Treebank Evidence for the Analysis of PP-Fronting

Gosse Bouma
Computational Linguistics
Rijksuniversiteit Groningen

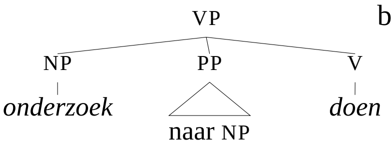
1 Introduction

A long-standing discussion in Dutch linguistics is concerned with the status of the PP in sentences like (1). In (1-a), a full PP appears in sentence initial position, and in (1-b), the initial pronoun *daar* is interpreted as the object of the preposition *naar*.

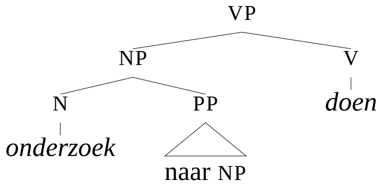
(1) a. **Naar** deze gebeurtenis wordt nu nader **onderzoek** gedaan.
Into this event is now further research done
Further research on this event is now done
b. **Daar** is echter nauwelijks **onderzoek naar** verricht
There is however hardly research into carried out
However, hardly any research on this has been carried out

The PP can be seen as a dependent of the noun *onderzoek* or of the main verb, thus the basic structure for sentence (1-a) could be either as in (2-a) (V+PP) or as in (2-b) (N+PP).

(2) a.



b.



The proper analysis of examples like those in (1) has been the topic of a heated debate (in Klein and van den Toorn (1977, 1979), and Kooij and Wiers (1979), among others), initiated by the observation in Bach and Horn (1976) that, according to the then current version of Transformational Grammar, extraction from NPs should not be possible in languages like Dutch. Thus, they are forced to adopt analysis (2-a) for both (1-a) and (1-b).

A similar issue arises in German. De Kuthy (2000) notes that sentences such as (3) (which she refers to as 'NP–PP *split*') have often been analyzed as involving extraction out of NP, but also as involving extraction out of a VP (perhaps after reanalysis).

(3) Über Syntax hat Hans sich ein Buch ausgeliehen
about syntax has Hans self a book borrowed
Hans borrowed a book on syntax

The N+PP analysis is intuitively plausible, as there seems to be a strong semantic relation between the noun and preposition. Furthermore, N+PP may precede the finite verb in main clauses, and thus clearly forms a constituent in some cases. Also, when N is preceded by certain definite determiners, fronting of the PP is almost impossible. This suggests PP-fronting is subject to a constraint on extraction from NP, something which seems highly problematic for a V+PP analysis. The V+PP analysis, on the other hand, is supported by the fact that PP-fronting seems to occur only with certain verbs. Furthermore, some nouns clearly select a PP, but do not allow fronting of this PP.

When constructing a treebank for Dutch, one frequently encounters examples such as (1) and a decision has to be made as to how to annotate these. The syntactic annotation of the Corpus of Spoken Dutch (CGN) (Moortgat, Schuurman and van der Wouden 2000) adopts an N+PP analysis for the following type of example:

(4) **daar** heb ik helemaal geen **zin in**
there have I totally no desire for
I have no desire for that at all

The Alpino-treebank of written Dutch,[1] on the other hand, has opted for the V+PP analysis. As one of the design goals of the Alpino-treebank was to produce output compatible with CGN, it seems that the annotation guidelines for either Alpino or CGN need to be reconsidered.

In this paper, we investigate to what extent corpus data can be used to decide on this matter. A corpus-based approach seems appropriate for at least two reasons. First, the claim that certain determiners block PP-fronting as well as the claim that PP-fronting occurs only with certain verbs, can be verified using corpus data. Second, there has been considerable disagreement between authors on the status of examples that were crucial in arguing for one or the other position. Examples considered ungrammatical in one paper were considered to be acceptable by authors arguing for a different analysis.[2] Coppen (1991) notes that the examples in

[1] see van der Beek, Bouma, Malouf and van Noord (2002) and `www.let.rug.nl/~vannoord/trees`

[2] See Klein and van den Toorn (1977, p. 432), Klein and van den Toorn (1979, p. 105) and Kooij and Wiers (1979, p. 488).

his paper show varying acceptibility, and that linguistic intuitions with respect to these data even seem to change over time.

In section 2, we describe the construction of a syntactically annotated corpus. Next, we investigate the role of the verb and the determiner in PP-fronting. We conclude that the verb plays an essential role in PP-fronting and that the preference for indefinite NPs PP-fronting may be related to this. We also observe that PPs may be included in relatives modifying the noun. This seems highly problematic for an N+PP analysis. A number of patterns which have been used as arguments for a particular analysis, are practically absent in the corpus. We conclude that the corpus data suggest that the V+PP analysis is more likely than the N+PP analysis, and that these expressions are best analyzed as phrasal verbs involving a prepositional complement.

2 Treebank Construction

We used the newspaper sections of the Twente News Corpus[3] (TWNC) as our initial corpus. The corpus contains text from major Dutch newspapers in the period 1994-2001, and has a size of approximately 300 million words. We believe that, at least for the phenomenon we are interested in, this corpus is representative for Dutch in general.

The discussion referred to in the introduction has focused on N+PP combinations displaying a strong semantic relation between the noun and the PP. Our first goal was to identify a number of such nouns in the corpus. To find N+P pairs with strong collocational properties, we ranked all N+P bigrams from the corpus using the *log-likelihood-test* of Dunning (1993). From the resulting list, we selected 16 bigrams as suitable candidates for our research (see table 1). Highly ranked bigrams which we discarded were parts of names (*ministerie van (ministry of)*), parts of complex prepositions (*(in) tegenstelling tot ((as) opposed to)*), and bigrams which did not occur in PP-fronting sentences.

Next, we automatically constructed two treebanks. The **general** corpus consists of sentences containing the relevant N+P combination. From the TWNC, we initially extracted 10.000 sentences per N+P collocation containing both N and P. 155.000 sentences were selected in total (as some bigrams did not occur 10.000 times). The *dependency tree* for each sentence was computed using the Alpino-system.[4] From the resulting treebank, we selected[5] those cases where the NP headed by N and the PP headed by P are both dependents of the same verb, or the

[3] wwwhome.cs.utwente.nl/~druid/TwNC/TwNC-main.html

[4] www.let.rug.nl/~vannoord/alp

[5] using the XML-tool for searching dependency trees described in Bouma and Kloosterman (2002).

behoefte aan	*need for*	kritiek op	*critique on*
belangstelling voor	*interest in*	onderzoek naar	*investigation into*
bezwaar tegen	*objections against*	protest tegen	*protest against*
contact met	*contact with*	relatie tussen	*relation between*
discussie over	*discussion about*	twijfel aan	*doubt about*
gebrek aan	*lack of*	verhaal over	*story about*
gesprek over	*conversation about*	verschil tussen	*difference between*
informatie over	*information about*	vraag naar	*demand for*

Table 1: Selected N+P collocations

PP is a dependent of N, and the NP headed by N is a dependent of a verb (i.e. and not part of a PP or other non-verbal constituent). Almost 51.000 sentences (containing 2.000 to 4.000 examples per N+P collocation) satisfy the syntactic selection criteria.

The second, **split**, corpus, consisted of PP-initial sentences and sentences containing a discontinuous PP consisting of an R-pronoun and a preposition (i.e. as in (1-b)). All sentences from the TWNC containing both N and P, but where P was also the first word in the string, were collected. Initially, this set consisted of almost 6.000 sentences. After syntactic analysis and selection, this was reduced to approximately 2.400 cases. The corpus was extended with all cases from the *general* corpus, containing a discontinuous PP. This gave rise to another 1.100 cases. Thus, the *split* corpus contains about 3.500 sentences (containing between 35 and 615 examples per N+P bigram).

Using an automatically constructed treebank, rather than raw or POS-tagged text, is essential for our purposes for two reasons. First, not all sentences containing N and P are actually valid instances of the pattern we are interested in (P might be heading a PP containing an NP headed by N, or the PP might be part of another NP, for instance). Second, we want to investigate which verbs co-occur with these N+P collocations. Therefore, we must be able to determine which verb actually selects for the NP headed by N. As we are interested in investigating the status of the PP, we need to consider both the case where the PP is analyzed as a dependent of N and the case where the the PP is analyzed as a dependent of V.

The main reason why we opted for using an automatically constructed treebank, instead of a manually annotated treebank, is size. Eventhough we used the full 300 million word TWNC as our source, the *split* corpus is relatively small, often containing less than 100 examples per N+P bigram. The largest manually annotated treebank for Dutch, CGN, contains only 1 million words, and contains only a few

N+P bigram occurring more than 3 times in a *split* configuration. Still, one might wonder whether automatic analysis is sufficiently reliable to create a representative corpus. Allthough automatic analysis is not completely error-free,[6] the effect it has on the task at hand seems small. Automatic analysis does reliably filter cases where the NP is not a dependent of a verb, or where the NP and PP are dependents of a different verb. Also, the main verb selecting NP or both NP and PP is identified reliably. Finally, note that manually annotated corpora are not error free either. For the Alpino-corpus, for instance, at the end of the project an inter annotator agreement of 94.6% was achieved. Thus, the difference in reliability between manually and automatically annotated data is a gradient, rather than absolute. Nevertheless, errors do sometimes occur, and thus we did manually inspect many of the results found in the experiments below, especially cases involving small numbers.

3 The role of the verb

The idea that fronting of a PP or a discontinuous PP is possible only with certain verbs, has been used as argument for the V+PP analysis. In this section, we investigate whether corpus-data confirm this intuition.

Using the information provided by automatic syntactic analysis, as described in the previous section, we we counted how often a specific verb occurs with a specific N+P collocation in the *general* and in the *split* corpus. In particular, we counted the verbs with a dependent NP headed by N and containing a PP (headed by P) functioning as a dependent of the verb or the noun. To avoid inclusion of (verbs functioning as) auxiliaries and modals, verbs with a VP-dependent were excluded. If the possibility of a *split* configuration is determined by the verb, only a limited number of verbs should be found in the *split* corpus, and in *split* these verbs should occur more frequently than in the general corpus. In table 2, we present an overview of verbs found more than once in *split* and *general*, for 4 representative N+P combinations.

Table 2 shows that the combination *behoefte aan* mainly occurs with *hebben* en *zijn*. There are significant differences in the distribution of these verbs between *split* and *general*, however. For allmost all N+P collocations we investigated, statistically significant differences in distribution can be observed for the most frequent verbs. In some cases, significant differences for low frequent verbs can be observed as well.

The verbs *hebben, zijn* and *bestaan* are special in that they seem to allow *split* with almost all investigated N+P combinations. The role of *bestaan* is remark-

[6] Malouf and van Noord (2004) report that the Alpino-system identifies dependency relations with an accuracy of 87.8% on a representative 500 sentence subset of the TWNC.

		Split	Gen
behoefte aan	N=	583	5699
hebben (*have*)	●	73.6	53.8
zijn (*be*)	●	19.0	24.5
bestaan (*exist*)		4.6	4.6
blijken (*turn out to be*)		0.5	0.5
blijven (*remain*)		0.5	0.6
toenemen (*increase*)	●	0.3	1.7
belangstelling voor	N=	428	5124
hebben (*have*)	●	33.4	28.4
zijn (*be*)	●	31.8	23.5
bestaan (*exist*)	●	20.6	5.2
tonen (*show*)	○	4.9	7.4
komen (*come*)		1.4	1.1
blijken (*turn out to be*)		0.9	0.8
verwachten (*expect*)		0.7	0.4
ontstaan (*come up*)		0.5	1.1
blijven (*remain*)		0.5	0.5

		Split	Gen
discussie over	N=	164	3857
zijn (*be*)	●	40.2	15.3
bestaan (*exist*)	●	10.4	0.7
voeren (*be engaged in*)		10.4	7.3
gaan (*go*)		9.8	7.4
hebben (*have*)	●	5.5	2.0
woeden (*rage*)		4.3	2.4
ontstaan (*come up*)		4.3	3.7
losbarsten (*burst out*)		2.4	1.9
ontbranden (*ignite*)		1.2	0.6
houden (*hold*)	●	1.8	0.6
gebrek aan	N=	303	2574
zijn (*be*)	●	62.4	32.8
hebben (*have*)	●	28.4	9.1
bestaan (*exist*)	●	2.6	0.6
liggen (*lay*)		1.0	0.3
heersen (*rule*)		1.0	0.5
lijken (*seem*)		0.7	0.5

Table 2: Distribution of verbs for several N+P collocations. Differences marked with ● (○) are significant according to the χ^2 test at p=0.05 (p=0.10).

able: this otherwise rather infrequent verb occurs frequently with 10 of the 17 investigated N+P combinations. In Haesereyn *et al.* (1997), Broekhuis (2004) and Loonen (2003) (non-exhaustive) lists of phrasal verbs involving a PP-complement are given. A considerable number of V+N+P combinations in the *split* corpus are presented as phrasal verbs in at least one of these sources, e.g. *een gesprek voeren met (be engaged in a conversation with), informatie verstrekken over (provide information on), een onderzoek instellen naar (start an investigation into), een onderzoek loopt naar (an investigation is being carried out into), protest rijst tegen (protest is raised against), een verhaal gaat over (a story is about), een verhaal doet (de ronde) over (a story goes around about)*, en *(er) zit een verschil tussen (there is a difference between).*

The V+PP analysis also predicts that for some verbs, PP-fronting should be impossible. This prediction is hard to test, as the absence of a verb in *split* might be due to lack of data. Nevertheless, in table 3 we provide a list of verbs missing in *split* which occur with more than 1% of the relevant N+P example sentences in the general corpus. All verbs listed for *gebrek aan* seem to resist PP-fronting. In other cases, fronting seems marked (*aan* NP *groeit er behoefte (for* NP *grows*

behoefte aan		Split=583 G=5699
onstaan (*come up*)	●	1.4
groeien (*grow*)	●	1.3
belangstelling voor		Split=428 G=5124
wekken (*wake*)	●	1.1
discussie over		Split=164 G=3857
beginnen (*start*)	●	2.7
aanzwengelen (*start up*)	○	1.8
volgen (*follow*)		1.0
aangaan (*engage in*)		1.0
krijgen (*get*)		1.0
gebrek aan		Split=303 G=2574
verwijten (*blame*)	●	8.7
compenseren (*compensate*)	●	2.4
opbreken (*stumble over*)	●	1.6
noemen (*mention*)	●	1.6
leiden (*lead to*)	●	1.5
vinden (*find*)	●	1.3
spelen (*play*)	●	1.3
hekelen (*criticize*)	○	1.2
worden (*become*)	○	1.0
kritiek op		Split=215 G=4077
toenemen (*increase*)	○	1.3
onderzoek naar		Split=228 G=3570
leiden (*lead*)	●	1.8
willen (*want*)	●	1.8
gelasten (*demand*)	○	1.5
eisen (*demand*)	○	1.2
aankondigen (*announce*)		1.1
twijfel over		Split=154 G=1714
uiten (*utter*)	●	2.5
wegnemen (*take away*)	○	1.9
groeien (*grow*)		1.6
verschil tussen		Split=130 G=3925
bedragen (*amount to*)	○	2.7
worden (*become*)		2.0
weten (*know*)		1.8
kennen (*know*)		1.5

Table 3: Frequent N-P-V-combinations in the general corpus (G), absent in *split*. Differences marked with ● (○) are significant according to the χ^2 test at p=0.05 (p=0.10).

the demand), naar NP *leidt/eist* NP *een onderzoek (into* NP, NP *demands an investigation), naar* NP *kondigt* NP *een onderzoek aan (into* NP, NP *announces an investigation), over* NP *nam* NP *alle twijfel weg (on* NP, NP *took all doubts away)*). For other verbs and N+P combinations, it seems that fronting is at least theoretically possible. The limited size of the *split* corpus might be the reason why these are absent in our data.

The corpus data clearly suggest that the verb plays a role in the possibility of PP-fronting and discontiuous PPs. The distribution of verbs in *split* and *general* shows large differences for most investigated N+P combinations. For frequent verbs, these differences are often statistically significant. Furthermore, there seem to be a number of verbs which easily combine with certain N+P combinations, but which do not allow *split* configurations.

	Split	**Gen**
determiner N=	3.601	50.892
geen (*no*)	30.7	8.0
NULL	27.7	31.8
een (*a*)	14.4	16.5
veel (*many/much*)	7.7	2.1
de/het (*the*)	7.3	32.9
weinig (*few/little*)	3.8	0.7
enkele (*some*)	2.1	0.8
meer (*more*)	0.8	1.0
minder(*less*)	0.6	0.2

Table 4: Frequency of determiners preceding the relevant noun in *split* and the general corpus.

4 The role of the determiner

It has been argued that so-called *specified subjects* within the NP block extraction:

(5) a. Over Piet herinnerde hij zich een verhaal.
About Piet remembered he REFL a story
He remembered a story about Piet
b. *Over Piet herinnerde hij zich Jans verhaal.
About Piet remembered he REFL Jan's story

An NP contains a specified subject if its determiner is a genitive NP or a possessive pronoun. The existence of a constraint like this would be a strong argument for the N+PP analysis. In this section, we investigate whether there is a relationship between the distribution of determiners and PP-fronting.

In table 4, a comparison of the frequencies in *split* and *general* is made of the most common determiners preceding the relevant noun. The indefinite determiners *geen, veel* and *weinig* occur relatively frequently in *split*, whereas the definite determiner *de/het* is relatively infrequent in *split*.

We believe that the difference in distribution of determiners in *split* and the *general* corpus can be explained to a large extent by the fact that the verbs in *split* and *general* have a very different distribution (as shown in the previous section). If we restrict attention to N-P-V combinations that contain a verb which is relatively frequent in *split*, we see that the definite determiner is much less frequent in *general* as well. This is illustrated in table 5.

At first sight, the corpus seems to confirm the observation that PP-fronting requires an NP which does not contain a 'specified subject' in the form of a possessive pronoun or genitive NP. Table 4 does not contain any of these determiners. Genitives are in fact absent in *split*, while possessives are scarce, and restricted to the N+P combinations *verhaal over* en *twijfel over*:

N+V+P	N=	determiners
behoefte hebben aan *have need for*	3001	NULL 60.4, geen 25.5, ...,de 1.0
behoefte zijn aan *be need for*	1051	NULL 71.6, een 10.5, geen 5.6, ..., de 2.1
behoefte bestaan aan *exist need for*	259	NULL 52.1, een 18.9 geen 11.6, de 5.8
belangstelling hebben voor *have interest in*	1343	NULL 70.3, geen 12.9 ..., de 0.5
bezwaar hebben tegen *have objection against*	1431	geen 53.9, NULL 34.2 ..., het 0.0
contact zoeken met *seek contact with*	462	NULL 93.9, geen 4.1 het 1.1
discussie zijn over *be discussion about*	257	NULL 36.6, de 16.3 geen 14.0, een 12.8
gesprek voeren met *be engaged in discussion with*	250	een 90, het 4.8 geen 1.6

Table 5: Frequency of common indefinite and definite determiners in the general corpus for frequent N-P-V-combinations in *split*.

(6) a. Over die worsteling gaat mijn verhaal
about that struggle goes my story
My story is about that struggle

b. Over adverteren in de verzorgingssfeer heeft hij zijn twijfels
About advertising in the health sector has he his doubts
He has his doubts about advertising in the health sector

Only the phrase *twijfels hebben over* is relatively frequent in *split*.

One might argue that the absence of genitives and the apparently highly restricted use of possessives, is evidence for the claim that PP-fronting is blocked for certain NPs. It should be noted, however, that NPs introduced by a possessive pronoun or genitive are not very frequent in the *general* corpus either: 2.1% of the relevant NPs in *general* contains a possessive pronoun and 0.9% a genitive NP. Furthermore, those verbs which do occur with this type of NP seem to be highly infrequent in *split*.

The preference for indefinite determiners in the *split* data correlates strongly with the preference for indefinite determiners in the general corpus, if one restricts attention to those verbs which are frequent in *split*. Furthermore, the absence of NPs introduced by a genitive and the restricted possibilities for using possessive pronouns seems to be a consequence of the fact that these are scarce in general, especially if one also takes the verb into account. It seems therefore that the differences in determiner distribution are for the most part a consequence of the differences in the distribution of the verbs in both corpora.

5 Related Corpus Observations

In this section we briefly discuss various corpus observations that are relevant for the analysis of PP-fronting.

We encountered one construction which has not been discussed in the literature but which seems problematic for an N+PP analysis. In relative clauses modifying the noun, the PP is sometimes clearly embedded in the relative clause ((7)). For PPs which are unambiguously part of the NP (and which cannot be fronted) this is not possible ((8)).

(7) a. De **kritiek** die hier **op** het boek wordt uitgeoefend
the critique that here on the book is offered
the critique on the book which is offered here

b. de **belangstelling** die Eduard **voor** het nazisme toonde
the interest which Eduard for the nazism showed
the interest which Eduard showed for Nazism

(8) *de **demonstratie** die **tegen** de hoge werkdruk in chaos ontaardde
the demonstration which agains the high work-load in chaos ended

Thus, the possibility of a PP to appear inside a relative clause is evidence for the fact that the PP can be interpreted as a dependent of the verb.

In the general corpus, for most of the N+P combinations we investigated, several examples can be found where the PP is included in a relative clause. This seems problematic for a N+PP analysis. Under such an analysis, it seems that the relative pronoun would have to inherit the selection or subcategorization properties of the noun it modifies. Furthermore, a mechanism needs to be established which allows the PP to appear in a position non-adjacent to the relative pronoun (i.e. head-movement, remnant movement, or argument transfer from the pronoun to the verbal head). We believe the syntactic literature does not provide evidence for assuming that such processes are at work here.

One argument for the V+PP analysis has been the suggestion that one also finds cases of ‘NP-fronting’, where the PP occupies a position in the ‘Mittelfeld’:

(9) Een **roman** heb ik **van** Vestdijk gelezen
A novel have I of Vestdijk read
I have read a novel by Vestdijk

Such examples are practically absent (i.e. we were able to find only 3 convincing examples) in the general corpus. The difference in frequency between PP-fronting and 'NP-fronting' is puzzling.

Another argument for the V+PP analysis has been the claim that the NP and PP may be separated from each other within the Mittelfeld. In the general corpus, we did not find a single example of an NP-XP-PP word order, however. We found only 4 examples of PP-XP-NP word order. On the other hand, PP-NP orders, as in (10), are relatively common in the general corpus (with 10-50 examples per N-P combination, except for *protest tegen*, for which we found only a single example):

(10) De Marokkanen hebben **aan** groepsvorming geen **behoefte**.
The Maroccans have on group formation little interest
the Maroccans have little interest in such group formation

Although this pattern seems equally problematic for an N+PP analysis as PP-XP-NP order, it has not been mentioned as such in the literature.

6 Concluding Remarks

Corpus investigation suggests that PP-fronting and discontinuous PPs are best analyzed as involving a PP which is a dependent of the verb. Certain verbs are far more frequent in *split* sentences than in general sentences containing the relevant N+P combination. The difference in the distribution of determiners in the *split* and *general* corpus seems to be mainly a consequence of the difference in distribution of verbs in both corpora. The corpus also contains a fair number of sentences containing PPs within relative clauses of the noun. Such examples seem problematic for an N+PP analysis. A number of patterns which have been used as argument for a specific analysis of PP-fronting are hardly encountered in a large corpus, except for PP-NP patterns.

The strong semantic relation between the noun and the preposition suggests that the examples we have investigated are examples of *phrasal verbs*, involving a verb with a more or less fixed NP-complement and a PP-complement. A similar conclusion was reached by Coppen (1991), who argues for an analysis which treats the PP as an argument selected by the combination of NP+V, i.e. a (pseudo) phrasal verb.

References

Bach, E. and Horn, G.(1976), Remarks on conditions on transformations, *Linguistic Inquiry* **7**, 265–299.

Bouma, G. and Kloosterman, G.(2002), Querying dependency treebanks in XML, *Proceedings of the Third international conference on Language Resources and Evaluation (LREC)*, Gran Canaria.

Broekhuis, H.(2004), Het voorzetselvoorwerp, *Nederlandse Taalkunde*.

Coppen, P.-A.(1991), Over vooropstaande PP's is het laatste woord nog niet gesproken, *Gramma* **15**(3), 209–225.

de Kuthy, K.(2000), Splitting PPs from NPs, *in* T. Kiss and D. Meurers (eds), *Constraint-Based Approaches to Germanic Syntax*, CSLI Publications, Stanford University, pp. 25–70.

Dunning, T.(1993), Accurate methods for the statistics of surprise and coincidence, *Computational linguistics* **19**(1), 61—74.

Haesereyn, W., Romijn, K., Geerts, G., De Rooy, J. and Van den Toorn, M.(1997), *Algemene Nederlandse Spraakkunst*, Martinus Nijhoff Uitgevers Groningen / Wolters Plantyn Deurne. Tweede, geheel herziene druk.

Klein, M. and van den Toorn, M.(1979), Van NP-beperking tot XP-beperking; een antwoord op Kooij en Wiers 1978, *De Nieuwe Taalgids* **72**, 97–102.

Kooij, J. and Wiers, E.(1979), Beperkingen en overschrijdingen: een antwoord aan Klein en Van den Toorn, *De Nieuwe Taalgids* **72**, 488–493.

Loonen, L.(2003), *Stante pede gaande van dichtbij langs AF bestemming @*, PhD thesis, Universiteit Utrecht, Utrecht.

Malouf, R. and van Noord, G.(2004), Wide coverage parsing with stochastic attribute value grammars, *IJCNLP-04 Workshop Beyond Shallow Analyses - Formalisms and statistical modeling for deep analyses*, Hainan.

Moortgat, M., Schuurman, I. and van der Wouden, T.(2000), CGN syntactische annotatie. Internal Project Report Corpus Gesproken Nederlands, see http://lands.let.kun.nl/cgn.

van der Beek, L., Bouma, G., Malouf, R. and van Noord, G.(2002), The Alpino dependency treebank, *Computational Linguistics in the Netherlands (CLIN) 2001*, Twente University.