
GUIDELINES FOR THE DATIVE ALTERNATION

Methods of extraction and data filtering

Melanie Röthlisberger —last updated October 2, 2014

Contents

1	General introduction	3
2	Corpora	3
2.1	The <i>International Corpus of English</i>	3
2.2	Characteristics of the different corpora	4
3	Parser comparison	4
4	Extraction of the data	6
4.1	The verb list	6
4.1.1	The pre-final verb list	6
4.1.2	The final verb list	7
4.2	The perl script	10
4.2.1	Data format	10
4.2.2	First extraction	10
4.2.3	Verification against ICE-CUP	11
4.2.4	Adjustments to the perl script	14
4.3	The first dataset —raw numbers before filtering	15
5	First selection and filtering of the data —Syntactic constraints	15
5.1	Introduction	15
5.2	Exclusion of double object and prepositional dative constructions	17
5.2.1	General exclusions	17
5.2.2	Specific exclusions in double object constructions	24
5.2.3	Specific exclusions in prepositional dative constructions	28

5.3	Object extraction and verification	30
6	Selection and Filtering of the data —the choice context	31
6.1	Already excluded with the perl script	32
6.2	Methodological constraints	33
6.3	Semantic constraints	33
6.3.1	Fixed expressions & Idioms	33
6.3.2	Verified non-fixed expressions	38
6.3.3	To-To constructions	39
6.3.4	Predicative PP	40
6.3.5	Spatial goals	40
6.3.6	Beneficiary constructions	40
6.4	Intervening prepositional phrases and adverbials	41
7	Delimitation of the objects	42
7.1	Postmodifications	42
7.1.1	Prepositional phrases	44
7.1.2	Nonfinite clauses	46
7.1.3	Finite clauses	48
7.1.4	Postposed adverbs and adjectives	50
7.1.5	General extenders	51
7.1.6	Incomplete constituents	51
7.2	Ellipsis	52
7.3	Intervening constituents —hesitators	54
7.4	Self-corrections and repetitions	55
8	Mark-up of ICE corpora	57
9	Other issues resolved	58
9.1	Spelling conventions	58
10	Appendix —The C7 Tagset	59

1 General introduction

These guidelines document the steps and approaches taken in the extraction and filtering processes when analysing the DATIVE ALTERNATION in varieties of English. They form part of the project *Exploring probabilistic grammar(s) in varieties of English around the world* and will be incorporated to some extent into the final PhD thesis “The Dative Alternation in Varieties of English”. The aim of these guidelines is to present an all-encompassing overview including an extensive list of examples. They remain subject to further additions.

2 Corpora

Since it is the aim of the study to explore the underlying probabilistic constraints that drive syntactical differences across various varieties of English, the use of comparable corpora from different varieties of English presents itself. For this reason, the study makes use of the *International Corpus of English* (ICE) series, more specifically the following components: Great Britain, India, Hong Kong, Ireland, Jamaica, New Zealand, Philippines, and Singapore.

2.1 The *International Corpus of English*

The *International Corpus of English* project started off in 1988 with a proposal submitted by Sidney Greenbaum:

We should now be thinking of extending the scope for computerized comparative studies in three ways: (1) to sample standard varieties from other countries where English is the first language, for example Canada and Australia; (2) to sample national varieties from countries where English is an official additional language, for example India and Nigeria; and (3) to include spoken and manuscript English as well as printed English. (Greenbaum 1988; taken from <http://www.ucl.ac.uk/english-usage/projects/ice.htm>)

Greenbaum (1996) —in his introduction to the book “Comparing English Worldwide” —called it “an ambitious project”. The principal aim of it was to “provide the resources for comparative studies of the English used in countries where it is either a majority first language (for example, Canada and Australia) or an official additional language (for example, India and Nigeria)” (Greenbaum 1996). Currently, 14 complete ICE corpora of different varieties of English are available for comparative research, and many more are still on their way

(ice-corpora.net/ice/, accessed June 27, 2014). In order to qualify for an ICE component, English has to be used in those countries as the major language for communication in government administration, educational institution but also among its speakers and in creative writing. Each ICE-component consists of approximately 60% spoken and 40% written data, a total of 500 texts with 2'000 words each making up a 1-million-word corpus for every variety of English. The spoken material consists of roughly 300 texts from dialogues (180) as well as monologues (120) covering a wide range of genres from face-to-face-conversations, to scripted and unscripted speeches. The written material consists of roughly 200 texts from printed and non-printed sources, covering such genres as letters, student essays, academic writing, creative writing, popular writing and reportage (Nelson 1996, see figure 1). The texts are written in educated or standard English, based on the criterion of the language producer: To be included in the corpus, speakers/writers need to have received formal education and completed secondary school or have an appropriate public status (e.g. as politicians, writers, etc.) (see Greenbaum 1996: 6).

However, due to the restrictions and difficulties that some ICE teams faced in other varieties of English (e.g. Nigeria, Fiji) and especially because times have changed, certain registers were hard to come by (e.g. letters) and were supplemented or replaced by similar text types (e.g. emails). Also, in some cases one register had to be extended quantitatively to compensate for the lack of enough texts in another genre. Each ICE team provides a manual to accompany the publication of their corpus, whereby possible derivations from the general layout of the ICE project are indicated.

2.2 Characteristics of the different corpora

ICE-GB (1990–1998) has been grammatically annotated and the output was manually checked. The parsed version is available through a software called ICE-CUP. The syntactic annotations of the parsed texts follow the grammar of Quirk, Greenbaum, Leech, and Svartik (1985). (...more info on ICE corpora here ...) More information on the various corpora in relation to the current study can be found in a separate documentation.

3 Parser comparison

Before starting on the project, I verified the possible use of a parsed version of the corpus data. Since it is my aim to analyse two syntactic constructions, the use of parsed text would render

Guidelines for the Dative Alternation

SPOKEN 300	Dialogues	180	Private	100	Face-to-face conversations	90	s1a
					Phonecalls	10	
	Public			80	Classroom lessons	20	s1b
					Broadcast Discussions	20	
					Broadcast Interviews	10	
					Parliamentary Debates	10	
					Legal cross-examinations	10	
					Business Transactions	10	
	Monologues	120	Unscripted	70	Spontaneous commentaries	20	s2a
					Unscripted Speeches	30	
					Demonstrations	10	
					Legal Presentations	10	
	Scripted			50	Broadcast News	20	s2b
					Broadcast Talks	20	
					Non-broadcast Talks	10	
WRITTEN 200	Non-printed	50	Student Writing	20	Student Essays	10	w1a
					Exam Scripts	10	
					Social Letters	15	w1b
	Printed	150	Academic writing	40	Business Letters	15	
					Humanities	10	w2a
					Social Sciences	10	
					Natural Sciences	10	
			Popular Writing	40	Technology	10	
					Humanities	10	w2b
					Social Sciences	10	
					Natural Sciences	10	
			Reportage	20	Technology	10	
					Press news reports	20	w2c
	Instructional writing			20	Administrative Writing	10	w2d
					Skills/Hobbies	10	
					Persuasive writing	10	w2e
					Press editorials	10	
	Creative writing			20	Novels & short stories	20	w2f

Figure 1: The general layout of the ICE corpora.

the whole enterprise much more feasible. An online platform developed in Zurich —Dependency Bank 2.0 (Lehmann & Schneider) —made it possible to access all nine ICE corpora as parsed versions. As recall has to be as high as possible, I made a direct comparison between the manually verified parsed version of ICE-GB as available through ICE-CUP and the output by the Dependency Bank 2.0 (henceforth called DB). Six verbs were selected (*give*, *bring*, *send*, *lend*, *hand*, *tell*) and the output of the two parsers compared.

The results were quite sobering. Depending on the verb, recall for the double object construction was as low as 44% for *tell* to 76% for *give*. For the prepositional dative construction, it was comparatively better but not satisfying either (42% for *hand* and 100% for *tell*). The parser comparison identified a small number of parsing and tagging mistakes in ICE-CUP. Additionally, it pointed out that the DB parser mostly missed out on heavy or more complex noun phrases or had parsing problems due to intervening interjection or adverbs (mainly in spoken material). In conclusion, we decided to rely on the tagged version and extract the dative constructions with the help of a perl script.

4 Extraction of the data

In the extraction of the data, more precisely, the double object (DO) and prepositional dative constructions (PD), I adhered to previous procedures. The extraction methods by Bresnan, Cueni, Nikitina, and Baayen (2007) and Bresnan and colleagues was followed as closely as possible in order to facilitate comparability with their research results. In a first step, I compiled a list of interchangeable verbs, which was then used to extract all syntactic constructions that featured this verb in any of its forms followed by two objects (pronoun or noun). For the PD, an intervening *to* was allowed. The complete extraction process is explicated in detail in what follows.

4.1 The verb list

4.1.1 The pre-final verb list

First, I compiled a pre-final list of dative verbs that could occur in both the ditransitive and prepositional dative construction in all 9 ICE corpora. For this, I made use of the DB and the parsed version of all nine corpora made available on that platform. The parser used is Pro3gres, a dependency parser developed by Gerold Schneider at the University of Zurich (Schneider 2008). The double object constructions were found by filtering for two dependencies (object, second object) with right alignment and direct links for each dependency (while Head1=Head2). The prepositional dative constructions were found by filtering for three dependencies (object, PP attached to verb, preposition) with right alignment and direct links for the first two dependencies and all-open for the last dependency. The dependent of the third dependency was restricted to *to* and Dep2=Head3, while Head1=Head2.

This list was increased with verbs that were found to occur in both variants diachronically (Wolk, Bresnan, Rosenbach, and Szmrecsanyi 2013: 389-90), and the ditransitive verbs denoted as such in the parser Pro3Gres (Gerold Schneider, p.c.) and in Mukherjee (2005). Following Theijssen (2012), additional verbs were taken from Levin (1993)'s verb classification.¹ I further consulted and verified the verb list against Cueni (2004)'s collection of verbs based on the SWITCHBOARD corpus, Bresnan, Cueni, et al. (2007)'s work which made use of the same corpus with the addition of written data, and studies that examined the dative alterna-

¹As suggested further below in the text, more exotic varieties may not have the same preferences when it comes to verbs occurring in either DO and PD. Thus, I also included non-alternating verbs (only "to"-datives) in my pre-final verb list as well as a number of ditransitive verbs (DO) from Mukherjee (2005)

tion in one or more of the English varieties under consideration in this study (see De Cuypere and Verbeke 2013: 171; Mukherjee and Hoffmann 2006; Schilk, Bernaisch, and Mukherjee 2012; Mukherjee and Schilk 2008). As Mukherjee and Hoffmann (2006) have pointed out, new varieties such as Indian English can display a different range of ditransitive verbs than, say, British English does. This includes verbs like *gift* or *advise* (Mukherjee and Hoffmann 2006: 164). Such verbs were necessarily included in my pre-final verb list. The pre-final verbs thus came from a diachronic, a syntactic, a variational as well as a semantic perspective. To my knowledge, no research has been conducted in verb complementation patterns in CanE, IrE, NZE, PhilE, and JamE so far. It remains to be part of future research to explore the extent to which verb complementation patterns in the aforementioned varieties deviate from British or any other English variety. For the current study, I work with the assumption that these VoE do not contain any alternating verb that occurs so frequently as to bias the results. If they include such a verb - so the hypothesis - it would be markedly frequently used and thus commented upon as a lexicogrammatical feature by previous linguists (e.g. in Kortmann, Schneider, Burridge, Mesthrie, and Upton 2004).

The pre-final verb list consisted of 225 verbs and was then further delimited. Previous studies have started off with a much smaller verb list at the beginning before delimiting the list due to the non-occurrence of certain verbs in the required constructions. However, Krifka (2003: 13) has pointed out that to rely only on verbal semantics to restrict possible variable context for the two dative constructions can be misleading as there are cases where other parameters influence the choice of constituent order in a way that could be called ‘untypical’ for certain verbs: “[...] there are also cases where information structure appears to override the semantic restrictions argued for above” (see also Bresnan, Cueni, et al. 2007). Hence, I work with the assumption that even though some verbs have been shown to be limited to one variant, for instance, in British English, there is a possibility that it is used in both variants in the more exotic varieties. The delimitation of the pre-final verb list was thus very tentative and based on corpus data whenever possible.

4.1.2 The final verb list

In a second step, I delimited the verbs for the final verb list (see Bresnan, Cueni, et al. 2007; Theijssen 2012; Cueni 2004). I only included verbs that occurred in either of the two dative variants in the dataset thus eliminating infrequent verbs. Since the aim of the project is to work with the 9 ICE corpora, each verb was looked up in all ICE corpora and subsequently

deleted from the verb list if it did not occur in the dataset in the form of either a double object or prepositional dative construction. The instances in ICE were verified for interchangeable context only, thus disregarding any particle verbs, spatial goals, etc. (see section 6 below).

In cases where only one variant was found in ICE or where the amount of data was not feasible for a manual search in a certain amount of time, the CORPUS OF GLOBAL WEB-BASED ENGLISH (GloWbE) was used to verify the alternation (see <http://corpus2.byu.edu/glowbe>). Mark Davies' GloWbE consists of 1.9 billion words and covers 20 varieties of English around the world. As this corpus is pos-tagged, the search queries were adapted to find prototypical instances; the threshold for inclusion (of the verb) was set to 5 for each individual variety. For the double object construction this included the search queries (a) [verb].[v*][P*], restricting the search for verb lemma followed by pronoun. In cases where not enough instances were found, this search was extended to (b) include a possible noun within the next 4-9 words to the right (collocation: [N*]). In cases where the verb occurred within the double object construction and was followed by a pronoun, the threshold was lowered to 3 instances for the verb to be included in the final verb list. In those cases where not a sufficient number of pronominal instances were found, search queries were further expanded to include (c) [verb].[v*] [N*] whereby the final [N*] could also be put into the collocation slot in order to allow for possible interrupting material between the verb and the noun (such as adjectives, tags, etc.), or (d) [verb].[v*] [N*] [N*] whereby wildcards (*) were used to allow for intervening material between the two noun phrases and the recipient and verb phrase. For the PD construction, an intervening to.[II] was added, leading to the search queries (e) [verb].[v*] [P*] to [N*] and (f) [verb].[v*] [N*] to [N*].

However, in cases where the verb didn't occur with a pronominal recipient in a double object construction, I assumed that, based on previous studies on the dative alternation, the probability of a verb occurring 5 times with a noun but not a pronoun in the recipient slot was next to null (see Bresnan, Cueni, et al. 2007: 80f.; NZE scheme: "pronominality of the recipient favours double object construction"). Hence, I didn't pursue the search any further. Whenever the verb was infrequent enough to allow for a feasible search through all its occurrences, the search was sometimes restricted to (g) [verb].[v*] to allow all possibilities. However, for some verbs this proved to be rather impracticable and the search had to be limited based on the restrictions outlined above.

If the verb had been shown to occur in one variant exclusively (e.g. verbs denoted as "to-datives only" in Levin 1993 or ditransitive verbs from Mukherjee 2005), the other variant

was looked for in GloWbE. The threshold was set to 3 instances with pronouns followed by a short noun phrase (noun with determiner) in at least one of the 9 ICE varieties that are available in GloWbE. Also, when the parsing results of the dependency bank had shown a verb to alternate (and alternation was exclusively found in the parser), the parsing results were manually verified in the data and the verb excluded or included depending on the evidence (e.g. some verbs have been shown to occur in the double object construction but after manual verification it became clear that all those cases were in fact misparsed constructions). Verbs that allow for a third variant (e.g. *provide him a paper*, *provide him with a paper*, *provide a paper to him*) have so far been included in the dataset. Here, variety-specific investigation into the matter of probability of the alternations is necessary.

With this more practicable approach it was possible to erase a total of 136 verbs, while keeping verbs that occurred only once or twice in one construction and/or one variety. The remaining 89 verbs were then used to find possible instances in the ICE-corpora. The list of verbs compiled is as follows:

accord, advise, allocate, allot, allow, answer, appoint, ask, assign, assure, award, bequeath, bid, bring, call, carry, cause, cede, charge, concede, convey, cost, deal, deliver, demonstrate, deny, develop, drop, entrust, explain, extend, feed, flick, flip, forward, get, gift, give, grant, guarantee, hand, impart, inform, issue, keep, lease, leave, lend, loan, lose, mail, name, offer, owe, pass, pay, permit, play, pose, post, prescribe, present, promise, propose, provide, quote, read, recommend, refuse, render, sell, send, serve, set, show, sing, slip, submit, suggest, supply, take, teach, tell, throw, toss, vote, wish, write, and yield.

It was necessary to apply such a wide range of known alternating verbs in the query search on the grounds that we deal here with new varieties of English that may show a difference in the use of their verbs with regard to their occurrence in double object and prepositional dative constructions (see for instance Mukherjee and Hoffmann 2006). As a previous study on Indian English has exemplified, verb complementation patterns vary across varieties and it is conceivable that one alternating constructions occurs more frequently in one variety of English than it does in another (Olavarra de Ersson and Shaw 2003: 156). Thus, a verb such as *gift* which has been shown to occur in the ditransitive pattern in Indian but not British English, may also vary in other varieties of English under consideration in this study (see Mukherjee and Hoffmann 2006: 163). Alternating verbs with a low frequency will not be captured by

my method in the compilation and filtering processes of the verb list. The ICE-corpora are generally too small for low frequency verbs and larger corpora will be needed in the future to analyse the dative alternation in different varieties of English with a focus on individual verbs (e.g. GloWbE, see Gries and Mukherjee 2010). However, it is not the aim of my thesis to cover all dative alternations that could possibly occur in one variety of English but to propose models of dative alternation into which all these other possible dative alternations could be fitted.

4.2 The perl script

4.2.1 Data format

The data had been cleaned and the tagging brought into a uniform appearance by Benedikt Heller. For ICE-GB, the parsing material was deleted, the raw text extracted and the corpus retagged with the tagger CLAWS 7 (see appendix). Thus, the cleaned tagged version of the nine ICE corpora now share identical denominations for the text file names, in each text file, every new utterance starts on a new line, and the markup for the text units are identical. The master copies of the ICE corpora are available on the QLV server “robin” and have been downloaded to a local folder. The text files of each corpus were copied and pasted into a separate folder to exert the perl extraction file upon them.

4.2.2 First extraction

Benedikt Heller generously shared his perl script for the genitive extraction with me. I then adapted those steps that were relevant for the dative extraction and deleted or commented irrelevant inputs out. In a first verification process, the dative variants were only extracted from ICE-GB, as I wanted to improve the perl script before using it on other corpora as well.

My set of 89 verbs was used to find the dative variants and a lemma list was additionally incorporated in the search query in order to include all possible instances of, for example, *give*, namely *give*, *gives*, *giving*, *gave*, *given*. Noun phrases were in a first step defined as specified in Benedikt Heller’s perl script for the genitives. A noun phrase is characterised by an optional determiner, optional pre-modifier and a noun as based on the CLAWS 7 tags, thus allowing for any word followed by the respective tag. Additionally, optional intervening interjection (UH) were allowed within the noun phrase structure. Pronouns were defined and incorporated based on their tag set.

The script used a regex search that looked for any string containing a verb from the verb list (in any form), followed by either noun phrase or pronoun followed by space followed by either noun or pronoun. For the prepositional dative construction, an intervening *to* was added between the two post-verbal constituents.

The search string looks as follows:

```
/\b${re_verbs}_V\w+\s($noun_phrase|$pronoun)\s (to_II)
\s($noun_phrase|$pronoun)/i
```

With this search query, I extracted a total of 3144 instances of double object constructions and 370 prepositional dative constructions from ICE-GB. Each variant was then filtered and false positives dismissed; unsure alternations were flagged and inserted on a separate excel sheet. For the double object construction, most false positives contained compound nouns or constituted monotransitive clauses followed by a relative clause without relativizer as in (1). Other unwanted instances include monotransitive clauses followed by an adverbial phrase (e.g. *next year*). The query search for prepositional datives was better in that respect. However, the dependency of the *to*-phrase was not always very clear and those instances where the second constituent attached to the first noun instead of the verb had to be discarded subsequently, e.g. (2).

(1) Actually it's not necessarily such a bad thing because the chances are that if they **leave it they'll** never do it. <ICE-GB s1a-005>

(2) I have **suggested two changes to the draft flyer**. <ICE-GB w1b-025>

After this first filtering step, I was left with a total of 290 prepositional and 580 double object constructions. These were then manually verified against the output of ICE-CUP to look for discrepancies in my perl script.

4.2.3 Verification against ICE-CUP

To verify the results and the accuracy of the perl script, the extracted double object and prepositional datives were compared with the output of the parsed ICE-GB, using the accompanying software ICE-CUP, which had been used for my previous parser comparison. Only the instances of the verb *give* were compared as *give* is the most frequent verb and I assumed that all errors that I find in the perl script for *give* cover most of the errors for the other verbs.

The problematic issues that I encountered in my perl script were the following:

1. **Tagging mistakes**

6 tagging mistakes were discovered in the prepositional datives of *give*, 23 in the double object construction. The parsed version of ICE-GB indicates a total of 423 double object and 117 prepositional dative constructions for *give*; thus, the amount of missed instances due to tagging mistakes are 5,1 % for the prepositional dative and 5,4 % for the double object construction.

2. **Capital letters for the verb**

These were initially not found but then the script was changed to include those instances as well, e.g. (3).

(3) Give it to me as an income. <ICE-GB s1a-079>

3. **Incomplete noun phrases**

Quite a few instances that weren't found by the script include what I termed as "incomplete noun phrases". They are incomplete in the sense that the head noun is missing. According to the tags, they consist mostly of quantifiers (tag DA) and optional determiners and adverbial modifiers. Thus, instances such as (4), (5), (6) and (7) were not found by the script.

(4) ... gives it to the _AT few _DA2. <ICE-GB w2b-013>

(5) ... given me a _RR21 lot _RR22 <ICE-GB s1a-002>

(6) ... given this _DD1 careful _JJ consideration <ICE-GB w1b-030#135>

(7) ... give the hours to those _DD2 that _CST are going to the regulars <ICE-GB s1a-083#64>

The inclusion of the tag _DA in the definition of head nouns improved the results in some ways but not completely. In accordance with Benedikt Szmrecsanyi, we resolved that the tagging mistakes, with which we missed around 5% of all variants, weigh more heavily than unrecovered incomplete noun phrases. As it is and cannot be the aim of this study to find *all* possible dative variants, an approximation should be sufficient. This was achieved by the subsequent adaption of the perl script (see below).

4. *Single letters*

The tag `_ZZ1` was at first not included in the definition of noun heads and thus instances such as (8) were not caught.

(8) ...given the `U_ZZ1 N_ZZ1` figures of civil ... <ICE-GB s2b-018#013>

5. *Special characters in word*

Special characters such as “&” and “;” had to be included in the word query to catch instances like (9):

(9) ...give you a better `&persand;_NULL` more relaxed ... <ICE-GB w1b-014#045>

6. *Intervening phrases*

Some instances were not found because of intervening adverbials (one word), adverbial clauses (e.g. *this weekend*) or prepositional phrases and clausal interjections (e.g. *you know*). These instances are few in numbers. Some examples include:

(10) ...giving way I suspect to other ... <ICE-GB s1b-054#065>

(11) ...giving it mainly to first aiders ... <ICE-GB w1b-018#020>

(12) ...giving way this weekend to the grimmer ... <ICE-GB w2e-002#018>

(13) ...gave me `sort_RR21 of_RR22` like a dinner ... <ICE-GB s1a-039#071>

Intervening adverbial phrases would need to be added to the normal query search with the string `[\w_RR\d*]` in order to find the above examples. However, and in accordance with the project by Joan Bresnan, Sali Tagliamonte and collaborators, we agreed to exclude a specific search string for adverbials. In cases where they were still found by the perl script, they were moved to a separate file and the adverb extracted in a distinct column. Adverbial noun phrases as in (12) were found in the search query for double object constructions and the sentence was then accordingly flagged as possible PD (see section 5.1).

7. *Complex noun phrases*

Complex noun phrases (e.g. genitives or prepositional phrases within one constituent) had not been incorporated in the script for the first extraction but were subsequently added in the verification process. This includes instances such as:

- (14) ... gave rise to plenty of the schoolboy ... <ICE-GB s2b-026#077>
- (15) ... giving a lot of trouble to these Australians. <ICE-GB s2a-004#179>
- (16) ... give one Saturday's contemplation to the moment of the failure. <ICE-GB s2b-028#003>
- (17) ... gave us more than <,> more than ... <ICE-GB s1a-029#125>
- (18) ... give us a_AT1 great_JJ many_DA2 books_NN2 ... <ICE-GB s1b-046#081>

8. *Missing parts of files*

Due to the retagging process of ICE-GB and because the annotators of the original corpus used special characters for their mark-up, some text was lost. These were noted and indicated as such to Benedikt Heller to allow him to retag ICE-GB again.

4.2.4 Adjustments to the perl script

While comparing the output of the perl script with the parsed ICE-GB dative variants for *give*, I subsequently adjusted the perl script to allow for those variants that were previously not found. These included (as indicated above) mostly complex noun phrases as in (14) to (18), special characters, capital letters for the verb, and single letters as in (8); complex determiners and pre-modifiers were allowed by adding conjunctions and the preposition *of* (CC, IO, CSN) in-between determiners and pre-modifiers; complex noun phrases were allowed by letting several nouns be interrupted by optional coordinating conjunctions, the preposition *of*, genitives with apostrophe 's (see ex. (16)) and general prepositions (tags CC, IO, GE, II). To allow for non-canonical word order in the pre-modification slot, the script was changed to search also for determiner - pre-modifiers - determiner-sequences as in ex. (18).

In order to account for and extract so-called “incomplete noun phrases” (incomplete in that the required head noun is missing), the set of head nouns was extended to encompass interjections (tagged UH), single letters (tagged ZZ1), after-determiner or post-determiner capable of pronominal function (tagged DA1/DA2), indefinite pronoun singular “one” (tagged PN1), and any other determiner which can be used in a pronominal function according to the C7 tag set (i.e. before determiner or pre-determiner capable of pronominal function, tagged DB; singular and plural determiners, tagged DD1/DD2). Hence, examples like ex. (7) can now be found by the perl script.

4.3 The first dataset —raw numbers before filtering

After final adjustments, I run the perl script over all nine ICE corpora and extracted the double object and prepositional dative constructions. Table (1) shows the raw numbers of all instances found by the perl script, including false positives. These were then weeded out in the next step. In the table, the raw numbers are called ‘double object constructions’ and ‘prepositional dative constructions’ respectively, as these were the constructions that the perl script aimed to find. However, as the script was rather widely cast to get as much recall as possible, the amount of *real* double object constructions is comparably small (see table 2).

Variety	double objects	prepositional datives
GB	3938	426
CAN	4152	424
HK	5612	662
IND	4465	720
IRE	3777	393
JAM	4033	485
NZE	4842	487
PHI	3985	513
SIN	4520	509
TOTAL	39'324	4'619

Table 1: Raw numbers of all instances before the first filtering step

5 First selection and filtering of the data —Syntactic constraints

5.1 Introduction

Because the perl script rendered a large amount of false positives, especially for the double object construction, an equally large amount of manual filtering is required. This is done separately for each corpus in a first filtering step. Whenever two variants occurred within the same utterance, the utterance was copied and pasted underneath in order to have one instance per row. This first filtering step of the raw dataset had nothing to do with the possible alternating context. Rather, the aim was to filter out any construction that wasn't a dative and

which didn't consist of a recipient and a theme. For syntactic decisions, I will refer to Quirk et al. (1985) for the definition of grammatical terminology and any other decisions regarding the syntactic category of specific elements (see section 5.2 for a list of exclusions).

So far, only ICE-HKE, ICE-CAN, ICE-GB, ICE-IRE, ICE-JAM and ICE-IND have completely undergone the first filtering step. In the "false-positive"-dataset, red was used to indicate *to*-datives (when filtering DO), yellow highlighting of the textID to indicate constructions with clausal objects, framed sentences are example constructions (used below), and dark olive green indicates object complements (see Quirk et al. 1985: 1198). When filtering the dataset of double object construction, only those instances of prepositional dative constructions were marked as red in the "non-dative set" that did not contain a double object construction, since these had to be kept in the filtered dataset of double object construction. In the filtered dataset itself, yellow highlighting of the textID was used for further discussion and light green for possible non-choice context.

Corpus external material, indicated by <X> and </X>, was excluded from analysis. In ICE-HKE and ICE-JAM, for instance, corpus external material is indicated within the speaker unit, making it easy to detect those instances. However, in some cases the speaker unit is indicated as corpus external material while the speaker is not (usually speaker 'Z' for external corpus material) or vice versa. Such instances were double checked and it was determined out of the context of the whole text whether the extracted instance was part of external corpus material (indicated by a preceding <X> and a subsequent </X>) or not. The mix-up was generally due to a missing \$ or because the POS-tagger had interpreted "<\$Z> laughs" as the next speaker being <\$Z>. Also, some instances had to be double checked with the original POS-tagged file due to the deletion of special characters in the perl script, for instance breaks marked as <,> or quotes marked as <quote>. Thus, sentences like (19), (20) and (21: self-correction) were excluded after second considerations when compared with their original POS-tagged speaker unit.

- (19) a. I well I as I as I've told you earlier I'm I've I haven't told you the tape but I told you earlier. <ICE-HKE s1a-040#313>
- b. <#313:1:A> I_PPIS1 <,> well_RR I_MC1 as_CSA I_ZZ1 as_CSA I_PPIS1 've_VH0 told_VVN you_PPY earlier_RRR I_PPIS1 'm_VBM I_PPIS1 've_VH0 <,> I_PPIS1 have_VH0 n't_XX told_VVN you_PPY <,> the_AT tape_NN1 but_CCB <,> <{1> <[1> I_PPIS1 </[1> told_VVD you_PPY earlier_RRR <{2> <[2> <,> </[2>

- (20) a. No I just have to give this a question. <ICE-CAN s2a-059#111>
 b. <#111:1:A> No_UH I_PPIS1 just_RR have_VH0 to_TO <,> give_VVI this_DD1 a_AT1
 <\$ZZ> <O> question_NN1 </O>
- (21) So either she has n't gone home or else she 's lost her my address. <ICE-GB s1a-093#222:1:B>

5.2 Exclusion of double object and prepositional dative constructions

In the first filtering step I followed previous approaches adopted by Bresnan and collaborators. The perl script was written in such a way as to already neglect unwanted instances (see section 6). The following is a list of constructions that were systematically excluded from the dataset, first grouping exclusions for double object and prepositional dative constructions together and then giving some details on variant specific exclusions (see Wolk et al. 2013; Bresnan and Nikitina 2009; Bresnan, Cueni, et al. 2007; Grimm and Bresnan 2009; Cueni 2004).

5.2.1 General exclusions

1. *Sentences lacking two overt constituents*

In such sentences, termed incomplete noun phrases, the head noun or the pronouns or anything that counts as head element of the noun phrase is missing;

- (22) I'll tell you <two words>. <ICE-CAN s1a-085#188>
- (23) Give you a bit of a <unclear words>. <ICE-GB s1a-065#89:1:C>
- (24) I have send uh almost eighty letters to uh. <ICE-HKE s1a-077#543>
- (25) I used to get rid of reporters you know cos I did n't want to feel as if I had to pay lip service to their <unclear word>. <ICE-GB s2a-050#33:1:A>
- (26) And h how about oral the oral examiners ask you some Chinese. <ICE-HKE s1a-042#906:1:A>
- (27) Something like that that 's why he was giving it a giving a party you know. <ICE-IND s1a-038#21:1:A>

- (28) It was supposed to start from today but the managing committee has asked yes that **give some more time to <one word>**. <ICE-IND s1a-083#74:1:A>

To define whether the head element was missing (e.g. indicated by <words> or <unclear words>), I concentrated on each individual speaker unit and did not consider the ensuing speaker units, even if these subsequent speaker units were by the same speaker. However, if the head element is incomplete but started or if the complements of the elements are given as in example (29), I included the whole construction. In the last sample, *cert* stands for the unfinished word *certificate* as indicated in the following speaker unit. As both speaker units are by the same speaker, I included the theme as *certificate* in my dataset.

- (29) <ICE-HK:S1A-079#65:1:A> Uh they will give me a <.> cert </.> <{2> <[2>
<,> </[2>
<ICE-HK:S1A-079#66:1:A> A certificate that uh to say that I have take this course
<{3> <[3> and </[3> things like that <{4> <[4> <,> </[4>

2. *Elliptic structures*

In elliptic structures, one or more of the three elements of the dative construction are missing or have to be inferred (see section 7.2 below). For the purpose of this study, the first complete structure was included in the dataset and the elliptic structure filtered out.

- (30) Probabilistic approach is to **assign probabilities of failure to fire protection measures** and frequencies of occurrence to unwanted events and to demonstrate that the likelihood of a given event occurrence is within acceptable limit. <ICE-HKE w2a-038#140:2>

3. *Constructions with non-canonical word order of the constituents*

I followed previous procedures and research in that, if encountered in the dataset, constructions with non-canonical word order were filtered out. These include instances with heavy NP shift (word order [V PP NP]), with a missing *to* and/or reversed order of the constituents (see (33));

- (31) And there 's a very recent decision of the Supreme Court of Canada and I 'll **read to you two paragraphs** to give you a little more detail as to what is maybe a word meant by a sexual assault. <ICE-CAN s2a-066#54>

- (32) In order to get the full support of the HKADC in HKAC 's application especially for programmes/projects which can not be self supporting through box office income or commercial sponsorship, we consider it vital to **bring to the attention of the Council members the present business and financial position of the HKAC**. <ICE-HKE w1b-017#101:5>
- (33) Well if you **give me it** tomorrow I might be able to do some tomorrow [] morning before I go to the rugby. <ICE-GB s1a-038#181:1:C>
- (34) I think we should **get one him** for his birthday and one for Christmas. <ICE-CAN s1a-098#51:1:B>
- (35) Will they **sold it people** that people at the Hollywood Road the customers. <ICE-HKE s1a-031#367>
- (36) And I find I **gave it some of my M Phil students** who had problems with usage and all that and they said it has been useful. <ICE-IND s1a-078#149:1:A>
- (37) Now which brings me my second point. <ICE-IND s1b-028#107:1:D>
- (38) And I do n't think that it 's something that two-or-three-words and I think the more people that participate to the blood donation programme the more blood hospitals have and the more they can give it people who need it. <ICE-JAM s2b-017#52:1:H>

4. *Constructions with a particle verb*

Constructions with particle verbs were excluded from the search query and not included in the perl script. However, some instances were still found and had to be manually discarded.

- (39) Actually I sent Tim Stark there on a mission once to **bring me back a black and white cookie** and he could n't find em them or he went to the wrong neighbourhoods and got them this these Italian cookies that were n't right. <ICE-CAN s1b-005#117>
- (40) I must remember to **give you your linguistics books back**, Laura. <ICE-GB w1b-009#124:3>

(41) He is in a that 's what he was telling me today see I want your uh uh draft like draft draft by January by the month of January by the end of January so that I I might uh rectify it and then I will do it uh I will **give it back to you** by mid Febraury so that you can get it final draft by uhm by the end of Febraury. <ICE-IND s1a-059#127:1:A>

(42) If you pay the average consumption we give you back some of the money. <ICE-JAM s1b-048#117:1:A>

5. *UH, YEAH as constituents*

Discarded sentences include constructions where the interjections *uhm*, *uh* or *yeah* with the POS-tag _UH constitute one constituent;

(43) Yeah you **told me yeah**. <ICE-CAN s1a-019#11>

(44) But many times those things are **presented uh uh to the public** and the public do n't understand what they 're about anyway. <ICE-CAN s1a-086#179>

(45) One was that I was being **given the opportunity to uhm** learn and develop uhm physical skills. <ICE-GB s1a-001#34:1:B>

(46) A real sir uh according to me uh this religion uh do not **tell us uhm** does not **tell us uhm** that uh we should uh kill anybody. <ICE-IND s1a-036#140:1:B>

(47) Even if we do n't go to their places they make it a point to get the sweets in the school or wherever it is otherwise they 'll pack it and they will **give you yeah**. <ICE-IND s1a-065#147:1:A>

(48) Sir uh I teach uh to FYSYTY and MA classes. <ICE-IND s1a-067#113:1:B>

(49) Actually he makes the whole thing uhm for your Voltas also have he has **given yeah to them**. <ICE-IND s1b-073#117:1:B>

6. *Answer/Question as constituents*

These sentences include constructions with an answer or question as constituent, such as *yes* or *no*;

- (50) Now if you **tell me no connection at all** Mary what a stupid question. <ICE-CAN s1b-024#141:1:A>
- (51) That 's right I wouldna would n't have went on the cart if she woulda would have **told me no**. <ICE-CAN s1b-061#165:1:B>
- (52) I mean she she **answers yes to everything**. <ICE-GB s1b-010#55:1:A>
- (53) Timothy Steele says the posters promise once Quebec **votes yes to sovereignty** the constitutional wrangling with the rest of Canada will be over once and for all. <ICE-CAN s2b-002#73:2:F>
- (54) I just **tell them Le Salon Mary**. <ICE-HKE s1a-014#6:1:A>
- (55) And I I I read uhm I just uhm read through the perspectives and then uhm the receptionist or or may be the teacher ask me uhm what area what area uhm I was s s uhm studying in then I telled I **told him Arts** but him he has misunderstood misunderstood me as Art. <ICE-HKE s1a-037#451>
- (56) Now uh if I **ask anybody What is in income**. <ICE-IND s2a-036#22:1:A>
- (57) History **tells us most emphatically yes**. <ICE-IND s2b-041#36:1:A>

7. *Adverbial phrases as constituents*

These are constructions where one of the noun phrases is an adverbial phrase (mostly denoting time) or a quantitative measure expression;

- (58) If it lasts for several weeks it could **cost more than a billion dollars a day** and lead to more than half a million [...]. <ICE-HKE s2b-018#97>
- (59) She **told me last night**. <ICE-CAN s1a-001#42>
- (60) So you **get six that way**. <ICE-GB s1a-010#146:1:B>
- (61) So we **get all the way to Molly**. <ICE-HKE s1a-096#495>
- (62) You might **ask yourself to what end** why she was involved in this from one at all. <ICE-HKE s2a-067#11>

- (63) Maya, is also **given freedom to a certain extent**. <ICE-CAN w1a-016#71:1>
- (64) In respect of the counter-claim My Lord that the defendant has made they obviously have not displayed or given any evidence to that respect. <ICE-JAM s2a-070#62:1:A>
- (65) In that case we **send bills every two min uh every two months**. <ICE-IND s1a-018#95:1:B>
- (66) I I **told him point blankly**. <ICE-IND s1a-094#296:3:A>
- (67) He was **telling one day to my friends** he had uh applied for this LIC LIC for two lakhs. <ICE-IND s1a-017#40:1:A>
- (68) Rakshit **playing square to Pasha**. <ICE-IND s2a-008#135:1:C>
- (69) It is ludicrous to believe that a university education should be offered free of charge to Jamaican students. <ICE-JAM w1a-004#46:3>

8. *Quotes/Titles from other sources*

These are sentences that are indicated and marked as quotes from another (clearly older and/or corpus-unrelated) source or as titles, etc.;

- (70) “If God **shows one of of this elect a chance of profit** he must do it with purpose.” said Richard Baxter.⁷ <ICE-HKE w1a-003#24>
- (71) In the postscript to another novel, published in serial form in 1846, Patrice La-combe ‘s La Terre paternelle, we read: ‘Let ‘s **leave their blood-soaked novels to the old countries** spoiled by civilization, and paint the child of the land the way he really is ...’. <ICE-CAN w2a-005#13>
- (72) Students occupying a central square chanted “Slobo is Saddam” and were led by a rock band in singing “**Give peace a chance**”.
- <ICE-GB w2c-019#10:1>
- (73) “The lock-out is illegal. <ICE-IND:W2C-010#X17:1> They are denying us our just rights.” </quote> <ICE-IND:W2C-010#X18:1>

- (74) I ‘ve been reading stories that apparently people are being arrested in America for wearing T-shirts some guy was picked up in a shopping mall for wearing a t-shirt with **Give Peace a Chance**. <ICE-IRE s1b-037#41:D>

9. *Indigenous words*

These are constructions where one constituent is an indigenous word, indicated as such in the markup (further annotation would be impossible in those cases), or where variety-specific elements (e.g. *him* instead of *his*) lead to a dative construction although it is not;

- (75) I told him very point blank that if you do n’t **give the commission** <indig>accha</indig> I’m going to tell ELPRO not to give you money again. <ICE-IND s1a-094#310:3:B>
- (76) Today we are **showing you** <indig>bharadwaja asana</indig>. <ICE-IND s2a-055#2:1:A>
- (77) Therefore the SCA would **send uhm the** <indig>muih jai</indig> **to the SCA** uh sorry uh the SCA would send the victims to the Po Leung Kuk where they would be interviewed. <ICE-HKE s2a-039#62:1:A>
- (78) She best way to uhm best way to **keep him uhm attention** is to him go on train and everything and stuff like that. <ICE-JAM s1a-066#168:2:A>

10. *Constructions with have got instead of get*

- (79) So you’re going to get Wall you’ve **got Wall Street duration estimates**. <ICE-CAN s1b-003#144>
- (80) He **’s got Hughes to his right**. <ICE-GB s2a-003#54:1:A>

11. *Problematic parsing*

These are constructions where several noun phrases are aligned and a correct parsing/understanding of the sentence is improbable —mostly with regard to the successful definition of the object boundaries and whether it is one or two NPs;

- (81) You just **read news the history of Hong Kong**. <ICE-HKE s1a-056#394:1:A>

- (82) And why is it so good for this woman to give a grandson to her mother-in-law because the grandson will **bring you life to you** and give you security in your old age. <ICE-HKE s1b-002#254:1:A>
- (83) Only you you **buy or sell some people sharing the goods to your friend one thousand point value**. <ICE-HKE s2a-053#102:1:A>
- (84) Just to **tell you the change is trans in transportation**. <ICE-HKE s1a-019#177:1:A>
- (85) Exam system what do you feel that uh you are **getting your rights the exam present kind of exams**. <ICE-IND s1a-011#101:1:D>

12. *Tagging mistakes*

Obviously, such constructions are generally not extracted by the perl script as the order of tags in such sentences does not correspond with the order of tags that the perl script queries for. However, in some instances tagging mistakes led to the inclusion of certain constructions where the tag correspond to my query search and which had to be manually discarded.

- (86) He said he had been presented with a demand **asking_VVG him_PPHO1 to_II hand_NN1** over power. <ICE-IND s2b-015#11:1:A>
- (87) And we **asked_VVD them_PPHO2 to_II rate_NN1** this on a scale of one to seven. <ICE-JAM s2b-029#40:1:A>

5.2.2 Specific exclusions in double object constructions

1. *Clausal constituents*

Following previous research, I also discarded sentences that include constituents (generally the last) realized as a clause. Among others, these were extracted by the script if the subordinate clause lacked the subordinate conjunction (*that*), or if the first constituent was followed by a relative pronoun (e.g. *who*) or a relative determiner (e.g. *what*) introducing a subordinate clause;

- (88) They didn't **tell me what** what I should go. <ICE-HKE s1a-074#85>

- (89) Mm hmm I ‘m gonna <going to> call and try to see uhm if uhm <French comment> and Easter Sunday <word> Do you wanna <want to> call Jean and Denille and **tell them we** ‘re coming. <ICE-CAN s1a-005#144>
- (90) But in order for them to trust you and to let them know **show you who** they are as a person you have to gain you have to establish that base of trust okay. <ICE-CAN s2a-026#9>
- (91) Mark I was **telling Rachel the deaconess** introduced you to Jean. <ICE-GB s1a-028#1:1:A>
- (92) So I asked him I **asked him what treatment** does is he going to give. <ICE-IND s1a-037#10:1:A>

Clausal objects were excluded from analysis even though it seems grammatically correct to say *I told what I did to the students*. However, as we wanted to stay as near as possible to the approach and procedures taken by Bresnan and collaborators, non-noun phrase constituents and the instances they occurred in had to be discarded.

Apart from Theijssen (2012), none of the previous studies on the dative alternation explicitly exclude clausal objects (see Quirk et al. 1985: 1056 on a definition). However, their definition of the dative alternation always involves the syntactic structure of VP NP NP (for the DO) and VP NP *to*-PP (for the PD). So even though they do not discard clausal objects from their dataset, they do *not* include them in their data extraction.

2. *Complex transitive clauses*

Complex transitive clauses entail object complements that refer to the same entity (mostly with *call*), and accordingly do not feature both a recipient and a theme;

- (93) We we do n’t **call it pubs**. <ICE-HKE s1a-005#X446>
- (94) We **called ourselves the Golden Group**. <ICE-HKE w2b-002#55>
- (95) They told the public it was a constitutional change that would **make Mr Tung a dictator**. <ICE-HKE w2b-015#125>
- (96) In the meantime Keith Joseph had **appointed me a part time Special Adviser to him** at the Department of Education. <ICE-GB w2b-012#142:1>

(97) I **consider it a privilege** and feel greatly honoured for the invitation to give the Convocation Address. <ICE-IND s2b-042#3:1:A>

3. *2 nouns = 1 entity*

Because the perl script searched for NP + NP constructions (as possible DO), it also retrieved compound nouns and appositions. These had to be discarded;

(98) And according to her it was at a time when she had indicated her intension to **leave Mr Hau**. <ICE-HKE s2a-065#20>

(99) Excessive intake of sodium can **cause heart problems**, including high blood pressure and heart attacks —not to mention water retention and bloating. <ICE-HKE w2b-029#121>

(100) And the first guy will **get all her commission** <laugh ends>. <ICE-CAN s1a-011#77>

(101) Unless you've **got hands the size of tractors** and you can <laugh> roll it over. <ICE-CAN s1a-013#141>

(102) Seventeen years ago I met for a day uhm a young man **called uh Nat David Schwartz**. <ICE-GB s1a-004#104:1:B>

(103) You should **ask George the security guard** how can you become an authorised person. <ICE-GB s1a-017#224:1:A>

(104) But suppose you **get a different kind of wife a nagging wife** from the some one who just ca n't adjust with you. <ICE-IND s1a-024#167:1:A>

(105) So in that way I **teach all the faculties Science Arts and Commerce**. <ICE-IND s1a-067#110:1:A>

(106) So first the writer has **given the fable the story about the ant and the grasshopper**. <ICE-IND s1b-012#45:1:A>

4. *Vocatives*

These include personal names as well as titles;

(107) You sing it you sing it you **sing it Roger**. <ICE-GB s1a-026#314:1:A>

(108) I mean one thing I would sort of **ask you Miriam** is this. <ICE-GB s1b-008#127:3:A>

(109) **Tell us Mr Shekhar** ah how do you attribute the right of regional po parties in India 's body politics. <ICE-IND s1b-044#4:1:A>

(110) Yes sir I can **tell you sir**. <ICE-IND s1b-060#32:1:C>

5. *Monotransitive clauses followed by another clause*

These are constructions with VP + NP followed by another clause that starts with a pronoun (e.g. (111)), a noun (e.g. (114));

(111) And uhm and she said and and where does it **get him you** know. <ICE-GB s1a-010#261:1:B>

(112) Then you just **give a topping you** give uh cheese. <ICE-IND s1a-007#204:1:A>

(113) Here even if you are ready to **pay them they** will not uh. <ICE-IND s1a-046#66:1:C>

(114) If in **taking this position the government** was acting on Beijing's instruction Hong Kong's autonomy would be only an illusion. <ICE-HKE s2b-031#107>

6. *Monotransitive constructions with a coordinated NP or genitives*

Due to the nature of the perl script (inclusion of optional coordinator, preposition e.g. *of*), these monotransitive clauses were also extracted;

(115) We will also **suggest a taxonomy of our own**. <ICE-CAN w2a-018#7>

(116) To celebrate Valentine 's Day a number of Yellowknife writers, inspired by Cupid and familiar faces in the audience, stood up in the museum theatre on the weekend and **read both the prose and poems of famous authors and their own more modest compositions**. <ICE-CAN w2e-004#111:7>

7. *Repetitions*

These repetitions including corrections where both NPs refer to the same entity;

(117) So I **ask the guy the guard**. <ICE-HKE s1a-074#154>

(118) Oh that's right **Get some sealer some wood sealer**. <ICE-CAN s1a-012#213>

8. *Reflexive constituents*

These are construction without a recipient as in (119) where the pronoun, e.g. *yourself*, refers reflexively back to the subject. They are distinct from “real datives”, as in (123);

(119) You **brought it yourself**. <ICE-HKE s1a-032#X349>

(120) Or maybe I'll **keep it myself**. <ICE-CAN s1a-029#67>

(121) So effectively you 'll be **paying ten pounds each**. <ICE-GB s1a-030#35:1:A>

(122) So you then get three thousand a year for the children and **make some contribution yourself**. <ICE-GB s1b-072#28:1:A>

(123) And you also have to court the muses of of manufacturing So you **get yourself a position** where you have enough of perhaps commercial success or enough of a name or enough credibility. <ICE-CAN s1b-043#109>

9. *constructions with home*

Semantically, “home” denotes a spatial goal but doesn't share the form of other recipients (with *to*);

(124) Did you **bring any home** with you. <ICE-CAN s1a-013#171>

(125) And Martin Brundle will **bring Jaguar number three home** in third place. <ICE-GB s2a-012#109:6:A>

5.2.3 Specific exclusions in prepositional dative constructions

1. *Pied-piped to*

This involves monotransitive constructions followed by a (relative) clause that starts with the preposition *to*;

(126) The centres have UI agents and employment counsellors whose jobs are to make certain that you **get all the benefits to which** you are entitled, while helping you get back into the work force. <ICE-CAN w2d-001#8:1>

(127) That is a very encouraging development and does very clearly **demonstrate the extent to which** employers have increasingly accepted responsibilities in the field of short-term sickness. <ICE-GB s1b-058#72:1:C>

(128) I wanted to **ask you to what** do you owe this gratification and love from them. <ICE-IND s1b-048#112:1:A>

2. *Stranded to*

This involves monotransitive constructions ending on *to* followed by a clause that starts with a pronoun or noun;

(129) A final thing is obviously with addresses like that you have a problem of if they 're very long or you 've got several people to **send messages to you** can get a lot of typing. <ICE-GB s2a-028#110:2:A>

(130) It just seems to be kind of ah a combination census and and kind of a list of who who should we **allocate these funds to who** 's gon na going to get this money. <ICE-CAN s1b-015#185>

(131) But police who you talk to and deal with and **given your statement to you** do n't recognise. <ICE-JAM s1b-070#102:1:A>

3. *Dependency*

These are constructions where the second (*to*)-constituent is dependent on the first constituent instead of the verb;

(132) Repeatedly governments has **denied access to public funded dental care** by the public on the grounds of gigantic cost. <ICE-HKE s1b-053#17>

(133) I was **refused entry to Beijing** by having my visa application rejected. <ICE-HKE s2b-023#69>

(134) Crime prevention Bureau officers have been **suggesting modifications to the layout** to make the bank less attractive to robbers. <ICE-HKE w2c-013#13>

(135) It 's a disturbing trend but according to the World Watch Institute in Washington D C the number of farm pests **developing resistance to pesticides** is growing rapidly. <ICE-CAN s2b-031#89:3:A>

- (136) So when the notion of Euclid and dimension poses a conceptual problem in understanding certain system the concept of fractal dimensionality **gives a manageable and convenient answer to such questions.** <ICE-IND s2a-045#79:1:A>

In the case of unsure dependencies, i.e. when it is not clear whether the last constituent is a complement to the verb or the noun, these are to be kept in the dataset. Unsure dependencies mostly occurred with regard to locatives (which were deleted from the dataset in the second filtering step anyway).

- (137) One may also **take a trip to Auroville**, which is about 8 / 9 kilometres from Pondicherry. <ICE-IND w1b-004#61:2>

4. *From-To*

These are constructions with a preceding or subsequent *from* before the *to*-constituent;

- (138) They have **changed from stem family, extended or joint family to the nuclear family of nowadays.** <ICE-HKE w1a-012#3>
- (139) On the same line we demanded in this house that we must have some a pipe line which can **carry the Bombay High gas from Maharashtra to southern India.** <ICE-IND s1b-053#100:2:B>
- (140) The bee expends a fair amount of its energy in **bringing pollen to a flower from a distant source.** <ICE-IND w2b-025#101:1>

5.3 Object extraction and verification

After the completion of the first filtering step, the verb, recipient and theme of each construction was extracted and copied to a separate column for each in the spreadsheet. An additional check was performed to verify again the inclusion of the individual variant in the dataset after the first filtering. Instances that had been previously marked yellow (instances that could maybe be thrown out in the first step) were verified with the original text, and non-alternating instances highlighted green. A further column added information on the reason why a particular instance was non-choice context (e.g. coordinated, fixed, beneficiary, etc.). Additionally for the prepositional datives, the discarded instances from the list of double object constructions (non-datives), which had been highlighted dark red to indicate prepositional datives,

were compared with the list of PD constructions and any missing *to*-dative was added. Of these previously discarded prepositional datives only those instances were added that would have passed the first filtering step. With this method, I was able to add around 15% of the PDs in each variety.²

Although I performed the object extraction before or simultaneously with the denotation of variable alternations, the next chapter elaborates first in detail the second filtering step (choice context) before expanding on the delimitations of the objects. This presentation order is logically conceivable in so far as the final determination of alternating variants should precede any further annotation, while the exact delimitation of the object boundaries remains secondary and only important for the annotation of length measurements. However, for practical and efficiency-related reasons, object extraction and choice context are in practice specified at the same time. Furthermore, it was possible with this approach —especially with regard to the prepositional datives —to define the choice context when focusing on the objects only.

At the same time that I defined object demarcations and identified alternating variables (or rather non-alternating variables as these were the ones to filter out), I used an online google spreadsheet, shared with Jason Grafmiller and Benedikt Szmrecsanyi, to discuss questionable instances / delimitations of constituents. Due to semantically ambiguous postmodification, the delimitation of objects proved to be particularly problematic.

Out of the total number as presented in table (1) I retrieved —after the first filtering step —a total of xxx number of variants.

6 Selection and Filtering of the data —the choice context

In order to analyse the alternation between the double-object construction and its prepositional equivalent, only constructions were taken into account where the syntactic alternation is semantically equivalent and grammatically correct. This is based on the assumption that grammatical alternation occurs within semantic equivalence and semantic differences between the alternatives do not exist (see Bresnan and Ford 2010: 170f. for discussion; see also Sankoff 1988). Note also that “the differences in the two constructions are preferences, not categorical regularities” (Bresnan and Ford 2010: 172). In order to facilitate comparabil-

²PD instances that occurred within the same utterance as a DO (and vice versa) could not be verified systematically. However, DOs present in PD instances were compared with the list of DOs, separately for each variety.)

Variety	double objects	prepositional datives
GB	747	424
CAN		
HK	1028	625
IND	753	
IRE		
JAM	804	413
NZE		
PHI		
SIN		
TOTAL		

Table 2: Number of instances after first filtering step

ity with previous studies conducted within the probabilistic framework, this work adopts the approach taken by Wolk et al. (2013); Bresnan and Nikitina (2009); Bresnan, Cueni, et al. (2007); Grimm and Bresnan (2009), Cueni (2004) and Theijssen (2012) with regard to the extraction and (possible) exclusion of relevant dative constructions in choice context.

After having extracted the verb, recipient and theme in the spreadsheet, I went through all instances again, filtered out the instances that had already been highlighted in light green and verified my dataset against the list of non-alternates as used in previous research by the aforementioned authors. The exclusion of non-alternates was based on methodological as well as semantic grounds and are further elaborated below.

6.1 Already excluded with the perl script

The following constructions, which are to be excluded for the choice context, were not found by the perl script. However, some instances did occur and were manually filtered out (first filtering step).

- non-canonical word order
- phrasal verbs
- constructions that lack two overt constituents
- passive phrases

6.2 Methodological constraints

Based on the methodological constraints as listed below, I filtered out instances that made annotation or further evaluation of the data impossible (see section 7.2 on ellipsis).

- if there is more than one verb, theme or recipient, e.g. coordinated themes or recipients;

(141) Beneficial may sell or assign all or part of the balance of your Account to **another person or entity** at any time without prior notice. <ICE-CAN w2d-010#75:1>

6.3 Semantic constraints

As Bresnan, Cueni, et al. (2007) have demonstrated, one's own linguistic intuition as to what constitutes a possible alternation does not always fully agree with the intuition of speakers in other varieties of English (especially since we are working with more exotic varieties of English). Hence, I only excluded instances that fell into one of the categories below. The exclusion of such invariable constructions assures that no other additional factors may influence the choice of variant (see Olavarra de Ersson and Shaw 2003).

6.3.1 Fixed expressions & Idioms

In order to verify whether an unsure case of a construction is fixed or constitutes an idiom, I entered the variant form in google and GloWbE (taking 5 instances as a threshold for inclusion). Excluded were:

- *ask them what: ...* : verified in GloWbE, not variable
- *ask NP to + event*, e.g. *ask them to lunch*: verified in GloWbE, not variable

(142) She sort of keeps asking them to things but that was twice she 's asked them to things has n't she. <ICE-GB s1a-019#363:1:A>

(143) but I did ask a few people I did **ask a few people at work to my birthday party** so. <ICE-GB s1a-081#57:1:B>

- *bring/get NP to their/his/her feet*: verified in GloWbE, not variable

(144) And Ah Some Toy his name disappeared off the infield one-word board quite quickly before they even got him to his feet now let us have a look what is going on over there. <ICE-HKE s2a-004#28:1:B>

- *bring tears to the eyes/cheeks*: verified in GloWbE, not variable

(145) it 's bringing tears to my cheeks. <ICE-GB w1b-001#80:2>

- *bring NP to justice*: verified with GloWbE, not variable

(146) He stresses that it may take longer to collect evidence but that also means a better chance of bringing criminals to justice. <ICE-HKE s2b-012#30:1:B>

- *bring it to the boil*: verified in GloWbE, not variable

(147) Well they bring it to the boil and whip it off the stove. <ICE-GB s1a-009#184:1:A>

- *This brings us to the question*: verified in GloWbE, not variable

(148) this will bring us to some questions about what can reasonably be expected from organised government operating within a framework of nations in which nationhood is regarded as incomplete unless it entails sovereignty. <ICE-GB s2b-041#63:2:A>

- *bring to life*: verified in GloWbE, not variable

(149) Although they are examples of the 'recreational' style which he does not usually support, they can be made to flutter when the central ridge is depressed, **bringing them to life** and so earning them a place in his renowned and charismatic teaching performances. <ICE-GB w2d-019#33:1>

- *bring matters to a head*: verified in GloWbE, not variable

- *bring to a halt*: verified in GloWbE, not variable

(150) Josh told his father of the miners who had brought his coach to a halt. <ICE-GB w2f-007#58:1>

- *bring NP to fruition*: verified in GloWbE, not variable

(151) It took more than Lloyd and Farrally to bring these dreams to fruition. <ICE-CAN w2b-001#18>

- *bring face to face*: verification not possible as this expression constitutes an idiom

(152) To be filled with that sense of wonder and perspective as one always is when **brought face to face** with the majesty of creation. <ICE-GB w1b-006#86:2>

- *bring matters to a head*: verification not possible as this expression constitutes an idiom

(153) The deadline was an arbitrary device, sought by Washington to **bring matters to a head** and is now serving that purpose only too well. <ICE-GB w2e-007#42:2>

- *bring/draw NP to the attention*; phrases like these always allow for the reverse order with *to*, i.e. *bring/draw the attention to NP*
- *cost + NP/pronoun + quantity of money or other value*, e.g. *it cost me \$10*; *cost him a frame*: verified in GloWbE, not variable
- *give way/rise to + NP*: verified in GloWbE, not variable; although *give way to NP* doesn't appear on first sight to be interchangeable, a verification with GloWbE turned out the following sentence, among others:

(154) And then these people also expect cars standing in the proper queue to give them way and accommodate them. <GloWbE, Indian English, <http://www.team-bhp.com/forum/street-experiences/3580-bad-drivers-how-do-you-spot-em-306.html>>

However, the number of such instances is very small, generally contains a pronoun in the recipient slot, and the alternative construction entails a slight difference in semantic properties. Thus, after careful consideration, constructions with *give way* are to be excluded from the dataset.

(155) The Chinese Communist Party Chief Jiang Zemin reportedly said when he granted an audience this morning to law-makers from the Cooperative Resources Centre but there seems to be a slim chance China would **give way to a faster pace of democracy**. <ICE-HKE s2b-015#5:1:B>

(156) but when you follow that through you 've got the means to **give rise to a change** in the method of accounting that 's adopted within a company. <ICE-GB s2a-037#122:1:A>

- *keep it to himself/herself/themselves*: verified in GloWbE for variation with *keep himself/herself/themselves* followed by another pronoun. However, sentences such as *keep*

himself a pet entail a different meaning than the *to*-phrase would. Furthermore, no variation with the personal reflexive pronouns was found in GloWbE.

- *leave it to* + NP + *to* VP, e.g.

(157) I ‘m actually disappointed that he did not go for the sales tax this year and have
to **leave it to his successor to deal with**. <ICE-HKE s1b-022#47:1:A>

: NOT VERIFIED YET

- *leave NP to their own devices*: NOT VERIFIED YET
- *leave it to your discretion*: verified with GloWbE, not variable; other verification with *leave* seems to restrict this verb to *messages*, etc. only

(158) I ‘ll leave it to your discretion. <ICE-GB s1a-091#145:1:B>

(159) Uhm I ‘d ‘[] better leave you to your conversations. <ICE-GB s1a-092#259:1:B>

- *lose ground to* + NP: verified in GloWbE with the alternative form and pronouns, i.e. *lose* + pronoun + ground, and only 2 instances found in total. This expression was hence considered as fixed. NOT VERIFIED IN GOOGLE
- *lose NP to NP* doesn’t seem to alternate in general, apart from (as JG suggests) maybe with words such as *war*, *game*, *competition*, etc. This needs to be verified though. For the moment, constructions with *lose* are generally filtered out in the second filtering step. The problematic issue with *lose* lays also in the realm of semantics: *This lost him the war* and *Lose the war to him* do not have the same meaning.
- *who paid them what*: 9 instances found in GloWbE for *pay what to whom*, but none found for *pay what to them* as these tend to render a main clause followed by a relative clause with missing relative pronoun, introduced by *what*: e.g. *They had to pay what to them seems a small sum of money*. So, no complete verification of alternation was possible in that sense and the item in question was regarded as invariable. As the structure of this instance shares similarities with *tell you what*, this decision can be justified.

(160) The Gazette says look we ‘ve got new technology we ‘ve got computers that can allow us to check up on what people claim they are being paid and uh compare it with what they actually received and who paid them what

- *pose national security/NP + NP* (e.g. *a risk, a threat*): verified in GloWbE, not all NP + NP constructions with *pose* are invariable. Hence, I subsequently verified all constructions with *pose* in GloWbE and google (i.e. *pose a problem to NP*) and left them in the dataset if alternating variants were found (threshold of 5 instances). For instance, *pose it a problem* or *pose him a threat* were generally found via google although not always in GloWbE. Other instances, such as *pose a risk to*, was found only once in GloWbE (*pose themselves a risk*) and thus discarded. In general, only four instances could be found with pronouns in the recipient slot of the double object construction and *a threat* as the theme. Encountering such low numbers, I verified possible alternation of constructions including *threat* via google and used the constituents as they stood. This way, *pose severe threats to the environment* and *pose considerable threats to health* had to be discarded as it did not turn up in google search (even if the search query included only the head nouns).

(161) Such colonisation programmes are carried out in Amazonia but pose severe threats to the environment. <ICE-GB w1a-013#68:2>

(162) Even in an affluent country such as Australia, mosquito-borne disease may pose considerable threats to health which will probably increase as a result of the greenhouse effect. <ICE-GB w2b-024#16:1>

(163) Uh and if the central government chose not to inform the Hong Kong government that this person in their opinion poses a risk to national security. <ICE-HKE s1b-035#47:1:A>

- *pose a danger to + NP*: checked in GloWbE —no evidence for variation found; in accordance with the guidelines above for *pose* I verified the occurrence in google and found 5 instances. Thus, this possibly fixed expression was considered as variable for my work.
- *read “a book” to her-/himself*: verified in GloWbE, not variable
- *submit NP to arbitration*: verified in GloWbE/google, not variable
- *take + NP/pronoun + time measurement*, e.g. *it took me one day*: verified in GloWbE, not variable
- *take exception to NP*, e.g.

(164) I think many people took exception to it it uh at the time. <ICE-HKE s1b-031#21:1:A>

: verified in GloWbE, not variable

- *take NP to heart*; verified with GloWbE, not variable

(165) The administration takes this policy to heart. <ICE-HKE s1b-054#86:1:C>

- *take NP to court*; verified IN GloWbE, not variable

(166) Surely there is something your government can do by taking the other party to the World Court for example. <ICE-HKE s1b-051#77:1:C>

- *To tell you the truth; I('ll) tell you what*;
- *tell me + date*, e.g. *tell me the year*: verified in GloWbE, not variable

(167) Which year could you tell me the year. <ICE-IND s1a-009#10:1:A>

6.3.2 Verified non-fixed expressions

These expressions were considered possibly “fixed”. Verification in GloWbE and google turned out, however, that they are variable and were thus left in the dataset.

- *ask you this*: followed by quote/question: verified in GloWbE and variation is possible (with pronouns). The same is true of *tell*, followed by a statement/question or *this*.

(168) Asked you this uhm when the new mem uh when the members were holding this meeting was there a stage Cantonese translation Recording fast-forwards. <ICE-HKE s2b-063#562:4:A>

- *tell you this*: even if the content of “this” follows after, verification on google and GloWbE turned out that this expression alternates with *tell this to you*
- *give it a try* = *give a try to it*; found on google
- *give it a shot*: CHECKED in GloWbE and there are a few (definitely more than 5) instances with *give a shot + NP*, e.g. *give a shot to power naps*, *give a shot to its USP*, *give a shot to photography*, etc. Verification with google also turned up more than 5 results (restricted to US domain). Hence, *give it a shot* was included in the dataset.

- *give birth to NP*: verified in GloWbE, and variation seems possible (e.g. *give me birth*)

(169) In 1947, my mother gave me birth, but she could not remember the exactly day. <GloWbE, Hong Kong English, <http://2007.tibetmagazine.net/english/2007-3/194C4B5A12237DB80F88D872E5F65029.html>>

- *give it another go*: CHECKED in GloWbE and only one instance found where *a go* occurs in the prepositional dative construction. A verification in google turned up several thousand instances with *give a go to* + NP, also numerous ones for *give a go to this* but none was found for *give a go to it*. Hence, I considered the instances found with *give a go* as variable.

(170) Adriana saw the Horizon documentary and was inspired to give another go to a less regimented version of the fasting diet. <GloWbE, GB, <http://www.dailymail.co.uk/femail/article-2195408/Starve-day-pig-Is-perfect-diet.html>>

(171) That 's right and the Olympic dream was one of the the reasons why uhm Heather Houston decided to **give it another go** this year. <ICE-CAN s2a-012#71:1:C>

6.3.3 To-To constructions

Prepositional datives are defined as “To-To constructions” when the reverse order of recipient and theme would still entail a *to*. These include phrases such as:

- *bring them to justice* = *bring justice to them*
- *draw this to the attention of the governor* = *draw the attention of the governor to this*
- *bring this matter to our attention* = *bring our attention to this matter*;

(172) Thank you again for taking the time to bring this matter to our attention. <ICE-HKE w1b-028#154:8>

(173) Mr. Sun is to be credited with having called our attention to it. <ICE-HKE w2a-007#75:1>

6.3.4 Predicative PP

Some kind of predicative PP, where the last constituent makes some changes to the first constituent (so far called ‘metaphorical locative’ for lack of an appropriate term; see Cueni (2004));

(174) I read that he turns in million-word manuscripts and that his editor, Maxwell Perkins, has to cut them in half in order to **get them to book size**. <ICE-CAN w2b-002#40:1>

(175) But Mr Sharon ‘s imprisoning of Mr Arafat, an elected leader, in his compound, and its destruction of the Palestinian infrastructure has **taken the conflict to a step beyond reason**. <ICE-HKE w2e-001#69:4>

6.3.5 Spatial goals

Prepositional dative constructions that included unambiguous spatial goals as recipients were also excluded from the dataset; whenever it was unclear whether the to-constituent referred to a goal or an entity, organization and such, the instance was included in the dataset (thus following the procedure taken by Cueni 2004).

(176) My Mom would n’t let me **bring my presents to school**, in case they got broken or dirty, but Santa came to my house too, and I can show you pictures of everything I got. <ICE-CAN w2f-001#143:1>

(177) And while they ‘d be quite happy to send their daughter to a boys ‘ school if they had the chance very few of them would send their son to a girls ‘ school. <ICE-GB s1a-012#223:1:C>

6.3.6 Beneficiary constructions

Double object constructions with verbs such as *get* that can be transformed into a beneficiary construction but not a prepositional construction were also excluded from the dataset. In general, it became clear from the context when the double object construction involved a beneficiary instead of a recipient. In the examples below, the two constructions follow each other closely, the first one being a double object and the second a prepositional construction with *for*, involving similar themes and the same beneficiaries. In such ambiguous contexts, the double object construction with *get* was discarded in the second filtering step.

- (178) a. We **get them uh typed photo copies** uhm uhm just a few of them. <ICE-IND s1a-060#114:1:B>
- b. With some great difficulty last year we could **get uh a couple of books for some students**. <ICE-IND s1a-060#135:1:A>

6.4 Intervening prepositional phrases and adverbials

Constructions with an intervening prepositional phrase (mostly starting with *with*) or with adverbs were extracted by accident in the extraction process. Such instances would allow more variability between the ordering of the constituents and in some cases do not even allow for the alternative variant. In accordance with the approach adopted by Tagliamonte (p.c.), sentences with intervening adverbials and prepositional phrases are flagged out and moved to a separate file.

- (179) The Prime Minister requested the Andhra Pradesh Chief Minister Dr M Chenna Reddy seated beside him at the press conference to **send his Finance Secretary with him to New Delhi** so that he could release the funds immediately. <ICE-IND s2b-003#19:1:A>
- (180) I have already **sent those cassettes by 1 st class mail to Dr Jerry Nelson**, the same day I also rang him up spoke to him. <ICE-IND w1b-010#170:1>
- (181) I'm reading it Treasure Island at the moment to my son. <ICE-GB s1a-013#65:1:D>

However, in cases where the prepositional phrase occurred at the end of the utterance, the variant was retained in the dataset. Here, the possible postmodification of the last constituent had to be considered (see next section).

- (182) I hope that you will **give the kind co-operation to the Dept. with the same vigor and involvement**, in future too. P.T.O. <ICE-IND w1b-017#137:1>

In contrast to previous guidelines (Cueni 2004), I retain imperative verbs, constructions with an object that is relativized upon (as I include in general all sentential postmodifications), and constructions with subsequent adjective phrases. Cueni (2004) also excludes instances of the verb *make* with a theme that is not *promise* or *offer*, and instances of the verb *do* that do not involve anarthrous nominal collocations (e.g. *do harm to someone*). However, *make* and *do* are already excluded from our search as they don't form part of the verb list.

7 Delimitation of the objects

When delimitating the object I was not only confronted with the task of determining the exact boundaries of each constituent (which proved to be somehow problematic in case of postmodifications) but also with the general difficulties one faces when working with spoken data, namely the issue of hesitators (*uhm*), self-corrections and repetitions. All of these issues I aim to address in the present section.

7.1 Postmodifications

The selection process of the choice context includes the extraction of the objects as mentioned above. The extraction is usually straightforward. However, in some cases the clear demarcation of the final object is rather problematic. Most final objects are followed by some kind of postmodification which needs to be either included or excluded within the object demarcations according to its semantic properties. To identify the postmodification accordingly, I differentiate between restrictive and non-restrictive modification: `RESTRICTIVE MODIFICATION` modifies the head of the noun phrase whereas the head can only be identified through the modification. `NON-RESTRICTIVE MODIFICATION` (adjuncts) modifies the whole VP and is not essential for the identification of the head (see Quirk et al. (1985)). Restrictive modification is always considered to form part of the object while non-restrictive modification is excluded. In addition to that, I also exclude any form of postmodification that modifies the whole sentence instead of just the preceding noun (e.g. temporal adverbials). To accommodate ambiguous instances where it is not clear whether the modification is restrictive or non-restrictive, extra columns denoted `REC_SHORT`, `REC_LONG`, `THEME_SHORT`, `THEME_LONG` are used to allow for inclusion and exclusion of the ambiguous postmodifying element. In those cases where the postmodification was considered restrictive and thus included within the object demarcations, a corresponding comment was entered into the column labelled `POSTMOD_THEME`, `POSTMOD_REC` and the classification of postmodification given (see table (3))

According to Quirk et al. (1985), the following postmodifications exist in the English language: prepositional phrases, nonfinite clauses, finite clauses (including relative clauses and restrictive appositive clauses), complementation, and postposed adverbs and adjectives. I will also add a short elaboration on general extenders in this section as these were encountered in the dataset as well. As with all other grammatical specifications, I rely on Quirk et al. (1985) for detailed elaboration.

Abbreviation	Type of postmodification
pp	nominal prepositional phrase, e.g. <i>all about it</i> , <i>admission to the college</i>
of-gen	of-genitive (not including partitives)
pp-clause	clausal prepositional phrase
pp-ing	pp with <i>-ing</i> clause
nonfin-ing	non-finite <i>-ing</i> clause
nonfin-to	non-finite <i>to</i> clause
nonfin-ed	non-finite <i>-ed</i> clause
restrRC	adnominal relative clause with overt relative pronoun (<i>that</i> , <i>which</i> , <i>who(m)</i>)
zeroRC	adnominal relative clause with zero relative pronoun
appthat	appositive clause with expressed <i>that</i>
appzero	appositive clause without expressed <i>that</i>
appnominal	restrictive appositions in noun phrase
postadv	postposed adverb, e.g. <i>here</i>
postadj	postposed adjective
ge	general extender

Table 3: Types of postmodification used in annotation

7.1.1 Prepositional phrases

According to Quirk et al. (1985), prepositional postmodifications are by far the most common in the English language. Most of the objects in our dataset are modified by *of*-phrases or by a prepositional phrase that has the preposition *about*, *over*, etc. as head. As with other postmodifications, prepositional postmodification can be subdivided into restrictive and non-restrictive postmodification. When non-restrictive, prepositional postmodifiers share the characteristics of adverbial expressions (see Quirk et al. 1985: 1286) and can be positioned at the beginning of the sentence. Apart from some partitive expressions (e.g. *a few*), *of*-phrases are generally restrictive. Prepositional postmodification can also include sentential clauses such as (189). In the case of *-ing* clauses within prepositional phrases, prepositional postmodification can also be paraphrased by an appositive structure (e.g. *chances of winning = chances that he will win*), while this is not the case for postmodification with an infinitive clause (see below). These can be paraphrased, on the other hand, as prepositional phrases with *of*: *the power to vote = the power of voting*; but not *the power that he can vote* (e.g. (192)). Hence, when encountering prepositional phrases with *-ing* clauses one has not only to consider the prepositional head but also the classification restrictive versus non-restrictive.

- **Restrictive prepositional clauses**

(183) I 'll tell you **a funny story about working class** later. <ICE-GB s1a-037#11:1:A>

(184) Tell tell me a little bit about this priest thing. <ICE-CAN s2a-025#77:1:B>

(185) Well I 'll tell you all about it because he 's taking me laugh. <ICE-CAN s1a-036#199:1:A>

(186) I did n't tell you **the subject of my lecture in the law school into which I walked**. <ICE-GB s2a-044#76:1:A>

(187) A driver can raise the car which is useful in conditions like these or engage sport mode which drags the car down on its haunches giving it **a torque stance through the bends**. <ICE-GB s2a-055#47:2:A>

(188) Meanwhile, a visiting professor from Ithaca College, Inar Holmes offered him **admission to the Ithaca College** on condition that he join his troupe which had performed all over the world. <ICE-IND w2c-011#98:3>

(189) With this commercial photograph it gives us **a very good idea of how two out of the three temples at Paestum are actually laid out in relation to the rest of the city.** <ICE-GB s2a-024#91:1:A>

(190) Above all, it needed a vision of the future which would unite management and workers to give them **a sense of belonging to an enterprise** that had challenge, inspiration and a sense of purpose. <ICE-GB w2b-016#12:1>

(191) Uh but what what has been saved or salvaged gives us **some idea of what can be done with some of these tools.** <ICE-IND s2a-30#60:1:A>

(192) ... give him a a chance of regaining the <ICE-IND w2c-017#38>

- **Ambiguous restrictive prepositional clauses**

(193) There are two two directories that I can direct you to uhm which will give you **the first lead on that.** <ICE-GB s1a-035#132>

(194) ...I wish you **success in finding a suitable career opening.** <ICE-GB w1b-019#51:4>

(195) ...to give AI full credit **for being** a science or technology ... <ICE-GB w2a-035#12:1>

- **Unambiguous non-restrictive prepositional clauses**

As with the subsequent postmodifying clauses, prepositional postmodification can also attach to the first instead of the last constituent, in which case the whole clause/object is interrupted by the second object. Here, the postmodification is included in the object demarcation (and hence entered within the same column), while an additional column labelled ENDMOD preserves the information of split modification.

(196) No matter how specific the criterion is, any need-based selection tends to give **a good deal of power** to the government officials **over the lives of potential recipients.** <ICE-IND w2b-015#24:1>

7.1.2 Nonfinite clauses

Postmodification by nonfinite clauses can take the form of either *-ing* participle (e.g. (200)), *-ed* participle (e.g. (202)) or infinitive clause (e.g. (197)) (Quirk et al. 1985: 1263). With regard to the *-ing* and *-ed* participle clauses, these correspond to relative clauses where the relative pronoun is the subject, and are usually restrictive. Infinitive clauses, on the other hand, correspond to relative clauses where the relative pronoun is subject, object, adverb (e.g. (203)) or complement. These latter are the ones commonly found in the dataset of the dative constructions with such antecedents as, for instance, *time*, *opportunity*, *chance*, *option*, *power*, *ability*, *right*, and so on. Each of these three nonfinite clauses can also occur in non-restrictive context but only if they correspond to relative clauses with a subject relative pronoun (e.g. *the man, crossing the street; the light-bulb, invented by Edison; these students, to be found in the library . . .*). Non-restrictive relative clauses can also be moved into sentence-initial position without change of meaning (Quirk et al. 1985: 1270f.).

Appositive postmodification by nonfinite clauses is further possible with either infinitive clauses or the *-ing* participle clauses (e.g. *the appeal to give blood*): In infinitive clauses the subject has to be inferred from the context (e.g. *the appeal (for us) to give blood*), while *-ing* participle clauses are generally complements of prepositional phrases and are discussed in the section above (Quirk et al. (1985)). In appositive postmodification, *-ing* participle clauses can be paraphrased as appositive clause with *that*, for instance *the hope of winning the game = the hope that X wins the game*, while infinitive clauses can only be paraphrased as prepositional phrases with the corresponding *-ing* clause but not with a *that*- clause, e.g. *the attempt to win this game = the attempt of winning this game*, but not *the attempt that he wins this game* (e.g. (195))

Similarly to the approach taken with all postmodifications, I included unambiguous and ambiguous nonfinite clauses within the object demarcation, while discarding non-restrictive nonfinite ones. The objects with ambiguous nonfinite postmodification are further entered into a separate column without their postmodification (THEMESHORT, RECSHORT).

- **Restrictive nonfinite clauses**

(197) . . . allowing all council tenants **the right to buy their own properties** . . . <ICE-GB w2c-017#13:1>

(198) I think Opera North jumped the bicentenary gun a couple of years ago because they gave the musical world **the opportunity to reassess La Finta Giardiniera**.

<ICE-GB s1b-044#44:1:B>

(199) He was already hurling the vehicle around the first bollard, the four-wheel drive giving him **the ability to force his way through the carefully prepared chicane** much faster than an ordinary vehicle. <ICE-GB w2f-012#136:1>

(200) If you do n't need glasses the optician will give you a statement **saying** this. <ICE-GB w2d-001#41:1>

(201) Starting tonight TVB News will bring you a special series **called** Hong Kong 's Olympic Dream. <ICE-HKE s2b-009#96:1:B>

(202) A $10 million fund would also be established to provide immediate interest-free loans to employees **injured** in work-related accidents. <ICE-HKE w2c-011#113:5>

- **Ambiguous restrictive nonfinite clauses**

(203) ... it 's very hard and it takes me a lot of time to study. <ICE-HKE s1a-017#220:1:A>

(204) He just gave his friend **enough time to get out of the area** and told me to get out there and not to bother coming back until I came back with some money. <ICE-GB s1b-049#177:1:B>

(205) It si **showed you him sitting** in restaurants with the mobile telephone. <ICE-IRE s1a-045#241:A>

(206) You recently sent us a letter **informing** us of unacceptable performance of our Telecom Security intruder alarm. <ICE-GB w1b028#106:9>

- **Unambiguous non-restrictive nonfinite clauses**

(207) ... she can give me some mark to improve my English. <ICE-HKE s1a-037#9:1:A>

There are, however, also examples where the nonfinite clause is not directly attached to the last object, for instance in examples (208) and (209). In those cases, as in all the other cases of end-modification, the nonfinite clause was added to its corresponding object in the column RECLONG or THEMELONG if restrictive and an supplementary comment was added in the column labelled ENDMOD

(208) Give chances to them to find out. <ICE-HKE s1a-060#395:1:C>

(209) Well the the the the uh uh decision of the MPC gives the power to the preparatory committee to establish the first legislature. <ICE-HKE s1b-050#172:2:A>

(210) Not the Irish, that 's for sure: Uncle Mike would eat his own shoes if you sat him down at the table and served them to him covered with ketchup. <ICE-HKE w2f-018#6:1>

7.1.3 Finite clauses

Quirk et al. (1985) distinguish two major types of finite clauses, namely RELATIVE CLAUSES and APPOSITIVE CLAUSES. **Relative clauses** can be further subdivided into three different types, two of which we encounter when extracting objects and their postmodification: ADNOMINAL relative clauses postmodifying a noun, SENTENTIAL relative clauses modifying a clausal antecedent, and NOMINAL relative clauses of the type *What surprises me is that he likes her* which we do not encounter in the dataset. Quirk et al. (1985) further distinguish adnominal relative clause that are restrictive and those that are non-restrictive. The restrictive relative clauses are of importance for this study as they form part of the object and are included in the object demarcations. Non-restrictive clauses are, on the other hand, excluded from the dataset.

Relative pronouns in restrictive adnominal relative clauses can either function as the subject (211), object (213), complement or adverbial (or complement in a prepositional phrase that functions as adverbial, e.g. (126)) of the relative clause. This pronoun can take several forms, namely *that*, *who(m)*, *which* or zero. Adnominal relative clauses with zero relative pronoun are always restrictive.

- **Restrictive relative clauses**

(211) I sent them **a questionnaire containing** a checklist of symptoms of depression often only requiring a simple yes no answer whether they had had those symptoms or not. <ICE-GB s2a-033#45:2:A>

(212) If you have two methodologies in any area of science when you have a response and a quotes dose that means a concentration or amount of a substance and you have a methodology A which gives you **that dose response curve shown here** and that methodology B which gives you that dose response curve in many. <ICE-GB s2a-042#114:1:A>

(213) Here ‘s uhm I ‘ll just show you **the actual setup we ‘re [] going to be using**. <ICE-GB s2a-035#96:3:A>

- **Ambiguous (non-)restrictive relative clauses**

(214) Now, I send you one of the photos which was taken during the Geography field camp last month. <ICE-HKE w1b-002#113:3>

- **Unambiguous non-restrictive relative clauses**

(215) Once it is in the stall, an automatic dispenser may give the cow a ration of concentrated food, which may be calculated to meet the needs of each individual cow. <ICE-GB w2a-033#66:1>

The other type of finite postmodifications is the **appositive clause**. The appositive clause resembles the restrictive relative clauses in that it can be introduced by *that* as in . . . *the fact that I had seen him yesterday* (Quirk et al. (1985)). According to Quirk et al. (1985) this resemblance is motivated by the similarity in loose connection between the apposition/relative clause and the sentence, the possibility of expanding an appositive into a relative clause, and the requirement for coreference between the apposition/relative clause and the head noun. Examples, such as *Anna, my best friend* “may be compared with nonrestrictive postmodification, in particular nonrestrictive relative clauses” (Quirk et al. 1985: 1301). Like the adnominal relative clause, the appositive clause can thus be divided into restrictive and non-restrictive appositive clauses. Irrespective of this division, each appositive clause is introduced by the conjunction *that* (or zero) and has an abstract noun as antecedent (e.g. *fact, claim, answer, idea, hypothesis*, etc.)

After having discussed appositive clauses, appositive nonfinite clauses and appositive prepositional phrases, a few words are due on the nature of apposition itself. In addition to appositive clauses, one also encounters nominal apposition whereby two nouns share coreferentiality. In non-restrictive apposition, the two linguistic units are separated by intonation or punctuation—they contribute relatively independent information (Quirk et al. 1985: 1304). In restrictive apposition, the two linguistic units share the same information unit and they are both necessary for identification (e.g. (222)). In accordance with the approach taken above, non-restrictive appositive clauses such as the (bold) ones in the examples below, are discarded from the object demarcations.

(216) Yiu was born in Xinhui, a Guangdong village where her father gave her a boy ‘s name, **Yick-cheung**, as she was the family ‘s only daughter. <ICE-HKE w2b-023#101:1>

(217) Now could you tell me something about your family background **your parents and how many children are there in the family and so on**. <ICE-IND s1a-009#99:1:A>

Included, however, are restrictive appositive clauses such as:

(218) He give me **an impression that he has a strong passion** towards the motherland and was a very resourceful person too. <ICE-HKE s2b-024#7:1:A>

(219) Many scout leaders believe using mahjong to raise money would give many youngsters **the impression gambling was an acceptable past-time pastime**. <ICE-HKE w2c-020#17:2>

(220) The results seem to lend support to **the hypotheses that a typological transfer of topic-comment structure so commonly found in Chinese varieties is attested**, especially those with lower proficiency of English, [...]. <ICE-HKE w2a-006#40:1>

(221) We should n’t give the people of Northern Ireland the **impression that we may return to a golden age**. <ICE-IRE s1b-026#31:C>

(222) The Labour leader Neil Kinnock has been in Westminster this morning and he **gave his reaction to our political correspondent Stephen Richards**. <ICE-GB s2b-018#80:2:A>

(223) And the Latin word for those huge lengths of papyrus glued together which you have to unroll in other words a scroll is a volumen so this **gives us our word volute**. <ICE-GB s2a-024#50:1:A>

(224) I lent a book to Jill]] Yeats ‘s daughter. <ICE-GB s1a-013#151:1:C>

7.1.4 Postposed adverbs and adjectives

Postposed adverbs and adjectives follow the head noun instead of preceding it, e.g. *the way back*, *a man big as a giant*, *this table here* or *a man taller than John*, *someone bigger*. These are included in the dataset in cases where they are restrictive.

(225) Unless you’ve got **hands the size of tractors** and you can <laugh> roll it over. <ICE-CAN s1a-013#141>

- (226) If you have two methodologies in any area of science when you have a response and a quotes dose that means a concentration or amount of a substance and you have a methodology A which gives you **that dose response curve shown here** and that methodology B which gives you that dose response curve in many. <ICE-GB s2a-42#114:1:A>

7.1.5 General extenders

General extenders are short phrases such as *and something*, *and stuff*, or *that kind of thing* that follow the clause. They can either modify the whole clause or the noun phrase immediately proceeding it. General extenders are included in the dataset in the column RECLONG, THEMELONG.

- (227) laugh They might they might not appreciate that during the holidays No I ‘m just kidding Uhm ya I ‘ll **send him a card or something** ... <ICE-CAN s1a-076#108:1:A>
- (228) Lend them **a couple of one-legged children or something** you know. <ICE-GB s1a-095#34:1:A>
- (229) Now tell me something about yourself your achievement and so on and so forth. <ICE-IND s1a-045#99>

7.1.6 Incomplete constituents

As previously mentioned, incomplete noun phrases, i.e. those without a head noun, were filtered out in the first filtering step. There are, however, cases of incomplete constituents, where modifying elements are missing. We resolved to include them in the dataset as they —so far —only constitute a problem for weight measurements. Indication of missing complements was only included within the object demarcations if the elements clearly constitute complements. However, in cases as (232), the object demarcations were limited to the head noun only.

- (230) They are the restitution made because an economy no longer **offer each and every citizens the opportunity to earn a <one-word> decent livelihood**. <ICE-HKE s1b-053#104:1:C>
- (231) They will laugh at me and I will not talk about this but in at this moment I can **tell you this about**. <ICE-HKE s1a-054#170:1:A>

(232) And the the old man said oh I **teach you a method** <one-word>. <ICE-HKE s1a-082#211:1:A>

7.2 Ellipsis

Ellipsis is a grammatical category used to define grammatical omission in cases where words are missing from the grammatical structure and their meaning is understood or implied. In contrast to substitution, where one word is replace by another, elliptic constructions are grammatically defective as “some normally obligatory element of a grammatical sentence is lacking” (Quirk et al. 1985: 885). There are certain criteria under which circumstances an omission can be called an *ellipsis* (Quirk et al. 1985: 884). In the dative constructions, we encounter mostly elliptic second constructions, where the verb is not repeated but implied, as in (233). Elliptic structures had to be excluded as they do not contain all three required elements overtly (verb, recipient, theme) and at least one of the elements has to be inferred. On the other hand, sentences with substitution instead of ellipsis were included in the dataset (e.g. (236)).

(233) In the Winter 1995 semester Giovanni assumed a number of duties in the Faculty ‘s Computer Graphics Laboratory, **providing** fundamental software maintenance on that facility ‘s workstations as well as **instructional assistance to its student users**. <ICE-CAN w1b-030#56:7>

(234) The government also plans to introduce an environmental bill of rights **giving people** the power to sue polluters and **a safe drinking water act**, he said. <ICE-CAN w2c-008#89:7>

(235) As required by law, the new owner should **deliver** the original of the notice to the Transport Department, and **the former owner the duplicate**. <ICE-HKE w2d-009#62:1>

(236) He ‘ll take his paperback out of his pocket &; and read a little while &; and not seem to **pay any attention to Velma** while she **pays none to him**. <ICE-CAN w2f-004#158>

(237) In 1980 censuses **gave** Brazil a population of 119 million, **Mexico 67 million**, Argentina 28 million and Colombia 26 million. <ICE-GB w2a-019#7:1>

(238) Just like the Agni has been launched in the space and I ‘m sure sir sir this government will be able to **give** a new thrust to our <one word> **a new thrust to our economic**

structure a new thrust to our industrial strength so that there is economic prosperity in this country then the peaceful the political situation in this country. <ICE-IND s1b-059#42:1:A>

(239) In a statement issued in Johannesburg the A N C leader **extended** deepest condolences to the bereaved family and **sympathies to the people of India**. <ICE-IND s2b-020#78:1:A>

(240) In Canada we allow serial murderers we allow convicted rapists we allow repeat offenders we allow petty thieves the full protection of the court system and the full protection of the law. <ICE-CAN s2b-008#90:2:D>

Whenever such elliptic structures occurred after a valid example of a dative variant, the first variant was included in the dataset and the elliptic structures dismissed. Hence, in example (237), I included *gave Brazil a population of 119 million* but excluded the other four constructions; in example (238), the first variant includes a missing constituent (<one word>) and thus none of the constructions was included in the dataset; while in example (240) the complete dative construction follows the elliptic constructions and is placed at the end of the whole sentence. Such instances had to be excluded as well as the choice of construction is heavily biased by the preceding elliptic constructions.

In general, elliptic structures have to be distinguished from coordinated objects where two or more NPs constitute one object, while in elliptic structures, the elements constitute separate objects. Thus, I had to consider a distinct treatment for instances with repeated *to*-phrase, because, apart from elliptic constructions with a missing verb, some instances in the prepositional dative construction also contain a second *to*-phrase and appear different from simple coordinated noun phrases, for instance, *I give roses to Mary and Sally* (together) versus *I give roses to Mary and to Sally* (separate). According to Quirk et al. (1985), coordination requires strict ellipsis rather than standard ellipsis; ellipsis is thus also possible in coordination. Furthermore, according to Quirk et al. (1985), coordination and ellipsis are complementary approaches to analyse syntactic structures with coordination focusing on sentence structure and ellipsis on intersentential relations (Quirk et al. 1985: 942). Finally, purely coordinated NPs can be found in cases where the attempt to fill the elliptic gaps (supplement words that were left out in the elliptic construction) renders a semantically different sentence, e.g. *The national flag of Japan is red and white*. Assuming an elliptic structure, we would have a sentence like *The national flag of Japan is red, and the national flag of Japan is white*, which is different

from the first sentence. Hence, in the examples below, the recipients are semantically and syntactically not coordinated and are considered elliptic constructions.

- (241) Could she describe uhm whether any of the negotiations and the European countries involved ever consulted any one of their Parliaments or or **presented any document whatsoever to any of their national parliaments or indeed to the European Parliament** during the course of these negotiations. <ICE-GB s1b-054#25:1:C>

In sum, I considered cases of PDs with a repeated *to* as containing a valid instance (for the dataset) and a (to be discarded) elliptic structure.

On the other hand, two NPs are considered to be coordinated if a premodifier modifies them both and/or if they are considered to constitute one entity.

- (242) I will give more I will **give more uh paperwork and group work to them**. <ICE-HKE s1a-084#17:1:A>

- (243) Later PC one nine two seven two showed the items he found in the trousers and the and the trousers and the blue t-shirt **to Mr Chin and Madam Pau**. <ICE-HKE s2a-062#169:1:A>

In the case of comparative clauses with elliptic elements, I always took the first variant found in the construction, leaving out the comparative elements. For instance, in a sentence such as *I gave him more apples than her*, I included the verb *give*, the recipient *him* and the theme *more apples* in the dataset. → SUGGESTION: the comparative modification could be included as postmodifying phrase (Check: Quirk et al.). So far, no grammatical explanation for *than NP* was found and I added the information about it being a comparative clause in a separate comment column.

7.3 Intervening constituents —hesitators

The constituents were sometimes accompanied by a hesitation marker (*uh*, *uhm*). In such cases, and for the object extraction, the intervening material was always counted towards the following constituent and left in the column REC_{LONG} or THEME_{LONG}. Since they indicate some sort of linguistic processing which then leads up to the next element in the clause, their inclusion in the subsequent constituent can be justified. In prepositional dative construction, any such element following the theme (and even if occurring before *to*) are counted towards the recipient. Any hesitation markers that follow *after* the complete construction are not

included. Columns REC_{SHORT} and THEME_{SHORT} are provided in the spreadsheet to enter the object without any of these pragmatic elements.

(244) So once we feed a sound or any sound segment to the machine it **gives us uh the visual impression of the sound** on the monitor. <ICE-IND s1b-076#41:1:C>

7.4 Self-corrections and repetitions

For the object demarcations in case of self-corrections and repetitions, we concluded to always go for the first completed overt prepositional or double object construction (thereby including any repetitions of parts of the intended object).

In the case of SELF-CORRECTIONS and in accordance with BS (see PhD meeting no. 4), we use and annotate always the corrected version. For instance, in example (245) the recipient is *yourself* and not *yourselves*, while in example (246), the complete (corrected) dative construction is *give them four exams*. Also, in example (247), we ignored the beginnings of the sentence with *give me* and *tell* as these do not constitute whole dative constructions. Instead, we only extracted the last dative *gave you that feeling*. The same goes for example (248), where the theme is *any sense of security*, not involving the repeated but incomplete first theme constituent *any security*. The same approach was taken when extracting the objects in example (249). Here the extracted construction is verb = *give*, recipient = *people in in the countryside*, and theme = *access*.

(245) Or sort of book yourselves **give yourselves give yourself** some complete free days. <ICE-CAN s1a-047#25:1:A>

(246) Uh one strange story that that came out in a cegep is say for example you **give your students an exa give them four exams** during the semester and we have this student who gets he's just passing gets six on each exam. <ICE-CAN s2a-027#71:2>

(247) Okay okay you **give me you tell part of the mo part of the movie** which particularly gave you that feeling then I'll relate that. <ICE-HKE s1a-039#220:1:A>

(248) They do n't give me **any security any sense of security** yeah. <ICE-HKE s1a-057#487:2:A>

(249) it 's quite clear that farmers are very happy uhm or most farmers are are very happy I 'm glad to say to **give access to the country to people in in the countryside**. <ICE-GB s1b-037#68:1:D>

More complicated cases are utterances like example (250), where the speaker starts with *a min*, corrects himself and says *a sec* and then finally completes the constituent by saying *a minute*. So even though we have a repetition here (*a a a min minute*), due to the self-correction (*a sec*), I only counted *a minute* as the theme. The same is true for example (251), where I took the intended, though repeated theme *the the the cassette tapes*. The speakers first utterance *the radio tape* is corrected to *the cassette tapes* preceded by some repetitions of *the*. Hence, in all cases of self-corrections and especially in those cases where premodification is added to the object (e.g. *any sense of, people in*), I always went for the final intended object.

(250) Now I just give you **a a a min a sec a minute**. <ICE-HKE s1b-003#124:1:A>

(251) But uh uh I think after they ca they came back after they come back from the census holiday I am going to play **the radio tape the the the cassette tapes** to the whole class.
<ICE-HKE s1a-016#239:1:A>

Cases of REPETITIONS, on the other hand, are always all included in the constituent strings if they form part of the first dative construction. Repetitions of parts of words are also included as they count towards the length. In example (252), the speaker begins with a false start and repeats him-/herself without pronouncing the complete word (here “ex” instead of “example”). Hence, I included the whole *a an ex a common example* as counting towards the theme. In example (254), the speaker repeats the whole dative construction, with all three elements repeated but not embedded in the same syntactic environment. While in example (255), the speaker repeats the theme twice. In these cases, I only take one variant, as the speaker obviously produces the same utterance again because of some kind of disfluency.

(252) I mean I I ‘ll give you **a an ex a common example**. <ICE-GB s1b-070#23:1:B>

(253) Uh Mr Jiang as NPC uh uh uh uh uh member I think at least you should given him **the freedom** the right to speak on whatever issues. <ICE-HKE s1b-038#334:1:B>

(254) It just **gives you the output gives you the output** only will not format wil uh the output but for the uh Report Writer it ‘ll give you the format. <ICE-HKE s2a-059#101:1:A>

(255) And it also **give us the target the target** at that time should be hundred and one [...].
<ICE-HKE s2b-041#74:1:A>

In sum, as well as in accordance with the approach taken for dealing with ellipsis, I always go for any first completed fully overt PD or DO (disregarding any following material), include any repetitions whatsoever (of elements of the constituents) and only in cases of self-correction go for the intended, maybe final, object. Thus, in cases such as (256), the first dative construction is *show them the procedures film*, and does not include the *to*-phrase *to all the new people* even though this refers back to the pronoun *them*. The same approach was settled on in any case where the objects are not coordinated but where the last element constitutes some nominal exemplification/specification, e.g. *give him some advice guidance*.

(256) So maybe this year what we do is actually show them the procedures film first to all the new people and then by ten o'clock. <ICE-HKE s1a-053#502:1:A>

The same approach is taken with regard to the repetitions of whole constituents, in cases where one full PD or DO has already been uttered and where the additional element does not modify the preceding constituent in any restrictive way (see section (7.1.3) on apposition). Hence, instances such as the first below only include *a boy's name* within the object demarcations, while in a phrase such as *the scissors, the two-edged weapon* both noun phrases are included as the latter constitutes some kind of modification to the head noun which is elementary for its identification.

(257) Yiu was born in Xinhui, a Guangdong village where her father gave her a boy 's name, **Yick-cheung**, as she was the family 's only daughter. <ICE-HKE w2b-023#101:1>

8 Mark-up of ICE corpora

Manual mark-up of the ICE corpora follows the general guidelines “Markup Manual for Spoken/Written texts” and includes compilers’ annotation with regard to, for instance, normalisation of spelling mistakes, indication of pauses, untranscribed text (e.g. coughs, laughs), foreign and indigenous words, changed names and words, unusable characters, unclear words and uncertain transcriptions, incomplete and discontinuous words, etc. (for a full list see “Markup”). Not all corpus compilers seem to have adhered to the same procedures, however, and inconsistencies within the same corpus can still be found. Despite these possible differences, we concluded to “interpret” as little as possible from the corpus data and always adhere to the compilers’ suggestions. That means, in cases of normalisation of spelling mistakes, we go for the normalised spelling (e.g. (261)); in cases where the spelling is not normalised

(despite obvious spelling error), we retain the original spelling, e.g. (258) and (259).

Whenever the markup was used for unusable characters, such as é in *caché* (which are spelled as *cachée* in the dataset), I transformed them back into their original spelling in our dataset (so also “pound sign” to “č” in ICE-IRE). Unusable characters can be identified by their SGML character description. As self-corrections are also signalled by markup (or at least should be), we can rely on the compilers’ work in questionable cases.

(258) [...] And uh uhm she ‘s teaching English to unversity students in in China [...].
<ICE-CAN s1a-057#163:1:B>

(259) tell u a good news, <ICE-HKE w1b-015#207:10>

(260) It causes me a lot of <->incoveniences</-> <+>inconveniences</+>. <ICE-HKE w1b-005#6:1>

(261) In such a cases the owner or occupier of the land on which it originated is held liable to pay compensation to the <->aggreviated</-> <+>aggrieved</+> person. <ICE-IND w1a-009#104:1>

9 Other issues resolved

9.1 Spelling conventions

Due to differences in the input variety of English (AmE in the case of Philippine English, BrE in all other cases) and because of the heavy influence of the American culture nowadays, spelling of words may vary across and within the nine ICE corpora. As these spelling differences are only relevant for weight issues and only with regard to the measuring of length of constituents in graphemes, they are so far disregarded: Spelling is always retained as it is used in the original text.

10 Appendix —The C7 Tagset

APPGE	possessive pronoun, pre-nominal (e.g. my, your, our)
AT	article (e.g. the, no)
AT1	singular article (e.g. a, an, every)
BCL	before-clause marker (e.g. in order (that), in order (to))
CC	coordinating conjunction (e.g. and, or)
CCB	adversative coordinating conjunction (but)
CS	subordinating conjunction (e.g. if, because, unless, so, for)
CSA	as (as conjunction)
CSN	than (as conjunction)
CST	that (as conjunction)
CSW	whether (as conjunction)
DA	after-determiner or post-determiner capable of pronominal function (e.g. such, former, same)
DA1	singular after-determiner (e.g. little, much)
DA2	plural after-determiner (e.g. few, several, many)
DAR	comparative after-determiner (e.g. more, less, fewer)
DAT	superlative after-determiner (e.g. most, least, fewest)
DB	before-determiner or pre-determiner capable of pronominal function (all, half)
DB2	plural before-determiner (both)
DD	determiner (capable of pronominal function) (e.g any, some)
DD1	singular determiner (e.g. this, that, another)
DD2	plural determiner (these, those)
DDQ	wh-determiner (which, what)
DDQGE	wh-determiner, genitive (whose)
DDQV	wh-ever determiner, (whichever, whatever)
EX	existential there
FO	formula
FU	unclassified word
FW	foreign word
GE	germanic genitive marker —(‘ or ‘s)
IF	for (as preposition)
II	general preposition

IO	of (as preposition)
IW	with, without (as prepositions)
JJ	general adjective
JJR	general comparative adjective (e.g. older, better, stronger)
JJT	general superlative adjective (e.g. oldest, best, strongest)
JK	catenative adjective (able in be able to, willing in be willing to)
MC	cardinal number, neutral for number (two, three..)
MC1	singular cardinal number (one)
MC2	plural cardinal number (e.g. sixes, sevens)
MCGE	genitive cardinal number, neutral for number (two's, 100's)
MCMC	hyphenated number (40 –50, 1770 –1827)
MD	ordinal number (e.g. first, second, next, last)
MF	fraction, neutral for number (e.g. quarters, two-thirds)
ND1	singular noun of direction (e.g. north, southeast)
NN	common noun, neutral for number (e.g. sheep, cod, headquarters)
NN1	singular common noun (e.g. book, girl)
NN2	plural common noun (e.g. books, girls)
NNA	following noun of title (e.g. M.A.)
NNB	preceding noun of title (e.g. Mr., Prof.)
NNL1	singular locative noun (e.g. Island, Street)
NNL2	plural locative noun (e.g. Islands, Streets)
NNO	numeral noun, neutral for number (e.g. dozen, hundred)
NNO2	numeral noun, plural (e.g. hundreds, thousands)
NNT1	temporal noun, singular (e.g. day, week, year)
NNT2	temporal noun, plural (e.g. days, weeks, years)
NNU	unit of measurement, neutral for number (e.g. in, cc)
NNU1	singular unit of measurement (e.g. inch, centimetre)
NNU2	plural unit of measurement (e.g. ins., feet)
NP	proper noun, neutral for number (e.g. IBM, Andes)
NP1	singular proper noun (e.g. London, Jane, Frederick)
NP2	plural proper noun (e.g. Browns, Reagans, Koreas)
NPD1	singular weekday noun (e.g. Sunday)
NPD2	plural weekday noun (e.g. Sundays)

NPM1	singular month noun (e.g. October)
NPM2	plural month noun (e.g. Octobers)
PN	indefinite pronoun, neutral for number (none)
PN1	indefinite pronoun, singular (e.g. anyone, everything, nobody, one)
PNQO	objective wh-pronoun (whom)
PNQS	subjective wh-pronoun (who)
PNQV	wh-ever pronoun (whoever)
PNX1	reflexive indefinite pronoun (oneself)
PPGE	nominal possessive personal pronoun (e.g. mine, yours)
PPH1	3rd person sing. neuter personal pronoun (it)
PPHO1	3rd person sing. objective personal pronoun (him, her)
PPHO2	3rd person plural objective personal pronoun (them)
PPHS1	3rd person sing. subjective personal pronoun (he, she)
PPHS2	3rd person plural subjective personal pronoun (they)
PPIO1	1st person sing. objective personal pronoun (me)
PPIO2	1st person plural objective personal pronoun (us)
PPIS1	1st person sing. subjective personal pronoun (I)
PPIS2	1st person plural subjective personal pronoun (we)
PPX1	singular reflexive personal pronoun (e.g. yourself, itself)
PPX2	plural reflexive personal pronoun (e.g. yourselves, themselves)
PPY	2nd person personal pronoun (you)
RA	adverb, after nominal head (e.g. else, galore)
REX	adverb introducing appositional constructions (namely, e.g.)
RG	degree adverb (very, so, too)
RGQ	wh- degree adverb (how)
RGQV	wh-ever degree adverb (however)
RGR	comparative degree adverb (more, less)
RGT	superlative degree adverb (most, least)
RL	locative adverb (e.g. alongside, forward)
RP	prep. adverb, particle (e.g. about, in)
RPK	prep. adv., catenative (about in be about to)
RR	general adverb
RRQ	wh- general adverb (where, when, why, how)

RRQV	wh-ever general adverb (wherever, whenever)
RRR	comparative general adverb (e.g. better, longer)
RRT	superlative general adverb (e.g. best, longest)
RT	quasi-nominal adverb of time (e.g. now, tomorrow)
TO	infinitive marker (to)
UH	interjection (e.g. oh, yes, um)
VB0	be, base form (finite i.e. imperative, subjunctive)
VBDR	were
VBDZ	was
VBG	being
VBI	be, infinitive (To be or not ... It will be ...)
VBM	am
VBN	been
VBR	are
VBZ	is
VDO	do, base form (finite)
VDD	did
VDG	doing
VDI	do, infinitive (I may do ... To do ...)
VDN	done
VDZ	does
VH0	have, base form (finite)
VHD	had (past tense)
VHG	having
VHI	have, infinitive
VHN	had (past participle)
VHZ	has
VM	modal auxiliary (can, will, would, etc.)
VMK	modal catenative (ought, used)
VV0	base form of lexical verb (e.g. give, work)
VVD	past tense of lexical verb (e.g. gave, worked)
VVG	-ing participle of lexical verb (e.g. giving, working)
VVGK	-ing participle catenative (going in be going to)

VVI	infinitive (e.g. to give . . . It will work . . .)
VVN	past participle of lexical verb (e.g. given, worked)
VVNK	past participle catenative (e.g. bound in be bound to)
VVZ	-s form of lexical verb (e.g. gives, works)
XX	not, n't
ZZ1	singular letter of the alphabet (e.g. A,b)
ZZ2	plural letter of the alphabet (e.g. A's, b's)

References

- Bresnan, J., Cueni, A., Nikitina, T., & Baayen, H. (2007). Predicting the dative alternation. In G. Boume, I. Kraemer, & J. Zwarts (Eds.), *Cognitive foundations of interpretation* (pp. 69–94). Amsterdam: Royal Netherlands Academy of Science.
- Bresnan, J. & Ford, M. (2010). Predicting Syntax: Processing dative constructions in American and Australian Varieties of English. *Language*, 86(1), 168–213.
- Bresnan, J. & Nikitina, T. (2009). The gradience of the dative alternation. In L. Uyechi & L. Hee Wee (Eds.), *Reality exploration and discovery: pattern interaction in language and life* (pp. 161–184). Stanford: CSLI Publications.
- Cueni, A. (2004). *Coding notes*, Stanford University. Manuscript.
- De Cuypere, L. & Verbeke, S. (2013, June). Dative alternation in Indian English: A corpus-based analysis. *World Englishes*, 32(2), 169–184.
- Greenbaum, S. (1996). *Comparing English worldwide: the International Corpus of English*. Oxford, New York: Clarendon Press.
- Gries, S. T. & Mukherjee, J. (2010, November). Lexical gravity across varieties of English: An ICE-based study of *n*-grams in Asian Englishes. *International Journal of Corpus Linguistics*, 15(4), 520–548.
- Grimm, S. & Bresnan, J. (2009). *Spatiotemporal variation in the dative alternation: a study of four corpora of British and American English*. Paper presented at the 3rd International conference on Grammar and Corpora, Mannheim, 22-24 September 2009.
- Kortmann, B., Schneider, E., Burridge, K., Mesthrie, R., & Upton, C. (Eds.). (2004). *A handbook of varieties of english*. Berlin: Mouton de Gruyter.
- Krifka, M. (2003). *Semantic and pragmatic conditions for the dative alternation*. Seoul, Korea.
- Levin, B. (1993). *English verb classes and alternations: A preliminary investigation*. Chicago: University of Chicago Press.
- Mukherjee, J. (2005). *English Ditransitive Verbs: Aspects of Theory, Description and a Usage-based Model*. Language and Computers. Amsterdam, New York: Rodopi.
- Mukherjee, J. & Hoffmann, S. (2006). Describing verb-complementational profiles of New Englishes: A pilot study of Indian English. *English World-Wide*, 27, 147–173.

- Mukherjee, J. & Schilk, M. (2008). Verb-complementational profiles across varieties of English: Comparing verb classes in Indian English and British English. In T. Nevalainen, I. Taavitsainen, P. Pahta, & M. Korhonen (Eds.), *The Dynamics of Linguistic Variation: Corpus Evidence on English past and present* (pp. 163–181). Amsterdam/Philadelphia: John Benjamins.
- Olavarra de Ersson, E. & Shaw, P. (2003). Verb complementation patterns in Indian Standard English. *English World-Wide*, 24, 137–161.
- Quirk, R., Greenbaum, S., Leech, G., & Svartik, J. (1985). *A Comprehensive Grammar of the English Language*. London/New York: Longman.
- Sankoff, D. (1988). Sociolinguistics and syntactic variation. In *Linguistics: the cambridge survey* (pp. 140–161). Cambridge: Cambridge University Press.
- Schilk, M., Bernaisch, T., & Mukherjee, J. (2012). Mapping unity and diversity in South Asian English lexicogrammar: verb-complementational preferences across varieties. In M. Hundt & U. Gut (Eds.), *Mapping unity and diversity world-wide: corpus-based studies of new englishes* (pp. 137–165). Amsterdam: John Benjamins.
- Schneider, G. (2008). *Hybrid Long - Distance Functional Dependency Parsing* (Doctoral dissertation, Faculty of Arts, University of Zurich, Zurich).
- Theijssen, D. (2012). *Making Choices: Modelling the English dative alternation* (Doctoral dissertation, Radboud Universiteit Nijmegen, Nijmegen).
- Wolk, C., Bresnan, J., Rosenbach, A., & Szmrecsanyi, B. (2013). Dative and genitive variability in Late Modern English: Exploring cross-constructual variation and change. *Diachronica*, 30(3), 382–419.