A corpus investigation of PP-fronting in Dutch

Gosse Bouma
Rijksuniversiteit Groningen

Abstract
A long-standing discussion in Dutch syntax concerns the question whether PP dependents of a noun may be fronted. Although examples which apparently illustrate this pattern can be easily found, it is difficult to come up with linguistic arguments which show once and for all that the fronted PP is actually a dependent of a noun, and not of a verb. In this paper, we investigate to what extent corpus data can be used to decide on this matter, and conclude that the data suggest that the PP is a dependent of the verb.

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 is interpreted as the object of the preposition naar.

(1) a. Naar deze pijnlijke gebeurtenis wordt nu nader onderzoek gedaan.
    Further research on this painful event is now further research done.

b. Daar is echter nauwelijks onderzoek naar verricht
    There is however hardly any research into carried out

Bach and Horn (1976) argued that extraction from Dutch NPs should not be possible, and thus, that neither fronting of an NP-internal PP nor fronting of an object of a PP contained within an NP, should be possible. At first sight, the examples in (1) clearly seem to falsify this claim. However, arguments that the PP in (1-a) and the preposition in (1-b) are actually (heading) a dependent of the verb have also been put forward.

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.

The proper analysis of examples like those in (1) has been the topic of a heated discussion (in Klein and van den Toorn (1977, 1979), and Kooij and Wiers (1979), among others). Coppen (1991) reviews most of the arguments presented earlier, and concludes that the most convincing arguments point towards a $V+PP$ analysis.

Although many of the theoretical assumptions which played a role in the original discussion have changed, the question whether fronting of PP-dependents from an NP is possible, is still relevant. The syntactic annotation of the Corpus of Spoken Dutch (CGN) (Moortgat, Schuurman and van der Wouden 2000), for instance, adopts an $N+PP$ analysis for the following type of example:

(2) a. daar heb ik helemaal geen zin in there have I totally no desire for $I$ have desire for that at all

b. m-hu wij uh zullen daar goede nota van nemen Uh m uh shall there good notice of take We shall take good notice of that

The Alpino-grammar (van der Beek, Bouma and van Noord 2002), on the other hand, does not allow extraction out of NPs, and thus has opted for the $V+PP$ analysis. As one of the design goals of the Alpino system was to produce output compatible with CGN, it seems that either the Alpino-grammar should be modified, or the CGN-annotation guidelines 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 marked with a star in one paper were considered to be acceptable by authors arguing for a different analysis.1 Coppen notes that the examples in 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 and annotation of the corpus. Next, we investigate the role of the verb in PP-fronting. We find that some verbs are far more frequent in PP-initial sentences containing the relevant $N+P$ combination than in general sentences with this $N+P$ combination. In section 4, we look at the determiner preceding the noun which apparently selects for the PP. There is a strong preference for indefinite determiners in PP-initial sentences, while possessive pronouns and genetive NPs are almost absent. We argue that this is not necessarily evidence for a constraint on extraction from NP. Instead, it seems that the verbs

---

1 I.e. see Klein and van den Toorn (1977, p. 432), Klein and van den Toorn (1979, p. 105) and Kooij and Wiers (1979, p. 488).
which easily admit PP-fronting, also have a preference for indefinite NP-dependents. In section 5, we observe that PPs may also be included in relatives modifying the noun, as in:

(3) de enige relatie die er tussen haar en Van Kooten bestaat

the only relation that exists between her and Van Kooten

This seems highly problematic for an N+P analysis.

In section 6, we note that a number of patterns which have been used as arguments for a particular analysis, are practically absent in the corpus.

The data suggest that the verb plays an important role in the question whether PP-fronting is possible, and that the PP is thus best seen as a dependent of the verb. In section 7, we note that many of the nouns combining with a prepositional complement in CGN behave similar to the examples we investigated and that many of the frequent N+P+V combinations found in PP-initial sentences have been identified as phrasal verbs taking a prepositional complement by other authors.

We conclude therefore 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 Corpus Construction

We used the newspaper sections of the Twente News Corpus\(^2\) (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.

2.1 Selection of relevant N+P combinations

Sentence-initial PPs in many cases are clearly dependents of a verb, and in many other cases, could equally well be seen as a dependent of a verb or a noun. 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.

In principle, one might try to find nouns selecting for a PP by looking at frequent N+P bigrams in the corpus. However, raw frequency by itself is not a very good indication of the fact that the nouns actually selects for a PP (rather than just co-occurring regularly with a modifier PP headed by this P). Therefore, we used the log-likelihood-test of Dunning (1993) to find promising candidates. From the list of N+P bigrams with strong collocational properties according to this test, we manually selected 20 bigrams as suitable candidates for our research.\(^3\) Three

\(^2\)wwwhome.cs.utwente.nl/~druid/TwNC/TwNC-main.html

\(^3\)Highly ranked bigrams which we discarded among others were parts of names (ministerie van (ministry of)), and parts of complex prepositions ((in) tegenstelling tot ((as) opposed to)).
bigrams turned out to be very infrequent in PP-initial sentences, or gave rise to a large number of errors in automatic syntactic analysis (see below). These were discarded, and thus we used 17 bigrams (given in table 1) for further research.

2.2 Construction of Analyzed Subcorpora

Using the N+P collocations in table 1, we constructed a syntactically analyzed general corpus as follows:

1. From the TWNC, we initially extracted per N+P collocation maximally 10,000 sentences containing both N and P. 155,000 sentences were selected in total (as some bigrams do not occur 10,000 times).

2. The dependency tree for each sentence was computed using the Alpino-system.\(^4\).

3. From the syntactically analyzed sentences, we selected\(^5\) the examples which satisfied either one of the criteria below:

   • The NP headed by N and the PP headed by P are both dependents of the same verb, or
   • The 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).

\(^4\)www.let.rug.nl/~vannoord/alp

\(^5\)using the XML-tool for searching dependency trees described in Bouma and Kloosterman (2002).
Syntactic analysis is important 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 PP is analyzed as a dependent of V. The selection step in 3 retrieved 56,000 examples containing between 2,000 and 4,000 examples per N+P collocation on average.

The general, 56,000 sentence, corpus described above was contrasted with a second corpus, which consisted of PP-initial sentences only. This ‘P1’ corpus was constructed by exhaustively searching the twnc for sentences containing both N and P, but where P was also the first word in the string. Initially, the P1 corpus consisted of almost 10,000 examples. After syntactic analysis and selection, this was reduced to approximately 5,000 cases. The corpus is dominated by sprake van, which occurred no less than 3,872 times in P1. For the other collocations, between 30 and 326 examples were found per collocation.

One might wonder whether automatic analysis is sufficiently reliable to create a representative corpus. Even though automatic analysis is not completely error-free,\(^6\) the effect it has on the task we are interested in 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. 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 is possible only with certain verbs, has been used as argument for the V+PP analysis. A problematic aspect of this argument is the fact that linguistic intuition alone does not seem to be sufficient to draw the line between those verbs that do allow PP-fronting and those that do not. In this section, we investigate whether corpus-data provide a clearer answer.

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 P1 corpus. In particular, we counted the verbs with a dependent NP headed by N and containing a PP or with a dependent NP headed by N and a dependent PP headed by P. To avoid inclusion of (verbs functioning as) auxiliaries and modals, verbs with a VP-dependent were excluded. If PP-fronting is determined by the verb, only a limited number of verbs should be found in the P1 corpus, and in P1 these verbs should occur more frequently than in the general corpus. In table 2, we present an overview of verbs found more than

---

\(^6\)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.
once in P1 and in the general corpus, for the first 7 N+P combinations on our list and the (idiomatic) sprake van.

Table 2 shows that the combination behoefte aan mainly occurs with hebben en zijn in P1, but the same is true in the general corpus. The verb bestaan appears significantly more often with behoefte aan in P1 than in general. The absolute number of occurrences of other verbs cooccurring with behoefte aan is very small, so the differences in distribution are not statistically significant. For almost all N+P collocations we investigated, statistically significant differences in distribution can be observed for the most frequent verbs.

The verbs hebben, zijn and bestaan are special in that they seem to allow PP-fronting with almost all investigated N+P combinations. The role of bestaan is remarkable: this otherwise rather infrequent verb occurs frequently with 10 of the 17 investigated N+P combinations. There are also a number of verbs in P1 which clearly seem to form a phrasal verb with the N+P combination, 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 P1 might be due to lack of data. Nevertheless, in table 3 we provide a list of verbs missing in P1 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 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 P1 corpus might be the reason why these are absent in our data.

The corpus data clearly suggest that the verb plays a role in PP-fronting. The distribution of verbs in P1 and the general corpus 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 PP-fronting.

4 The role of the determiner

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

\[\text{(4)}\]
\[
\begin{align*}
a. & \quad \text{Over Piet herinnerde hij zich een verhaal.} \\
& \quad \text{About Piet remembered he REFL a story} \\
& \quad \text{He remembered a story about Piet} \\

b. & \quad *\text{Over Piet herinnerde hij zich Jans verhaal.} \\
& \quad \text{About Piet remembered he REFL Jan’s story}
\end{align*}
\]
<table>
<thead>
<tr>
<th></th>
<th>P1</th>
<th>Gen</th>
<th></th>
<th>P1</th>
<th>Gen</th>
</tr>
</thead>
<tbody>
<tr>
<td>behoefte aan</td>
<td>N= 241</td>
<td>5699</td>
<td>discussie over</td>
<td>N= 137</td>
<td>3857</td>
</tr>
<tr>
<td>hebben (have)</td>
<td>56.0</td>
<td>53.8</td>
<td>zijn (be)</td>
<td>42.3</td>
<td>15.3</td>
</tr>
<tr>
<td>zijn (be)</td>
<td>26.6</td>
<td>24.5</td>
<td>voeren (be engaged in)</td>
<td>12.4</td>
<td>7.3</td>
</tr>
<tr>
<td>bestaan (exist)</td>
<td>11.2</td>
<td>4.6</td>
<td>bestaan (exist)</td>
<td>12.4</td>
<td>0.7</td>
</tr>
<tr>
<td>blijken (turn out to be)</td>
<td>1.2</td>
<td>0.5</td>
<td>woorden (rage)</td>
<td>5.1</td>
<td>2.4</td>
</tr>
<tr>
<td>toenemen (increase)</td>
<td>0.8</td>
<td>1.7</td>
<td>ontstaan (come up)</td>
<td>5.1</td>
<td>3.7</td>
</tr>
<tr>
<td>blijven (remain)</td>
<td>0.8</td>
<td>0.6</td>
<td>gaan (go)</td>
<td>4.4</td>
<td>7.4</td>
</tr>
<tr>
<td></td>
<td>N= 326</td>
<td>5124</td>
<td></td>
<td>3.6</td>
<td>2.0</td>
</tr>
<tr>
<td>zijn (be)</td>
<td>37.4</td>
<td>23.5</td>
<td>hebben (have)</td>
<td>26.9</td>
<td>9.1</td>
</tr>
<tr>
<td>bestaan (exist)</td>
<td>27.0</td>
<td>5.2</td>
<td>ontbranden (ignite)</td>
<td>1.5</td>
<td>0.6</td>
</tr>
<tr>
<td>hebben (have)</td>
<td>19.3</td>
<td>28.4</td>
<td>houden (hold)</td>
<td>1.5</td>
<td>0.6</td>
</tr>
<tr>
<td>tonen (show)</td>
<td>6.4</td>
<td>7.4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>komen (come)</td>
<td>1.8</td>
<td>1.1</td>
<td>zijn (be)</td>
<td>63.6</td>
<td>32.8</td>
</tr>
<tr>
<td>blijken (turn out to be)</td>
<td>1.2</td>
<td>0.8</td>
<td>hebben (have)</td>
<td>26.9</td>
<td>9.1</td>
</tr>
<tr>
<td>verwachten (expect)</td>
<td>0.9</td>
<td>0.4</td>
<td>bestaan (exist)</td>
<td>2.7</td>
<td>0.6</td>
</tr>
<tr>
<td>ontstaan (come up)</td>
<td>0.6</td>
<td>1.1</td>
<td>liggen (lay)</td>
<td>1.0</td>
<td>0.3</td>
</tr>
<tr>
<td>blijven (remain)</td>
<td>0.6</td>
<td>0.5</td>
<td>heersen (rule)</td>
<td>1.0</td>
<td>0.5</td>
</tr>
<tr>
<td></td>
<td>N= 163</td>
<td>3772</td>
<td></td>
<td>0.7</td>
<td>0.5</td>
</tr>
<tr>
<td>bezwaar tegen</td>
<td>N= 380</td>
<td>35.9</td>
<td>gesprek met</td>
<td>N= 82</td>
<td>2067</td>
</tr>
<tr>
<td>hebben (have)</td>
<td>32.5</td>
<td>36.6</td>
<td>hebben (have)</td>
<td>39.0</td>
<td>34.3</td>
</tr>
<tr>
<td>maken (make)</td>
<td>12.3</td>
<td>13.0</td>
<td>voeren (be engaged in)</td>
<td>20.7</td>
<td>12.8</td>
</tr>
<tr>
<td>aantekenen (register)</td>
<td>8.6</td>
<td>1.1</td>
<td>zijn (be)</td>
<td>6.1</td>
<td>7.4</td>
</tr>
<tr>
<td>bestaan (exist)</td>
<td>4.9</td>
<td>8.0</td>
<td>worden (become)</td>
<td>3.7</td>
<td>0.6</td>
</tr>
<tr>
<td>zijn (be)</td>
<td>1.2</td>
<td>0.1</td>
<td>verlopen (develop)</td>
<td>3.7</td>
<td>0.5</td>
</tr>
<tr>
<td>aangaan (engage in)</td>
<td></td>
<td></td>
<td>aangaan (engage in)</td>
<td>3.7</td>
<td>4.0</td>
</tr>
<tr>
<td></td>
<td>N= 304</td>
<td>3645</td>
<td></td>
<td>2.4</td>
<td>0.8</td>
</tr>
<tr>
<td>kontakt met</td>
<td>N= 56.2</td>
<td>29.3</td>
<td>kontakt met</td>
<td>N= 63</td>
<td>3492</td>
</tr>
<tr>
<td>hebben (have)</td>
<td>9.2</td>
<td>7.1</td>
<td>informatie over</td>
<td>30.2</td>
<td>16.8</td>
</tr>
<tr>
<td>zijn (be)</td>
<td>5.6</td>
<td>3.8</td>
<td>geven (give)</td>
<td>17.5</td>
<td>4.8</td>
</tr>
<tr>
<td>houden (keep)</td>
<td>4.9</td>
<td>13.2</td>
<td>verstreken (provide)</td>
<td>4.8</td>
<td>9.7</td>
</tr>
<tr>
<td>zoeken (search)</td>
<td>3.9</td>
<td>4.3</td>
<td>hebben (have)</td>
<td>7.9</td>
<td>4.4</td>
</tr>
<tr>
<td>onderhouden (maintain)</td>
<td>3.6</td>
<td>5.0</td>
<td>zijn (be)</td>
<td>7.9</td>
<td>5.6</td>
</tr>
<tr>
<td>leggen (lay)</td>
<td>1.6</td>
<td>2.3</td>
<td>krijgen (get)</td>
<td>4.8</td>
<td>9.7</td>
</tr>
<tr>
<td>opnemen (take up)</td>
<td>3.6</td>
<td>14.2</td>
<td>vinden (find)</td>
<td>3.2</td>
<td>3.7</td>
</tr>
<tr>
<td>maken (make)</td>
<td>2.6</td>
<td>2.9</td>
<td>verschaffen (provide)</td>
<td>3.2</td>
<td>3.3</td>
</tr>
<tr>
<td>krijgen (get)</td>
<td>1.0</td>
<td>0.8</td>
<td>ontbreken (lack)</td>
<td>3.2</td>
<td>0.5</td>
</tr>
<tr>
<td>verliezen (lose)</td>
<td>1.0</td>
<td>3.2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>herstellen (reestablish)</td>
<td>1.0</td>
<td>0.8</td>
<td>zijn (be)</td>
<td>98.1</td>
<td>98.6</td>
</tr>
<tr>
<td>verloren (decrease)</td>
<td>0.7</td>
<td>0.7</td>
<td>lijken (seem)</td>
<td>1.3</td>
<td>0.9</td>
</tr>
<tr>
<td>komen (come)</td>
<td>0.7</td>
<td>0.4</td>
<td>blijven (turn out to be)</td>
<td>0.6</td>
<td>0.4</td>
</tr>
</tbody>
</table>

Table 2: Distribution of verbs for several N+p collocations. Differences marked with • (○) are significant according to the chi-square test at p=0.05 (p=0.10).
<table>
<thead>
<tr>
<th>Behoefte aan N=</th>
<th>Onstaan (come up) o 1.4</th>
<th>Groei (grow) o 1.3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Belangstelling voor N=</td>
<td>Leiden (lead) o 1.8</td>
<td>Willen (want) o 1.8</td>
</tr>
<tr>
<td>Wekken (wake) o 1.1</td>
<td>Gelasten (demand) o 1.5</td>
<td>Eisen (demand) 1.2</td>
</tr>
<tr>
<td>Discussie over N=</td>
<td>Aankondigen (announce) 1.1</td>
<td>Verschil tussen N=</td>
</tr>
<tr>
<td>Beginnen (start) 2.7</td>
<td>Twijfel over N= 1714</td>
<td>Bedragen (amount to) o 2.7</td>
</tr>
<tr>
<td>Brengen (bring) 1.9</td>
<td>Uiten (utter) 2.5</td>
<td>Woorden (become) 2.0</td>
</tr>
<tr>
<td>Aanzwengelen (start up) 1.8</td>
<td>Uitspreken (pronounce) 2.3</td>
<td>Weten (know) 1.8</td>
</tr>
<tr>
<td>Volgen (follow) 1.0</td>
<td>Wegnemen (take away) 1.9</td>
<td>Kennen (know) 1.5</td>
</tr>
<tr>
<td>Aangaan (engage in) 1.0</td>
<td>Groeien (grow) 1.6</td>
<td></td>
</tr>
<tr>
<td>Krijgen (get) 1.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gebrek aan N= 2574</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 3: Frequent N-p-v-combinations in the general corpus, absent in P1. Differences marked with • (o) are significant according to the chi-square test at p=0.05 (p=0.10).

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 first investigate whether there is a relationship between the distribution of determiners and PP-fronting. Next, we discuss how the scarcity of genitives and possessive determiners in the P1 corpus might be explained.

### 4.1 Definite and indefinite NPs

In table 4, a comparison of the frequencies in P1 and general is made of the most common determiners preceding the relevant noun.\(^7\) Table 4 suggests that the indefinite determiners geen, veel and weinig occur relatively frequently in P1, whereas the definite determiner de/het is relatively infrequent in P1.

We believe that the difference in distribution of determiners in P1 and the

---

7\(^{\text{The idiomatic sprake van (talk of) was not included, as it is far more frequent than the other combinations, and would distort the results too much. However, in the general corpus we find that 80% of the examples consists of NULL sprake van whereas 20% is geen (no) sprake van. In P1 sentences this is 70% for geen sprake van and 30% for NULL sprake van.}}\)
Table 4: Frequency of determiners preceding the relevant noun in P1 and the general corpus.

<table>
<thead>
<tr>
<th>determiner</th>
<th>P1</th>
<th>Gen</th>
<th>determiner</th>
<th>P1</th>
<th>Gen</th>
</tr>
</thead>
<tbody>
<tr>
<td>geen (no)</td>
<td>30.7</td>
<td>8.0</td>
<td>weinig (few/little)</td>
<td>3.8</td>
<td>0.7</td>
</tr>
<tr>
<td>NULL</td>
<td>27.7</td>
<td>31.8</td>
<td>enkele (some)</td>
<td>2.1</td>
<td>0.8</td>
</tr>
<tr>
<td>een (a)</td>
<td>14.4</td>
<td>16.5</td>
<td>meer (more)</td>
<td>0.8</td>
<td>1.0</td>
</tr>
<tr>
<td>veel (many/much)</td>
<td>7.7</td>
<td>2.1</td>
<td>minder(less)</td>
<td>0.6</td>
<td>0.2</td>
</tr>
<tr>
<td>de/het (the)</td>
<td>7.3</td>
<td>32.9</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The general corpus can be explained to a large extent by the fact that the verbs in P1 and general have a very different distribution (as shown in the previous section). If we restrict our attention to N-P-V combinations that contain a verb which is relatively frequent in P1, we see that the definite determiner is much less frequent in the general corpus as well. This is illustrated in table 5.

The combination verhaal vertellen over is one of the few examples where the definite determiner is relatively frequent in the general corpus. In this case, the P1-data show an even stronger preference for the definite determiner: 25 out of the 42 cases of PP-fronting with verhaal over contain a definite NP.

The conclusion to be drawn from these data is that the preference for indefinite determiners in the P1 data correlates strongly with the preference for indefinite determiners in the general corpus, if one restricts attention to those verbs which are frequent in P1. 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.

4.2 Possessive pronouns and genitive NPs

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 P1, while possessives are scarce, and restricted to the N+P combinations verhaal over en twijfel over:

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

     b. Over adverteren in de verzorgingsfeer 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 P1.

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 the general corpus 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 P1. The absence of NPs introduced by a genitive and the restricted possibilities for using possessive pronouns can therefore also be attributed to properties of the N+P+V combination as a whole.

5 PPs in relative clauses

It has been argued that pronominalization provides an argument for the V+PP analysis of PP-fronting. If a PP is a dependent of the noun, pronominalization of that noun requires the PP to disappear as well:

(6) a. Hij heeft artikelen tegen die stelling gelezen
   He has articles against that thesis read
He has read articles against that thesis

b. *Hij heeft ze tegen die stelling gelezen
   He has them against that thesis read

If the PP can remain without giving rise to ungrammaticality, this shows that the PP can also be interpreted as a dependent of the verb

(7)  a. Hij heeft alle boeken van Vestdijk gelezen
    He has all books by Vestdijk read
    *He has read all books by Vestdijk
    b. Hij heeft ze van Vestdijk (allemaal) gelezen
    He has them by Vestdijk all read
    *He has read them (all) by Vestdijk

As this seems to be the case for nouns that allow PP-fronting, it suggests that the PP in those cases is actually a dependent of the verb. A problem with this argument, from our point of view, is that pronominalization of the nouns investigated here is scarce, and hard to locate reliably, even in a syntactically analyzed corpus.

However, there is a related construction, involving relative clauses, which also provides evidence that the PP can be interpreted as a dependent of the verb. In relative clauses modifying the noun, the PP is sometimes clearly embedded in the relative clause:

(8)  a. Een bezwaar dat je tegen deze boeken zou kunnen aanvoeren
    An objection that one against these books should can raise
    *an objection that one might raise against these books
    b. de belangstelling die Eduard voor het nazisme toonde
    the interest which Eduard for the nazism showed
    *the interest which Eduard showed for Nazism

For PPs which are unambiguously part of the NP (and which cannot be fronted) this is not possible:

(9)  a. *een demonstratie, die tegen de hoge werkdruk in chaos
    a demonstration which against the high work-load turned into chaos
    turned
   against the high work-load turned a demonstration into chaos

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 te N+P combinations we investigated, several examples can be found where the PP is included in a relative clause. Some more examples are given below.
For a few combinations (twijfel over, vraag naar, verhaal over), only exhaustive search in the Twente News Corpus provided us with some examples. For sprake van and gebrek aan, we did not find examples where the PP is unambiguously part of a relative clause.

The fact that most PPs which can be fronted also may occur within a relative clause 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.

6 Infrequent and Missing Patterns

As argument for the v+PP analysis, it has been suggested that there are word orders which seem incompatible with the idea that the PP is a dependent of a noun. In this section, we observe that these patterns are practically absent in even a large corpus of Dutch. However, the word order PP+NP is relatively frequent in the Dutch Mittelfeld, and seems problematic for an N+PP analysis.

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

(11) 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 in the general corpus. The most convincing cases are given below:

(12) a. Verder contact is er met Den Haag niet geweest
Further contact is there with The Hague not been
There has not been further contact with The Hague
b. Dat gesprek zou vandaag ook met de ouders van de
That conversation should today also with the parents of the
Today, that conversation with the parents of the potentially abused children shall be conducted.

c. Meer informatie kunt u er over krijgen bij een notaris

You can get more information about this at the notary.

A few other cases involving an NP containing the relevant noun in first position and a PP in the Mittelfeld, had to be discarded as false positives, as they most likely involved a verb selecting for a PP-complement (i.e. gaan over (is about)). Examples such as (13), which are more frequent, have to be discarded as well, as they may be the result of extraposing a PP-complement of NP in initial position.

(13) Hoeveel behoeftet er in de toekomst nog aan de diverse joodse zorg-, onderwijs- en welzijnsorganisaties, ....?

How much need is there in the future for the various Jewish care, education, and welfare organizations,....?

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, one of which is given below:

(14) Gek genoeg bestond er voor zijn oorlogsfoto’s tot voor kort weinig belangstelling.

Curiously, until quite recently there was little interest in his war photography.

On the other hand, PP-NP orders, as in (15), 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):

(15) De Marokkanen hebben aan dergelijke groepsvering geen interesse.

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.
7 Related Work

As was noted in the introduction, the Corpus of Spoken Dutch (CGN) frequently annotates nouns as taking a prepositional complement. We collected all N+P combinations which are annotated at least three times in this way in CGN. The resulting list of 35 N+P combinations contained 8 combinations which were also selected as collocational N+P combinations in section 2. For the other combinations (e.g. vertrouwen in (confidence in), problemen met (problems with), inzicht in (insight in), waardering voor (appreciation for), etc.) we checked in the Twente News Corpus whether they allow PP-fronting. This was true for almost all N+P combinations. For a few combinations (verschil met/in (difference with/in), kans op (chance on), deelname aan (participation in), parodie op (mockery of), interpretatie van (interpretation of)) no examples of PP-fronting could be found. This suggests that the majority of the N+P combinations annotated as involving a prepositional complement in CGN are comparable with the expressions we investigated, and thus are probably best analyzed as V+PP in cases where the PP has been fronted. For the remaining cases, the CGN-annotation (i.e. N+PP) is most plausible.

Our corpus based analysis suggests that PP-fronting is best analyzed as involving a PP which is a dependent of the verb. 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. This analysis is compatible with observations in the ANS (Haesereyn et al., 1997), Broekhuis (2003) and Loonen (2003). These works all provide (non-exhaustive) lists of phrasal verbs involving a PP-complement. A considerable number of V+N+P combinations which we found in the P1 corpus are mentioned by these authors as well. Also, a large number of the N+P combinations annotated as involving a noun with a prepositional complement in CGN, is listed with one or more verbs as a phrasal verb in one of these sources.

8 Conclusion

In this paper, we have reported on research in which a number of claims with respect to PP-fronting in Dutch were checked against corpus data. The results can be summarized as follows:

1. Certain verbs are far more frequent in P1 sentences than in general sentences containing the relevant N+P combination.

2. There is a difference in the distribution of determiners in P1 and in the general corpus. The difference seems to be mainly a consequence of the difference in distribution of verbs in both corpora.

3. PPs that occur sentence initially may also appear within relative clauses.

4. A number of patterns which have been used as argument for a specific analysis of PP-fronting are hardly encountered in a large corpus.
Conclusions 1-3 suggests that the v+PP analysis is most plausible. This confirms the conclusion in 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


