Abstract
This paper presents arguments from a corpus study that confirm the claim by Jack Hoeksema that extraction from PPs and subjects is possible in Dutch. Hoeksema presents data involving relative clauses, but we show that similar cases can be found in WH-questions as well. The suggestion that these are restricted to PP-arguments of a noun is problematic, however, as there are no clear tests for distinguishing PP-arguments of a noun from PP-adjuncts.

Keywords: Dutch, syntax, corpus, treebanks, PP-fronting

1. Introduction
Is extraction from inside an NP possible in Dutch? Examples such as (1a) suggest that the answer is yes, as the PP can be interpreted as forming a constituent with the NP (1b).

(1) a. Waarvan is Anne Frank het symbool?
what-of is Anne Frank the symbol?
‘What is Anne Frank the symbol of?’

1 Center for Language and Cognition Groningen. Rijksuniversiteit Groningen
Corresponding author: Gosse Bouma, g.bouma@rug.nl
ISSN 0165-9200. Published by University of Groningen Press, Copyright © by author(s)
How to cite this article: Bouma, G. & Noord, G. van (2024). Van wie heeft u een foto op uw boksbal hangen? Extraction from NPs in Dutch. TABU Festschrift for Jack Hoeksema. 219-234. https://doi.org/10.21827/tabu.2023.41260
This article is licensed under Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License (CC BY-NC-ND)
b. Dit symbool van onderdrukking hoort niet thuis in het openbare leven
   this symbol of oppression belongs no place in the public life
   ‘This symbol of oppression does not belong in public life’

A further question is whether extraction from NP is possible if that NP functions as the subject (contrary to what is suggested by the Specified Subject Condition of Chomsky (1973)). Again, corpus data such as (2a) suggests that this is the case for PPs that can be seen as being part of the NP (2b).

(2) a. Ach, van welk volk zijn de mannen wel trouw?
   Well, of which nation are the men indeed faithful?
   ‘Well, the men of which nation are faithful?’

b. De 10 mannen van Aziatische afkomst zitten vast op het bureau
   The 10 men of Asian origin are locked at the precinct
   ‘The 10 men of Asian origin are held in detention’

Hoeksema (2022) presents results from the Lassy Large corpus that suggest that extraction from NP is possible (including cases where that the NP is the subject). In these cases, the noun clearly selects for a PP. The main finding is that there are several classes of nouns (derived from verbs with a prepositional complement and relational nouns like father) that select for a PP dependent that can be fronted in relative clauses.

Yet, following Bach and Horn (1976), some researchers have argued that extraction from an NP is not possible in Dutch (Klein and van den Toorn 1977, Klein and van den Toorn 1979, Coppen 1991, Bouma 2004) and that examples such as (1a) and (2a) are better analyzed as involving a PP that is a dependent of the verb. Support for this second analysis comes from the fact that (a) there are cases where a PP that is a dependent of a noun cannot be fronted, (b) in some cases PPs that may be seen as dependent of a noun can also occur in the Mittelfeld, separated from the noun, suggesting that they are in fact not a dependent of that noun, and (c) the acceptability of some examples depends on the governing verb which can be seen as an argument for an analysis in which that verb selects for the PP.

Here, we address two aspects of the analysis of Hoeksema (2022): (a) the corpus data was restricted to relative clauses, and we investigate to what extent similar data can be observed in WH-questions, and (b) we investigate the claim by Hoeksema that the unacceptability of fronting certain PP dependents is explained by the fact that fronting is restricted to arguments
of the noun. We searched the Lassy Large corpus for WHquestions involving a fronted PP dependent of a noun, and we evaluated whether these positive cases can all be seen as arguments of a noun. We found that the observations of Hoeksema for relative clauses can be extended to WH-questions, i.e. we found examples where the fronted PP is best analyzed as being part of a NP, including cases where this NP can be a subject. However, Hoeksema’s claim that such cases are restricted to argument PPs is problematic, as the few tests that have been proposed to distinguish between arguments and adjuncts of a noun are inconclusive for such cases, as we will argue below.

2. Hoeksema 2022

Hoeksema (2022) selects 500 occurrences of the pronominal adverb *waarvan* from the Lassy Large corpus. Of these, 494 are relative clauses headed by *waarvan*, and within this group, 314 are cases of adnominal PPs, i.e. cases where *waarvan* is analyzed as a fronted dependent of a relational noun, as in (3). Relational nouns are nouns derived from a transitive verb (*uitwerking*, ’implementation’), inherently relational nouns (*eigenaar*, ’owner’), and part-of nouns (*helft van*, ’half of’).

(3) ... restaurant Sardegna, waarvan hij eigenaar was

... restaurant Sardegna, of-which he owner was

‘... restaurant Sardegna, which he was the owner of’

The grammatical function of the NP that *waarvan* is part of, is that of subject (217 cases, eg. (4)), predicate (44 cases) or direct object (53 cases).

(4) De Nationale Haringtest, waarvan de resultaten via internet zijn te raadplegen

the national herring-test, of-which the results on internet are to inspect

‘the national herring-test, results of which can be seen on the internet’

Contrary to Bouma (2004), who investigated a small set of collocational noun-preposition cases (*behoefte voelen aan*, ’feel the need for’, *belangstelling hebben voor*, ’have an interest in’), Hoeksema also finds that there is a wide range of governing verbs in the data. This indicates that the analysis of Bouma for such collocational noun-preposition cases is not available. Bouma analysed cases such as:
(5) Ik heb in die tijd echt behoefte gevoeld aan een bisschoppelijke figuur

I have in that period really need felt for an episcopal figure

‘I really felt the need for an episcopal figure at that time’

as involving a verb (gevoeld, ‘felt’) which selects for both an NP involving behoefte (‘need’) and a PP headed by aan (‘for’). In case the PP is extracted, that PP, therefore, is not extracted out of an NP. This analysis is convincing only for cases in which the verb is collocational with the preposition. However, Hoeksema finds many examples where this collocational analysis is problematic. A good example is (6):

(6) ... redeneringen waarvan je de kwaadaardigheid minder snel herkent
... argumentations of-which you the evilness less fast recognize

‘argumentation which you will less easily recognize as evil’

The conclusion is that fronting of PPs that are a dependent of a noun, is clearly possible in Dutch. Furthermore, there is no evidence that this is not possible where the noun is heading a subject NP. The fact that the nouns involved are all relational in nature, supports the claim that fronting is restricted to cases where the PP is an argument of the noun. Only a very small part of the data (6 of 500) are WH-questions. This could be due to the fact that questions in general are far less frequent in the corpus than relative clauses, but it also raises the question whether fronting a PP is equally wide-spread in questions. We investigate this in the next section.

3. WH-questions with extraction from NP

We selected examples of WH-questions with potential extraction from NP, by searching in the Newspaper part (WR-P-P-G) of the Lassy Large corpus. This part of the corpus contains about 15 million sentences. As in the investigation by Hoeksema, we focused on prepositional phrases headed by van. For reasons of reproducibility, we provide the actual XPath query:

//node[@rel="whd" and (@lemma="waarvan" or (@cat="pp" and node[@rel="hd" and @lemma="van")))) and
It should be noted that the corpus is automatically annotated with syntactic constituency and dependency information with the Alpino parser (van Noord 2006). Obviously, the automatic analysis sometimes is wrong, but in this specific case a further problem arises because the grammar of Alpino will not analyse fronted constituents as part of an NP, but rather the fronted constituent PP is annotated as a dependent of the verb. As a result, the true governor of the fronted PP cannot be identified reliably, and all hits of the search pattern were validated and classified by hand (by both authors, and disagreements between the two annotations were resolved after discussion). The result is listed in Table 1.

<table>
<thead>
<tr>
<th>extraction from NP</th>
<th>45</th>
<th>subject</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>other extraction</td>
<td>232</td>
<td>predicative</td>
<td>25</td>
</tr>
<tr>
<td>wrong parse</td>
<td>261</td>
<td>direct object</td>
<td>10</td>
</tr>
<tr>
<td>total hits</td>
<td>538</td>
<td>total</td>
<td>45</td>
</tr>
</tbody>
</table>

Table 1: Fronted van-PP cases in questions, in the newspaper part of Lassy Large. The left table indicates for the query whether the hit involved a potential extraction from NP, or extraction from another constituent, or whether the hit constituted a wrong parse. The table on the right breaks down the potential extraction from NP cases by indicating the grammatical role of the relevant NP.

In 15 million sentences, we only found a little over 500 hits, but about half of these hits were - upon inspection - wrong (in most cases, this involved relative clauses). In only 45 sentences, the fronted PP could potentially be analysed as a dependent of a noun (subject, object or predicative complement). These 45 cases are broken down further by taking into account the grammatical relation of the NP. In 10 cases the NP is a direct object, and in another 10 cases the NP is the subject (4 of these involve the subject of a passive construction). The majority of cases involves predicative complements, i.e., examples such as (8b). It appears, therefore, that the grammatical role of the NP is not a decisive factor for the acceptability of extraction of PP out of NP.

The positive cases involve combinations of a verb with a direct object such as *prijzen loslaten* (‘to release prices’), *vader vermoorden* (‘to murder father’), *foto hangen* (‘to hang up
a photo’), as well as collocational expressions such as *spijt hebben* (’to have regret’), *gezelschap krijgen* (’to get company’). Thus, the data shows the same variation in verb choice that was also found by Hoeksema for relatives. Furthermore, the cases where a noun heads a subject NP show that in WH-questions fronting of PPs that are part of the subject is possible.

Some of the examples which appear to be true cases of extraction out of NP are the following, including examples where the relevant NP functions as the subject:

(7) a. De beroepsgroep mag zelf aangeven van welke behandelingen de prijzen worden losgelaten

The professional-group may itself indicate of which treatments the prices will be released

‘The professional group may indicate itself for which treatments the prices will be released’

b. Van wie heeft u een foto op uw boksbal hangen?

Of whom have you a picture on your punching-bag hang?

‘Who do you have a picture of on your punching bag?’

c. Ik weet niet van welke kerk de leer bedoeld wordt

I do not know from which church the doctrine is meant

‘I do not know from which church the doctrine is meant’

d. Van welk staatshoofd verschenen vertrouwelijke brieven op een uitwisselingsdienst op het internet?

Of which head-of-state appeared confidential letters on an exchange-service on the Internet?

‘From which head of state did confidential letters appear on an Internet exchange service?’

Some typical examples which potentially could be analyzed as collocational expressions (as in Bouma (2004)) are the following:
In conclusion, fronting of a PP that is a dependent of a noun in contexts other than relative clauses is not very frequent in the corpus, but the examples above indicate that positive examples indeed occur, and that these are not restricted to collocational expressions.

4. Do nouns select for PP-arguments

Following Broekhuis and Keizer (2015), Hoeksema assumes that only argument PPs can be fronted. To verify this claim, one needs independent criteria to distinguish between argument and adjunct prepositional dependents of a noun. For verbs, following Tesnière (1959), it is widely assumed that such a distinction can be made, although it should be noted that some have argued against it (Przepiórkowski 2016b, Przepiórkowski 2016a, Przepiórkowski and Patejuk 2018), and an annotation framework such as Universal Dependencies (De Marneffe et al. 2021), for instance, marks dependents of verbal and nominal heads as $OBL^{19}$ and $NMOD^{20}$, but does not
adopt a distinction between arguments and adjuncts. It remains to be seen, therefore, whether there are independent criteria for deciding that a prepositional dependent of a noun is an argument or adjunct.

The Algemene Nederlandse Spraakkunst (ANS) (Haesereyn et al. 1997) (section 14.6.2, Voorzetselconstituuteten\(^{21}\)) presents cases such as (9a) and (9b), involving either a noun derived from a verb that has a clear PP-complement or a noun that combines with a PP headed by a specific preposition, where one could consider the PP to be a argument of the noun.

(9) a. de schreeuw om hulp
   ‘the shout for help’

b. de toegang tot de vallei
   ‘the entrance to the valley’

However, the ANS also states that no clear distinction between adjuncts and arguments of nouns exists, and that PP-arguments of a noun have the same word order distribution as PP-adjuncts, although they observe that arguments tend to precede appositions and relative clauses modifying the noun.\(^{22}\)

The discussion in Broekhuis and Keizer (2015, section 2.2.1) is more detailed, and provides four tests for distinguishing prepositional arguments of nouns from adjuncts: (a) obligatoriness, (b) occurrence in postcopular predicative position, (c) er-pronominalization, and (d) fronting. Criterion (a) has many exceptions, and (d) is the subject of this study, which leaves tests (b) and (c). It should also be noted that Broekhuis and Keizer observe that fronting is not a good test for argument-status, as they say about test (d):

...this test provides us with the least clear results, which furthermore often conflict with those of the three tests discussed earlier. We therefore tend to dismiss this test as a good test for determining complement/adjunct status of PPs within the noun phrase.

The reasons for being sceptical about this test is that not all argument PPs can be fronted (although results can improve given the right context or focus), and vice versa, while fronting of adjuncts seems impossible in general, exceptions can be found with the right context. Nevertheless, it would be informative if independent tests for the argument status of a PP exists, as it would give us a test to see whether the corpus data confirms that fronted PPs are always arguments.
Test (b) of Broekhuis and Keizer (2015) states that in copular sentences with the NP as subject, only *van*-adjuncts, but not *van*-arguments may occur in postcopular, predicative, position:

\[(10) \quad \begin{align*}
a. & \quad \text{Het huis is van Jan} \\
& \quad \text{‘The house is of John’}
\end{align*} \]
\[(12) \quad \begin{align*}
a. & \quad \text{De soundtrack is van ‘The Jewel of the Nile’} \\
& \quad \text{‘The soundtrack is of ‘The Jewel of the Nile’}
\end{align*} \]

Even if we accept these acceptability judgements, it appears that for some of the examples in our dataset, paraphrases involving the PP in postcopular position are not fully ruled out. The examples in (11) are instances of fronted PPs as found in the corpus, while (12) are paraphrases with the relevant noun as subject and the *van*-PP as predicate.

\[(11) \quad \begin{align*}
a. & \quad \text{Van welke film was ‘When the Going Gets Tough’ van Billy Ocean de soundtrack?} \\
& \quad \text{‘Of which movie was ‘When the Going Gets Tough’ by Billy Ocean the soundtrack?’}\,^{23}
\end{align*} \]
\[(12) \quad \begin{align*}
a. & \quad \text{Van welke.Inject the missing word here to complete the example.}
\end{align*} \]
Cases with the noun *foto* (‘picture’) occur relatively frequently in the corpus, with the *van-*PP referring to the photographer, but there are also cases where the PP refers to the visual referent of the picture:

(13) a. de allereerste erotische foto was van een tepeltje in een doorkijkblouse
    ‘the very first erotic picture was of a nipple in a see-through blouse’

28

b. De foto is van de Mazda Axela this picture is of the Mazda Axela
    ‘This picture is of the Mazda Axela’

29

c. De oudste foto is van een vrouw die om het leven kwam
    ‘The oldest picture is of a woman who died’

30

d. De winnende foto is van het voetpad in de De Saedeleerstraat
    ‘the winning picture is of the footpath in the De Saedeleerstreet’
‘The winning photo is of the footpath in the De Saedeleerstraat’\footnote{31}

Thus, at least for so-called picture nouns the test is inconclusive. That is, the examples with a van-PP (where the PP refers to the visual referent of the picture) follow the copula in examples such as (13b), which indicates that these PPs must be adjuncts, not arguments. But then examples such as (11c), where the PP is fronted indicate that the PP is an argument, not an adjunct (if fronting is restricted to arguments, as claimed by Hoeksema).

The pronominalization test (c) is based on the observation that argument PPs can be realized by pronominal adverbs consisting of an R-pronoun and a preposition (i.e. ervan (‘of it’), ermee (‘with it’), eraan (‘on it’), etc.) but that this is impossible for adjunct PPs.

(14) a. Ik heb de verwoesting van de stad/ervan meegemaakt
I have the destruction of the city/of it witnessed

‘I have witnessed the destruction of the city/of it’

b. een laken van satijn/ervan

‘a sheet of satin/of it’

We searched the newspaper section of the Lassy Large corpus for occurrences of NPS consisting of a head noun followed by ervan. This gave over 8000 hits, with a large diversity of nouns. The most frequent cases are gebruik (‘use’), deel (‘part’), gevolg (‘consequence’), uitvoering (‘implementation’), inhoud (‘content’), waarde (‘value’), effect (‘effect’), kwaliteit (‘quality’), karakter (‘character’), belang (‘importance’). The list contains many cases of nouns derivedtitoepassing (‘application’), bestaan (‘existence’), invoering (‘introduction’), werking (‘functioning’), maker (‘creator’), aanleg (‘construction’), ...). This confirms the observation by Broekhuis and Keizer (2015) that pronominal adverbs only occur with argument PPs. Note though that the opposite statement is not always true since er-pronominalization is only possible with inanimate referents. So from the absence of cases such as vader ervan (‘father of it’), we cannot conclude that vader (‘father’) does not take a van-PP complement.

There is a further construction which appears to involve extraction of part of a prepositional phrase out of a noun phrase. In this construction, an R-pronoun is extracted from a prepositional phrase, possibly out of a noun phrase:
Note that the fronted R-pronoun can be topicalized, as in (15), but it can also end up in a different location to the left of the NP:

(15)  a.  Daar zijn veel voorbeelden van

There are many examples of
‘there are many examples of this’

b.  Hier kan ik veel voorbeelden van verzinnen

Here can I many examples of invent
‘I can think of many examples of this’

This construction appears to be much more frequent than the WH-extraction cases which we investigated in section 3. However, Broekhuis and Keizer (2015) claim that such examples do not involve extraction of the er-word from the complement of the noun but from an independent adverbial phrase. This analysis is based on the observation that this option is unavailable for PP-complements headed by other prepositions (this example is ungrammatical under the intended reading):

(17)  a.  *Ik heb er een uitbreiding mee kunnen tegenhouden.

I have there an extension with can stop.

‘I have been able to stop an extension with it’

The conclusion from this discussion is that explicit criteria for distinguishing between argument and adjunct PP-dependents of a noun are not readily available, and that the tests that are provided in Broekhuis and Keizer (2015) are not without exception either. Furthermore, these authors also observe that the generalisation that only argument PPs can be fronted, is
problematic if one considers a wider range of constructions and/or situations where context influences grammaticality judgements. The latter is in line with the observations in Abeillé et al. (2020), who observe that acceptability judgements of similar fronting constructions in English and French are gradient and are strongly influenced by discourse constraints.

5. Conclusion

In this paper, we found corpus evidence in Lassy Large for WH-questions involving PPs that are dependents of a noun. In line with the observations in Hoeksema (2022) for relative clauses, we find that such cases occur, although rarely, and that these also occur where the NP is functions as a subject. The explanation for when this type of fronting is possible remains problematic however, as the hypothesis by Hoeksema that this is restricted to argument PPs is hard to verify, given the lack of robust tests for distinguishing between PP-arguments and PP-adjuncts of a noun in the first place.

Endnotes

1 Examples are from the newspaper part (15 million sentences) of the Lassy Large corpus, https://taalmaterialen.ivdnt.org/download/tstc-lassy-groot-corpus/ (van Noord et al. 2013). In some cases, examples are simplified but we provide the Lassy Large identifiers of the relevant full sentences.

2 WR-P-P-G-0000092467.p.10.s.3
3 WR-P-P-G-0000040034.p.11.s.5
4 WR-P-P-G-0000158073.p.15.s.1
5 WR-P-P-G-0000026529.p.3.s.4
6 WR-P-P-G-0000384439.p.2.s.2
7 WR-P-P-G-000018987.p.1.s.1
8 WR-P-P-G-0000116565.p.9.s.4
9 WR-P-P-G-0000090296.p.3.s.4
10 In the newspaper part of the Lassy Large corpus there are about five times as many relative clauses than questions.
11 WR-P-P-G-0000155917.p.1.s.2
12 WR-P-P-G-0000606241.p.31.s.1
13 WR-P-P-G-000001746.p.1.s.3
Their example (6) (een boek over Diana, de ongelukkige vrouw van prins Charles, a book about Diana, the unhappy wife of prince Charles) illustrating the position of appositions is incorrect, however, as this is an example where the apposition is a dependent of the NP inside the PP.

References


van Noord, Gertjan (2006), At last parsing is now operational, in Mertens, Piet, Cedrick Fairon, Anne Dister, and Patrick Watrin, editors, *TALN06. Verbum Ex Machina. Actes de la 13e conference sur le traitement automatique des langues naturelles*, pp. 20–42.