Starting a sentence in Dutch
Bouma, G.J.

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2008

Link to publication in University of Groningen/UMCG research database

Citation for published version (APA):

Copyright
Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

Take-down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): http://www.rug.nl/research/portal. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.
Chapter 2

Preverbal Behaviour

This dissertation deals with word order variation in the left-peripheral, directly preverbal domain. The position that is investigated is known as the Vorfeld (‘prefield’, German) – a term that comes from a traditional model of grammatical description. In this chapter, I will give an overview of word order variation in the preverbal domain, introduce necessary terminology, and delimit the range of investigated constructions. At the end of the chapter, I will present predictions about Vorfeld occupation that will be tested against a corpus in Chapter 4.

I will begin by introducing the descriptive model of the Dutch clause in Section 2.1. The four sections that follow will then give an overview of variation in Vorfeld occupation. In Section 2.2, I will show data in which the Vorfeld is occupied by constituents of different grammatical functions and of different syntactic categories. Section 2.3 elaborates on the special relation between the Vorfeld and pronouns. In general, the Vorfeld contains exactly one constituent, however, in Section 2.4, I will discuss data with more than one preverbal element. The information structural properties of the Vorfeld are discussed in Section 2.5.

After the formal and functional overview of material in the preverbal domain, Section 2.6 discusses existing results on word order variation of Dutch, German and English. On the hypothesis that the word order trends discussed are caused by global trends in word order, I will formulate expectations for Vorfeld variation that mirror these global trends. Other predictions regarding Vorfeld occupation will come from the functional properties of the Vorfeld discussed earlier in the chapter.
2.1 Topological fields

Dutch allows for a fair amount of word order variation throughout the sentence, provided a couple of core positions are taken by certain elements. Consider the twelve variations on a theme in (1). Each example is a declarative main clause expressing the same propositional content: Tom Boonen (subject) would have beaten Jens Voigt (object) if the two cyclists would have sprinted against each other.\footnote{This interpretation is available for (1i)–(1l), with the direct object Voigt in initial position. Native speakers may find they need to pronounce Voigt with stress, and deaccent Boonen to get the reading. Also, it might be worth pointing out that although Jens Voigt is an excellent cyclist, he is no match for specialist Tom Boonen when it comes to sprinting.}

(1) a. Boonen zou Voigt in de sprint geklopt hebben.
   Boonen would Voigt in the sprint beaten have
   ‘Boonen would have beaten Voigt in the sprint.’

b. Boonen zou Voigt in de sprint hebben geklopt.

c. Boonen zou in de sprint Voigt geklopt hebben.

d. Boonen zou in de sprint Voigt hebben geklopt.

e. In de sprint zou Boonen Voigt geklopt hebben.

f. In de sprint zou Boonen Voigt hebben geklopt.

g. Boonen zou Voigt geklopt hebben in de sprint.

h. Boonen zou Voigt hebben geklopt in de sprint.

i. Voigt zou Boonen in de sprint geklopt hebben.

j. Voigt zou Boonen in de sprint hebben geklopt.

k. Voigt zou Boonen geklopt hebben in de sprint.

l. Voigt zou Boonen hebben geklopt in de sprint.

Across these twelve sentences, there is variation in the place and order of the arguments, the position of the PP, and the order of the non-finite verbs. The constant skeleton in all of these sentences is formed by the finite verb zou in second position, and the non-finite verbs hebben & geklopt, which are clustered towards the end of the sentence. Dutch can be considered to be a verb-second, verb-final language. The other constituents in the sentences in (1) occur in the three ‘fields’ that are to the far left of, in between, and to the far right of these verbal positions.

It is common to describe the clauses and clause types for German and Dutch in terms of topological fields, characterizations of which can already be found in 19th century German grammars (Höhle, ms). Figure 2.1 shows topological fields for the description of Dutch used in the reference grammar Haeseryn et al. (1997).

The template in Figure 2.1 gives the names of the topological fields, as well as some typical inhabitants. For main declarative clauses like (1) – ignoring the fields lead and tail for a moment – the Vorfeld spans from the left edge of the clause until the left bracket,
and is occupied by exactly one constituent.\(^2\) The *left bracket* contains the finite verb in main clauses. The *Mittefeld* is found between the two brackets and may contain any amount of material. In the *right bracket*, or verbal cluster, the remaining verbs in the clause are found and possibly a small amount of non-verbal material. Dutch allows for some variation in the ordering of verbal material in the verb cluster. Finally, extraposed material, such as CP complements, or PP complements and adjuncts occupy the *Nachfeld*.

The fields *lead* and *tail* are not a common part of clause topology, and may be considered as not belonging to the clause proper. They are reserved for material that is more loosely associated with the clause such as hanging topics, afterthoughts and vocatives. An example sentence with material in the tail is (2). The Nachfeld-tail border is marked with a ‘|’.

(2) Ze had een mooie naam, | mijn gids.
  she had a pretty name my guide
  ‘My guide had a pretty name.’

The topological field template is not only used to describe main declarative clauses, but it is also used to describe subordinate clauses and interrogative clauses. Examples of different clause types in the template can be found in Table 2.1, p22. Clauses generally follow the V2 generalization. However, in the case of a polar interrogative (example e in Table 2.1), the Vorfeld is empty, resulting in a V1 clause. In a subordinate clause (f), all verbs are in the verb cluster, and the complementizer is assumed to be in the left bracket.

Topological field templates for clauses are found in descriptive grammars of Dutch such as (Haeseryn et al., 1997) and German (Kunkel-Razum and Münzberg, 2005). In many theoretical and empirical linguistic studies, topology is used in a purely descriptive manner, although some researchers have given topological fields theoretical status, too (Kathol, 2000).

As a descriptive model, topology is very useful, but it is not perfect. There are cases in which it is not easy to decide how a given sentence should fit into the template. The identification of topological fields relies on V2 to separate the Vorfeld from the rest, and (other) verbs clustering towards the end of the sentence to form a Mittelfeld and a Nachfeld. If there is, say, more than one constituent before the finite verb, it is not clear how to assign these to topological fields: The constituents could be forced into

\(^2\)But see Section 2.4 for exceptions to this generalization, that is, declarative main clauses with more or less than one constituent in the Vorfeld.
Table 2.1: Examples of different constructions in the clause template

<table>
<thead>
<tr>
<th>Vorfeld</th>
<th>left bracket</th>
<th>Mittelfeld</th>
<th>right bracket</th>
<th>Nachfeld</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. subject initial, declarative main clause (V2):</td>
<td></td>
<td>Boonen zou Voigt in de sprint geklopt hebben.</td>
<td>Boonen would Voigt in the sprint beaten have</td>
<td></td>
</tr>
<tr>
<td>b. preposed PP, declarative main clause (V2):</td>
<td>In de sprint zou Boonen Voigt geklopt hebben.</td>
<td>In the sprint would Boonen Voigt beaten have</td>
<td></td>
<td></td>
</tr>
<tr>
<td>c. object initial, declarative main clause (V2):</td>
<td>Voigt zou Boonen geklopt hebben in de sprint.</td>
<td>Voigt would Boonen beaten have in the sprint</td>
<td></td>
<td></td>
</tr>
<tr>
<td>d. constituent question, main clause (V2):</td>
<td>Wie zou Boonen geklopt hebben in de sprint?</td>
<td>Who would Boonen beaten have in the sprint</td>
<td></td>
<td></td>
</tr>
<tr>
<td>e. polar interrogative, main clause (V1):</td>
<td>Zou Boonen Voigt in de sprint geklopt hebben?</td>
<td>Would Boonen Voigt in the sprint beaten have</td>
<td></td>
<td></td>
</tr>
<tr>
<td>f. subordinate clause (V-final):</td>
<td>dat Boonen Voigt geklopt zou hebben in de sprint.</td>
<td>that Boonen Voigt beaten would have in the sprint</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

the Vorfeld (breaking V2), they could be analyzed as one constituent (respecting V2), or one constituent could be assigned to the Vorfeld, and the others to the lead (possibly respecting V2). In this dissertation, I will confine myself to sentences with exactly one preverbal constituent, that is, to one constituent in the Vorfeld, though brief discussion and examples of ‘multiple Vorfeld occupancy’ can be found in Section 2.4.

Another case in which it is unclear how to apply the topological field template is when there is no material in the verbal cluster to separate the Mittelfeld from the Nachfeld. In (3a) the Mittelfeld and Nachfeld are separated by the main verb, and in (3b) a verb particle marks the right bracket, but in (3c) there is nothing – apart from the analogy with (3a) – that tells us where the Mittelfeld ends and the Nachfeld begins. In the examples, field borders are indicated by ‘|’.

(3) a. Ella | had | Fitz | ontmoet | bij Gerald thuis.

   Ella had Fitz met at Gerald’s
Section 2.2: Vorfeld occupants

b. **Ella** kwam Fitz tegen bij Gerald thuis.
   Ella met Fitz **VPART** bij Gerald’s
c. **Ella** ontmoette Fitz bij Gerald thuis.
   Ella met Fitz **VPART** bij Gerald’s

‘Ella met Fitz at Gerald’s.’

Since this dissertation is concerned with the Vorfeld, the distinction between Mittelfeld and Nachfeld is not always relevant. In certain situations, to avoid making false or unfounded claims about the position of the elements, I will therefore use the terms *preverbal* (left of the left bracket) and *postverbal* (right of the left bracket) to refer more generally to the position of material in the clause. In the investigation of the influence of grammatical complexity on Vorfeld occupation, presented in Section 4.4, it will turn out that the distinction between Mittelfeld and Nachfeld is relevant. In that section, I therefore define further fields or positions distinguished by whether we can recognize the Mittelfeld from the Nachfeld or not.

The topological approach to clause description can also be found in the literature on Scandinavian languages, for instance in the reference grammars Teleman, Hellberg, and Andersson (1999, Swedish), and Faarlund, Lie, and Vannebo (1997, Norwegian). These languages are V2 languages like Dutch and German, which means that the Scandinavian Vorfeld is very similar to the Vorfeld in Dutch and German. However, since the Scandinavian languages are not verb final, Mittelfeld and Nachfeld in descriptions of those languages do not correspond well to the Dutch and German fields.

Now that we have the topological field model for Dutch in place, I will concentrate on word order variation that targets the Vorfeld. I will use the topological field template throughout this dissertation as a descriptive device.

2.2 Vorfeld occupants

In this section, I will present data to give a feel for the kind of material, in terms of grammatical function and syntactic category, that can appear in the Vorfeld. I will concentrate on subjects and objects in the Vorfeld, because they are the focus of this dissertation. However, I will briefly mention predicate fronting and Vorfeld adjuncts, too.

2.2.1 Vorfeld subjects

In the course of this chapter, we shall see that the subject is the unmarked Vorfeld occupant. In (4), the subject in the Vorfeld is alternatively a pronoun, a proper name or a definite full NP.
Quantificational and/or indefinite subjects can also appear in the Vorfeld, as shown in (5).

(5) Iedereen / Iemand / Niemand ligt bijna 7 minuten voor Twee renners liggen 7 minuten voor

‘Everybody / Nobody / Someone leads / Two riders lead by 7 minutes.’

Even though indefinite subjects are allowed in the Vorfeld, there is a strong tendency to use an existential construction (EC) when the subject is indefinite. In that case, the Vorfeld is typically occupied by an expletive subject er ‘there’, and the logical subject appears postverbally (6a). Dutch, unlike English and the mainland Scandinavian languages, allows for transitive ECs.

(6) a. *(Er) heeft iemand bijna 7 minuten voorsprong.
   EXPL has someone almost 7 minutes lead
   ‘Someone is leading by almost 7 minutes.’
   b. Gek genoeg had (er) iemand bijna 7 minuten voorsprong.
      funnily enough had EXPL someone almost seven minutes lead
      ‘Funnily enough, someone was leading by almost 7 minutes.’
   c. . . . weil (*es) niemand gearbeitet hat.
      because EXPL nobody worked has
      ‘. . . because nobody was working.’ (German, Meinunger, 2007, example 7a)

The EC in (6b) shows that the use of an expletive subject is not restricted to the Vorfeld in Dutch. This contrasts with German, where expletive es in an EC is restricted to the Vorfeld. On the other hand, Dutch does not allow for universally quantified or definite logical subjects in ECs (7a), which German readily allows (7b).

(7) a. *Er komt de dood.
   EXPL comes the death
   ‘Death is coming.’
   b. Es kommt der Tod.
   EXPL comes the death
   ‘Death is coming.’
   (German)

Clausal subjects are allowed to be in the Vorfeld (8a), although there is a tendency for them to appear extraposed in the Nachfeld (8b). In case of extraposition, an expletive or ‘preliminary’ subject het ‘it’ appears earlier in the sentence.

(8) a. Dat de Harmonie ’s avonds sluit is misschien maar goed.
   that the Harmonie at night closes is maybe PART good
b. Het is misschien maar goed dat de Harmonie 's avonds sluit.
   ‘Maybe it is a good thing that the Harmonie closes at night.’

Clausal subjects in the Mittelfeld are only marginally possible in Dutch (9a). The extra-
posed variant (9b) is strongly preferred.

(9) a. Waarom is dat de Harmonie sluit misschien maar goed?
   why is that the Harmonie closes maybe PART good
   ‘Why would it be a good thing that the Harmonie-building closes?’
   b. Waarom is het misschien maar goed dat de Harmonie sluit?
      why is it maybe PART good that the Harmonie closes
      ‘Why would it be a good thing that the Harmonie-building closes?’

The expletive pronouns *er* (6a) and *het* (8b) are examples of so called reduced pronouns
in the Vorfeld. Dutch has a paradigm of reduced (or: weak) pronouns that are more
restricted in their distribution than full (or: strong) pronouns. One restriction on reduced
pronouns is that a reduced pronoun object is not allowed to appear in the Vorfeld. I will
discuss the relation between Vorfeld occupation and reduced pronouns in more detail
in Section 2.3. For reference, the two paradigms are given in that section, in Table 2.2,
p32. For now, however, it is worth pointing out that *er* and *het* demonstrate that reduced
pronoun subjects are allowed in the Vorfeld. This is an indication of the default status of
subjects as Vorfeld occupants.

### 2.2.2 Topicalization

Other arguments besides subjects can also appear in the Vorfeld. Direct objects, indirect
objects and oblique/prepositional complements can all be fronted. I will refer to placing a
non-subject in the Vorfeld as *topicalization* – the terms fronting and Vorfeld occupation
can apply to any constituent. Topicalization should be understood in a purely formal
sense. Although a topicalized constituent may be a discourse topic or a sentence topic, it
is not clear that topicalization is restricted to topics, or that it makes constituents topics.
The information structural properties of topicalization will be discussed in Section 2.5.

As a result of V2, topicalization forces the subject to appear in the postverbal domain.
Typically the subject in a sentence with a topicalized constituent appears directly to right
of the finite verb, although other configurations are possible. In (10), we can see examples
of definite NPs and pronouns in the Vorfeld in several non-subject functions.

(10) a. *direct object:*
    De koning van Frankrijk / Die / Hem kom ik vaak tegen op straat.
    The king of France DEM him meet I regularly VPART on street
    ‘I regularly meet the king of France / him in the street.’
b. *indirect object (experiencer in transitive):

De koning van Frankrijk / Die / Hem bevalt het uitstekend in Oslo.

The king of France / DEM / him pleases it splendidly in Oslo

‘The king of France / he likes it a lot in Oslo.’

c. *indirect object (recipient in ditransitive):

De koning van Frankrijk / Die / Hem geef ik geen geld.

The king of France / DEM / him give I no money

‘I do not give the king of France / him any money.’

d. oblique complement:

Tegen de koning van Frankrijk / hem zeg ik “U”

To the king of France / him say I “U”

‘I address the king of France / him with “U”.’

Examples (10a-c) show that full personal pronouns (in this case hem) and demonstrative pronouns (die) can be topicalized.\(^3\) The reduced variant of hem is ’m. The reduced pronoun ’m cannot be topicalized, as is shown in the counterparts of (10a-c) with reduced pronouns, given in (11).

(11) a. *’m Kom ik vaak tegen op straat.

b. *’m Bevalt ’t uitstekend in Oslo.

c. *’m Geef ik geen geld.

Topicalization is not restricted to definite NPs. Example (12a), contains a topicalized indefinite indirect object demonstrates. Example (12b) shows an indefinite (bare nominal) direct object in the Vorfeld.

(12) a. Een paar jongens heb ik een klap verkocht.

A couple guys have I a blow sold

‘I punched a couple of the guys.’

b. Raketsla heeft kapitein Picard alle dagen op het menu staan.

Garden rocket has captain Picard all days on the menu stand

‘Captain Picard eats garden rocket every day.’

A brief detour to topicalized predicates and verbs

In this dissertation, I investigate subject and object fronting. However, I wish to briefly mention here that predicates (whether they are adjectival or nominal), can be topicalized. Examples are given in (13).

\(^3\)Demonstrative refers to form alone. Demonstrative pronouns are frequently used as anaphoric pronouns in Dutch, and are not restricted to deixis.
These examples are somewhat idiomatic, but not all constructions involving Vorfeld predicates are. The reason to bring up topicalized predicates is the role that negation plays in facilitating these types of topicalization. For instance, the negation in (13c) is needed to make the variant with reduced pronominal subject *het* ‘it’ acceptable.

Birner and Ward (1998) ascribe the (nearly) obligatory presence of the negation for English counterparts of (13c) to a rhetoric device they call *proposition denial*. An example of propositional denial can be found in (14).

(14) The international hordes now streaming in from the west and south have, in contrast, no-nonsense ideas about what they want: a chance to work hard and make money. Laid back they are not. (Birner and Ward’s 71b, p68)

Birner and Ward also note that fronted predicate APs or NPs are often found in an explicit contrast that involves a negation, as in (15):

(15) Pretty they aren’t. But a sweet golden grapefruit taste they have. (their 43b, p48)

Topicalizing non-finite verbs appears to be facilitated by negation, too. The example in (16a) even features an explicit contrast in polarity. Example (16b) shows that topicalizing the non-finite verb is marginal without a negation.

(16) a. Geregend heeft het niet, wel gedauwd.
Rained has it not, AFF dewed
‘It hasn’t rained, but it has dewed.’ (Kruisinga, 1938, 4d, p66)

b. ??Geregend heeft het.

I will not be able offer an explanation of why negation has a positive effect on acceptability of topicalization. However, in Section 4.6 I do investigate whether the facilitating effect of negation is found with object topicalization, too. We will see that direct object topicalization is more frequent in the presence of certain sentence adverbs, than when these adverbs are not present. As far as I am aware, this is a hitherto unknown fact about direct object fronting.
2.2.3 Preposition stranding

Standard Dutch allows for preposition stranding provided the object of the preposition is one of the R-pronouns *daar, waar, hier, d’r en er* (‘there’, ‘where’, ‘here’ and two reduced forms of ‘there’), or one of the locatives *overal* (‘everywhere’) and *(n)ergens* (‘no-/somewhere’). When they are the object of a preposition, the R-pronouns, formally identical to locative adverbs, refer as regular pronouns do. Many speakers will even accept them with a human referent. The locatives *overal* and *(n)ergens* have denotations like quantified NPs, respectively *alles* ‘everything’ and *(n)iets* ‘no-/something’, when they are the object of a preposition.

In contrast to their regular counterparts, R-pronouns obligatorily precede their preposition. It is quite common for other material to intervene between the R-object and its preposition. In other words, the object may be extracted from the PP, and the preposition is stranded in its canonical position. Preposition stranding occurs with Mittelfeld scrambling as well as with topicalization to the Vorfeld (Bech, 1952; Van Riemsdijk, 1978; Haeseryn et al., 1997).

In (17), we see a regular personal pronoun as the object of a PP. The pronoun can appear in situ (17a), but not alone in the Vorfeld with the preposition stranded (17b). Fronting the complete PP is fine, as in (17c). The preposition and its object are in boldface.

(17) a. Ik kan goed *met hem* praten
    I can well with him talk

b. ??*Hem* kan ik goed *mee* praten
    Him can I well with talk.

    With him can I well talk
    ‘For me he is a good person to talk to.’

When the object is an R-pronoun, for instance *daar*, it appears left of the preposition (18). The object may even appear alone in the Vorfeld (18c). Like before, fronting the full PP is no problem (18d).

(18) a. *Ik kan goed *met daar* praten.
    I can goed with there talk.

b. Ik kan *daar* goed *mee* praten
    I can there well with talk

---

4In fact, Van Riemsdijk (1978) talks of *post*position stranding.

5The examples show two forms of the preposition translated as ‘with’. The prepositional form *met* is used when the object of the PP follows the preposition. The adverbial or independent form *mee* shows up in preposition stranding, as a postposition, or as a verb particle.
c. Daar kan ik goed mee praten
   there can I well with talk

d. Daarmee kan ik goed praten.
   There with can I well talk
   ‘For me he is a good person to talk to.’

With this behaviour, the Dutch standard language resides between English and most
Scandinavian languages on one side, and German on the other. English and the Scandi-
navian languages allow full fledged preposition stranding and Standard German allows
none. Some German dialects do however show the Dutch type of preposition stranding
(Fleischer, 2002).

Colloquial Dutch and certain dialects are more flexible, however, and may allow
preposition stranding with other NPs, too (De Vries, 1911; Jansen, 1981; Haeseryn et
al., 1997). The examples in (19) feature topicalized full NPs with preposition stranding.

(19) a. nou dubbel glas stap je niet zomaar doorheen hoort.
   PRT glass step you not just like that through PART
   ‘You know, you don’t just walk through double glazing.’ (Nl-a 676:321)

b. [playing Scrabble:]
   de C kun je makkelijk iets mee.
   the C can you easily something with
   ‘“C” is easy to use’ (Nl-a 491:117)

The constraint on reduced pronouns in the V orfeld can be observed for topicalization
of objects of prepositions, too. Since they are not subjects, these objects cannot appear in
the V orfeld in reduced form er or d’r (spoken Dutch). As a result, we have the contrasts
in (20a) and (20b).

(20) a. Ik kan hier / er / d’r niets mee
   I can here there.RED there.RED nothing with

b. Hier / *d’r / *er kan ik niets mee
   here there.RED there.RED can I nothing with
   ‘This / It is useless’ (lit.: ‘I cannot do anything with this/it’).

Preposition stranding is not restricted to oblique objects. Adverbal PPs and PP indirect
objects (by means of dative alternation) allow it as well.

---

6 Jansen (1981) presents an empirical study of several phenomena in spoken Dutch, including Vorfeld
occupation. Part of his book can thus be seen as an early precursor of the current work.

7 Sentences that are taken from the Corpus Gesproken Nederlands, the spoken Dutch corpus used in this
dissertation, are always marked with a region code, a component code, and a sentence number. ‘Nl-a’ refers to:
recorded in the Netherlands, component a. An overview of the corpus is given in Section 3.1. The region code
‘VL’ is used for Flanders.
2.2.4 Non-arguments in the Vorfeld

The Vorfeld may be occupied by adjuncts. The variety in this group is great, and I will not attempt a survey of all possibilities. Not all adjuncts front with equal ease. Intuitively, ‘speaker oriented’ sentence adverbs are good Vorfeld occupants, where manner adverbs, for instance, are not. Some examples are given in (21). Examples (21a) and (21b) involve a complex and a simple temporal adverbial, respectively. In example (21c) a conjunction occupies the Vorfeld. The Vorfeld in (21d) contains coordinated predicative adjuncts.

(21) a. Net als ik naar huis ga, loopt ze met me mee.
   just when I to house go walks she with mee VPART
   ‘Just when I am about to go home, she follows me.’

b. Gisteren ging ik naar de cinema.
   yesterday went I to the cinema
   ‘Yesterday, I went to the cinema.’

c. Daarom zijn de mensen zo moe.
   Therefore are the people so tired
   ‘That is why people are so tired.’

d. Dronken, dol en dwaas, beet ik in mijn bier.
   drunk crazy and foolish bit I in my beer
   ‘Drunk, crazy and foolish, I swigged at my beer.’

It is unclear whether fronting adverbials and fronting arguments are really alike. Many researchers have observed that fronting adverbials is not as restricted as fronting arguments. As said before, this dissertation focuses on subjects and objects in the Vorfeld, so I will not consider this issue much further. However, it is worth pointing out that the ban on non-subject reduced pronouns applies to certain adverbials, too (22).

(22) a. Ik heb in de disco / daar / er mijn koekje verloren.
   I have in the disco there there.RED my cookie lost

b. In de disco / daar / *er heb ik mijn koekje verloren.
   in the disco there there.RED have I my cookie lost
   ‘I lost my cookie in the disco / there.’

In (22), the fronted adverbial was a locative PP or a locative adverb. Similar data can be given for at least causative and instrumental PPs, and extraction out of such PPs.

Under the header ‘non-arguments in the Vorfeld’, one should also mention *split topicalization and extraction out of NP*. In split topicalization (23a), it seems as if the head word of an argument NP (here: onweersbuien) is topicalized alone, leaving other material of the NP (here: eentje) behind. The parts of the split-NP need not be morphological identical with their counterparts in an in-situ realization (23b), however (Van Hoof, 1997, Salverda, 2000; for German: Kuthy, 2002, Féry, 2006).
2.3 Subjecthood and Vorfeld pronouns

The subject restriction on weak pronouns in the Vorfeld – and a similar restriction in German regarding *es ‘it’ – has received a lot of attention in the literature. Some reduced and full personal pronouns have already figured in examples in this chapter. Table 2.2, p.32, gives a more complete overview of the full and reduced personal pronominal paradigms (based on Haeseryn et al., 1997; Van Eynde, 1999)." The possessive pronouns show a similar full-reduced variation, but these are not relevant in the current discussion.

The reduced paradigm in Table 2.2 is reduced in several ways. First, most reduced forms can be considered to be phonological reductions of the strong forms. Secondly, but related to the first point, the full pronouns can be realized with prosodic prominence, but need not be, whereas reduced pronouns can never receive such prominence. Finally, the reduced paradigm shows fewer semantic distinctions, using one form for subject and object in over half of the cases, whereas the full forms in the table show this distinction throughout, except for *jullie.* Compared to the full forms, the reduced forms are restricted in their distribution. We have seen this in relation to Vorfeld occupation. Another example of the restricted distribution is that reduced pronouns cannot generally be conjoined or modified (Cardinaletti and Starke, 1996; Van Eynde, 1999). On the other hand, unstressed full pronouns are marginal with inanimate referents, and ungrammatical with an inanimate referent when stressed. The obligatorily unstressed reduced pronouns show no such
Table 2.2: Full and reduced paradigms for personal pronouns in Dutch.

<table>
<thead>
<tr>
<th>Function</th>
<th>Form</th>
<th>Person singular</th>
<th>Person plural</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>1st</td>
<td>2nd</td>
</tr>
<tr>
<td>subject</td>
<td>full</td>
<td>ik</td>
<td>jij</td>
</tr>
<tr>
<td></td>
<td>reduced</td>
<td>'k</td>
<td>je</td>
</tr>
<tr>
<td>non-subject</td>
<td>full</td>
<td>mij</td>
<td>jou</td>
</tr>
<tr>
<td></td>
<td>reduced</td>
<td>me</td>
<td>je</td>
</tr>
</tbody>
</table>

(demonstrative) - - die / dat - - die

Note: See footnote 8 for additional remarks.

restriction. In Table 2.2, I have also included the demonstrative forms dat and die. They behave a lot like the full personal pronouns, except for the fact that they have no animacy restrictions.

Observations about the non-topicalizability of het ‘it’ in Dutch can already be found in Kruisinga (1938). More recently, the asymmetry has been used by theoretical syntacticians to argue in favour of an asymmetry in syntactic structure between canonical (that is, subject initial) and topicalized sentences in German and Dutch. Canonical word order sentences would be IPs, and have their subjects in SpecIP. Topicalized sentences would be CPs, with SpecCP containing the topicalized element. The asymmetry in topicalizability of reduced and full pronouns could then be modeled by restricting reduced pronouns to certain positions. This approach can be contrasted with approaches that assume that main clauses are always CPs. There are other issues besides reduced pronoun fronting involved in this debate, and I refer the interested reader to Zwart (1997), Gärtner and Steinbach (2003) and Van Craenenbroeck and Haegeman (2007) for comparison and references. These theoretical syntactic issues will not occupy us further here. The term Vorfeld allows us to refer to the position of interest without having to decide whether it is SpecIP or SpecCP.

8 Some remarks on the paradigms in Table 2.2 are in order. The third person singular cells show different forms for gender (MASC, FEM, NEUTER), but the non-reduced paradigm lacks a neuter pronoun (like German es, which is also assumed to be weak/reduced). The form -ie (‘he’, reduced) behaves like a real enclitic (Weerman, 1989) – it requires attachment to a finite verb or a complementizer on its left. Furthermore, the table is biased towards (written) Northern Standard Dutch. Some omitted forms that are encountered in the Corpus Gesproken Nederlands are: -m (‘he’, enclitic, Southern Dutch), gij, ge, -de and u (‘you’, full, reduced and enclitic subject, full non-subject, Southern Dutch); ze (‘her’, non-standard, reduced, both Northern and Southern); and gijlen/gulle, wijlen and zulle (‘you.pl’, ‘we’ and ‘they’, full, Southern).

9 The exception jullie might be explained by its synthetic nature. One probable origin is that it formed out of second person pronoun je and noun lui (‘people’, cf. Eng. you people/all/ones/guys) (Philippa, Debrabandere, and Quak, 2003-9; Van der Sijs, 2004).
The data in (24) illustrates the by now familiar point that reduced pronouns cannot be topicalized. Example (24a) shows that reduced subject pronouns can occur in the Vorfeld, whereas (24b) shows that a reduced object pronoun cannot.

(24) a. Ze heeft Jan gekust.
   she.\red has Jan kissed
   ‘She kissed Jan.’
   \Zwart, 1997, p.35
b. *'r Heeft Jan gekust.
   her.\red has Jan kissed
   ‘Jan kissed her.’
   (ibid)

In contrast to reduced pronoun objects, full pronoun objects can occur in the Vorfeld. Example (25a) is to be read with the typical ‘hat pattern’ – a rising accent on \_haar\_, and the falling, nuclear accent on \_gekust\_, with either a low or a high level tone over \_heeft\_ Jan. In example (25b), nuclear accent falls on the initial constituent \_haar\_ and the verb and postverbal domain are deaccented. Slashes indicate rises and falls, capitals indicate nuclear stress. See Section 2.5 for a discussion of the information structural differences between (25a) and (25b).

(25) a. /\_Haar/ heeft Jan \_geKUST\_.
   her.\full has Jan kissed
b. \_\_\_\_\_HAAR\_ heeft Jan gekust.
   her.\full has Jan kissed
   ‘Jan kissed her.’

On the basis of observations like the ones above, and similar data from German and Yiddish, Travis (1984) offers the constraint in (26).

(26) \textit{Restriction on Topicalization:}

Unstressed pronouns may not topicalize.

The data in (24) and (25) seems to support this generalization. However, the question is what is meant by \textit{unstressed}. Gärtner and Steinbach (2003, citing Lenerz, 1994) offer (27a) and (27b) as counterexamples to (26). Nuclear stress is on the finite verb. The fronted element, even though it appears to have some prosodic prominence, does not contain nearly as much stress as needed to make the topicalization in (25a) acceptable. And yet, example (27c) shows that a reduced pronoun is not allowed, even with neighbouring nuclear stress.

(27) a. Dich KENN ich doch
   you know I \_\_\_\_\_PRT
   German
b. Jou KEN ik toch
   you.\full know I \_\_\_\_\_PRT
   Dutch
We might be tempted to recast the constraint in (26) in terms of reduced pronouns. Still, according to Gärtner and Steinbach (2003), this runs into problems in two ways. First, examples of reduced non-subject pronouns in the Vorfeld do exist in the literature. Second, Gärtner and Steinbach claim that reduced pronoun subjects also show a tendency to avoid the Vorfeld, but that this tendency is not nearly as strong as the one observed for reduced pronoun objects.

Apropos of the first point, Gärtner and Steinbach present examples of reduced pronoun objects in the Vorfeld from colloquial German and German dialects. Weerman (1989) presents a Dutch example (28).

(28) 't Hebben we 'm gisteren nog verteld

it.RED have him.RED yesterday PART told

'We told him that yesterday.' (Weerman’s judgement)

The problem with a sentence like (28) is that I find it only marginally acceptable – an intuition shared by other native speakers of Dutch when asked. However, I do agree that it is not nearly as bad as (24b). Gärtner and Steinbach argue that we should look for examples of topicalized reduced pronouns in colloquial and/or dialectal data. In Chapter 4 I will discuss whether the corpus that I used to investigate Vorfeld occupation contains instances of topicalized reduced pronouns.10

Gärtner and Steinbach’s second claim – that subject reduced pronouns also avoid the Vorfeld, but to a lesser extent – is something that is hard to establish on the basis of intuition data. However, frequency data would be ideal to evaluate such a claim. In Section 4.3.3, I investigate whether reduced pronoun subjects appear in the Vorfeld less often than full pronoun subjects do. Indeed, we will see that weak pronoun subjects are less likely to appear in the Vorfeld than full pronoun subjects. Gärtner and Steinbach propose that there is a phonological reason for the dissociation between the Vorfeld and reduced

10There is a class of examples with reduced pronoun er in the Vorfeld, when this er is, for instance, the object of a preposition (i). The preposition and its object are in boldface.

(i) Er zit een ondeugdelijk slot op

there.RED is a bad lock on

‘The lock that is on it is no good.’

Examples like (i) have been analyzed as er actually being a expletive subject in a existential construction from early on (Bech, 1952; Bouma, 2000). The examples do therefore not necessarily form counterexamples to the reduced object pronoun restriction. Purported German counterexamples in Meinunger (2007), with Vorfeld object es can be analyzed in the same way. They all involve a pronoun that is homophonous with the expletive subject, and meet the requirements for the German existential constructions.
Section 2.4: Violations of V2

Until now, I have assumed that there is exactly one preverbal constituent: the Vorfeld occupant. This assumption is justified when we consider a sentence where a constituent is

pronouns; however in the course of this dissertation, we will see two further possible explanations. The first alternative explanation refers to information structure: Weak pronouns avoid the Vorfeld because the information structural properties of the Vorfeld are incompatible with personal pronouns in general, and reduced personal pronouns in particular (Sections 2.5 and 2.6). Section 4.3.3 offers a comparison of the phonological account and the information structural account with the help of corpus data. The second alternative explanation is of a very different kind. In Chapters 5 and 6, I argue that the ease with which grammatical function assignment can be recovered from a string has an influence on freedom of word order. The effect this has on the placement of personal pronoun subjects will be spelled out and investigated in Section 6.2.1.

Matters surrounding the Vorfeld pronouns are further complicated by the availability of demonstrative pronouns in third person (also in Table 2.2, p32). Examples are given in (29).

(29) a. Die heeft hij geKUST
    DEM has he kissed
    ‘He kissed her/him/it/them.’

b. Dat WEET ik niet
    DEM.N know I not
    ‘I don’t know.’

The demonstratives in (29) can be anaphoric. The demonstrative in (29a) may have a human referent. The mechanisms behind the choice between the demonstrative and the personal forms are still not very well understood, but it seems that demonstratives can be used to refer to a discourse salient referent which is not at that point the established or continuing discourse topic (Comrie, 2000; for German: Bosch, Katz, and Umbach, 2007). A typical examples is a demonstrative pronoun that refers to an entity that was recently (re-)introduced (and thereby salient, but not topical), or to abstract objects such as events, propositions, etcetera. Topicalized demonstrative pronouns do not have to be stressed. In fact, the demonstrative in (29b) need hardly receive any prosodic prominence, and can be reduced to [da] or even [d].

I will return to the issue of pronouns in the Vorfeld in Section 2.5, when I discuss the information structural properties of topicalization in Dutch. Investigating the distribution of pronominal elements in the Vorfeld fits in naturally with the investigation of the corpus predictions discussed in Section 2.6.
topicalized, but the subject does not move to the Mittelfeld. The result is ungrammatical, as illustrated in (30).

(30) *Die ik heb niet gezoend
   DEM I have not kissed
   ‘I haven’t kissed him/her.’

In (30), there are two preverbal constituents, die and ik, that cannot be analyzed as one. As a result, the finite verb is not in second position. The V2 generalization is a simplification of the facts, however. There are numerous exceptions to V2 in declarative main clauses – at least at a superficial level. The finite verb may occur directly at the left edge (V1), or there may be more than one constituent before it (V3, etcetera). The discussion of non-V2 examples that is to follow serves to indicate the boundaries of the constructions that I will investigate in this dissertation.

2.4.1 No elements in the Vorfeld

Under certain circumstances, a Vorfeld element may be dropped. The result is a declarative sentence with no Vorfeld constituent. Subjects and direct objects can be dropped, as well as the objects of oblique arguments that have left their preposition behind (31). Object drop is more common than subject drop (Jansen, 1981). As the data in (32) illustrates, dropping indirect objects is less acceptable than direct objects or subjects (Thrift, 2003).11

(31) a. subject
   (ik) Was inmiddels getrouwd.
   I was by then married
   ‘By that time I had married.’

b. direct object
   (dat) Wil ik wel geloven.
   that want I PRT believe
   ‘I can believe that.’

c. object of a oblique argument
   (daar) Heb ik veel aan gehad.
   there have I much on had
   ‘It’s been useful to me.’

(31) (all from Jansen, 1981)

---

11Thrift (2003) bases her claims on data collected with a questionnaire. Thrift also considers demonstrative pronoun indirect objects ungrammatical in Dutch. She connects this alleged ungrammaticality of demonstrative pronoun indirect objects with the reduced dropfähigkeit of indirect objects. In my opinion, demonstrative indirect objects are fine, but the results in Section 4.3 will show that they are very infrequent.
The sentences in (31) and (32) are recognizable as topic drop sentences, because they miss an argument. However, adjuncts can sometimes be dropped, too, which results in a sentence without a ‘hole’. Such a sentence (33) will look exactly like any other V1 construction, for instance like a polar interrogative or an the antecedent clause in a conditional.

(33) (Toen) Ben ik in dienst gegaan.
then am I in service went
‘Then I went into military service.’ (from Jansen, 1981)

There is a interesting paradox concerning object pronouns in the Vorfeld and topic drop. Pronominal objects that appear in the Vorfeld cannot be reduced, as we have seen. They have to be stressed personal pronouns, or demonstrative pronouns. However, in many of those cases dropping the pronoun completely is unproblematic. The sentences in (31b) and (32) are cases in point: They would be ungrammatical with a reduced personal pronoun.

The differences between subjects, direct objects and indirect objects, and the paradoxical reduced pronouns and topic drop, mean that topic drop would be a very interesting subject for a corpus study. However, the large scale studies that I present in this dissertation are not suitable for this. The size of the corpus is such that manual investigation is too time consuming. Automatically establishing when something is dropped or missing is already slightly problematic. Automatically figuring out what has been dropped would appear to be nearly impossible. Hence, I do not look at topic drop in this dissertation.

### 2.4.2 Left dislocation and hanging topics

There are two common constructions in Dutch in which the verb does not appear to be in second position, but in third. It is not clear to which extent the extra elements actually appear in the clause, or outside of it (in terms of the topological field template: in the lead). The two constructions are (contrastive) left dislocation (CLD) and hanging topic (HT). An example of CLD is given in (34) and in (35) is an example of HT. Capitals indicate nuclear accent, ‘∥’ indicates a not further specified prosodic discontinuity.
Superficially, the two constructions differ only in the place of the resumptive pronoun (preverbal in CLD, unrestricted in HT), and the type of the pronoun that corefers with the initial constituent (demonstrative in CLD, unrestricted in HT). However, it has long been observed that the two constructions are in fact not very similar. Apart from different discourse functions (Frey, 2005, for German, which has basically the same distinction as Dutch), the two constructions differ in ‘connectedness’. The left dislocated constituent in CLD (34) behaves like part of the clause in terms of intonation and binding, while the initial element in HT (35) does not (Van Riemsdijk and Zwarts, 1997; Zaenen, 1997; German: Altmann, 1981; Grohmann, 2003; Frey, 2005). The tighter connection of contrastive-left-dislocated material with the clause is also demonstrated in the contrast in (36). When there are two constituents ‘before’ the Vorfeld, it is the left-most (that is, outermost) one that is the hanging topic.

The ungrammaticality of (36b) is due to the fact that the resumptive pronouns do not agree under the interpretation that is forced by the order of the preverbal elements. The outermost element has to be the hanging topic, the innermost element the contrastive-left-dislocated constituent. However, the demonstrative preverbal pronoun agrees with the outermost element, and the personal postverbal pronoun agrees with the innermost.

An important question in the context of this dissertation is whether we can safely ignore these constructions in our corpus investigation. Although it may be clear that HT does not really resemble topicalization, the same is not true for CLD. As a matter of fact, topicalization has been analyzed as a combination of CLD and topic drop (for instance Zwart, 1997; also Odijk, 1998, for clausal complements).13

12For many other languages two constructions have been observed that are similar and different in much the same ways as Dutch and German CLD and HT for Dutch. See Grohmann (2003), for a comparison with Romance data.

13Hanging topics can also be combined with topic drop, but the construction is clearly distinguishable from a topicalized sentence by prosody and other measures of connectedness, and possibly by a hole in the argument structure.
However, CLD is only possible in a subset of the cases that topicalization is allowed in. For instance, contrastive left dislocation is not allowed with certain quantified NPs, reflexive pronouns or foci.

(37) a. Geen enkele film van Godard (*die) heeft hij gezien.
    No single film by Godard DEM has he seen.
    ‘He hasn’t seen a single film by Godard.’
    (Zaenen, 1997)

b. Zichzelf, (?die) respectEERT hij niet.
    himself, DEM respects he not
    ‘He doesn’t respect himself.’
    (Zwart, 1997)

c. Wie zou je wel ‘s willen ontmoeten?
    ‘Who would you like to meet?’
    ELvis / ??ELvis die / *Elvis DIE zou ik wel ‘s willen ontmoeten
    Elvis Elvis DEM Elvis DEM would I PART want meet
    ‘I would love to meet ELVIS sometime.’

It seems that contrastive left dislocation, although perhaps formally related or similar to topicalization, does not have the same function as topicalization has (Frey, 2005, for a much more thorough analysis but with the same conclusion for German). I therefore deem it safe not to include contrastive left dislocation in the corpus study in this dissertation. However, the comparison of the two constructions would be an interesting issue for future research (see Snider and Zaenen, 2006, on English topicalization and left dislocation).

### 2.4.3 Multiple elements in the Vorfeld

Other cases of multiple constituents in the Vorfeld cannot as easily be analyzed as containing clause external constituents. Whether these cases should actually be considered to be instances of multiple Vorfeld occupation or whether the multiple constituents should be analyzed as one at some level is up for debate.

In the syntactic annotation of the corpus that is used in this dissertation (see Chapter 3), preverbal focus particles are not attached to a fronted NP, but rather to the sentence as modifiers. As a result, these constructions contain two constituents in the Vorfeld, resulting in V3. The finite verb is in boldface, the two preverbal constituents are shown within brackets.

---

14Note that I am assuming a left dislocation prosody here. Pronouncing the answer as a hanging topic, with a break between the initial constituent and the Vorfeld, and with a nuclear accent on both the initial constituent and the demonstrative as in (i), is unproblematic.

(i) ELvis. || DIE zou ik wel eens willen ontmoeten.
An analysis of focus particle attachment in German with the same consequence is given in Büring and Hartman (2001). Bouma, Hendriks, and Hoeksema (2007) argue in favour of an analysis that treats a focus particle and the XP that contains the associated focus as one constituent, partly on the basis of wishing to retain the V2 generalization.

However, there are other particles that may appear before the finite verb together with another constituent for which a one constituent analysis is not as likely. Example (39a) contains the discourse particle *dus* before the Vorfeld. Interestingly, *dus* may occur as a single Vorfeld constituent as well (39b), and can surface in several positions throughout the sentence (39c).

(39) a. [dus] inzake preventie is er niets geweest.
   ‘So nothing was done to prevent it.’
   (Vl-j 600065:60)
   b. Dus is er niets geweest inzake preventie
   ‘As a result nothing was done to prevent it.’
   c. Inzake preventie (dus) is er (dus) niets geweest (dus).

It is unclear whether all examples in (39) have the same interpretation. Sentence (39b), where *dus* occurs in the Vorfeld alone, only allows for a causal reading. Examples (39a) and (39c) also have readings where *dus* is speaker oriented. Meinunger (2004) suggests that in German V3 surfaces in cases like (39a) to force a ‘speech act reading’ of adverbs like *ehrlich* ‘honestly’ (German) vs a sentence internal predicate reading. This contrast between V2 and V3 will briefly be discussed in the context of interpretation constraints on word order variation in Section 5.7.

Combinations of arguments and something else in the Vorfeld are rare. The sentence in (40) has received an annotation in the corpus used in this dissertation in which the subject and a PP that modifies the sentence are preverbal. This annotation is (presumably) motivated by the preferred interpretations of the sentence: The PP does not specify the location of the kitchenette, but it specifies where the kitchenette has the mentioned price.

(40) een kitchenette in den Hubo kost zevenduizend frank en een douche ook.
   ‘A kitchenette costs BEF 7000 at Hubo and a shower, too.’
   (Not: ‘A kitchenette located at Hubo is BEF 7000’)
However, changing the order of the preverbal NP and PP is not possible, which suggests that the PP is part of the NP after all. More common is a combination of modifiers, for instance locative and temporal. They function as complex frame-setting adverbials, indicating both time and place, or manner and time, etcetera, of the event at once. In these cases the order of the constituents could be reversed.

(41) a. [in de Brandpunt] [met uitgaan] hadden ze ’m dus ook al. in de Brandpunt with going out had they him.RED PRT PRT PRT. ‘So, in De Brandpunt, when we/they were going out, they already had it.’ (Nl-a 389:258)

b. [straks] [op Klara] is er Lut Van Der Eycken en ook . . . soon on Klara is EXPL Lut Van Der Eycken and too ‘In a few moments you can listen to LVDE, and . . . , too, on Klara.’ (Vl-f 600840:50)

Multiple Vorfeld occupancy of the type illustrated in (40) and (41) is also observed in German. Müller (2005) presents a wide range of German data, and proposes a HPSG analysis. In his analysis, the Vorfeld is occupied by one verbal constituent with an unpronounced head, so that the multiplicity is only apparent and V2 is respected.

### 2.4.4 Two left brackets

In spoken Dutch, it is not uncommon to find sentences in which the finite verb and possibly some other material is repeated after a part of the Mittelfeld. Effectively, the sentence contains two left brackets, meaning that the topological field template given in Figure 2.1 will not fit as neatly anymore. Haeseryn et al. (1997) describe the construction as involving two overlapping templates.

The construction is called a herhalingsconstructie ‘repeating construction’, but also spiegel- ‘mirror-’ (Huesken, 2001, unseen; Van der Wouden et al., 2002) or overloopconstructie ‘overflow construction’ (Haeseryn et al., 1997). In the example the finite verbs are in boldface.\(^{15}\)

(42) we zijn min of meer zijn we ’t ermeen eens we are more or less are we it.RED with it agreed ‘We kind of agree with it.’ (Nl-n 61:26)

Note that it is typically, but not necessarily, the case that the two finite verbs are the same. Informal inspection of corpus examples also shows examples in which there are changes

---

\(^{15}\)To give you an idea of the frequency of this construction, I note that in the Corpus Gesproken Nederlands, \(~1.2\% (943/77316)\) of SMAINs (roughly: declarative main clauses) contains more than one finite head verb. This means that the construction is as frequent as, for instance, the presence of an indirect/dative object in a sentence.
in agreement *gaan...ga* (plural...second singular ‘go’), aspect *is...wordt* (‘is’ & ‘becomes’), modality *kan...gaat* (‘can go’...‘goes’), and tense *heet...heette* (‘is called’...’was called’), or which the verbs are near synonyms. An in depth analysis of the construction can be found in Huesken (2001, unseen).

### 2.4.5 Summary

We have seen a range of examples where the V2 generalization does not clearly hold, and only additional assumptions about the syntactic structure can decide whether V2 is respected or not. These examples were discussed to show where the boundaries of the investigation in this dissertation lie. As said in the introduction, I constrain the data in the corpus investigation to sentences that contain exactly one preverbal constituent. This constituent is be referred to as the Vorfeld occupant. A definition of Vorfeld occupant in terms of the annotation that is available in the corpus that is used will be given Section 3.3. Topic drop, contrastive left dislocation, hanging topics, multiple Vorfeld occupants and mirror constructions are not be investigated.

This choice is partly made on methodological grounds. In this dissertation I investigate properties of the Vorfeld occupant such as grammatical function, definiteness and grammatical complexity (see Section 2.6). Defining these properties on the Vorfeld occupant, when in fact there are multiple or none, is tricky. Furthermore, we saw that some constructions differ in functional and formal properties to the extent that they warrant a separate investigation, that is, topic drop, left dislocation, hanging topics and mirror constructions.

### 2.5 Topicalization and information structure

I have used the term topicalization to refer to the placement of non-subject arguments in the Vorfeld. What distinguishes topicalization from canonical subject initial word order, is that topicalization is restricted with respect to information structure. A subject in the Vorfeld is not restricted. In this section I will discuss the information structure restrictions on Vorfeld objects. Following Gundel (1974), we can distinguish two main types of topicalization: focus topicalization and topic topicalization. I will discuss these two in turn. We shall see that it is difficult to give a clear unifying principle of all examples of topicalization. I will invoke the concept of informational importance to explain the observations (Givón, 1988; Gundel, 1988). Although this concept is vague, we will be able to draw some plausible parallels between it and pronominal form. Informational importance can thus offer an explanation of the special relation between reduced pronouns and the Vorfeld.
2.5.1 Focus topicalization

The part of the sentence that is in focus contains material that is ‘informative, newsy, dominant, or contrary-to-expectation’ (Vallduví and Engdahl, 1996, p462). I will refer to non-focussed material as background material. We may induce a focus/background division using a constituent question (43), or an explicit contrast (44).

(43) A Wie heb je gisteravond gekust?
   ‘Who did you kiss last night?’
   B Ik heb gisteravond [focus Grace \ KELLY\] gekust.
   I have last night Grace Kelly kissed.
   ‘I kissed Grace Kelly last night.’

(44) A Je hebt Gene Kelly gekust!?
   ‘You kissed Gene Kelly!?’
   B Nee, Ik heb [focus \ GRACE\ Kelly] gekust!
   No, I have Grace Kelly kissed
   ‘No, I kissed Grace Kelly!’

In Dutch, focus is associated with prosodic prominence. In the examples in (44), the nuclear accent falls within the focussed constituent.

Focus topicalization puts the focussed constituent in the Vorfeld. The following sentences are responses to (43A) and (44A), respectively. As a result of focus topicalization, the direct objects are in the Vorfeld.

(43) B’’ [focus Grace \ KELLY\] heb ik gisteravond gekust.
(44) B’’ Nee, [focus \ GRACE\ Kelly] heb ik gekust!

Because of focus topicalization, nuclear accent resides in the Vorfeld. The rest of the sentence, the finite verb and the postverbal domain, is deaccented. The example of a topicalized stressed personal pronoun (25b), in Section 2.3, is a case of focus topicalization.

Jansen (1981) observes that focus topicalization in Dutch is rare. In a corpus he collected through interviews with native speakers, he found only a few cases of focus topicalization. Amongst these is the example in (45). The focus brackets were added by me.

(45) A What did you make in handicraft class?
   B [focus zo’n Engelse themuts] heb ik onder andere eens gemaakt
   such an English tea pot cover have I amongst other once made
   ‘Amongst other things I made one of those English tea pot covers.’

In spite of its rareness, the example in (45B) feels natural. It is in my ears completely equivalent in meaning to a canonical, subject-initial realization (45B’).
Recall from Section 2.3 that reduced pronouns in Dutch cannot receive prosodic prominence. They can therefore not be focussed (46B). And, since focus on a reduced pronoun is not possible, focus topicalization is not compatible with a reduced pronoun either (46B’).

Thus, focus topicalization will not lead to a reduced pronoun in the Vorfeld.

### 2.5.2 Topic topicalization

In topic topicalization, a non-subject constituent moves to the Vorfeld, just like in focus topicalization. However, in topic topicalization, there is a focus after the Vorfeld. Nuclear stress falls on a syllable inside the focussed material. Therefore, the finite verb and postverbal domain are not deaccented in a topic topicalization. What kind of material appears in the Vorfeld in a topic topicalization may differ. We may distinguish the cases with contrastive material in the Vorfeld, and cases with non-contrastive material in the Vorfeld. Let me start with the former.

In contrastive topic topicalization, the material in the Vorfeld is prosodically prominent. An example is given in (47). There are now two prominent accents in the sentence: an accent in the contrastive topic in the Vorfeld, and nuclear accent in the postverbal focus. The initial accent receives a marked rise, possibly preceded by a slight dip, and the nuclear accent is a fall. The result is the so called hat pattern. See Jacobs (2001), Braun and Ladd (2003), and Féry (2006) for different characterizations of the rising accent in German contrastive topics. See ’t Hart (1998) for a general description of the Dutch hat pattern.

The topicalized stressed pronoun in (25a), Section 2.3, is also an example of contrastive topic topicalization.

There is a clear interpretation effect involved in using the contrastive topic construction. The utterer of (47) is either going to say something about, or is implying something about, alternatives of the Vorfeld constituent (in this case: other drinks). This inherent contrast means that contrastive topic is similar to focus as described above. Indeed, it has been
argued that contrastive topics are foci (Van Hoof, 2003), that they are semantically related to foci (Büring, 2003), or that they contain a focus (Steedman, 2000).

Roberts (1996), and later Büring (2003), argue that contrastive topic constructions like (47) should be understood as strategies to answer complex questions. For (47) we can imagine a complex question ‘who is drinking what?’ One way to answer this question is to start listing the drinks, and ask for each drink who is drinking it. This scenario is made explicit in (48).

(48) In a bar, the waiter brings drinks to a table, and wonders: who is having what?

| Who is having beer?           | Het biertje gaat naar Gillis. (SVO) |
| Who is having Dutch gin?     | De klare krijgt Hjallis. (OVS)      |
| Who is having red wine?      | De rooie wijn is voor Gunnis. (SVO) |

In each case there are prominent accents on both the initial element (the contrastive topic) and the postverbal element (the focus). This is independent of whether the elements are objects or subjects. The choice for a specific answering strategy is not fixed by the question. In the case of the complex question answered in (48), one might as well have started with the drinkers and revealed for each drinker what they are drinking.

Like (ordinary) focus, contrastive topic is associated with prosodic prominence. Therefore, like focus, contrastive topic cannot fall on a reduced pronoun, independent of whether the contrastive topic is put into the Vorfeld or not. Contrastive topic topicalization will not lead to a non-subject reduced personal pronoun in the Vorfeld.

Not all cases of topic topicalization are contrastive. I will now turn to examples of non-subject Vorfeld occupants that do not involve contrast. These cases are very hard to characterize, and I will only be able to give an intuitive characterization. An example of a topic topicalization that does not involve contrast of the topicalized element is given in (49). Note that both sentences contain direct speech, and that the second sentence is a self-quotation (indicated in the translation).

(49) A   hij zegt ja je kent Margarita toch wel
       ‘He says: “surely you know Margarita?”’

       B   ja die ken ik wel maar
            yes DEM know I AFF but
            ‘(I said:) “Sure I know her, but…”’

Sentence (49B) starts with the demonstrative die, referring to Margarita. Nuclear accent in (49b) falls on ken, which makes (49B) an example of topic topicalization. However,
the referent of the demonstrative pronoun is not contrasted with anything. There is no indication that the speaker will go on to talk about other people besides Margarita. Sentence (49B) is not a contrastive topic topicalization. A prosodic difference between contrastive topic topicalizations and the topic topicalization in (49) is that the topicalized die does not have to receive much prosodic prominence in (49).

The demonstrative pronoun in the Vorfeld feels very typical of a construction like (49). Still, in the context of (49A), it would have been possible to refer to Margarita in (49B) with a definite full NP, a proper name, or a full personal pronoun (49B'). A reduced personal pronoun is ungrammatical. When the direct object is not topicalized and realized postverbally, a reduced personal pronoun is available as a referential form, too (49B').

(49) B' Dat meisje / Margarita / Haar / *'r ken ik wel.
     that girl Margarita her.FULL her.RED know I AFF but
     B'' Ik ken dat meisje / Margarita / haar / 'r wel.

This presents us with a problem, though. In the case of contrastive topic topicalization, and in the case of focus topicalization, we could explain the fact that a reduced personal pronoun could not be topicalized by appealing to the fact that the context did not allow a reduced personal pronoun to be used at all. However, this will not explain the data in (49B'), since (49B'') shows that a reduced pronoun can be used in the context. On a more general note, we can ask ourselves whether this is any connection between examples like (49B) and contrastive topic topicalization, apart from the fact that nuclear focus does not fall on the topicalized constituent.

In their study of non-canonical word order in English, Birner and Ward (1998) argue that being related to a contextually available referent through so called partially-ordered-set relations is a necessary condition for topic topicalization. Birner and Ward thus propose a weakened givenness requirement. According to them, a given NP may always be topicalized. An (indefinite) NP realizing a new referent may be topic-topicalized as long as the referent can be related to a given referent – where related to a referent means: to be a subset or part of it, to be of the same type as it, or to come before or after it in some given order, etcetera. This criterion is met by both contrastive topic topicalization, and by the examples of Vorfeld demonstratives in (49B). In the contrastive topic examples above, each contrastive topic (a drink) came from a contextually available set (the drinks on the waiter’s tray). In the case of (49B), the demonstrative coreferred with Margarita, who was mentioned in the sentence before. However, this relaxed givenness requirement does not explain why the constituent referring to Margarita in (49B) and (49B') cannot be a reduced pronoun. The status of the referent as given does not change, only the way the referent is referred to does. As a result, being related to a contextually provided referent cannot be more than a necessary condition.\(^\text{17}\)

\(^{16}\)In my ears the version with haar ‘her.FULL’ sounds stilted.
Other researchers have proposed that appearing at the left periphery is indicative of being an aboutness-topic, and especially of being a changed – or shifted – aboutness-topic (Gundel, 1974; Reinhart, 1982; Givón, 1983; Gundel, 1988; Jacobs, 2001; Frey, 2006; Frascarelli and Hinterhölzl, 2007, amongst others). An aboutness-topic is the entity about which a sentence can be understood to convey information. We speak of a shifted aboutness-topic when the aboutness-topic of a sentence is not that of the preceding sentence. The problem with aboutness is that it is hard to pin down. For instance the sentence (49B) could intuitively be understood as supplying information about the speaker (ik ‘I’), as well as about Margarita (see McNally, 1998, for a critique of aboutness-topics; see Ward, 1988, Prince, 1998 for a discussion of the applicability of aboutness-topic tests). I will not try to answer the question of whether the concept of aboutness-topic is useful for the analysis of word order in Dutch. This will have to remain a question for further research.

However, even without fully understanding the concept of aboutness-topic, we can see an interesting parallel between shifted aboutness-topics on one hand, and focus and contrastive topic on the other: Shifted aboutness-topics are not new or contrastive as such, but they are new in their role of aboutness-topic. On the assumption that aboutness-topics have a tendency to remain the same over a stretch of discourse, a shifted aboutness-topic is less predictable than a continuing one. Gundel (1988) lets the cross-linguistically observed tendency to put foci, contrastive topics and shifted topics at the front of a sentence follow from the principle stated in (50).

(50) *First-things-first:*

Provide the most important information first.

Gundel considers information that is new, unpredictable, contrastive or emphasized to be important. A very similar principle is proposed by Givón (1988) as *attend first to the most urgent task*, citing low predictability as a cause of urgency. Also see Herring (1990) for a comparative overview and discussion.

---

17 In fact, Birner and Ward present data that suggests this, too. The contrast in (i) shows that a fronted constituent is subject to additional restrictions (the demonstrative determiner), even when the its referent is given.

(i) I have a recurring dream in which... I can’t remember what I say. I usually wake up crying.

a. This/#The dream I’ve had maybe three, four times.

b. I’ve had this/the dream maybe three, four times. (their 284, p226)

They do not offer an explanation for this additional restriction.

18 The first four cited works are of a general and/or cross-linguistic nature. The second three works are about German word order or partly about German word order. The works may show considerable differences in the details, but they share that the left periphery is related to (shifted) aboutness-topics. More references on the subject can be found in the cited works.

19 I hope that the results presented in the rest of this thesis can be of help in this research.
I propose that we consider the Vorfeld, as a left-peripheral position in the Dutch clause, a place for important material. Topicalized constituents have to be in some sense important. In the case of focus topicalization and contrastive topic topicalization, the fronted material is important because it is new or contrastive. However, the importance of the fronted material in (49B), repeated below, still needs to be established.

(49) B die ken ik wel
DEM know I AFF
‘I know her.’

Recall that the fronted material is not focussed, nor contrastive. What would justify calling it important, then? I think that the behaviour of demonstrative pronouns in discourse is key. Personal pronouns pick up referents that have typically been a) mentioned continuously in preceding discourse, b) realized as subjects and c) pronominalized before. Personal pronouns realize referents of which it is highly predictable that they will be mentioned again (for English Arnold, 1998; see also Arnold, 2006, for an overview and implications of this perspective on pronominal use). Demonstrative pronouns however, pick up referents that are either recently introduced, have not been repeatedly mentioned, not realized as a subject or not pronominalized before (for Dutch: Comrie, 2000, Kaiser and Trueswell, 2004; for German: Bosch, Katz, and Umbach, 2007). Compared to personal pronouns, demonstrative pronouns realize referents of which is it less predictable that they would be mentioned again. According to Gundel (1988) and Givón (1988), this difference in predictability translates into a difference in importance. The lower predictability of mention of a referent of a demonstrative pronoun means it qualifies as a Vorfeld occupant. This predictability should be understood in general terms, and is not completely a case of the context, though. For instance, that Margarita would be mentioned again in (49B) is rather likely. As (49B′) illustrated, the speaker of (49B) could have chosen to use a reduced personal pronoun. However, by presenting the referent as so highly predictable, and therefore unimportant, one loses the option of realizing it in the Vorfeld.

Non-subject material in the Vorfeld is characterized by its (relative) importance. Focus topicalization and topic topicalization involve material that is new, contrastive or relatively unpredictable. The relation between important material and the left periphery is made by Gundel’s first-things-first principle, based upon cross-linguistic investigation of word order.

Even though I think it is appealing at an intuitive level, the characterization of the Vorfeld as containing important material leaves open many questions that will have to be answered in future research. The notion of importance is very vague, and for it to be used as a robust linguistic concept, one should try to find ways to measure importance. Givón (1988) gives several corpus based measures of predictability, and thereby of
importance. Future work should try to extend this work on larger corpora and with modern statistical methods (see Snider, 2005 for some results for English). By assuming that the lack of ability to realize important material is what separates personal pronouns, and especially reduced personal pronouns, from all other forms of NPs, we explain why they should behave differently when it comes to topicalization. Even if I am not able to measure importance at this point in the corpus that I will be using, I will be able to see whether the corpus data supports a split between personal pronouns and everything else.

The appeal to the first-things-first principle does make a further prediction that I have not mentioned yet. I have focussed on topicalized constituents, and I have tried to argue that they have to be important in order to appear in the Vorfeld. However, Gundel’s statement of the first-things-first principle does not mention topicalization or non-canonical word order anywhere. In contrast, the principle is meant to be a universally applying (but violable) principle. Now, if the Dutch Vorfeld is preferably a place for important material, subjects should also be sensitive to this. Non-important subjects should be repelled from the Vorfeld, too. In particular, reduced pronoun subjects should show a tendency of avoiding the Vorfeld. This tendency cannot be as strong as with objects, because, obviously, reduced pronoun subjects are grammatical in the Vorfeld. However, we may see it as a statistical trend in the corpus. Thus the assumption that the Vorfeld is a place for important material makes the same prediction that Gärtner and Steinbach (2003) made on prosodic grounds. Section 4.3.3 compares the two approaches on the basis of corpus results.

2.6 Word order trends

In the preceding four sections, we have seen many sides of the preverbal domain in Dutch, and especially of material in the Vorfeld. We have seen what kind of material can occupy the Vorfeld, in terms of form and function. I have also contrasted clauses with exactly one Vorfeld occupant with clauses that deviate from this, in order to show which constructions will be investigated in the dissertation. The chapter thus far has been of a descriptive nature, although we have come across the ‘accidental’ question to be investigated quantitatively. In the brief discussion of the relation between negation predicate topicalization, I announced that the relation between sentence adverbs and direct object topicalization is investigated in Section 4.6. The discussion of the special status of pronouns in the Vorfeld brought up a claim put forward by Gärtner and Steinbach (2003), which is that reduced pronouns show a tendency to avoid the Vorfeld across the board. This prediction is particularly interesting because it can be made on the basis of the discussion of the information structural properties of the Vorfeld, too.

In this section, I will be more systematic in generating expectations or hypotheses to investigate quantitatively. The aim of this dissertation is to gain insight into the choice
of a Vorfeld occupant. In particular, we want to learn more about fronting of subjects, direct objects and indirect objects. The first step towards understanding this choice is to investigate which properties of potential Vorfeld occupants influence whether they end up in the Vorfeld or not. Put differently, we seek to answer the question: How do constituent properties relate to the chance that this constituent is fronted?

Before we can begin to generate expectations for the corpus research, we must decide what kind of properties we can expect to play a role in the choice of a Vorfeld occupant. We have already seen one property that (at least non-subject) constituents must have if they want to appear in the Vorfeld; they have to be important. Importance is not a concept we can directly use in a corpus study. However, I proposed that we could link NP form to importance, and use this to investigate whether the Vorfeld indeed prefers important material over unimportant material.

Importance is a property that links a constituent directly to the Vorfeld. However, other properties may cause a constituent to appear in the Vorfeld because these properties are associated with early realization in the sentence in general. In that case, Vorfeld occupation by a constituent is a result of the fact that the Vorfeld ‘happens’ to be the earliest possible position. Similarly, some properties may be associated with late realization in the sentence. A constituent with such a property may be prevented from appearing in the Vorfeld. Properties that promote early or late realization in the whole sentence form global word order trends. The general hypothesis that there are global word order trends that also affect Vorfeld occupation is a rich source of concretely testable hypotheses about trends in fronting. There is a large literature on word order in the Mittelfeld of German and Dutch, and on word order in the postverbal domain of English. The results from this literature can be used to generate predictions about the relation between certain constituent properties and Vorfeld occupation.

In the rest of the section, I will look at three factors in postverbal word order. On the basis of each, I will formulate concrete questions that can be answered quantitatively. In each case, the concrete questions are the result of applying the global word order hypothesis to the respective factor. The three factors are canonical argument order, definiteness, and grammatical complexity. These factors have been extensively discussed in the literature on Dutch, German, and English. They also have the advantage that they can be fairly robustly measured or annotated, which facilitates large scale corpus analysis.

### 2.6.1 Canonical argument order

Results about the canonical or unmarked order of subject, indirect object, and direct object depend on what is considered to be canonical or unmarked. Lenerz (1977) defines markedness of German Mittelfeld argument order in terms of distribution. Of two word orders, the one that appears only under special circumstances, whereas the other can be
used freely, is the more marked word order. The special circumstances may be related to
information structure, definiteness, pronominality, etcetera. Let me paraphrase Lenerz’
conception of markedness as follows: On the assumption that we have identified the
factors favouring a certain word order that are not related to grammatical function,
canonical word order is the allowed word order that is not favoured by any of those
factors (see also Dryer, 1995). That is, canonical argument order is the word order that
does not have an ‘excuse’, the order that cannot be derived from other principles.

The prevailing opinion in the literature on German is that the unmarked order in the
Mittelfeld is subject before object (S≺O). However, when we look at the order of indirect
and direct object there are different claims. Lenerz (1977) and Uszkoreit (1987) conclude
that IO≺DO is the unmarked order. In contrast, Müller (1999) takes DO≺IO to be the
underlying order in German, on the basis of word order data and binding data. Müller is
however careful to distinguish underlying from unmarked word order, describing the latter
in terms of typicality or acceptability of a surface word order. Other researchers have
proposed that different verbs carry different unmarked word orders for their arguments
(Haider, 1993; also see Müller, 1999 for discussion and references). Support for claims
that different verbs prefer different argument orders can be found in corpus based work
(Kurz, 2000a; Kempen and Harbusch, 2004; Heylen, 2005). Some ditransitive verbs were
found to prefer DO≺IO, and others IO≺DO. Dative-selecting transitives frequently allow
O≺S (breaking the S≺O generalization).

In the Dutch Mittelfeld, S≺O is not only default, but even near-categorical. For instance,
German readily allows pronominal objects to precede non-pronominal subjects. However,
this is rare in Dutch. Important exceptions are formed by dative experiencers (51a) and
certain reflexive arguments (51b), which may precede their subjects (Den Besten, 1985;
Shannon, 2000; Haeseryn et al., 1997).

(51) a. Eigenlijk is me de accommodatie niet zo goed bevallen.
   Really is me the accommodation not so well pleased
   ‘I was not happy with the accommodation, to be honest.’

   b. Bij dergelijke zinnen doen zich de volgende gevallen voor: . . .
   With such sentences happen REFL the following cases VPART
   ‘Amongst such sentences we can find the following cases: . . .’
   (both from Haeseryn et al., 1997)

The reference grammar Haeseryn et al. (1997) takes the default order of two nominal
objects in the Mittelfeld to be IO≺DO throughout, as do Jansen (1981) and Van der
Beek (2005).

20In his paper Müller formulates a number of constraints on linear order that account for the difference
between the underlying and surface word orders. We may therefore interpret Müller’s results as meaning that
DO≺IO is canonical word order in German in our sense, since it is the word order without any excuses.
How should these findings about the Dutch and German Mittelfeld be transferred to the Vorfeld, if we assume that canonical argument order is a global word order trend? Obviously, S≺O in Dutch is not near-categorical when the Vorfeld is considered. We have already seen many examples in this chapter where one of the objects preceded the subject by occupying the Vorfeld. The directly observable relation between canonical argument order and Vorfeld occupation is not the same as the relation between canonical argument order and word order in the Mittelfeld. This does not mean that we have to conclude that canonical argument order is confined to the Mittelfeld, however. Canonical argument order was the argument order that surfaced when other influences on word order have been controlled for. These other influences may be different for the Mittelfeld than for the Vorfeld. The preference that the Vorfeld has for important material is an example of such an influence. Canonical argument order S≺O for Dutch predicts that ceteris paribus subjects have a stronger tendency to appear in the Vorfeld than objects do. This would also explain why the subject is information structurally unrestricted as Vorfeld occupants, and objects are not. As a first indication of whether we can expect to find this tendency, we can look at Jansen’s (1981) corpus study of Vorfeld occupation. Jansen indeed reports that subjects front more often than other arguments.

Applying the same reasoning to IO≺DO, we predict that indirect objects show a stronger tendency to appear in the Vorfeld than direct objects do. There are results in the literature that suggest that this prediction will not be borne out. Thrift (2003) concludes from questionnaire data that putting an indirect object in the Vorfeld is only marginally possible in Dutch. Lamers (2001) presents experimental data comparing Vorfeld occupation and word order in the Mittelfeld. A questionnaire study of the understandability of different word orders revealed no difference in understandability of subject or direct object in the Vorfeld. Indirect objects in the Vorfeld were however slightly harder to understand. In the Mittelfeld, subject and indirect object initial word orders were easiest to understand, and direct object initial word order hardest. The studies of Thrift and Lamers therefore suggest that the prediction that having an indirect object in the Vorfeld is less marked than having a direct object in the Vorfeld will not be borne out.

I have until now used qualitative terms like ‘unmarked word order’ and ‘has a stronger tendency to front’. These terms can receive a quantitative interpretation which can be used in a corpus investigation. The interpretation of canonical word order that I started out with can be related in a natural way to statistical analysis. The idea that of two word orders, canonical argument order is the one that can be encountered in most circumstances has a straightforward statistical interpretation. It means that if we control for other influences on word order in our statistics, the unmarked or canonical of two possible word orders should be the most frequent one.21

Canonical argument order gives rise to the following prediction about Vorfeld occupation, that is tested against the corpus in Chapter 4: Elements that are higher on the
grammatical function scale in (52) front more often after we correct for the influence of other factors.

(52) subject ≺ indirect object ≺ direct object

In the statistical models used to investigate the data in Section 4.5, canonical argument order is a factor treated on a par with other factors on Vorfeld occupation. I will now turn to the discussion of those other factors: definiteness and complexity.

### 2.6.2 Definiteness

The second factor in word order that is investigated is definiteness. The tendency to realize definite material early in the Mittelfeld, and indefinite material late in the Mittelfeld, has been observed for Dutch as well as for German. In the discussion of these effects below, I will make two assumptions. First, I will treat definiteness as a formal property, purely on the basis of surface form. This does not mean that I deny the possibility that there are forces underlying definiteness effects on word order, such as givenness/anaphoricity, referentiality, or intonation. However, treating definiteness as NP form allows us to investigate definiteness in our corpus. The corpus does not have annotation for any of the properties mentioned, but automatically categorizing NPs after form is fully feasible. Secondly, I will distinguish three main levels of definiteness: pronoun, definite full NP, and indefinite full NP. As we will see below, these three levels of definiteness have different effects on word order: Pronouns have a strong tendency of being realized early in the Mittelfeld, and indefinite NPs have a strong tendency of appearing later in the Mittelfeld. The behaviour of definite full NPs falls somewhere in between.

In the Dutch and German Mittelfeld, arguments can take part in two types of word order variation: **scrambling over argument**, referring to the reordering of arguments with respect to canonical word order, and **scrambling over adverb**, which refers to positioning an argument left of an adverb in the Mittelfeld. Definiteness influences word order in both types of scrambling.

In Dutch, scrambling over adverb is generally impossible with indefinite full NPs (but see exceptions below), optional with definite full NPs, and obligatory with pronouns (Haeseryn et al., 1997; de Hoop, 2003). The examples in (53) illustrate this partition. The adverb is in boldface.

---

21 Lenerz (1977) himself opposes the idea that one can derive canonical word order from corpus frequencies. However, the examples that he uses to argue his point are examples where the corpus linguist fails to control for variables that the linguist using intuition data does control for. In my opinion, the conclusion of this should not be that canonical word order and quantitative analysis are incompatible, but that a linguist using quantitative analysis should not rely on simple counting alone.
(53) a. We moeten het / de was / *een voorstel maar doen.  
   ‘We should do it / do the laundry / make a proposal’ (De Hoop’s 9–11)
b. We moeten maar *het / de was / een voorstel doen.

Van der Does and de Hoop (1998) and de Hoop (2003) also demonstrate that deaccented, anaphoric definite NPs scramble optionally, although they might have a stronger tendency to scramble than non-anaphoric definite NPs. Accented pronouns are also free to appear right of the adverb. Specific indefinite full NPs form an exception to the ban on scrambling of indefinite full NPs. They are allowed (but not forced) to scramble left of an adverb.22

Scrambling over argument does not appear to be influenced by the definite/indefinite full NP distinction in Dutch. Pronouns, however, show a clear preference for early realization in the Mittelfeld. For instance, the basic order IO≺DO in a Dutch ditransitive is not respected when the direct object is a pronoun, and the indirect object a full NP (Van der Beek, 2005). The indirect object is in boldface.

(54) a. Kees reikt Koos ??’m / de kaasschaaf aan.  
   Kees hands Koos itMASC.RED / the cheese knife over  
   ‘Kees hands over the cheese knife/it to Koos.’
b. Kees reikt ’m / *de kaasschaaf Koos aan.

Apparently, in the Mittelfeld, the order pronoun≺full NP is preferred over the canonical argument order IO≺DO.23

To summarize, we see that in the Dutch Mittelfeld, pronouns tend to appear early, and indefinite full NPs appear late. These are tendencies, and not categorical statements, as we have seen that there are exceptions to the generalization.

Similar trends as those illustrated above for Dutch have been found for the German Mittelfeld. The main difference between Dutch and German is that the definiteness effects are more pronounced in German, which is presumably related to the fact that word order in the German Mittelfeld is freer than in the Dutch Mittelfeld. Pronominal arguments have a strong tendency to move leftward in the German Mittelfeld, and may scramble over subjects, even when accusative. Scrambling over adverb behaves as in Dutch. In addition, scrambling over argument shows a sensitivity to the definite-indefinite full NP distinction that is not apparent in Dutch (see Lenerz, 1977, Uszkoreit, 1987, Müller, 1999, for the contrast pronoun-full NP and the contrast definite-indefinite full

22On some definiteness scales in the literature, for instance Aissen (1999), specific indefinite full NPs are ranked above non-specific indefinite NPs. This difference in ranking between specific and non-specific indefinites is motivated by independent distributional properties.
23I will gloss over many intricate details of Dutch and German word order here. One interesting detail that I am ignoring is that when both IO and DO are reduced pronouns, German and Dutch order these DO≺IO (Haeseryn et al., 1997; Müller, 1999). This, and other similar facts, does not contradict that there is a tendency to order pronouns before full NPs, and definite NPs before indefinite NPs, however.
NP; Büring, 2001, on definite-indefinite full NP; and Kempen and Harbusch, 2004 for corpus results).

The strategy I have taken when I formulated the expectations for the relation between grammatical function and Vorfeld occupation would lead to the following prediction about the relation between definiteness and Vorfeld occupation: Controlling for other factors, elements on the scale in (55) appear more often in the Vorfeld than elements to their right.

(55) pronoun ≺ definite full NP ≺ indefinite full NP.

As with canonical argument order, Vorfeld occupation need not be fully parallel to Mittelfeld word order. For instance, placing a full NP direct object before a pronominal subject in the Mittelfeld is ungrammatical in Dutch (56a). However, (56b) shows that placing a full NP direct object in the Vorfeld, while the pronominal subject is in the Mittelfeld, is fine (see Weber and Müller, 2004, for similar observations about German).

(56) a. *’s nachts houdt de buren ze wakker.
      at night keeps the neighbours she awake
      ‘She keeps the neighbours awake at night.’

b. de buren houdt ze ’s nachts wakker.

We have two different possible reasons to expect that reduced personal pronouns do not like to appear in the Vorfeld: Gärtner and Steinbach’s (2003) prosodic account, and the proposed status of the Vorfeld as a position for important material (Section 2.5). This contradicts with the predictions of the scale in (55). The corpus study will tell us whether the preference of the Vorfeld for important material is stronger than the global effect that definiteness has on word order. In any case, demonstrative pronouns should show a very strong tendency to front, since they are both important material and of the highest definiteness level.

The proposal that the Vorfeld prefers important material was based on a universal word order principle put forward by Gundel (1988): first-things-first. Gundel observes a conflict that is similar to the conflict between the Vorfeld’s preference for important material and the global effect of the definiteness scale on word order sketched above. Gundel (1988) proposes two possibly conflicting universal word order principles, given in (57).

(57) a. Given-before-new:
      State what is given before what is new in relation to it.

b. First-things-first:
      Provide the most important information first.

The principles conflict when the given material in a sentence is not important. The first-things-first principle then prefers that the new material be provided first, since new
material is always important. The given-before-new principle disfavours this word order. The conflict between first-things-first and the definiteness hierarchy is likely to make the relation between NP form and Vorfeld occupation in the corpus quite complex. The results from Jansen (1981) may give us an first impression of this relation. Jansen finds that pronouns front more often than other NPs, and that definite NPs front more often than indefinite NPs. This is compatible with the definiteness scale. In addition, Jansen reports that demonstrative pronouns front more often than personal pronouns, a difference predicted by the first-things-first principle.

Before I conclude this section about definiteness and Vorfeld occupation, I have to mention a factor in the distribution of subjects that may be unrelated to whatever underlies the definiteness hierarchy given above. In Section 2.2.1, I mentioned the Dutch existential construction (EC). A sentence with an indefinite logical subject (geen beeld ‘no statue’ in 58) is preferably not realized as a canonical sentence (58a), but as an EC (58b). In the EC in (58b), the expletive subject er is in the Vorfeld. Alternatively, other material like a locative PP can be placed in the Vorfeld, in which case the expletive er is optional (58c). Mikkelsen (2002, for Danish) and De Hoop and Krämer (2006) propose that the use of an EC is to repel a ‘poor’ subject from its canonical position. The optionality of er in (58c) suggests that in Dutch, a PP may also serve this purpose. The logical subject is in boldface.

(58)  a. ??geen beeld stond in de tuin
   ‘There was no statue in the garden.’
   no statue stood in the garden

   b.  er stond geen beeld in de tuin.

   c.  in de tuin stond (er) geen beeld.

The fact that (part of the) indefinite logical subjects will appear in a construction that moves them out of the Vorfeld means that we have an additional reason to expect to see a definiteness effect for subjects, where indefinite subjects are less likely to appear in the Vorfeld than definite or pronominal subjects. Note that the logical subject of an EC is not

---

24 It may be tempting to propose that the definiteness scale is a grammaticalized version of the given-before-new principle. Pronouns are most likely to realize given referents, and indefinites typically introduce new referents. However, by given and new Gundel (1988) means something more like aboutness-topic and comment, although she does recognize the correlation between aboutness-topichood and discourse givenness.

25 The sentence is acceptable under the partitive reading: None of the statues was in the garden.

26 It may be that this is a locative inversion, even though in a Dutch main clause they would be hard to distinguish from any other PP topicalization. The matter is irrelevant here, however, because all example (c) is supposed to show is that the expletive er is optional when a locative PP is in the Vorfeld. Whether (c) without er is an EC, a locative inversion, or something else is besides the point.
prohibited from appearing in the Vorfeld. Example (58d) shows that focus topicalization may be used to put the logical indefinite subject of an EC ‘back’ into the Vorfeld.

(58)  
\[\text{d. geen BEELD stond er in de tuin.}\]

However, we expect focus topicalization to be relatively rare (Section 2.5.1), so that the overall effect of EC on the statistics will be that indefinite subjects are less likely to be realized in the Vorfeld.

In summary, we expect to find the definiteness scale pronouns \(\prec\) definite full NPs \(\prec\) indefinite full NPs reflected in the Vorfeld statistics, with the possible exception of personal pronouns. This prediction can be explained as arising from the interaction between two tendencies: a) the tendency to realize definite/given material early on in the sentence, b) the tendency to use the Vorfeld for important, unpredictable or highlighted information. The existential construction in Dutch may be an independent factor to negatively influence Vorfeld occupation by indefinite subjects.

### 2.6.3 Grammatical complexity

The classic statement about the effect of grammatical complexity on word order is due to Behaghel (1909). One of the word order laws formulated by Behaghel is the Gesetz der Wachsenden Glieder (‘Law of increasing constituent size’), which states that when possible, shorter constituents precede longer ones. A similar principle can be found in Haeseryn et al. (1997) for Dutch. Their Complexity Principle states that more complex, heavier or longer constituents have a tendency to appear towards the end of a sentence. For instance, the sentence in (59) is ungrammatical with a demonstrative pronoun. A direct object is not allowed to appear to the right of the adverbs in the Mittelfeld when it is a demonstrative pronoun. This word order is allowed, however, with the more complex demonstrative full NP.

(59)  
\[\text{Ik heb gisteren eindelijk *dat / dat artikel gelezen.}\]

\[\text{I have yesterday finally that that article read}\]

\[\text{‘I finally read that / that article yesterday.’ (judgement from Haeseryn et al., 1997)}\]

Note that the predictions of the complexity principle for (59) overlap with the effect of pronominality on word order that we discussed above. Without making explicit reductionist claims, Haeseryn et al. (1997) suggest that other word order facts could also be understood in this fashion. For instance the order object-before-oblique in Dutch (not discussed here) could be reduced to the complexity principle, because NPs are generally shorter than PPs. See also the discussion of Hawkins (1994), below.
Empirical work on English has shown the influence of grammatical complexity in a large variety of constructions. In each case, the conclusion was that there is a preference to put more complex constituents more towards the end of a phrase or clause. For instance, in the postverbal domain in English, this effect has been found in *heavy NP shift*, particle-verb constructions, order of PP sentence modifiers, and the dative alternation (Hawkins, 1994; Wasow, 2002; Arnold et al., 2004; Bresnan et al., 2007, to name but a few). Bresnan et al. present the examples in (60). The contrast in grammaticality of (60a) and (60b) suggests that the ditransitive *give X the creeps* does not show dative alternation. Only the NP NP construction is allowed. However, in (60c), the NP PP realization is allowed. Bresnan et al. convincingly argue that the NP PP realization in (60c) is allowed because it lets the complex constituent *people whose . . . follow the simpler* the creeps.

(60) 

a. That movie gave [me] [the creeps]
b. *That movie gave [the creeps] [to me].
c. Stories like these must give [the creeps] [to people whose idea of heaven is a world without religion] . . .

The effects of grammatical complexity have also been found in the NP domain, for instance in the genitive alternation (Rosenbach, 2005). In each of these cases, the researchers have found that postponing complex constituents is only one of the tendencies in English word order – there are additional factors besides complexity that influence word order.

In influential work, Hawkins (1994) proposes that many of the claimed word order universals (given before new, definite before indefinite, pronominal before full NP) fall out of the complexity effect. Hawkins constructs an intricate processing account of complexity effects. Summarized very concisely, Hawkins posits that languages prefer to minimize the domain needed to identify all sub-constituents of a constituent during processing. This leads to a preference for putting the heads of sub-constituents as close to the head of the main constituent as possible. In the English postverbal domain, this preference translates into the preference to move longer constituents to the back. However, for a consistently head final language, the account would predict that long constituents are fronted – a prediction that appears to be borne out for Japanese (Hawkins, 1994; Yamashita, 2002). The claim that other word order trends can be reduced to a complexity effect, however, cannot be maintained given the work on English above, and Dutch and German to be discussed below. In more recent work, Hawkins (2004) allows other factors besides a complexity effect.

Existing work on the effect of complexity on word order in German and Dutch deal almost exclusively with Mittelfeld phenomena and the results are not as clear as the English results cited above. Hawkins (1994) argues for the effect of complexity in the German Mittelfeld, and for a complexity effect on relative clause extraposition in German.
and Dutch. As mentioned, Haeseryn et al. (1997) assume a Complexity Principle for Dutch. Corpus evidence for the influence of grammatical complexity on word order in Dutch and German comes from Shannon (2000). Shannon shows that the word order pronominal object before subject is more common as the subjects become longer. Heylen (2005) presents a corpus study showing that constituent length has an effect on certain cases of German scrambling.

In contrast, other corpus studies have found no evidence for a tendency to place complex constituents at the end in the German Mittelfeld (Fanselow, 2000) or have found only a very weak preference dominated by, for instance, definiteness effects (Kurz, 2000b). Unlike the work on the English dative alternation, Van der Beek (2005) concludes that there is no length effect on the order of direct object and indirect object in Dutch.

To the conflicting results in the literature on the effect of complexity on word order in the postverbal domain in German, Dutch and English, we can add uncertainty about what the effect of complexity on topicalization should be. For instance, Birner and Ward (1998, pp24ff) consider topicalization in English to be pragmatically driven, and suggest that complexity plays a minor role at best. Hawkins (2004) allows for the interaction of complexity with other (grammaticalized) performance considerations that favour for instance putting a topic in sentence initial position. However, it is not clear what the effect of complexity on Vorfeld positioning in Dutch should be in Hawkins’ model.

There is only little empirical work on Dutch and German Vorfeld positioning that one could base a hypothesis on. Jansen and Wijnands (2004) present results of a small corpus study of Dutch newspaper articles. They find, amongst other things, that PPs and CPs avoid the Mittelfeld and occur either in the Vorfeld or the Nachfeld; that NPs that contain a relative clause appear mostly in the Nachfeld; and that NPs that contain PP modifiers appear as often in the Vorfeld as NPs without such modifiers. The examples in (61) illustrate the tendency of a CP (in boldface) to avoid the Mittelfeld (61b), but not the Vorfeld (61a) or Nachfeld (61c).

(61) a. **Dat het hoe dan ook gênant wordt,** is al vaak gebleken.
That it anyway embarrassing becomes is often proven
‘It’s been proven many times before that everything ends in embarrassment.’
b. ?Al vaak is **dat het hoe dan ook gênant wordt** gebleken.
c. Al vaak is gebleken **dat het hoe dan ook gênant wordt.**

The behaviour illustrated in (61) suggest that Vorfeld occupants may be more complex than Mittelfeld constituents. This contradicts the general Complexity Principle of Haeseryn et al. (1997) that would predict that Mittelfeld occupants are more complex than Vorfeld occupants because the Mittelfeld comes after the Vorfeld.

Looking at constituent length, rather than category or internal make-up, Jansen and Wijnands find that there is no significant difference in constituent length anywhere in
2.7 Conclusion

In this chapter I have introduced some aspects of word order in the Dutch main clause. I have described the direct preverbal position referred to as the Vorfeld in terms of form and function. I have also demarcated the area of investigation by contrasting Vorfeld occupation and topicalization with other constructions that involve the left periphery, like topic drop, left dislocation, hanging topic, the mirror construction, and the phenomenon of multiple Vorfeld occupation.

I have also presented predictions about Vorfeld occupation in a corpus. These predictions were generated from existing results on Mittelfeld word order in Dutch and German, and postverbal word order in English, on the hypothesis that global word order trends underly these domains and the Vorfeld alike. The relation of these global trends and Vorfeld occupation may be complicated by the status of the Vorfeld as a position for important material.

1. Canonical argument order In the Dutch Mittelfeld, canonical word order has been argued to be subject \( \prec \) indirect object \( \prec \) direct object. On the basis of this order, we expect that subjects have the highest chance of appearing in the Vorfeld, followed by indirect objects. Direct objects should have the lowest chance of occupying the Vorfeld.

2. Definiteness Scrambling over argument and adverb in the Mittelfeld in Dutch and German can be related to a scale pronoun \( \prec \) definite full NPs \( \prec \) indefinite full NPs. Elements on the scale show a stronger tendency to scramble than elements to their right. This leads to the prediction that pronouns have the highest chance of appearing in the
Vorfeld, and indefinite NPs the lowest. The relation between definiteness (NP form) and Vorfeld occupation is also influenced by the first-things-first principle. Because Vorfeld material is preferably important, (reduced) personal pronouns are not good Vorfeld occupants. We thus make conflicting predictions with respect to Vorfeld occupation by personal pronouns.

3. Grammatical complexity  It has been claimed that there is a general preference to order light and simple constituents before heavy or complex material. This trend has not been clearly established for Dutch, but there is solid evidence that complexity plays a role in the English postverbal domain. The prediction that will be investigated is that less complex constituents have a higher chance of appearing in the Vorfeld.

These predictions are independent of each other; that is, we predict to see each of the effects above after controlling for the other two factors.

In addition to the results of investigating the predictions above, I will also present results on the relation between the presence of sentence adverbials and direct-object–initial word order.

I will introduce the spoken Dutch corpus in Chapter 3. This chapter also contains a discussion of the methods and tools used in the investigation. A portion of the chapter is dedicated to the definition of the Vorfeld in terms of the corpus, which does not contain direct annotation for topological fields. The results of the corpus investigation itself are presented in Chapter 4.