

ARGUMENT ORDER ALTERNATIONS IN DUTCH

Leonoor van der Beek

Center for Language and Cognition (CLCG), University of Groningen

Proceedings of the LFG04 Conference

University of Canterbury

Miriam Butt and Tracy Holloway King (Editors)

2004

CSLI Publications

<http://csli-publications.stanford.edu/>

Abstract

We investigate the factors that influence argument order variations in Dutch, focusing on ditransitive verbs. Evidence from grammaticality judgments is complemented with evidence from the distributions of the alternants in corpora of spoken and written Dutch. We find that while the NP/PP alternation is influenced by weight, the direct object shift (DOS) in the dative alternation is bound to certain types of pronouns. Additional evidence for our account for the DOS is found in the Dutch AcI construction. Our findings are modeled within the framework of Optimality Theoretic syntax, allowing for violable and ranked constraints, as well as a stochastic interpretation of the analysis.

1 Introduction

Even a relatively fixed word-order language like Dutch allows for some word order variation. The scrambling data discussed in for example de Hoop (2003) are a well known example. In addition to the variable placement of objects with respect to adverbial phrases, Dutch permits some variation in the relative order of verb arguments. The examples (1)-(2) illustrate two of these argument order alternations.

- (1) Ditransitive Verbs
 - a. Jo gaf de student een boek.
Jo gave the student a book
 - b. Jo gaf een boek aan de student.
Jo gave a book to the student
- (2) Accusativus cum Infinitivo
 - a. Jo zag de student een boek lezen.
Jo saw the student a book read
 - b. Jo zag het de student lezen.
Jo saw it the student read

In this paper we try to answer the question which factors determine the choice of one argument ordering over the other and how we can capture the influence of those factors in a grammar model. We will take the ditransitive construction in (1) as our explanatory example, and turn to the AcI construction in (2) for additional evidence.

We try to identify the most important influences on word order by looking at the distribution of the various alternants in corpora of spoken and written Dutch. Not only does this provide us with real world data, but it also gives us information about the frequency of a particular realization, and the context in which an alternant most often occurs. We thus find that some relevant distinctions are (near) categorical, while others only give rise to preferences for one of the alternants. We model our findings in the framework of Optimality Theoretic (OT) syntax

The rest of this paper is structured as follows. We introduce the Dutch ditransitive constructions in section 2, as well as the differences between the Dutch and the English construction. Section 3 discusses the influences of several (morpho-syntactic) features on the dative alternation, based on the results of our corpus study. In this section we also

introduce the constraints that are employed to formalize these results in the OT syntax framework. We provide some additional evidence for an important part of our analysis based on the Accusativus cum Infinitivo (AcI) construction in section 4. Finally, we present our conclusions and discuss some open ends in 5.

2 The dative alternation in Dutch

2.1 Differences between Dutch and English

The dative alternation is by no means specific to Dutch. Much work has been done on the dative alternation in English, illustrated in (3) (Pinker, 1989; Levin, 1993; Krifka, 2001; Bresnan and Nikitina, 2003, for example).

- (3) a. Jo gave the student the book
b. Jo gave the book to the student

In Dutch, the alternation is more complex, though. In addition to the regular double object construction and the dative PP construction ((1-a) and (1-b), repeated here as (4-a) and (4-b)), we have two more variants: both the double object construction and the PP construction occur with non-canonical word orders. In (4-c) we find the direct object shifted in fronted of the indirect object. In (4-d) we see that the ‘dative’ PP is shifted and precedes the direct object. Both variations violate the canonical argument order $\text{SUBJ} < \text{OBJ2} < \text{OBJ1} < \text{OBL}, \text{XCOMP}$ for Dutch.¹

The Direct Object Shift (DOS) differs from object shift in Scandinavian languages in that only the direct object shifts, and the shift is independent of the position of the verb: it occurs both in V2 main clauses and verb final subordinate clauses. DOS differs from Wackernagel movement in German in that it does not allow ‘movement’ of OBJ2 and it does not allow movement over the subject.

- (4) a. Jo gaf de student een boek.
Jo gave the student a book
b. Jo gaf een boek aan de student.
Jo gave a book to the student
c. Jo gaf het de student.
Jo gave it the student
d. Jo vroeg aan de student het antwoord op de vraag wanneer WOII
Jo asked to the student the answer to the question when WWII
eindigde.
ended

¹Throughout this paper, we will use both ‘indirect object’ and OBJ2 to refer to the grammatical role to which the recipient argument is mapped, contrary to much work in Lexical Mapping Theory on English, where the recipient is assumed to map to OBJ1. There is reason to assume that English and Dutch differ in this respect, e.g. Dutch does not allow the recipient to be mapped onto the subject function in passive sentences and does allow passive sentences with theme subjects and recipient objects.

The existence of the non-canonical variants in Dutch teases apart two distinctions that are merged together in the English situation: where English has two variants which differ with respect to both the syntactic category and the order of the arguments, Dutch has both an alternation between NP and PP recipients and argument order variations, resulting in a total of four different realizations.

2.2 Previous approaches

Analyses of the dative alternation in English have employed both the difference in ordering and the difference in grammatical role. General alignment principles have been applied to explain the distribution of the dative alternation. For example, although the double NP construction is generally favored, heavy recipients may be realized as (right-aligned) PPs in order to avoid a violation of the general principle on word order that says that heavy constituent align right. On the other hand we find analyses of the dative alternation that focus on the NP/PP alternation specifically, arguing that the two constructions have a different semantics or are selected by different lexical items.

Representatives of the first class are Behaghel (1909/10), Wasow (1997) and Arnold et al. (2000), among many others. They all argued that long and complex phrases tend to occur at the right edge of a clause. Gundel (1988) and Prince (1992) showed that the same holds for new information: it prefers the right edge, following the old, topic information. In addition, Arnold et al. (2000) showed that although weight and givenness are not independent of each other, they do have distinct effects on word order.

The main representative of the first class in German linguistics is Uszkoreit (1987). He identified several word order principles for German, e.g. the unmarked word order is SUBJ<IOBJ<DOBJ, personal pronouns precede other NPs, definite NPs precede non-definite NPs and light constituents precede heavy constituents. These principles are rephrased for Dutch as the Inherence Principle (canonical word order), the Left-Right Principle (constituents that are rich in information align right) and the Complexity Principle (heavy constituents align right) (Haeseryn and others, 1997).

The second class has focused more specifically on the NP/PP alternation. For example, Krifka (2001) and Pinker (1989) have tried to identify distinct meanings for the two realizations of ditransitive verbs. According to this line of explanation, there is no dative alternation proper: the double object construction and the PP construction are not alternative ways of expressing the same meaning, but they are expressions of different meanings. Bresnan and Nikitina (2003) provide examples of alternating dative syntax in contexts of repetition, which form a challenge for this approach.

Others have tried to classify verbs into classes that select for one construction or the other (Levin, 1993). Although statistically significant differences in the frequencies of certain verbs occurring with the two constructions exist (see Lapata (1999) for corpus methods to test the empirical value of the semantic verb classes described by Levin (1993)), Bresnan and Nikitina (2003) convincingly showed that these are mere tendencies, indicating improbability, rather than categorical differences.

Finally, Bresnan and Nikitina (2003) argued that it is the recipient argument that puts constraints on the grammatical role that it is mapped to. They claimed that local person NPs should be realized as objects, not obliques. Therefore, local recipients will lead to double object constructions instead of dative PP constructions.

In Dutch, one might expect to find a differentiation between NP vs. PP recipients on the one hand and canonical vs. non-canonical argument order on the other hand, where construction specific constraints determine the grammatical role of the recipient argument, and general alignment constraints determine the order of the arguments.

The next section discusses some corpus data that show that the predicted differentiation between construction specific constraints and general alignment constraints is not borne out. Instead we find that both the grammatical function of the recipient and the order of the arguments are influenced by constraints that apply to other constructions as well. As we will see, the direct object shift in ditransitive constructions as well as AcI constructions is triggered by certain types of (direct) object pronouns and weight influences both the NP/PP alternation and word order in the dative PP construction. The analysis presented below does leave open the possibility to incorporate lexically encoded preferences of verbs for one construction or the other.

3 Distribution of the alternants: a corpus study

3.1 Preliminaries

Corpora contain valuable information about the distribution of different realizations of the dative construction. A potential problem is that the various alternants are specific and complex syntactic structures, which cannot be retrieved from corpora on the basis of simple pattern recognition. Therefore, we used syntactically annotated and automatically parsed data in our corpus study. Both annotated corpora, the annotated part of the Corpus of Spoken Dutch (CGN, about 1M words (Levelt, 1998)) and the Alpino Treebank (the annotated cdbl newspaper part of the Eindhoven Corpus, about 150K words (van der Beek et al., 2002)) are annotated with dependency structures (Moortgat, Schuurman, and van der Wouden, 2001).

When the annotated corpora proved too small for statistically relevant results, we used a corpus of about 75M words of newspaper text (CLEF) that was automatically parsed by the Alpino parser (Bouma, van Noord, and Malouf, 2001; van der Beek, Bouma, and van Noord, 2002). Alpino outputs the same dependency structures as those used in the annotated corpora. In particular, data sparseness occurred when looking for pronouns in double object construction. With a 85.5% parsing accuracy (measured over the dependency relations), the quality of the annotation in the automatically parsed corpus is lower than the manually annotated corpora. The handwritten Alpino grammar overgenerates, always allowing both the canonical and the shifted word order in double object constructions. The resulting OBJ1/OBJ2 ambiguity is resolved by a Maximum Entropy disambiguation model. Although the data will some contain parse errors and there is always a chance that the disambiguation system does not pick up on a certain phenomenon, manual inspection of (samples of) the data did not show signs of systematic errors.

The corpora were queried using DT_SEARCH (Bouma and Kloosterman, 2002), a tool which allows us to query the treebank on dependency relations, syntactic category and linear order.

	NP NP _{unshift}	NP NP _{shift}	NP PP	PP NP	TOTAL
CGN	226	33	63	8	334
Alpino	122	7	43	10	182

Table 1: Distribution of the three alternants of the dative alternation in Dutch

We excluded from our search all instances of (in)direct object topicalization, all (wh)relativizer direct and indirect objects and all clausal objects such as that-clauses because in these sentences, the order of the arguments is determined by other factors. Also excluded were passive sentences and instances of the *krijgen*-passive (the ‘get-passive’). The motivation for this is that the direct object (in the regular passive) or the indirect object (in the *krijgen*-passive) surfaces as the subject of the matrix clause, therefore the word order for subjects applies here. Finally, we excluded all instances of ‘split’ dative PPs. In these sentences, illustrated in examples (5), the recipient argument is third person, inanimate and singular and realized as a pronoun inside a PP. In these cases, a so-called R-pronouns (*er*, *daar*, *hier* (there, here)) is used instead of the regular third person neuter singular *het* (it) and this pronoun is often fronted. The preposition stays in position, resulting in a split PP. The alignment of *er* is a characteristic of R-pronouns, not a characteristic of the dative construction.

- (5) Ik geef daar geen les aan.
 I give there no class to
I won't teach those.

3.2 The general distribution

The four alternants are not represented equally in the corpus. In table 1, the distribution of the different realizations is given. As expected, the canonical argument orders (NP NP_{unshift} and NP PP) are much more frequent than the non-canonical variants. Furthermore, the double object construction is much more frequent than the PP construction. This corresponds to the idea that the PP construction is somehow marked.

In an Optimality Theoretic Syntax framework we can model the canonical word order by the f-precedence² constraint CANON ((6)). The preference for the double object construction is modeled by the markedness constraint *STRUCT, familiar from Bresnan and Nikitina (2003).

- (6) CANON: SUBJ_{<f} OBJ2_{<f} OBJ1_{<f} OBL

*STRUCT: avoid syntactic structure, here: PP.

²For two f-structures f_1 and f_2 , $F_1 f$ -precedes f_2 if and only if all the nodes that map onto f_1 c-precede all the nodes that map onto f_2 (Zaenen and Kaplan, 1995).

	NP NP _{unshift}	NP NP _{shift}	NP PP	PP NP	TOTAL
CGN	143	33	57	3	247
Alpino	45	6	21	3	83

Table 2: Distribution of dative alternation realizations with one word themes.

3.3 Direct Object Shift

This distribution changes drastically if we control for weight by restricting the object to one lexical item only (we do allow additional function words such as determiners). While the numbers for the shifted double object construction hardly change, the numbers for the unshifted and PP variants drop by 10-70%. This is caused by the fact that DOS almost exclusively occurs with direct object pronouns.

We did in fact find one example in which a full NP shifted (7), but here we find the archaic dative marking on the indirect object. We assume that it is this overt dative marking that makes available the freer word-order and that DOS is generally restricted to pronouns.³

- (7) [daar] heeft Paul Badura-Skoda het nieuwe pianoconcert van Frank Martin
 there has Paul Badura-Skoda the new piano_concert of Frank Martin
 den muzikale volke voorgesteld.
 the_{dat} musical_{dat} people_{dat} presented
there, Paul Badura-Skoda presented the Frank Martin's new piano concert to the musical people.

It is not the case that all direct object pronouns always shift. While the pronoun *het* (it) usually shifts irrespectively of the category of the indirect object, most other personal pronouns and the demonstratives shift if the indirect object is a full NP, but stay in their canonical position if the indirect object is a personal pronoun (8-a). First and second person pronouns do not shift. Made up examples of local pronoun DOS lead to ungrammaticality under the intended reading (in (9), the sentence is grammatical under the reading without DOS, i.e. the reading with a recipient *jou* (you)).

- (8) a. De student geeft dat de student.
 the student gives that the student
The student gives that to the student.
 b. De student geeft hem dat.
 the student gives him that
The student gives that to him.

³However, Zwart (1997) presents examples that show that NP-DOS with definite NPs is not impossible:

- (i) dat Jan het boek Marie terug gegeven heeft.
 that Jan the book Marie back given has
that Jan gave the book back to Marie.

No examples of this kind were found in the corpus. We suspect the exceptional definite NP shift to be a focus effect and leave this and other effects of focus on word order for future research.

Shifted		Canonical	
542	het (it)	372	dat (that)
45	dat (that)	83	dit (this)
21	't (it _{reduced})	51	het (it)
19	ze (them)	28	die (that)
7	dit (this)	24	hem (him/it)
4	u ⁴ (you _{honorific})	14	zich (himself/herself)
4	hem (him/it)	8	hetzelfde (it _{same})
4	die (that)	4	me (me)

Table 3: Direct object pronouns in constructions with two pronominal objects

- c. De student geeft het hem.
the student gives it him
The student gives it to him.
- (9) a. De student wijst 'm de student aan.
the student points him the student at
The student points him out to the student.
- b. %De student wijst jou de student aan.
the student points you the student at
The student points you out to the student.

Table 3 shows the most frequent direct object pronouns in double object constructions where both arguments are pronominal. The data are based on the automatically parsed CLEF corpus. The frequency lists confirm the intuition that *het* shifts while demonstratives usually do not shift in front of another pronoun. Importantly, the table shows that the distinctions are not categorical: we do find *het* (it) in the canonical object position, although ten times less frequently than in the shifted position. The one place where we would not expect any variation is with the local pronouns, as even made up examples were ungrammatical. Nevertheless, we do find four occurrences of *u* (you_{honorific}). Further inspection showed that these are the result of parse errors.

We conclude from the examples and the corpus data that pronouns prefer to align left. Furthermore, this tendency is stronger for *het* than for personal pronouns and demonstrative pronouns. A similar differentiation among the pronouns is found in German with respect to Wackernagel movement (Müller, 2001).

We model these restrictions with the constraints PRO_{it}-L and PRO-L, stating that *het* and other pronouns should align left (10) in the clause. The constraints are in competition with the constraint on canonical word order: only subject pronouns can simultaneously satisfy PRO-L and CANON. Although each constituent that separates the pronoun from the left edge of the clause incurs one violation, the tableaux show only the crucial violations.

A local constraint conjunction (Smolensky, 1995) of the constraint on canonical word

⁴All occurrences of the local pronoun *u* result from parse errors.

Input: <i>gives</i> (<SUBJ><OBJ1><OBJ2>)		*STRUC	*LOL	PRO _{it} -L	PRO-L	CANON
OBJ1='the book' OBJ2='de student' ex.(4-a)	☞ NP NP _{unshift} NP NP _{shift} NP PP PP NP	*! *!				*! *
OBJ1='it' OBJ2='de student' ex.(4-c)	☞ NP NP _{unshift} NP NP _{shift} NP PP PP NP	*! *!		*! *	* *	* *
OBJ1='it' OBJ2='him' ex.(8-c)	☞ NP NP _{unshift} NP NP _{shift} NP PP PP NP	*! *!		*! *	* *	* *
OBJ1='that' OBJ2='the student' ex.(8-a)	☞ NP NP _{unshift} NP NP _{shift} NP PP PP NP	*! *!			*! *	* *
OBJ1='that' OBJ2='him' ex.(8-b)	☞ NP NP _{unshift} NP NP _{shift} NP PP PP NP	*! *!			* *	*! *
OBJ1='you' OBJ2='the student' ex.(9-b)	☞ NP NP _{unshift} NP NP _{shift} NP PP PP NP	*! *!	*!		* *	* *

Table 4: Shifted vs. canonical double object constructions

order and the constraint on local objects (Aissen, 2003) models the fact that local objects do not shift. This constraint conjunction is a formalization of the intuition that local direct objects are an instance of 'the worst of the worst' (Lee, 2003): a combination of a marked category and a marked word order.

(10) PRO_{it}-L: the pronoun *het* (it) aligns left.

PRO-L: personal and demonstrative pronouns align left.

*LOCAL OBJECT LEFT (LOL): CANON & *OBJ1_{local}

Table 4 shows how the constraints interact to account for various example sentences. Pronominal direct objects will shift if the indirect object is a full NP, in order to avoid

a violation of the constraint on the alignment of pronouns, which is higher ranked than CANON. If both objects are pronominal (but not *het* or local), CANON is the highest ranked constraint that is violated by one argument order but not the other and thus determines the optimal candidate: OBJ2 precedes OBJ1. *Het*, on the other hand, will always shift, because a violation of PRO_{it}-L is worse than any other right aligned pronoun or a non-canonical word order. All alignment constraints are outranked by LOL, preventing local pronouns from shifting.

Our findings contradict the claim in Zwart (1996) that only reduced direct object pronouns can shift: the demonstratives were among the most frequently shifted pronouns and we also found non-reduced examples of third person pronouns. We do see a tendency, though, of the reduced pronouns *'m* (him, it) and *ze* (them) to group with *het* if the antecedent is inanimate. In this case, they tend to shift, even if the indirect object is a pronoun. We do not have enough data for a quantitative evaluation of this intuition, but integration of it in our model is straightforward if it proves correct.

Data sparseness also prevented further research into the relative ordering of two animate personal pronoun objects. In our model, the animate personal pronouns form one homogenous group. If both objects are from the same group, canonical word order is always predicted to be more optimal. In both annotated corpora, no sentences were found with two objects consisting of pronouns referring to humans. The unannotated part of the CGN corpus (9M words) was parsed to obtain more spoken language data. In this corpus and the 75M word automatically parsed CLEF corpus together, we found only three sentences, two of which were canonical and one of which was an instance of DOS. This lack of data is due to the fact that animate direct objects are marked and thus generally sparse. This tendency is even stronger in ditransitive sentences: even when we included all ditransitive sentences, with pronominal and with full NP indirect objects, we found no animate direct objects in either CGN or the Alpino Treebank. The situation is further complicated by the fact that people hesitate and disagree about their grammaticality judgments for this type of sentence. We leave this issue for future research.

In this model, the DOS is driven by the syntactic category of the objects: NPs, personal or demonstrative pronouns or *het*. Pronominality is not independent of syntactic weight: pronouns are the lightest possible NPs. Thus, the pronominal DOS is in line with the Complexity Principle and Uszkoreit's weight principle. But we did not differentiate between heavy NP recipients and light NP recipients, although the weight principles would predict the former to allow DOS more easily than the latter. Table 5 lists the average weight (in number of words) of the direct and indirect object in all four variants of the dative alternation, as well as the OBJ1/OBJ2 weight ratios. We see that the average weight of the indirect object in shifted double NP constructions (1.09 and 1.71) is *lower* than in the canonical double object construction (1.40 and 2.43), contrary to what the Complexity Principle would predict. We assume syntactic weight not to be of influence on the DOS.

3.4 The NP/PP alternation

Although syntactic weight does not have a direct effect on DOS, it does seem to influence the NP/PP alternation. The effect is not very clear if we look at the ratios of

				OBJ1	OBJ2	OBJ1/OBJ2
CGN	NP	NP _{unshift}	(N=231)	3.75	1.40	2.68
Alpino	NP	NP _{unshift}	(N=123)	5.87	2.43	2.42
CGN	NP	NP _{shift}	(N=33)	1.03	1.09	0.94
Alpino	NP	NP _{shift}	(N=7)	1.71	1.71	1.00
CGN	NP	PP	(N=63)	1.62	2.57	0.63
Alpino	NP	PP	(N=43)	3.70	5.21	0.71
CGN	PP	NP	(N=8)	5.63	1.63	3.45
Alpino	PP	NP	(N=10)	4.80	3.30	1.45

Table 5: Average weight per grammatical role in number of words.

			Weight
NP	NP _{unshift}	(N=126)	3.02
NP	NP _{shift}	(N=9)	1.89
NP	PP	(N=55)	5.00
PP	NP	(N=17)	3.71

Table 6: Average weight of non-pronominal indirect objects

the direct and indirect objects weight in table 5, but this ratio is distorted by the light, pronominal direct objects in the DOS. If we only look at the indirect objects, we see that the recipient arguments that are realized in PPs are much heavier than those that are realized as NPs.

This difference in weight may be a result of a constraint blocking pronominal recipients in a PP. This would increase the average weight of the PP variant in the same way the pronouns in the DOS lower the average object weight. We therefore looked at the average weight (in number of words, in both corpora together) of the indirect objects in the various ditransitive constructions excluding all pronominal recipients. The results are listed in table 6. Although the numbers are too small for drawing definite conclusions, we see that the PP recipients in the corpus are still heavier than their NP counterparts. We conclude that heavy recipient arguments prefer realization as a PP, even though obliques are generally more marked than objects.

Note also that in the Alpino Treebank, which consists of written language, the constituents are on average heavier than in CGN, which is a corpus of spoken Dutch. At the same time, the proportion of PP constructions is larger: 53 PPs and 130 NPs in the Alpino Treebank, versus 71 PPs and 264 NPs in CGN.

The question is what triggers the non-canonical dative PP construction which has both the marked grammatical function and the marked order. Looking at table 5 we see that the direct objects in the shifted canonical constructions are heavier than the direct objects in both construction where OBJ1 precedes OBJ2. This indicates another instance of the Complexity Principle, that states that heavy constituents align right. If the indirect object is realized as a PP (either because of its weight or because of other factors, which we will discuss in 3.5), this results in the non-canonical dative PP

Input: example (12-a)	HEAVY-PP	*STRUC	*LOL	PRO _{it} -L	PRO-L	HEAVY-R	CANON
NP NP _{unshift}	*!					*	
NP NP _{shift}	*!						
NP PP		*					
PP NP		*				*!	

Table 7: Optimization for ditransitives with heavy recipients

construction as in example (12-b).

To model both effects of weight, two constraints are introduced: HEAVY-PP and HEAVY-R. HEAVY-PP is violated by heavy recipients that are realized as NPs in a double object construction. In an implementation of this model, one would have to set a critical value, which indicates how many words a light NP may maximally have. Alternatively, one could envisage a stochastic OT syntax model that allows for cumulativity effects (Jäger and Rosenbach, 2003). In such a model, the heavier a recipient argument, the higher the probability that it is realized as an oblique argument.

HEAVY-R is the constraint that makes possible the shifted PP construction by saying heavy constituents should align right. However, the PP recipients are usually heavy, too. This means that both the canonical ordering and the shifted PP construction would violate this constraint once, in which case the canonical word order is optimal. This problem is circumvented by defining HEAVY-R in such a way that it applies to the heaviest argument only, or by having heavier constraints violate the constraint more often.

- (11) HEAVY-R: heavy constituents align right.

HEAVY-PP: heavy recipient arguments are realized as obliques

- (12) a. Ik vraag het aan iemand die in de Vlaamse Beweging actief is.
 I ask it to someone who in the Flemish Movement active is
I will ask it to someone who is active in the Flemish Movement.
- b. Niemand kan aan de Westduitse bondskanselier de heen- en
 nobody can to the West German president the to and
 terugreis voorschrijven.
 from_jjourney prescribe
Nobody can prescribe both ways of the journey to the West German chancellor.

We have seen how the Inherence Principle and the Complexity Principle (or canonical word order and syntactic weight) influence the dative alternation in Dutch. The third principle assumed to have an influence on word order is the Left-Right Principle that states that constituents that are rich in (new) information follow constituents that carry

Input: example (12-b)	HEAVY-PP	*STRUC	*LOL	PRO _{it} -L	PRO-L	HEAVY-R	CANON
NP NP _{unshift}	*!						
NP NP _{shift}	*!					*	*
NP PP		*				*!	
PP NP		*					*

Table 8: Optimization for ditransitives with both heavy recipients and heavy themes

			OBJ1	OBJ2	OBJ1/OBJ2
CGN	NP NP _{unshift}	(N=231)	40	163	0.25
Alpino	NP NP _{unshift}	(N=123)	2	32	0.06
CGN	NP NP _{shift}	(N=33)	32	27	1.19
Alpino	NP NP _{shift}	(N=7)	6	2	3.00
CGN	NP PP	(N=63)	30	13	2.31
Alpino	NP PP	(N=43)	4	1	4.00
CGN	PP NP	(N=13)	0	3	0.00
Alpino	PP NP	(N=10)	0	0	-

Table 9: Number of pronominal (in)direct objects

less new information. As (personal and demonstrative) pronouns are by definition given and indefinite NPs are by definition new, Uszkoreit's principles 'pronouns before full NPs' and 'definite NPs before indefinite NPs' both fall under the Left-Right Principle.

A first influence of the pronoun principle was seen in our account of the DOS, which is restricted to (certain types of) pronouns and can be modeled by constraints that are violated if pronouns are not aligned left. The question is whether there are similar alignment constraints on pronouns in the dative PP construction. Although table 9, listing the number of pronouns per grammatical function in the four alternants, shows the expected pattern of pronouns preferring the first argument position over the second argument position, it is hard to find evidence for *independent* influence of pronominality on the NP/PP alternation in Dutch. After all, pronouns are extremely light NPs, which are not expected to show up as PPs anyway because they do not fall under the scope of the HEAVY-PP constraint. Note that table 9 does not list any instances of third person singular inanimate pronouns in the PP alternant, because these are realized as R-pronouns, which we excluded from our search. In CGN, 6 R-pronoun obliques were found, in Alpino 1.

More reliable evidence of a Left-Right Principle effect may be expected from the definite/indefinite distinction. A first attempt at the identification of a definiteness effect was made by counting the number of direct and indirect objects with the indefinite article *een* (a) and those with the definite article *de* (the). For both corpora together,

we find a 1.68 indefinite/definite ratio (84/50) for the direct object and a 0.10 ratio (6/58) for the indirect object in the (unshifted) double object construction. For the (unshifted) PP construction, we find 1.38 (18/13) for the direct object and 0.09 (3/32) for the indirect object. These numbers do not differ significantly ($p=0.05$), but further research should be carried out to confirm and explain these preliminary results.

3.5 More factors in the dative alternation

We have discussed the influences of canonical word order, pronominality, weight and definiteness on the dative alternation in Dutch and we have identified several constraints on this alternation. We have no doubt that there are many more factors that co-determine which alternant is realized. First of all, we excluded various constructions from our research for the very reason that they would introduce other constraints that would interfere with the constraints on the dative alternation proper, such as the passive constructions and constructions with R-pronouns.

Secondly, we ignored lexical preferences. Bresnan and Nikitina (2003) showed that many verbs that were thought to categorically select for the NP or the PP construction, do in fact alternate. Nevertheless many verbs do show preferences for one realization over the other. In Dutch, the verb *verhuren* (to let) has a preference for the PP construction, while for example *aanwijzen* has a preference for the double object construction.⁵

Bresnan and Nikitina (2003) argued that the person feature influences the dative alternation in English through the constraint HARMONY(1,2), which penalizes local PPs and third person NP recipients. In English, it is unclear whether this is a constraint on the grammatical function of the local recipient or an alignment constraint, as English does not allow for non-canonical word orders. The Dutch data show that person does not effect the NP/PP alternation. Table 10 shows the distribution of local and 3rd person recipients over the four constructions. We restricted our search to pronominal recipients, because local recipients can only be realized by a pronoun and we want to measure an effect of person independent of the influences of weight and pronominality. The results for both corpora were combined to get more representative numbers and to generalize over the differences between spoken and written language (with local recipients generally being more frequent in spoken language). The distribution of local and third person pronouns is not significant ($p=0.05$).

The data on DOS in section 3.3 showed, on the other hand, that person does have an influence on argument order: the constraint conjunction LOL penalized shifted first and second person direct objects.

One may suggest that the relevant feature is not person but animacy. Unfortunately, none of the available corpora of Dutch is annotated with information about animacy. Within the restricted search space of the pronominal recipients, there were too few inanimate recipients to draw any conclusions. That being said, it does seem to be the

⁵These lexical preferences form a problem for OT systems. Two ways of implementing them are by language particular constraints, that block a particular construction for a particular verb (Bresnan and Nikitina, 2003) or as a lexical feature. The latter would save the principle of a universal set of constraints, but crucially depends on a lexicon friendly OT system, as in van der Beek and Bouma (2004).

	local	3rd person
NP NP _{unshifted}	101	52
NP NP _{shifted}	13	9
NP PP	7	3
PP NP	2	1

Table 10: Person features of pronominal indirect objects

case that with (marked) inanimate recipients, the DOS is ungrammatical and the PP-construction is preferred (example (13)).

- (13) a. Ik geef dit boek een tien.
 I give this book a ten
I give this book ten out of ten.
- b. ?Ik geeft dat geen enkel boek
 I give that no single book
I do not give that to any book.
- c. En toch geef ik dat wel aan dit boek.
 and still give I that indeed to this book
But I still do give that to this book.

Besides (morphosyntactic) feature driven constraints on the dative alternation, we also suspect some influence from the surface string. Among the sentences with PP recipients, for example, we find many that have proper name recipients, proper name agents and non-pronominal themes. As DOS is only available for pronouns, a double object construction would lead to two proper names in a row (example (14-a)). Realizing the recipient as a PP argument successfully avoids this sequence of proper names (example (14-b)).

- (14) a. Daar gaf volgens de overlevering God Mozes het gebod
 there gave following the tradition God Moses the commandment
 “Gij zult niet stelen”.
 thou shalt not steal
- b. Daar gaf volgens de overlevering God aan Mozes het
 there gave following the tradition God to Moses the
 gebod “Gij zult niet stelen”.
 commandment thou shalt not steal
Tradition has it that this is the place where God gave Moses the commandment “Thou shalt not steal”.

Finally, a radically different approach on word order is taken by Reinhart (1996). She argues that the sentence focus is determined by the position of the main stress: the focus of IP is a(ny) constituent containing the main stress of IP. Usually, main stress falls on the right edge of the middle field in Dutch. If the focus of IP is a constituent that does *not* contain the rightmost phrase in the middle field, there are two options:

stress shift or scrambling. Reinhart claims that scrambling is more economical than stress shift and therefore the preferred strategy for stress (and thus focus) assignment.

The non-canonical versions of the double object construction and the dative PP-construction could be regarded as scrambling and even the NP/PP alternation could be regarded as a means of avoiding stress shift. It would nicely explain why the phonologically weak pronoun *het* almost always shifts and why we find so few emphasized forms of the pronouns (e.g. *hijzelf*, 'he himself') in shifted position. However, the data in (15), on which Reinhart (1996) bases her theory, are not uncontroversial.

- (15) a. *Ik heb de krant nog niet gelezen, maar ik heb het boek al
 I have the newspaper yet not read but I have the book already
 wel gelezen.
 indeed read
- b. Ik heb nog niet de krant gelezen, maar ik heb al wel het
 I have yet not the newspaper read but I have already indeed the
 boek gelezen.
 book read
I haven't read the newspaper yet, but I did read the book already.

In any case, focus cannot be the full explanation for the DOS: the alternation between canonical and non-canonical orderings persists even if both arguments are reduced pronouns and therefore necessarily unstressed ((16)).

- (16) a. Jo wees ze 'm aan.
 Jo pointed them him on
Jo pointed them out to him.
- b. Jo wees me ze aan.
 Jo pointed me them on
Jo pointed them out to me.

4 Additional evidence: the AcI construction

In this section we illustrate that the various constraints for aligning different sorts of NPs can also be applied to other word order alternations. We show how these detailed constraints account for the distribution of embedded object shift (EOS) in the Accusativus cum Infinitivo (AcI) construction.

The AcI construction illustrated in examples (17) and figure 1 is headed by a sensory verb, the verb *laten* (to let) or the verb *helpen* (to help). The verb takes an object and an XCOMP. The embedded subject is functionally controlled by the object.

- (17) a. Ik zag Jo een boek lezen.
 I saw Jo a book read
I saw Jo reading a book.
- b. Ik zag jou Jo helpen zwemmen.
 I saw you Jo help swim

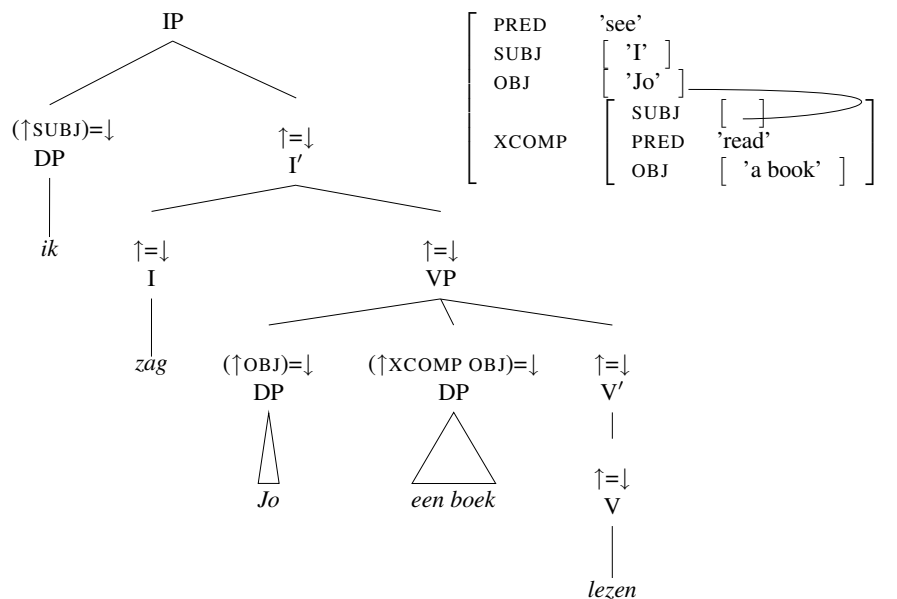


Figure 1: C-structure and f-structure for an AcI-construction in Dutch

I saw you helping Jo to swim.

Several LFG analyses of this construction exist, e.g. Bresnan et al. (1982), Zaenen and Kaplan (1995) and Kaplan and Zaenen (2003). All nominal arguments (also the embedded ones) are selected for in the VP, all verbal arguments in V', thus accounting for the crossing dependencies that occur when one AcI constructions is embedded in another, as illustrated in (17-b):

(18) C-structure rules for the AcI-construction (Kaplan and Zaenen, 2003)

$$VP \rightarrow \left(\begin{array}{c} NP^* \\ (\uparrow \text{XCOMP}^* \text{OBJ}) = \downarrow \end{array} \right) V'$$

$$V' \rightarrow V \left(\begin{array}{c} V' \\ (\uparrow \text{XCOMP}) = \downarrow \\ (\uparrow \text{XCOMP}^+ \text{OBJ}) \not\prec_f (\uparrow \text{OBJ}) \end{array} \right)$$

The order of the nominal arguments is restricted to the canonical word order in (17-a) and (17-b) by the f-precedence requirement $(\text{XCOMP}^+ \text{OBJ}1) \neg \prec_f (\uparrow \text{OBJ}1)$ (Kaplan and Zaenen, 2003). This constraint says that the constituent that maps onto the embedded OBJ1 in the f-structure cannot precede the constituent that maps onto the f-structure of the main clause direct object. However, under certain conditions, the embedded object can shift over the higher object (or embedded subject) (19). In other words: the f-precedence constraint is violable. The conditions under which we find EOS resemble the conditions on DOS. A difference is that DOS was only blocked with local person pronouns, while EOS is blocked with all animate pronouns. This is best illustrated with

animate and inanimate examples of the weak pronoun *ze* (them) ((20-b)-(20-a)). Note that inanimate objects are very unmarked. More marked objects have to stay in their canonical object position.

- (19) a. Ik zag 't Jo doen.
 I saw it Jo do
I saw Jo doing it.
- b. Ik zag dat haar ouders doen.
 I saw that her parents do
I saw her parents doing that.
- c. Ik zag ze dat doen.
 I saw then that do
I saw them doing that.
- d. Ik zag het ze doen.
 I saw it them do
I saw them doing it.
- (20) a. Ik heb ze Jo door zien slikken.
 I have them you seen swallow
I saw you swallowing them.
- b. %I heb ze Jo zien zoenen.
 I have them Jo seen kiss
I saw Jo kissing them.

We can model the restrictions on the argument ordering in the AcI with a set of OT constraints similar to the one used for the OS in the double object construction. The only difference is that we have to exclude third person animate pronouns and that CANON now applies to OBJ1 and XCOMP OBJ1, instead of OBJ1 and OBJ2. For this purpose, we adapt the definition of CANON and formulate the constraint conjunction LAX.

- (21) CANON: SUBJ<_f OBJ2<_f OBJ1<_f OBL, XCOMP OBJ1

*LEFT ANIMATE XCOMP OBJ1 (LAX): *OBJ1_{anim}&CANON

The rest of the analysis works as for the OS in the double object construction, as illustrated in table 11.

5 Conclusion and discussion

We investigated the influence of various alignment principles on the dative alternation: canonical word order, light precedes heavy and pronouns precede full NPs. We found the influence of word order reflected in the general distribution of the four realizations of the dative alternation. Weight proved an important factor for both the NP/PP alternation and the ordering of the arguments in the PP construction. The principle ‘pronouns precede full NPs’ was made more specific to account for the direct object shift in the double object construction and the AcI.

Input: <i>saw</i> (<SUBJ><OBJ1><XCOMP>)			LAX	PRO _{it} -L	PRO-L	CANON
OBJ1='Jo' XOBJ1='a book' ex.(17-a)	☞	OBJ1 XOBJ1 XOBJ1 OBJ1				*!
OBJ1='Jo' XOBJ1='it' ex.(19-a)	☞	OBJ1 XOBJ1 XOBJ1 OBJ1		*!	*	*
OBJ1='her parents' XOBJ1='that' ex.(19-b)	☞	OBJ1 XOBJ1 XOBJ1 OBJ1			*!	*
OBJ1='them' XOBJ1='that' ex.(19-c)	☞	OBJ1 XOBJ1 XOBJ1 OBJ1			*	*!
OBJ1='Jo' XOBJ1='them' ex.(20-b)	☞	OBJ1 XOBJ1 XOBJ1 OBJ1	*!		*	*

Table 11: Embedded Object Shift in the Acl

No evidence was found for independent influence of person or definiteness on the dative alternation or for independent influence of pronominality on the NP/PP alternation. This is contrary to the work of Bresnan and Nikitina (2003) for English (with regard to person) and the predictions of the Left-Right Principle (with regard to definiteness and pronominality). The extent and nature of the influence of several other factors was left for further investigation.

The model presented in this paper accounts for the most frequent patterns. The corpus data clearly showed, however, that variations on these patterns occur. One would need a stochastic implementation (Boersma and Hayes, 2001) of the constraint ranking to account for those less frequent outputs. It would be interesting to see whether such an implementation would predict the frequency distributions that we observed in the corpora.

References

- Aissen, Judith. 2003. Differential object marking: Iconicity vs. economy. *Natural Language & Linguistic Theory*, 21(3):435–483.
- Arnold, Jennifer E., Thomas Wasow, Anthony Losongco, and Ryan Ginstrom. 2000. Heaviness vs. newness: The effects of complexity and information structure on constituent ordering. *Language*, 76:28–55.
- van der Beek, Leonoor and Gerlof Bouma. 2004. The role of the lexicon in optimality theoretic syntax. In Miriam Butt and Tracy Holloway King, editors, *The Proceedings of the LFG '04 Conference*. CSLI Publications.

- van der Beek, Leonoor, Gosse Bouma, and Gertjan van Noord. 2002. Een brede computationele grammatica voor het nederlands. *Nederlandse Taalkunde*, 7(4):353–374.
- van der Beek, L., G. Bouma, R. Malouf, and G. van Noord. 2002. The alpino dependency treebank. In *Computational Linguistics in the Netherlands (CLIN) 2001*, Twente University.
- Behaghel, Otto. 1909/10. Beziehungen zwischen umfang und reihenfolge von satzgliedern. *Indogermanischen Forschungen*, 25:110–42.
- Boersma, Paul and Bruce Hayes. 2001. Empirical tests of the gradual learning algorithm. *Linguistic Inquiry*, 32(1):45–86.
- Bouma, Gosse and Geert Kloosterman. 2002. Querying dependency treebanks in XML. In *Proceedings of the Third international conference on Language Resources and Evaluation (LREC)*, Gran Canaria.
- Bouma, Gosse, Gertjan van Noord, and Rob Malouf. 2001. Alpino: Wide coverage computational analysis of dutch. In *Computational Linguistics in the Netherlands CLIN 2000*.
- Bresnan, Joan, Ronald M. Kaplan, Stanley Peters, and Annie Zaenen. 1982. Cross-serial dependencies in dutch. *Linguistic Inquiry*, 13:613–635.
- Bresnan, Joan and Tatiana Nikitina. 2003. On the gradience of the dative alternation. <http://www-lfg.stanford.edu/bresnan/new-dative.pdf>.
- Gundel, Jeanette K. 1988. Universals of topic-comment structure. In Michael Hammond, Edith A. Moravcsik, and Jessica R. Wirth, editors, *Studies in syntactic typology*. Amsterdam: John Benjamins, pages 209–39.
- Haeseryn, W. et al., editors. 1997. *Algemene Nederlandse Spraakkunst*. Nijhoff, Groningen.
- de Hoop, Helen. 2003. Scrambling in dutch: Optionality and optimality. In S. Karimi, editor, *Word Order and Scrambling*. Blackwell Publishers, Oxford, pages 201–216.
- Jäger, Gerhard and Anette Rosenbach. 2003. The winner takes it all - almost. cumulativity in grammatical variation. University of Potsdam and University of Dsseldorf.
- Kaplan, Ronald M. and Annie Zaenen. 2003. West-germanic verb clusters in lfg. In Pieter A.M. Sueren and Gerard Kempen, editors, *Verb Constructions in German and Dutch*. John Benjamins, pages 127–150.
- Krifka, Manfred. 2001. Lexical representations and the nature of teh dative alternation. Available on his webpage, November 9. University of Amsterdam.
- Lapata, Maria. 1999. Acquiring lexical generalizations from corpora: A case study for diathesis alternations. In *Proceedings of the 37th Meeting of the North American Chapter of the Association*, pages 397–404, College Park, MD.

- Lee, Hanjung. 2003. Prominence mismatch and markedness reduction in word order. *Natural Language & Linguistic Theory*, 21(3):617–680.
- Levelt, Pim. 1998. Corpus gesproken Nederlands. zie ook www.elis.rug.ac.be/cgn/.
- Levin, Beth. 1993. *English Verb Classes and Alternations. A Preliminary Investigation*. The MIT Press, Cambridge, Massachusetts.
- Moortgat, Michael, Ineke Schuurman, and Ton van der Wouden. 2001. CGN syntactische annotatie. internal report corpus Gesproken Nederlands.
- Müller, Gereon. 2001. Harmonic alignment and the hierarchy of pronouns in German. In Horst Simon and Heike Wiese, editors, *Pronouns - Grammar and Representation*. Amsterdam: Benjamins, pages 205–232.
- Pinker, Steven. 1989. *Learnability and Cognition. The Acquisition of Argument Structure*. The MIT Press, Cambridge, Massachusetts.
- Prince, Ellen. 1992. The ZPG letter: Subjects, definiteness, and information status. In Sandra Thompsons and William Mann, editors, *Discourse Description: Diverse analyses of a fundraising text*. Amsterdam: John Benjamins, pages 295–325.
- Reinhart, Tanya. 1996. Interface economy: Focus and markedness. In C. Wilder, H.M. Geartner, and M. Bierwisch, editors, *The role of economy principles in Linguistic Theory*. Akademie Verlag, Berlin.
- Smolensky, Paul. 1995. On the internal structure of the constraint component Con of UG. online. ROA-86-0000, <http://roa.rutgers.edu>.
- Uszkoreit, Hans. 1987. *Word order and constituent structure in German*. CSLI.
- Wasow, Thomas. 1997. Remarks on grammatical weight. *Language Variation and Change*, 9:81–105.
- Zaenen, Annie and Ronald M. Kaplan. 1995. Formal devices for linguistic generalizations: West Germanic word order in I-f. In Mary Dalrymple, Ronald M. Kaplan, John T. Maxwell III, and Annie Zaenen, editors, *Formal Issues in Lexical-Functional Grammar*. CSLI Publications, pages 215–239.
- Zwart, C. Jan-Wouter. 1996. Clitics, scrambling, and head movement in Dutch. In *Approaching Second: Second Position Clitics and Related Phenomena*. CSLI Publications, pages 579–612.
- Zwart, C. Jan-Wouter. 1997. *Morphosyntax of Verb Movement; A Minimalist Approach to the Syntax of Dutch*. Dordrecht: Kluwer.

Leonoor van der Beek
 vdbeek@let.rug.nl
 Center for Language and Cognition
 Faculty of Arts, University of Groningen
 P.O. Box 716, NL-9700 AS Groningen