McCarthy and Anna Korhonen. 1998. Detecting verbal participation in diathesis alternations. In COLING/ACL 1998, pages 1493–1495. Student Session.

George A. Miller, Richard Beckwith, Christiane Fellbaum, Derek Gross, and Katherine J. Miller. 1990. Introduction to WordNet: An on-line lexical database. *International Journal of Lexicography*, 3(4):235–244.

Steven Pinker. 1989. *Learnability and Cognition: The Acquisition of Argument Structure*. MIT Press, Cambridge MA.

James Pustejovsky, Sabine Bergler, and Peter Anick. 1993. Lexical semantic techniques for corpus analysis. *Computational Linguistics*, 19(3):331–358.

Adwait Ratnaparkhi. 1998. Unsupervised statistical models for prepositional phrase attachment. In *Proceedings of the 7th International Conference on Computational Linguistics*, pages 1079–1085.

Douglas Roland and Daniel Jurafsky. 1998. How verb subcategorization frequencies are affected by corpus choice. In COLING/ACL 1998, pages 1122–1128.

Sabine Schulte im Walde. 1998. Automatic semantic classification of verbs according to their alternation behaviour. Master's thesis, Institut f'ur Maschinelle Sprachverarbeitung, University of Stuttgart.

Sidney Siegel and N Castellan. 1988. *Non Parametric Statistics for the Behavioral Sciences*. McGraw-Hill, New York.