Chapter 2

Subject-object ambiguities

1 Introduction

This chapter will first show that Dutch has a number of constructions in which the order of subject and object is at least temporarily ambiguous (Section 2). Next, in Section 3, experimental evidence will be discussed showing that there is a general preference for an interpretation in which the subject precedes the object. This preference for a subject-object order is most likely to be driven by a syntactic generalization that abstracts away from the properties of the specific lexical items and the discourse context. In Section 4, several syntactic accounts for the subject-object preference will be discussed. In Section 5, it will be argued that in addition to this syntactic bias for the subject-object order, other factors may affect word order preferences. It will be shown how the specific properties of the NPs may support or counteract the syntactic bias. Which order is ultimately preferred and the strength of this preference may therefore also depend on the specific properties of the NPs used. Two hypotheses will be formulated that will be tested in the next chapters.

2 Subject-object ambiguities

In a number of clause types in Dutch, the subject can either precede or follow the object. Except for some pronouns, NPs in Dutch are not overtly case-marked. This means that when such clauses are processed from left to right, the order of the subject and object NP is temporarily or even fully ambiguous. For instance, the main clause in (1) can be paraphrased as (2a) corresponding to a subject-object (SO) reading, or as (2b), corresponding to the object-subject (OS) order.\(^1\)

(1) De dichter heeft de boer gegroet.
    the poet has the farmer greeted

(2) a. The poet greeted the farmer. \([SO]\]
    b. The poet, the farmer greeted. \([OS]\]

\(^1\) The object-subject interpretation is rather marked for declaratives. I will return to this in Section 5.2.
Main clause \textit{wh}-questions, embedded \textit{wh}-questions and relative clauses may be ambiguous in the same manner. \textit{Wh}-questions are questions starting with a \textit{wh}-phrase, such as ‘which’-N (Dutch: \textit{welke}-N), ‘who’ (\textit{wie}) or ‘what’ (\textit{wat}). An example of a main clause \textit{wh}-question is given in (3).

(3) Welke dichter heeft de boer gegroet? 
    which poet has the farmer greeted

In (3), either the first NP \textit{welke dichter} or the second NP \textit{de boer} can be the subject of the clause. Paraphrases are given in (4):

(4) a. Which poet greeted the farmer? \textit{[SO]}
b. Which poet did the farmer greet? \textit{[OS]}

\textit{Wh}-questions can also be embedded, as shown in (5). In embedded clauses, the finite verb appears after the two NPs in the final or penultimate position of the clause. However, the order ambiguity remains, cf. (6).

(5) Ik vroeg me af welke dichter de boer heeft gegroet.
    I asked me PART which poet the farmer has greeted

(6) a. I wondered which poet greeted the farmer. \textit{[SO]}
b. I wondered which poet the farmer greeted. \textit{[OS]}

Finally, relative clauses display the same ambiguity. The relative clause in (7) can be interpreted as (8a), in which the head of the relative, \textit{die}, is the subject of the clause; or as (8b) in which the head is the object. Note that also here the verb appears after the two NPs.

(7) De dichter die de boer heeft gegroet...
    the poet that the farmer has greeted

(8) a. the poet who greeted the farmer... \textit{[SO]}
b. the poet who the farmer greeted... \textit{[OS]}

The declarative main clauses, \textit{wh}-questions and relative clauses exemplified above are all fully ambiguous. Clauses may, however, be syntactically disambiguated by case marking on the NPs (pronominal NPs only, in Dutch), or by number agreement between the finite verb and the subject. This disambiguating information need not be present immediately at the first word of the clause, but may come in only later. Thus, NP-V-NP or NP-NP-V clauses in Dutch may be at least temporarily ambiguous between a subject-object and an object-subject interpretation.
3 A subject-object preference

In this section some experimental evidence will be discussed showing a general preference for the reading in which the subject precedes the object, even though both subject-object and object-subject orders are possible. There is also evidence that this preference for a subject-first reading is established rather quickly during on-line processing. In addition, a subject-first reading is preferred in the absence of, or even despite, potentially biasing semantic or contextual information. This suggests that the subject-object order is favored on syntactic grounds.

As only few experiments have been conducted in Dutch, I will also discuss some data from German. German is structurally quite similar to Dutch, and displays the same kind of order ambiguities as those outlined above. The German data are therefore relevant to the present discussion. One difference between Dutch and German is that in German, also non-pronominal NPs are case-marked. Word order ambiguities are therefore somewhat less abundant, as case information can be used to disambiguate the sentence. Another difference, presumably related to the presence of case marking, is that word order in German is even less rigid compared to Dutch. However, despite these differences, word order preferences found for German resemble those found for Dutch.

I will first discuss experimental evidence on main clauses (declaratives and wh-questions) in Section 3.1, and turn to verb-final embedded clauses (relative clauses and wh-clauses) in Section 3.2. For each subtype, Dutch and German data will be discussed separately.

3.1 Main clauses

3.1.1 Dutch data: declaratives

In general, temporarily ambiguous main clauses in Dutch are preferably assigned a subject-object interpretation. Evidence for this preference has been found by Frazier and Flores d’Arcais (1989), and by Lamers (1996). Frazier and Flores d’Arcais conducted an end-of-sentence grammaticality judgment experiment, using declarative main clauses like the ones in (9). For clarity, the point of disambiguation in this and following examples is printed in bold.

(9) a. De patiënt heeft de dokters bezocht.  [SO]
    the patient has-SG the doctors visited

   b. De patiënt hebben de dokters bezocht.  [OS]
    the patient have-PL the doctors visited

The sentences are disambiguated by number agreement between the verb and one of the NPs. In (9), the first NP is singular; the second is plural. In (9a), the finite verb, heeft, is singular, indicating that the first, singular NP is the subject; in (9b),...
the finite verb is plural, *hebben*, which forces the second, plural NP to be the subject of the clause.

Sentences were flashed on the screen, word by word, at a rate of 300 ms per word. At the end of the sentence, subjects had to indicate whether the sentence was grammatical or not. Frazier and Flores d’Arcais found that the subject-object versions were responded to faster and more accurately than the object-subject versions. Apparently, there is a preference for a subject-object interpretation of the clause, leading to processing difficulties when this analysis is not compatible with other information.

The Frazier and Flores d’Arcais experiment does not tell anything about the point at which processing difficulties occur during reading, however, since responses were only recorded at the end of the sentence. More information concerning on-line processing of temporarily ambiguous main clauses is provided by Lamers (1996) using Event-Related Potentials (ERPs). ERPs are obtained by recording the ongoing electrical activity of the brain via scalp electrodes while stimuli are presented. Next, the activity time-locked to the presentation of a certain stimulus is averaged across trials. In the ERPs, several components (positive or negative peaks or waves) can be distinguished. Previous research has shown that sentences containing syntactic violations or syntactically less preferred continuations elicit a positive going wave relative to grammatical or syntactically preferred base-line conditions (Osterhout and Holcomb, 1992; Hagoort, Brown and Groothusen, 1993; Mecklinger, Schriefers, Steinhauer and Friederici, 1995). A positivity with a comparable scalp distribution and latency can thus be interpreted as a sign of syntactic processing difficulty.

Lamers compared subject-object and object-subject main clauses, using case-marked pronouns to disambiguate the structures:

\[(10) \quad a. \ \text{De oude vrouw in de straat verzorgde} \ \textbf{hem} \ \text{elke dag} \quad [SO]
\]
\[
\text{the old woman in the street took-care-of him-ACC every day}
\]

\[b. \ \text{De oude vrouw in de straat verzorgde} \ \textbf{hij} \ \text{elke dag} \quad [OS]
\]
\[
\text{the old woman in the street took-care-of he-NOM every day}
\]

In (10a), the pronoun following the verb bears accusative case which indicates that the pronoun is the object and the sentence-initial NP is the subject of the sentence.\(^2\) In (10b), on the other hand, the pronoun is nominative and therefore the subject; the sentence-initial NP is the object.

At the disambiguating pronoun, Lamers found a number of different positive components for the object-subject relative to the subject-object condition. Some of these components were comparable to the positivities related to processing difficulty that are reported in the literature. The results therefore

\(^2\) In Dutch, the same pronominal form is used for accusative, dative and oblique cases. ‘Accusative’ is therefore equivalent to ‘non-nominative’.
suggest that people have more trouble with object-subject than with subject-object sentences. The difference between the two order conditions became apparent already at the point of disambiguation. This suggests that the first NP is preferred as the subject of the clause by the time the second NP is encountered.

3.1.2 German data: declaratives

Although word order ambiguities are less abundant in German because of case marking, the same preference for a subject-object order has been found in this language. Main clause NP-V-NP declaratives have been tested by Hemforth (1993) and Bayer and Marslen-Wilson (1992). Hemforth (1993) conducted an off-line acceptability rating, an on-line self-paced reading and an on-line grammaticality judgment study on subject-object and object-subject main clauses in German. Clauses were disambiguated by case information; the position of the disambiguating information was varied. The conditions resembling the Lamers stimuli are given in (11).

(11) a. Die kluge Tante besucht den kleinen Jungen.  
   the clever aunt visits the-ACC small boy.  [SO]

b. Die kluge Tante besucht den klugen Onkel.  
   the clever aunt visits the-ACC clever uncle.  [OS]

Here, the first NP is ambiguous between a nominative and an accusative; the sentence is disambiguated only at the determiner following the verb. As in the Dutch studies, a preference for a subject-object order was seen: more judgment errors were made on sentences like (11b) than on (a), and increased grammaticality decision and reading times were seen at the determiner of the second NP for (11b) relative to (11a). The effects did not reach significance in the reading times, however, but this may be due to the fact that only two items were used per condition.

Hemforth (1993) also provides some evidence that the difficulty of the object-subject order does not result only from ambiguity resolution. In one condition, the first NP was unambiguously marked nominative or accusative, e.g. (12):

(12) a. Der kluge Onkel besucht den kleinen Jungen.  
   the-NOM clever uncle visits the-ACC small boy.  [SO]

b. Den klugen Onkel besucht der kleine Junge.  
   the-ACC clever uncle visits the-NOM small boy.  [OS]

Here, longer decision times were seen at the first noun in the object-subject sentences (12b) compared to the subject-object sentences (12a). Reading times showed effects in the same direction. This suggests that the subject-object order is not only preferred in cases of ambiguity, but that this order is easier to process.
per se. The results are interesting in another respect: the preference for the subject-initial reading of main clauses was seen before the verb. This shows that the order preference is not driven by semantic or thematic information provided by the verb (cf. also Bader and Lasser, 1994; Section 4.1.4).

Results from Bayer and Marslen-Wilson (1992) suggest that order preferences are not driven by contextual information either. Object-initial declaratives are rather marked when presented in isolation, and one could claim that this is what may have caused the preference for the subject-object order. Bayer and Marslen-Wilson tested the processing of ambiguous and unambiguous main and embedded clauses as a function of contextual bias. Sentences were preceded either by a neutral or by a biasing context. The biasing contexts were constructed such that the subject of the test sentence could easily receive a contrastive interpretation and the object could refer to given information. Object-initial clauses are quite natural under these conditions. An example of a biasing context is given below.

(13) Neulich gab es einen Brand in der Innenstadt. In der Zeitung stand, dass ein Mann von Feuerwehrmännern aus seiner brennenden Wohnung befreit wurde. Später stellte sich aber das Folgende heraus:

‘Recently, there was a fire in the city. The newspaper reported that a man was rescued from his burning house by some firemen. Later, however, the following turned out:’

This context was followed by a test sentence. Examples of the unambiguous subject-object and object-subject main clause conditions are given in (14a) and (b), respectively:

(14) a. **Der** Hausmeister hat den Mann gerettet.  \[SO\]
    the-NOM janitor has the-ACC man saved.
b. **Den** Mann hat der Hausmeister gerettet.  \[OS\]
    the-ACC man has the-NOM janitor saved.

Thus, in the biasing context conditions, the object of the test sentence, *den Mann* ‘the man’, referred to an entity that had already been introduced. The subject NP in the test sentence, *der Hausmeister* ‘the janitor’ in (14), was new to the discourse, but could be interpreted contrastively with another entity in the previous context, *Feuerwehrmänner* ‘firemen’ in (13). In the neutral conditions, on the other hand, the test sentence did not express a contrast with the preceding discourse. The context thus rendered the object-subject clauses pragmatically rather odd.

The results of this self-paced reading experiment suggest that the processing of the test sentences is hardly affected by the preceding discourse. Reading times showed a preference for a subject-object order, irrespective of the
nature of the context. Apparently, the preference for a subject-object order is not, or at least not primarily, driven by contextual information.

3.1.3 Dutch data: wh-questions

Let us now turn to the processing of main clause wh-questions. Frazier and Flores d’Arcais (1989) and Read, Kraak and Boves (1980) provide some evidence that also these clauses show a subject-object preference.

Frazier and Flores d’Arcais (1989) tested order preferences for wh-questions in the same grammaticality judgment study that was discussed previously for declarative clauses. Together with the declaratives in (9), wh-questions like the ones given in (15) were tested. As in the declarative conditions, number information on the finite verb was used to disambiguate the sentence.

(15) a. Welke arbeiders hebben de voorman geprezen? [SO]
which workers have-PL the foreman praised
b. Welke arbeiders heeft de voorman geprezen? [OS]
which workers has-SG the foreman praised

End-of-sentence decision times for the wh-questions were similar to those for the declaratives: subject-initial clauses were responded to faster than object-initial questions. The size of the difference was comparable to that of the declarative cases. Also with respect to decision accuracy, performance was worse for the object-subject than the subject-object condition. The difference in accuracy between the object-subject and subject-object conditions was smaller for the wh-questions than for the declaratives, however. This suggests that the preference for a subject-object order is somewhat weaker for wh-questions than for declaratives.

A subject-object preference for wh-questions was also found in a study by Read et al. (1980). Results from this experiment suggest that the preference for this order is not driven by prosodic cues, although intonation may modulate the preference. Read et al. tested wie (‘who’)-questions that were structurally completely ambiguous between a subject-object and an object-subject reading. An example of the Read et al. materials is given in (16).

(16) Wie zoent de vrouw? who kisses the woman?

Sentences were presented visually or auditorily. In the auditory conditions, main stress was either on the verb or on the second noun phrase. Subjects read or listened to the sentence, and had to indicate whether they preferred a statement like (17a) or (17b) as the answer to the question.
Note that these response items, too, are ambiguous. However, the data for unambiguous filler questions suggest that the answers such as the ones in (17) were read as subject-object clauses. The data for the ambiguous questions showed an overall preference for a subject-object interpretation, that is, (17a) in response to (16). This preference was strongest when the second NP was stressed (86%), and weakest when the verb was stressed (72%). The visual condition scored in between (78%). The differences between the conditions were statistically significant. The Read et al. data therefore suggest that the subject preferably precedes the object in wh-questions, although intonation appears to affect the strength this preference.

3.1.4 German data: wh-questions

To date, no on-line studies have been carried out for wh-questions in Dutch. A number of on-line experiments have been conducted wh-questions in German. For main clause questions, a preference for a subject-object order has been found, but the robustness of this effect appears to depend on whether case or number information is used to disambiguate the clause.

First, Meng (1995) and Schlesewsky, Fanselow, Kliegl and Krems (to appear) report a subject-object preference when clauses are disambiguated by number information at the verb. Meng used materials like the ones given in (18):

(18)   a. Welche Lehrerin der Stadtschule **hat** die Eltern neulich
       which teacher of-the city school has-**SG** the parents recently
       angerufen? [SO] phoned?

       b. Welche Lehrerin der Stadtschule **haben** die Eltern neulich
       which teacher of-the city school have-**PL** the parents recently
       angerufen? [OS] phoned?

In (18), the first NP is singular, the second plural. In the subject-object version (a) the verb agrees in number with the first NP; in the object-subject version (b) it agrees with the second NP. Longer reading times were found for the object-subject clauses compared to the subject-object ones at and after the disambiguating auxiliary. This indicates that the subject-object order is preferred, and that this preference is established at, or even before, the finite auxiliary. This again shows that the order preference is not driven by semantic and thematic
information: such information is only provided by the lexical verb, *angerufen*, at the end of the clause. If order preferences of structurally ambiguous clauses were primarily based on thematic information, no increase in reading times would have been seen for the object-subject clauses before the lexical verb.

Facts are less clear when case information on the second NP is used to disambiguate the *wh*-clause, cf. e.g. (19) from Schlesewsky *et al.* (to appear).

(19) a. Welche Frau sah **den** Mann am Freitag? 
   which woman saw the-ACC man on Friday 
   \[SO\]

b. Welche Frau sah **der** Mann am Freitag? 
   which woman saw the-NOM man on Friday 
   \[OS\]

Here, the second NP is unambiguously marked accusative in (19a), and nominative in (19b), forcing a subject-object analysis in (a) and an object-subject reading in (b). Schlesewsky *et al.* report no differences in self-paced reading times at or immediately after the point of disambiguation. Reading times were longer for the object-subject clauses only at the end of the sentence. Farke (1994), on the other hand, reports shorter self-paced reading times for object-subject than subject-object sentences (collapsed over all word positions) using materials quite similar to (19). Finally, Meng (1995) reports no processing difficulty for object-subject questions disambiguated by case marking at the second NP. Meng compared sentences such as (20), among others:

(20) a. Welche Bewerberin hat **der** Chef heute früh gesagt, werde die Stelle bekommen? 
   which applicant has the-NOM boss today early said will the job get 
   \[OS\]  
   ‘Which applicant did the boss say early today, will get the job?’

b. Offenbar hat **der** Chef heute früh gesagt, diese Bewerberin werde die Stelle bekommen? 
   apparently has the-NOM boss today early said this applicant will the job get 
   \[-S\]  
   ‘Apparantly, the boss said early today, that this applicant will get this job.’

The sentence in (20a) contains a sentence-initial *wh*-phrase, which is temporarily ambiguous with respect to its syntactic function. The nominative case on the second NP signals that this first NP cannot be the subject of the matrix clause. The sentence in (b), on the other hand, does not contain such an ambiguous *wh*-phrase. No effects of processing difficulty were seen at the nominative NP for the ambiguous relative to the unambiguous condition. This suggests that there is no preference to interpret the *wh*-phrase as the subject of the current clause, or, at
least, that a preference for such a reading can be easily overridden by nominative case on the second NP.

### 3.1.5 Order preferences in main clauses: summary

To summarize the data on main clause declaratives and *wh*-questions: the experimental data on off- and on-line processing show a general tendency to assign a subject-object interpretation to an incoming main clause in both Dutch and German. This preference is established early in the sentence, and is present even if the clause is immediately disambiguated by case information at the first NP. Furthermore, a subject-object order is preferred despite potentially biasing contextual cues for an object-subject order. Prosodic cues and manner of disambiguation (case, number) have been shown to weaken the subject-object preference but, with the exception of Farke (1994), did not lead to a preference for the reverse order.

### 3.2 Embedded clauses

In both Dutch and German, embedded clauses differ from main clauses in the position of the finite verb. In main clauses, the finite verb immediately follows the sentence-initial constituent; in embedded clauses, it appears in the final or penultimate position of the clause, following the subject and object NPs. The data to be discussed below show that the position of the verb does not have any consequences for the preferred order. A subject-object interpretation is favored for relative clauses in both Dutch and German, even despite potentially biasing plausibility information. Dutch experiments on relative clauses will be discussed first. In Section 3.2.2 some German data on relative clauses will be dealt with. Embedded *wh*-questions have been investigated only for German and will be discussed in Section 3.2.3.³

---

³ As opposed to Dutch, German quite naturally allows an object-subject order in embedded declaratives:

(i) weil den Dichter der Bauer gesehen hat... 
   because the-ACC poet the-NOM farmer seen has
   'because the farmer saw the poet...'

However, also for German embedded declaratives, a preference for a subject-object reading has been found (cf. Bader, 1994, 1996; Bayer and Marslen-Wilson, 1992; Scheepers, Hemforth and Konieczny, 1996; Friederici, Steinhauer and Mecklinger, 1995).
3.2.1 Dutch data: relative clauses

Order preferences for relative clauses in Dutch have been investigated by Frazier (1987b) and Van Gompel (1995). Frazier (1987b) conducted a self-paced reading study using sentences containing relative clauses that were either ambiguous or unambiguous. In the unambiguous conditions, the order was disambiguated by number agreement on the clause-final verb, cf. (21).

(21) a. Karl hielp de mijnwerkers die de boswachter vonden. [SO]
   Karl helped the mineworkers who the forester found-PL
b. Karl hielp de mijnwerkers die de boswachter vond. [OS]
   Karl helped the mineworkers who the forester found-SG

In (21) the relative pronoun die refers to the plural head of the relative clause, de mijnwerkers; the second NP, de boswachter is singular. In (a), the verb in the relative clause is plural, signaling that the relative pronoun is the subject of the clause; in (b) the verb is singular, forcing an object-subject interpretation of the relative clause. In the ambiguous conditions in the experiment, the head, the first NP and the verb were all either singular or plural.

Frazier used a self-paced reading paradigm in which the full relative clauses and their heads were displayed in one single frame. Subject relatives were read faster than object relatives, but this effect just failed to reach significance. Subjects were also requested to answer comprehension questions that indicated how they had interpreted the sentence. The responses must be interpreted with caution, as the comprehension questions themselves were ambiguous. Nevertheless, the results are striking. The ambiguous relative clauses were interpreted as subject-object clauses in 74% of the cases. In the unambiguous conditions in which number information forced an object-subject reading, the first NP was incorrectly interpreted as the subject of the clause in 31% of the cases (as compared to 3.7% incorrect responses in the subject-object condition). So, there seems to be a preference to interpret the relative pronoun as the subject of the clause; this preference may even override morphological information that enforces an object-subject order.

A subject-object preference for relative clauses is also found when sentences are presented in a word-by-word fashion. Van Gompel (1995) describes a number of self-paced reading experiments investigating word order preferences in relative clauses. An example of the materials of one of the experiments is given in (22).
Chapter 2

(22) a. Halverwege het studiejaar besloot de professor, die de studenten gesproken heeft, naar het buitenland te gaan. [SO]
‘In the middle of the academic year, the professor, who spoke with the students, decided to go abroad.’

b. Halverwege het studiejaar besloot de professor, die de studenten gesproken hebben, naar het buitenland te gaan. [OS]
‘In the middle of the academic year, the professor, who the students spoke with, decided to go abroad.’

As in the Frazier materials, the relative clauses were disambiguated at the verb by number information. Reading times were longer at the disambiguating verb and the following position for the object-subject compared to the subject-object conditions. This suggests that the relative clause had initially been read as a subject-object clause.

The preference for this order could not have resulted from the semantic information provided by the lexical verb. First, the predicates used in the experiment cited were completely reversible as tested by off-line ratings. For instance, a situation in which a professor is speaking to students was rated to be as plausible as a situation in which students are speaking to a professor. Furthermore, the experiment also included conditions in which the order of the two NPs in the clause (professor and studenten in (22)) was reversed. The subject-object preference therefore cannot be due to a plausibility bias for such an interpretation.

Van Gompel presents additional evidence showing that a subject-object order is preferred independently of plausibility information: a subject-object preference was attested even if an object-subject reading was the most plausible. For instance, in (23a), the participle uitgefloten is semantically incongruent with a subject-object interpretation, but compatible with the object-subject order: a professor is more likely to be catcalled by students than to catcall students. The reverse situation holds for (23b): a professor is more likely to train students than vice versa. The participle opgeleid therefore biases a subject-object interpretation for (23b).
In-the-middle-of the academic-year decided the professor, who the

students *uitgefloten hebben*, naar het buitenland te gaan. *[OS]*

‘In the middle of the academic-year, the professor, who the students catcalled, decided to go abroad.’

b. *Halverwege het studiejaar besloot de professor, die de*

In-the-middle-of the academic-year decided the professor, who the

students *opgeleid heeft*, naar het buitenland te gaan. *[SO]*

‘In the middle of the academic-year, the professor, who trained the students, decided to go abroad.’

Word by word reading times showed the following pattern. An increase in reading times was seen at the participle when it was incongruent with a subject-object interpretation: reading times were longer at *uitgefloten* than at *opgeleid* in (23), although this effect was only significant in the analysis by subjects. This suggests that a subject-object interpretation is preferred even before the participle is encountered: in (a) *de professor* is initially taken as the subject and is interpreted as the agent of *uitgefloten*. At the participle the reading therefore becomes pragmatically implausible, which leads to an increase in reading times at this position. Plausibility information was not used to change the preference for a subject-object interpretation, however: at the disambiguating auxiliary and following word positions, subject-object clauses were read faster than object-subject clauses, irrespective of the nature of the preceding participle. This shows that a subject-object reading is preferred independently of, and even despite, the presence of potentially biasing plausibility information.

### 3.2.2 German data: relative clauses

The same preference for a subject-object interpretation and insensitivity to plausibility information has been found for relative clauses in German. Experiments have been conducted using self-paced reading techniques (Schriefers, Friederici and Kühn, 1995) and ERPs (Mecklinger, Schriefers, Steinhauer and Friederici, 1995; Friederici, Mecklinger, Steinhauer and Meyer, 1996).

In the Schriefers *et al.* (1995) study, number agreement at the auxiliary following the lexical verb was used to disambiguate the clause. An example of their materials is given below. In (24a), the finite auxiliary agrees in number with the first NP, enforcing a subject-object reading; in (b), it agrees with the second
NP, enforcing an object-subject interpretation. The auxiliary was either the sentence-final word or was followed by a subordinate clause.

(24)  
  a. Das ist die Managerin die die Arbeiterinnen gesehen hat... \[SO\]  
      that is the manager that the workers seen has-SG
  b. Das ist die Managerin die die Arbeiterinnen gesehen haben... \[OS\]  
      that is the manager that the workers seen have-PL.

In addition, plausibility was manipulated in the same way as in the Van Gompel study: the verb could either be neutral (24) or introduce a plausibility bias for a particular order (25). The plausibility bias could either be compatible (25a) or incompatible (25b) with the ultimate syntactic resolution of the ambiguity.

(25)  
  a. Das ist die Managerin die die Arbeiterinnen entlassen hat... \[SO\]  
      that is the manager that the workers fired has-SG
  b. Das ist die Managerin die die Arbeiterinnen entlassen haben... \[OS\]  
      that is the manager that the workers fired have-PL.

Results were comparable to those reported in the Van Gompel study on Dutch. Reaction times at the point of disambiguation, and some following positions, were longer for object-subject relatives than for subject-object relatives. The semantic bias introduced by the participle did not significantly affect this preference for the subject-object order.

Mecklinger et al. (1995) conducted an ERP experiment using similar materials. The participle was either semantically neutral, or biased towards the reading enforced by the number agreement. Significant effects of order were obtained only for subjects with fast response times on the comprehension questions following the test sentences. The data suggested, first, that these subjects had a preference for a subject-object interpretation even before the lexical verb was encountered; and second, that plausibility information was not used to change this preference.

First, the N400 component at the participle biasing the subject-object order was reduced relative to the neutral conditions and the object-subject bias condition. The N400 is a negative component with a maximum amplitude at approximately 400 ms after the onset of the word. The amplitude of this component has been argued to reflect the ease with which a word can be semantically integrated into the context (e.g. Kutas and Hillyard, 1984): the easier the integration, the smaller the amplitude. The smaller amplitude found at the subject-object biasing verb thus suggests that the subject-object reading is preferred before the verb: if the first NP is favored as the subject of the clause, a verb biasing a subject-object reading can be integrated into the context more easily than a verb that is semantically neutral or biases an object-subject reading.
Second, the ERP data suggested that the plausibility information does not affect the preference for an object-subject interpretation. At the disambiguating auxiliary, the object-relative clauses elicited an early positive component (P345) relative to the subject relatives. Mecklinger et al. claim that this positivity is possibly related to the P600 that is often found to be reflecting parsing difficulty (Osterhout and Holcomb, 1992; Hagoort et al. 1993). The semantic bias introduced by the preceding verb did not have any effect on the amplitude or latency of this component, however. Confirming the self-paced reading data reported in Schriefers et al. and Van Gompel, the Mecklinger et al. data thus show a preference for a subject-object interpretation of relative clauses, at least for a subset of the participants in the experiment. This subject-object preference appeared not to be affected by plausibility information, even though this information was available and noted before the point of disambiguation.

Another study on relative clauses in German using ERPs is Friederici et al. (1996). In this study the manner and point of disambiguation was varied. Relative clauses were disambiguated either by number marking on the final auxiliary as in (26), by case marking on the relative pronoun (27), or by case marking on the determiner of the second NP (28).

(26) a. Das ist die Direktorin die die Sekretärinnen gesucht hat. [SO]
   This is the director that the secretaries sought has-SG.
   [OS]
   This is the director that the secretaries sought have-PL

(27) a. Das ist der Direktor der die Sekretäre gesucht hat. [SO]
   This is the director that-NOM the secretaries sought has.
   [OS]
   This is the director that-ACC the secretaries sought have.

(28) a. Das sind die Sekretäre die den Direktor gesucht haben. [SO]
   These are the secretaries that the-ACC director sought have.
   [OS]
   These are the secretaries that the-NOM director sought hat.

In all conditions, object-subject clauses showed signs of processing difficulties relative to the subject-object clauses. However, the latency of the effects differed depending on the manner of disambiguation. When the clause was disambiguated by number information (26), subjects with a high working memory span showed an early positivity at the disambiguating verb, replicating the early positivity found by Mecklinger et al. (1995). In addition, a later positivity (P600) was attested. Low span subjects showed no effects of processing effort for the object-subject clauses in their ERPs. However, their performance on comprehension questions following the test sentences was relatively poor for both subject-object and object-subject conditions. This suggests that low span subjects had difficulties...
with either order when sentences were disambiguated by number information at the sentence-final verb.

When case marking was used to disambiguate the clause, both low and high span subjects showed effects of processing difficulty for the object-subject order. The fact that low spans now do show effects suggests that disambiguation by case early in the sentence is somewhat easier than disambiguation by number later in the clause. When the sentences were disambiguated immediately by case marking on the relative pronoun as in (27), a late positivity (P600) was seen at the disambiguating pronoun itself. This is in accordance with the findings by Hemforth (1993) showing that a subject-object order is preferred even if the clause is completely unambiguous. Finally, when sentences were disambiguated by case information at the second NP, cf. (28), no effects were seen at the point of disambiguation. Only at the end of the sentence an early positivity (P350) was seen. This patterns with Schlesewsky et al.’s findings for main clause wh-questions in German: when wh-questions were disambiguated by case information on the second NP, effects were seen only at the end of the clause.

In sum, the Friederici et al. data confirm the previous findings that object-subject clauses are harder to process than subject-object ones, even when they are completely unambiguous. In addition, the manner of disambiguation appears to affect the ease of ambiguity resolution: number disambiguation appears to be more difficult than disambiguation by case, and effects appear to be somewhat delayed when clauses are disambiguated by case at the second NP.

3.2.3 German data: embedded wh-questions

Order preferences in simple embedded wh-questions have not been investigated in Dutch to date. For German, Meng (in preparation) conducted a self-paced reading study and an end-of-sentence grammaticality judgment experiment using materials as given below. Clauses were disambiguated either by number information at the sentence-final auxiliary (29), or by case information at the wh-word (30).

(29) a. Alle waren neugierig zu erfahren welche Politikerin die Minister all were anxious to get to know which female-politician the ministers kritisiert hat. [SO] criticized has-8G

4 There have been a number of studies on long-distance extraction, in which the wh-phrase is the subject or object of an embedded clause but appears sentence-initially (for Dutch: Frazier, 1990b; Jordens, 1991; German: Farke, 1994; Farke and Felix, 1994; Schlesewsky et al., 1996; Italian: De Vincenzi, 1996). Since the present work is concerned with simple extraction only, these data will not be discussed here.
b. Alle waren neugierig zu erfahren welche Politikerin die Minister
all were anxious to get to know which female-politician the ministers
kritisiert haben. [OS]
criticized have-PL

(30) a. Alle waren neugierig zu erfahren welcher Politiker die Minister
all were anxious to get to know which-NOM politician the ministers
kritisiert hat. [SO]
criticized has
b. Alle waren neugierig zu erfahren welchen Politiker die Minister
all were anxious to get to know which-ACC politician the ministers
kritisiert haben. [OS]
criticized have

At the disambiguating auxiliary in (29) reading times were faster for subject-object compared to object-subject clauses. When, however, case information was used to disambiguated the clause, as in (30), no difference was found at the auxiliary. Reading times before the auxiliary were slightly longer for the object-subject clauses when the clause was disambiguated at the wh-phrase (30) than when it was not (29). The interaction of order and manner of disambiguation was not significant, however. These effects were replicated in an end-of-sentence grammaticality judgment task: response times were longer and more errors were made for the object-subject than for subject-object clauses, but only if they were disambiguated by number.

These findings suggest that the manner of disambiguation affects the strength of the order preference: object-subject wh-clauses are harder to process than subject-object clauses. The difference between the two orders is more robust when the clauses are disambiguated by number information at the end of the sentence than by case information earlier on.

3.2.4 Order preferences in embedded clauses: summary

In sum, word order preferences in embedded clauses pattern with order preferences for main clauses: a subject-object interpretation is favored. This preference appears to be unaffected by plausibility information. As in the German main clause wh-questions discussed in Section 3.1.4, the manner of disambiguation appears to have an effect: effects of processing difficulty for object-subject conditions are smaller when case rather than number information is used to disambiguate the clause.
3.3 A syntactic preference

The experimental evidence available to date shows that in both Dutch and German, a subject-object order is preferred to an object-subject, although the strength of this preference may vary. The data suggest furthermore that this preference is syntactic in nature, that is, based on information that abstracts away from the discourse context and the specific lexical items used. First, the preference is found across a number of different clause types: declarative main clauses, main and embedded wh-questions and relative clauses. These clause types differ in the type of the first NP (a definite NP, wh-phrase, relative pronoun) and the position of the verb. An account for the subject-object preference therefore must generalize across these different NP types.

Second, the preference is not driven by plausibility information. A subject-object preference is obtained even when reversibility of the predicate is explicitly controlled for (Van Gompel, 1995), or when plausibility information favors an object-subject reading (Van Gompel, 1995; Schriefers et al., 1995; Mecklinger et al., 1995). A subject-object preference is even shown before potentially biasing information becomes available: when the sentence is disambiguated immediately at the first NP (Hemforth, 1993; Friederici et al., 1996) object-subject clauses are harder to process than subject-object clauses. The subject-object preference is therefore not driven by plausibility information or other semantic-thematic information provided by the verb (cf. also Section 4.1.4).

Finally, the preference appears to be independent of contextual information: a subject-object reading is preferred even if an object-subject clause is a quite natural continuation of the preceding discourse (Bayer and Marslen-Wilson, 1992).

These data suggest that the subject-object order is favored on syntactic grounds. In the next section some syntactic approaches to word order preferences will be discussed.

4 Syntactic accounts

Several syntax-based approaches have been proposed to account for the subject-object preferences in Dutch and German. The common assumptions of all these approaches is that the subject-object order is the most basic order in some sense, and that an analysis corresponding to the base order is the most parsimonious in terms of processing. If the sentence processing mechanism opts for the most parsimonious solution, a subject-object interpretation is initially preferred. When other information (e.g. morphological, semantic, discourse) is incompatible with this interpretation, an increase in processing effort is the result. Theories differ with respect to why a subject-object order is more basic, and why an interpretation corresponding to this base order is the most parsimonious from a processing point of view. I will discuss three different proposals. The first two,
gap-filling and structure building approaches, are non-probabilistic, rule-based approaches. The third is a frequency-based approach.

4.1 Gap-filling approaches

Gap-filling approaches attribute word order preferences to the way in which NPs are related to their syntactic base position. The first and most influential approach has been the Active Filler Strategy (AFS). This principle was first applied to Dutch word order ambiguities by Frazier (1987b). The AFS was originally proposed to account for order preferences in English wh-questions (Crain and Fodor 1985; Tanenhaus, Carlson and Seidenberg, 1985; Stowe, 1986). I will therefore first discuss some cases of word order ambiguities in English, and show how the AFS captures these facts. Next, I will discuss some underlying assumptions concerning Dutch syntax, and show how the subject-object preference in Dutch can be accounted for by the AFS. Finally, some problems with this approach will be discussed.

There are a number of proposals which are related to the AFS in the sense that they account for the subject-object preference in terms of an economy requirement on the relation between the NPs and their functional or thematic base positions; they differ from the AFS in either structural or processing assumptions (e.g. De Vincenzi, 1991a; Gibson, Hickok and Schütze, 1994; Crocker, 1994; Phillips, 1995). These proposals will not be discussed here.

4.1.1 Word order ambiguities in English

Wh-phrases in English are ambiguous in approximately the same manner as sentence-initial NPs in Dutch and German. For instance, which poet in (31) is temporarily ambiguous: it can either be the direct object of the verb as in (31a), or the argument of the following preposition, as in (31b).

(31) I wondered which poet William...
    a. brought to town.
    b. brought a book for.

The syntactic function of the wh-phrase only becomes clear at or after the lexical verb. Syntactically, the wh-phrase is said to have its base position immediately after the verb brought in (31a) and after the preposition for in (b). This base position is explicitly filled in declarative sentences. Consider for instance the declarative equivalent of (31a), given in (32a). Here, the direct object immediately follows the verb. Wh-questions are assumed to be syntactically derived from this base order by moving the wh-phrase to a clause-initial position. This is illustrated for (31a) in (32b). To indicate the relation of the fronted element to the verb, a trace (t) is inserted in the base position. Via this trace, the wh-phrase can be
interpreted as the direct object of the verb, and receive a thematic role (the role of theme in (32)).

(32)  
  a. I know that William brought the poet to town.
  b. I wondered which poet William brought to town.

When a wh-question is processed, word by word from left to right, the wh-phrase needs to be related to a trace in order to receive a syntactic function and thematic role. A wh-phrase is also called a filler, which needs to fill a gap near a verb or preposition. One strategy for relating a filler to its gap is the Active Filler Strategy. Frazier and Flores d’Arcais (1989) give the following definition:

(33)  
Active Filler Strategy
Assign an identified filler as soon as possible; i.e. rank the option of a gap above the option of a lexical NP within the domain of an identified filler.

Briefly put, once a filler has been encountered, a gap is postulated in the first position in which it is syntactically allowed. The parser does this immediately, that is, without waiting to see whether another element is occupying the position instead. The underlying motivation of this strategy is processing efficiency: a filler that has not yet been assigned to a gap imposes a burden on working memory, as it is not yet functionally integrated into the current clause (Miller 1951). The faster a filler is assigned, the sooner the burden is cleared from memory. Evidence in favor of the Active Filler Strategy comes from filled-gap effects. For instance, Stowe (1986) conducted a self-paced reading experiment, contrasting, among others, the following conditions:

(34)  
  a. My brother wanted to know if Ruth would bring us home to Mom for Christmas.
  b. My brother wanted to know who Ruth would bring us home to for Christmas.

The difference between (34a) and (b) is that (b) contains the filler who, which is absent in (a). An increase in reading times was seen at the pronoun us for (b), relative to (a). This suggests that who is initially interpreted as the object of the verb bring. At the next word position this analysis is disconfirmed by the object NP, us. This results in an increase in processing difficulty. This is exactly what would be expected if the parser acts according to the Active Filler Strategy.

4.1.2 Dutch syntax: some underlying assumptions

Now let us see how the Dutch order preferences can be accounted for. The main idea is that in Dutch not only wh--phrases, but every NP starting a main or relative
Subject-object ambiguities

33

clause is a filler, and hence, subject to the AFS. The underlying syntactic assumption is that the base word order of Dutch is subject-object-verb, and that main clauses, relative clauses and wh-questions are derived from this base order (cf. Koster, 1975; Den Besten, 1990; but see Zwart, 1993). This base order is the order seen in embedded declarative clauses:

(35) a. Ik weet dat de dichter de boeren groet. [SO]
    I know that the poet the farmers greets
b. *Ik weet dat de dichter de boeren groeten. [OS]
    I know that the poet the farmers greet

In the embedded clause in (a) the subject *de dichter precedes the object *de boeren; in (b) the object precedes the subject. The sentence in (b) is generally perceived as ungrammatical (denoted by the star *).

The object-subject order in embedded declarative clauses is not totally excluded, but requires special restrictions. First, the object-subject sentence in (b) is rendered acceptable for at least some native speakers if the object is heavily stressed. Furthermore, the subject can be preceded by another NP in passives, cf. (36a), and, at least for some native speakers, in some psych verb predicates, cf. (37a). In both cases, the embedded clause subject is not the agent of the action expressed by the verb. Note, however that the subject-object order can always be used, cf. (36b) and (37b).

(36) a. Ik weet dat de dichter de boeken zijn gegeven. [OS]
    I know that the poet the books have been given
b. Ik weet dat de boeken de dichter zijn gegeven. [SO]
    I know that the books the poet have been given

(37) a. Ik weet dat de dichter de boeken hebben vermaakt. [OS]
    I know that the poet the books have amused
b. Ik weet dat de boeken de dichter hebben vermaakt. [SO]
    I know that the books the poet have amused

The object-subject order is thus far more restrictive than the subject-object one. This observation led to the assumption that the base order for Dutch is subject-object. Dutch is not unique in this respect: Greenberg (1966) noted that the number of languages in which the object-subject order is the most neutral is extremely small compared with the number of languages in which the subject precedes the object in a neutral clause. This, too, supports the idea that there is a subject-object base order, which is perhaps universal.

All clauses are assumed to be syntactically derived from this base order. In the previous section we have discussed how English wh-questions are assumed to be derived from the order seen in declaratives. Similarly, Dutch relative clauses, wh-questions and main clauses are taken to be derived from this subject-
object order in embedded declaratives. Let us first consider relative clauses. In relative clauses the clause-initial NP (the relative pronoun) has been moved from its base position to the clause-initial position, leaving a trace \( t \) in its original position. The syntactic structure of the relative clause in (38) is given in (39). (39a) corresponds to the subject relative interpretation; (39b) to the object relative.

(38) De dichter die de boer groet...
the poet who the farmer greets

(39) a. De dichter \( die \ t \ de boer \) groet \[SO\]
b. De dichter \( die \ de boer t \) groet \[OS\]

In (39a) the \( t \) is in the original subject position, cf. the declarative in (35); in (39b), the trace of \( die \) is in the canonical object position.

Embedded \( wh \)-clauses are derived in the same way. The \( wh \)-phrase is moved from its base position, leaving a trace in the original position. The representation in (40a) corresponds to a reading in which the \( wh \)-phrase is the subject of the clause; (40b) represents an object reading of the \( wh \)-phrase.

(40) I vroeg me af
I wondered
a. \( welke dichter t \ de boer \) groet. \[SO\]
b. \( welke dichter de boer t \) groeten. \[OS\]

Main clauses are also derived from the base order, although the derivation is a bit more complex. In main clauses, the verb appears sentence initially (41a) or after the first constituent (41b). If it appears in another position, the sentence is ungrammatical (41c):

(41) a. Groet de dichter de boeren?
greets the poet the farmers
‘Does the poet greet the farmers?’
b. De dichter groet de boeren.
the poet greets the farmers
c. *De dichter de boeren groet.
the poet the farmers greets

This is in apparent contrast with the order seen in embedded clauses. Here, the only order allowed is the one in which the verb follows the NPs, cf. (42):
Subject-object ambiguities

(42)  a. *Ik weet dat groet de dichter de boeren.
    I know that greets the poet the farmers
b. *Ik weet dat de dichter groet de boeren.
    I know that the poet greets the farmers
c. Ik weet dat de dichter de boeren groet.
    I know that the poet the farmers greets

In (42) the dichter is the subject, the boeren the object. The examples in (42) show that the verb must follow the NPs in embedded clauses, as in (c): any other order yields an ungrammatical sentence. According to some syntactic approaches (e.g. Koster, 1978; Den Besten, 1990), the main clause order is derived from the verb-final order seen in embedded clauses. The verb is therefore assumed to move to a higher position. The result of this movement is shown in (43). Again, a trace t is inserted in the original position of the moved element. For convenience, the trace and the moved element are annotated with a subscript to show that they are related.

(43)    \[ groet_v \[ de\ dichter\ de\ boeren\ t_v \] \]

This movement of the verb yields a yes/no question as in (41a). Subject-initial declarative main clauses like the one in (41b), are obtained by moving the subject to the sentence-initial position. This is illustrated in (44). The trace \( t_i \) is in the canonical subject position.

(44)    \[ De\ dichter,\ groet_v \[ t_i\ de\ boeren\ t_v \] \] \[ SO \]

Alternatively, the object can be fronted, yielding an object-initial declarative, with the trace in the canonical object position:

(45)    \[ De\ boeren,\ groet_v \[ de\ dichter\ t_i\ t_v \] \] \[ OS \]

Main clause wh-questions are derived in the same way, except that the fronted NP is a wh-phrase.\(^5\)

4.1.3 Deriving the subject-object preference

Now let us see how the subject-object preference follows from the AFS assuming these syntactic structures. The sentence processor uses syntactic knowledge to

\(^5\) This analysis of Dutch main clauses is somewhat controversial. According to Travis (1986) and Zwart (1993), for example, sentence-initial (non-wh) subjects do not occupy the same position as sentence-initial objects, cf. also Sections 4.2 and 5.2.3 for alternative analyses.
combine the incoming words to phrases and larger units. According to the syntactic analysis presented above, all clause-initial wh-phrases, relative pronouns and main clause initial NPs in Dutch are fillers: all have been moved from their base positions. This means that the AFS applies. When the sentence is analyzed from left to right, the sentence-initial NP is recognized as a filler, and immediately linked to the first possible gap position. Since the base order in Dutch is subject-object, this first gap position corresponds to the subject position. A subject-object analysis thus is initially preferred. When this analysis cannot be correct, e.g. because the second NP is marked nominative or the verb agrees with the second NP and not with the first, this analysis has to be revised. This leads to an increase in reading times or positivity in the ERPs, as shown in the experiments discussed previously.

The AFS is thus a purely structural strategy: a filler is identified as such because of its (non-base) position in the structure, and which positions are possible gap positions is determined on structural grounds only.

4.1.4 A problem for the AFS

The AFS is meant to be a universal strategy on gap-filling, based on the assumption that gap-filling is primarily guided by syntactic structure configuration. However, there is some experimental data in English that suggests that gap-filling is mainly driven by the properties of the lexical verb. This is in contrast with the Dutch and German data showing that the order preference is not based on verbal information. This suggests that order preference in English wh-questions and order preferences in Dutch and German are not driven by the same strategy.

4.1.4.1 Gap-filling is driven by lexical information

First, gap-filling has been found to be sensitive to transitivity preferences of the verb (cf. also Fodor, 1978; Stowe, Tanenhaus and Carlson, 1991; Clifton, Frazier and Connine, 1984; but see Clifton and Frazier, 1988). For instance, Stowe et al. compared sentences like the ones given in (46) and (47) below. In (46), the verb in the embedded clause, read, is preferably transitive. The filler which book is a plausible object of this verb, whereas which song is not.

(46)  a. The teacher didn’t know which book the students read quietly about.
    b. The teacher didn’t know which song the students read quietly about.

An increase in self-paced reading times was seen at the verb read in (46b) compared to (a). This suggests that the filler was immediately interpreted as the object of the verb, leading to a semantic anomaly at the verb read in (b). However, results were different when the verb was preferably intransitive, such as hurried in (47):
Subject-object ambiguities

(47)  
  a. The physical therapist didn’t know which doctor the orderly hurried quickly toward.
  b. The physical therapist didn’t know which bed the orderly hurried quickly toward.

In (47b), the filler *which doctor* is a plausible object of this verb, whereas *which bed* in (b) is not. According to the AFS, a gap has to be postulated immediately following the verb, as this is the canonical position for a direct object. This should lead to a semantic incongruency effect at the verb for (b) relative to (a). However, Stowe et al. did not find any significant increase in reaction times at the verb in (47b) versus (47a). This suggests that the subcategorization preferences of the verb are taken into account during gap-filling.

Second, there is some evidence that gap-filling can be delayed depending on the properties of the verb and the filler. Boland, Tanenhaus, Garnsey and Carlson (1995) investigated filled-gap effects using a continuous acceptability judgment task. In this task, people have to indicate at each word position whether or not the sentence still is acceptable. In one of their experiments they used materials like the ones in (48).

(48)  
  a. Which child did Mark remind them to watch?
  b. Which movie did Mark remind them to watch?
  c. Samuel asked whether Mark reminded them to watch the child.

The verb *remind* selects both an object NP (them in the example) and an infinitival complement (*to watch*). In (48a) the *wh*-phrase *which child* is a semantically plausible filler of the object role of *remind*, whereas *which movie* (b) is not. However, *which movie* is plausible as the object of the infinitival verb. The item in (c) is a control condition which does not contain any filler element. According to the AFS the *wh*-phrase is initially interpreted as the object of the verb *remind*, as this is structurally the first possible gap position. This will lead to an increase in the number of rejections at the verb *remind* for sentences containing a semantically implausible filler (b), relative to sentences containing a semantically plausible filler (a) or no filler at all (c). In addition, both (a) and (b) are expected to show an increase in the number of rejections relative to (c) at the pronoun *them*. At this position it becomes clear that the filler cannot be the object of *remind*, since this position is already filled.

However, Boland et al. report a different pattern of results. First, no effects of plausibility were seen at the verb at all. Second, clauses such as (48a) were generally judged unacceptable at the pronoun, whereas the (b) and (c) versions showed no such increase in the number of rejections at this position. This suggests that gap-filling can be suspended if the verb subcategorizes for an infinitival complement: rather than interpreting *which movie* as an implausible direct object of *remind*, gap-filling is delayed in the knowledge that an infinitival complement is coming up, which perhaps contains a more suitable gap-position.
This is in opposition to what is predicted by the AFS: according to the AFS, the filler should have been immediately related to the first possible gap position, which is the direct object position of *remind*.

These data therefore suggest that gap-filling in English is not completely driven by structural information and the AFS, but at least as much by the properties of the lexical verb and the nature of the filler.

### 4.1.4.2 Gap-filling is not driven by lexical information

The AFS however remains a possible explanation for the Dutch and German data. In contrast to the English order preferences, order preferences in Dutch and German are not driven by the properties of the verb: a preference for a subject-object order has been shown before the lexical verb (Hemforth, 1993; Friederici *et al.*, 1996). In addition, thematic information at the lexical verb may even be ignored, as is shown by Bader and Lasser (1994). Bader and Lasser compared the following German sentences in a self-paced reading paradigm.

\[ (49) \]

\[ \text{a. daß sie/er nach dem Ergebnis zu fragen tatsächlich erlaubt hat.} \]

\[ \text{that she-AMB/he-NOM for the result to ask indeed permitted has} \]

\[ \text{‘that she/he indeed has given permission to ask for the result.’} \]

\[ \text{b. daß sie/ihn nach dem Ergebnis zu fragen tatsächlich erlaubt worden ist.} \]

\[ \text{that she-AMB/him-ACC for the result to ask indeed permitted been is} \]

\[ \text{‘that permission indeed has been given to ask her/him for the result.’} \]

The clause-initial pronominal form was either the subject of the highest clause (a), or the object of the verb in the more embedded clause, *fragen* (b); the pronoun was either ambiguous between a nominative and an accusative form (*sie*), or unambiguous (*er /*ihn*). In the ambiguous conditions, the correct interpretation became syntactically disambiguated only at the sentence-final auxiliaries. Note that the first lexical verb encountered from left to right is *fragen*, which can assign a thematic role to the first NP. If word order preferences are indeed driven by thematic or other information provided by the lexical verb, the first NP should be interpreted as object of the verb in the embedded clause as in (b) rather than as the subject of the highest clause as in (a). However, the results showed otherwise: in the ambiguous conditions, the clause-final auxiliary *ist* in (b) was read more slowly than *hat* in (a). In the unambiguous control conditions, no difference in reading times was seen between the auxiliaries. This suggests that an ambiguous clause-initial NP is preferably read as the subject of the highest clause possible, and that this preference is not driven by thematic information provided by the first lexical verb.

As opposed to English, then, word order preferences in Dutch and German may be based on syntactic structure configuration. The AFS appears not to be the universal gap-filling strategy as it was originally intended: it applies to
word order ambiguities in Dutch and German, but does not, or at least not in a straightforward way, to the processing of wh-phrases in English.

4.1.5 Summary

Gap-filling approaches such as the AFS attribute the subject-object preference in Dutch and German to the efficiency with which the NPs can be related to their base position. The base order is the one in which the subject precedes the object. In subject-object clauses, the NPs can be more quickly related to their base position and thus be functionally integrated than in object-subject clauses. In cases of structural ambiguity, then, a subject-object interpretation is preferred.

4.2 A phrase structure approach

An alternative syntactic approach is proposed by Gorrell (1996; to appear). According to Gorrell, a subject-object ordering is not motivated in terms of the relation between the NP and its base position, but rather in terms of structural simplicity. Following Travis (1986), and partly, Zwart (1993), Gorrell assumes that the structural representation of subject-object clauses is in some ways a substructure of object-subject clauses. The syntactic representation of a subject-initial relative clause is given in (50a); (50b) represents the structure of an object-initial relative:

(50) a. De dichter [ die de boer groet ] [SO]
    b. De dichter [ die [ de boer t groet ]] [OS]

In the subject-object clause no constituent has been moved. In the object-subject clause, the object has been moved to a position in front of the subject. The subject itself occupies the same position as in the subject-initial clause. Subject-initial clauses are therefore structurally simpler than object-initial ones. The processing preference for a subject-object order now falls out of a parsimony condition on structure building. Assuming that the sentence processing mechanism always opts for the most parsimonious solution, the simplest structure is preferred. Subject-initial clauses involve the least structure. A clause-initial NP is therefore preferably assigned the function of subject. If case or number information signals that this structure is not the correct one, a more complex structure has to be built. Because of this revision and addition of structure, more processing difficulty is predicted for object-subject than for subject-object ordered clauses.

In this account, the subject-object preference does not result from a parsimony strategy relating an NP to its base position, but rather from a parsimony principle with regard to the number of phrase markers to be built (cf. the principle of Minimal Attachment mentioned in Chapter 1, Section 3.1). In this approach, subject-object ambiguities in Dutch and the ambiguities in English wh-questions need not be instances of the same phenomenon (cf. Section 4.1.4).
Whereas under the syntactic analysis assumed, subject-object ambiguities in Dutch always involve a difference in structural complexity between the alternative readings, this need not be the case for ambiguous wh-questions in English. If there is no difference in structural complexity, the structural parsimony principles do not apply. Interpretation preferences in English wh-questions therefore need not result from the same structural principles that account for the subject-object preferences, but may instead be driven by lexically specific information and other principles.

4.3 A frequency-based approach

In the accounts discussed above, the subject-object preference is derived from gap-filling or structure building principles operating on non-probabilistic syntactic knowledge. However, probabilistic accounts of the subject-object preference can be conceived of. According to frequency-based models of parsing the processing mechanism is sensitive to the frequency with which certain constructions occur in the language (cf. e.g. Mitchell, 1994). When confronted with an ambiguity, the parser opts for the solution that is the most frequently occurring.

Although at present no extensive frequency studies have been conducted in Dutch, it is very likely that subject-initial clauses are indeed more frequent than object-initial clauses. Some frequency data for subject- and object-initial main clauses can be found in Nieuwborg (1968). Nieuwborg used a corpus of 5,000 sentences drawn from ten Flemish and ten Dutch novels. The corpus contained 5,215 main clause declaratives; 3,124 (60%) started with a subject NP, whereas only 101 (2%) started with an object NP. When the counts were restricted to subject- and object-initial clauses only, 97% is subject-initial versus 3% object-initial. Although these counts concern main clause declaratives only, there are reasons to assume that the order in which the subject precedes the object is the most frequent in general. First, in embedded declaratives, the subject almost always precedes the object. Exceptions are the (rare) cases of psych verb predicates, passives and special stress, cf. Section 4.1.2. Second, wh-questions and relative clauses, which easily allow an object-subject order, are less frequent than main clause declaratives: main clause wh-questions occur less often than main clause declaratives (3,225 NP-initial declarative main clauses versus 33 NP-initial wh-questions in the Nieuwborg study); and embedded wh- and relative clauses generally occur within a main clause, and, hence, cannot be more frequent than main clauses. It is therefore unlikely that the number of object-initial wh- and relative clauses exceeds the number of subject-initial declaratives. Collapsing over clause types, then, the subject-initial order is likely to be the most frequent in Dutch (cf. also Chapter 5 of this book).

If the sentence processing mechanism is sensitive to frequencies of structures, a subject-object interpretation will be therefore preferred to an object-subject reading. This frequency-based subject-object bias can be implemented in several ways. For instance, comprehenders could make use of a set of prestored
Subject-object ambiguities

structures while analyzing the incoming sentences. This corresponds to Fodor, Bever and Garrett’s (1974) idea of a canonical sentoid structure. These structures may be ranked in terms of frequency, with the most frequent one being the most accessible. Equivalently, rules for generating these structures can be annotated with frequency information (Jurafsky, 1996). The most rules which are used the most frequently will then be applied first.

In a frequency-based approach, then, the subject-object order in Dutch is the more basic as it is the most frequent. This order is the most parsimonious in terms of processing either because a frequently occurring representation is easiest to access or construct, or because such an analysis has a higher chance of yielding the correct interpretation of an incoming clause than analyses that correspond to less frequent structures. A subject-object analysis will therefore be initially preferred for the Dutch ambiguities. When morphological or other information in the sentence is incompatible with this interpretation, this frequent structure is discarded and a more infrequent structure activated. This leads to an increase in processing difficulty.

4.4 Summary

Several syntactic accounts have been proposed to explain the preference for the subject-object order in Dutch and German. All accounts share the assumption that the subject-object order is the more basic one. Gap-filling and structure building approaches are motivated by the linguistic assumption that the subject-object order is the base order, and that all other orders are derived from this order. A subject-object interpretation is more efficient from a processing point of view, either because the NPs can be quickly related to their base position (gap-filling approaches), or because subject-object clauses are structurally less complex (structure building approaches). Instead of, or in addition to, these rule-based accounts, frequency can play a role. In Dutch, subject-object clauses are more frequent than object-subject clauses. Hence, if the processing mechanism is sensitive to this frequency information, a subject-object analysis is preferred to an object-subject reading.

Whatever the exact details of the various approaches, the general preference for a subject-object order can be accounted for in syntactic terms, that is, in terms of principles or generalizations that abstract away from the specific properties of the lexical items involved.

5 Other factors influencing word order preferences

Although a subject-object order is favored on syntactic grounds, the ultimate outcome of the ambiguity resolution process may be also determined by other factors. In the overview in Section 3, we have already seen that intonation and the manner of disambiguation (case or number information) may affect the strength of
the preference. This suggests that other information may play a role in addition to the syntactic bias.

In the present work, one such additional source of information will be investigated, namely the discourse-related information triggered by the type of NPs involved. This information can either support or counteract the syntactic bias for the subject-object order. Investigating word order preferences as a function of the type of NPs will help clarify the relative contribution of syntactic and other sources of information used in ambiguity resolution. Below I will first discuss some evidence from English showing that discourse inferences triggered by the nature of the NP can indeed affect ambiguity resolution, even when sentences are presented in isolation. Next, I will show how the nature of the NPs may affect word order preferences in Dutch.

5.1 Discourse inferences in isolated sentences

There is some evidence from English and other languages that ambiguity resolution is influenced by the discourse-related properties of particular lexical items, even if the experimental sentences are presented in isolation. One example is the influence of the focus operator only on the resolution of reduced relative/main clause ambiguities in English (e.g. Crain, Ni and Conway, 1994; and Ni, Braze, Conway, Crain and Shankweiler, 1996). Examples of the reduced relative/main clause ambiguity have been discussed in Chapter 1, Section 3.1. Another, well-known example of a reduced relative/main clause ambiguity is given in (51); the continuations in (a) and (b) are from Ni et al. (1996):

(51) The horses raced past the barn....
   a. but were unable to clear the jump.
   b. were unable to clear the jump.

The verb raced can either be interpreted as the main clause verb, as in (a), or as a passive participle introducing a reduced relative clause as in (b). There is a strong preference to take a verb like raced as the main clause verb. When a second verb appears in the sentence, such as were in (51b), an increase in processing difficulty is seen. For instance, Ni et al. (1996) reported a large increase in first pass reading times for were compared to but in sentences like (51). Numerous other studies have reported comparable effects (for an overview cf. e.g. MacDonald and Seidenberg, 1994).

Crain et al. and Ni et al. investigated to what extent this strong preference for main clause interpretation could be changed as a function of discourse presuppositions. They investigated this by inserting the focus operator only before the subject NP. The reasoning runs as follows. When used as a modifier of a subject NP, only introduces a contrast between the set denoted by the NP and another set already present in the discourse. Accordingly, when
sentences with *only* are used in isolation, this contrast set must be inferred. For instance in (52a), a set of horses is contrasted with a set of entities that are not horses. The identity of the latter set (the contrast set) is assumed to be known or inferable. A modifier directly following the N in the scope of *only* facilitates the identification of the contrast set. For instance in (52b), the contrast is between horses that were raced past the barn versus horses that were not.

(52)  
   a. Only horses were able to clear the jump.  
   b. Only horses that were raced past the barn were able to clear the jump.

If sentences with *only* are presented in isolated contexts, sentences with an NP modifier should be preferred to sentences without, as the contrast set can be more easily inferred if a modifier is present. The use of *only* should then favor the reduced relative interpretation in the case of a reduced relative/main clause ambiguity: this interpretation is the easiest in terms of discourse requirements. For instance in (53), the continuation in (b) should be preferred to (a).

(53) Only horses raced past the barn....

   a. but were unable to clear the jump.  
   b. were unable to clear the jump.

This is indeed what Ni *et al.* found: significantly longer reading times were obtained for *but* in (a) than for *were* in (b). This is the opposite of what was found for clauses starting with a definite determiner (51). Thus, these results show that the parser is sensitive to discourse-related, lexical information (the difference between *the* and *only*) and that the discourse inferences are drawn immediately, even if the sentences are presented out of context. Furthermore, the results show that a strong preference for a particular reading can be radically changed as a function of discourse information triggered by a particular lexical item.

There is some other evidence that the parsing mechanism is sensitive to specific properties of NPs and that these properties are made use of during ambiguity resolution. For instance, Crain and Steedman (1985) show that a reduced relative reading can be facilitated by using indefinite and generic first NPs rather than definite NPs. Spivey-Knowlton and Sedivy (1995) show that the definiteness of an NP may affect PP-attachment preferences; and De Vincenzi (1991a, 1991b, 1996) shows that *which*-N phrases and *who* phrases, which differ in terms of their discourse properties, are treated differently during the resolution of word order ambiguities in Italian. In all studies mentioned, effects were found when experimental sentences were presented in isolation. The results of these experiments therefore show that discourse information introduced by the use of specific lexical items may influence ambiguity resolution, even in the absence of a preceding context.
On the basis of the evidence cited above, it is likely that the resolution of word order ambiguities in Dutch, too, may be affected by discourse-related properties of the NPs in the clause. In this section I will discuss how the properties of the clause-initial NP may effect the order preference. The potential influence of the type of second NP will be discussed in Section 5.3.

5.2 The nature of the first NP

There are some reasons to expect that the preference for a subject-object order is stronger when the clause-initial NP is a definite NP than when it is a *wh*-phrase. Below I will show that subject-object and object-subject declarative main clauses differ not only in word order but also with respect to the restrictions they impose on the context. This may lead to additional processing difficulties for object-subject declaratives. In *wh*-questions, on the other hand, subject- and object-initial clauses do not differ with respect to discourse requirements. These differences between declaratives and *wh*-questions may also be reflected in their structural representations. The preference for a subject-object interpretation will therefore be stronger for main clause declaratives than for *wh*-questions.

5.2.1 Restrictions on object-initial declaratives

Object-subject declarative main clauses are much more restricted in their use than subject-object declaratives. A subject-object sentence such as (54a) is quite natural when used in isolation. The object-subject sentence in (b), however is not.

(54)  a. De dichter heeft de boeren gegroet.  [SO]  
      the poet has-SG the farmers greeted  
      ‘The poet greeted the farmers.’
   
   b. De dichter hebben de boeren gegroet.  [OS]  
      the poet have-PL the farmers greeted

In general, object-initial declaratives are only used (i) when the fronted element refers to an element that has just been mentioned; or (ii) when the fronted element expresses a contrast (cf. Kooij, 1978).

First, objects can be fronted when the referent has just been mentioned and is very salient in the discourse context. This is shown in (55). The question introduces the poet as a salient entity. This is the referent of the sentence-initial object in (55a-d). The object either is not expressed at all (55a), is a demonstrative pronoun (b), or an epithet with a demonstrative determiner (c) (see also Jansen, 1978). Repetition of the definite NP (d) is possible, but not often used in natural discourse (cf. Gordon, Grosz and Gilliom, 1993; Grosz, Joshi and Weinstein, 1995).
Subject-object ambiguities

(55) Waar is de dichter?
where is the poet
a. ø heb ik niet gezien.
   have I not seen
   ‘I haven’t seen him.’
b. die heb ik niet gezien.
   that have I not seen
c. die idioot heb ik niet gezien
   that idiot have I not seen
d. de dichter heb ik niet gezien
   the poet have I not seen

The referent of the fronted object can also be introduced by a definite NP in left-peripheral position, as in (56). In this case, the real object of the clause is the ‘d’-word *die* rather than the sentence-initial NP (e.g. Koster, 1978):

(56) De dichter, die heb ik niet gezien.
the poet, that have I not seen

‘The farmers did greet the poet, but the knight, they didn’t.’

An object can thus quite naturally appear main clause initially if it refers to an entity that already has just been mentioned, either in the preceding context, or by a left-peripheral NP.

Second, an object can appear sentence initially to express a contrast. For instance, (57) can be uttered in a situation in which the poet was greeted by the farmers, but the knight was not. To indicate this contrast, the first NP receives special stress, represented by the capital letters.

(57) De DICHTER hebben de boeren gegroet, maar de ridder niet.
the poet have the farmers greeted but the knight not

‘The farmers did greet the poet, but the knight, they didn’t.’

Such contrastive expressions are usually not uttered out of the blue, but in a context in which the referent of the stressed NP (*de dichter* in the example) and the entity it is in contrast with (*de ridder*) have been mentioned before or are inferable.

Object-subject declarative main clauses thus only occur in restricted contexts. When an object-subject sentence is presented in isolation, inferences have to be drawn to render the object-subject sentence pragmatically felicitous: either that the referent of the first NP has just been mentioned and is what the current discourse is about, or that the object NP expresses a contrast. In sentences like (54b) these inferences are hard to make, as deictic and contrastive cues on the first NP are lacking.

In contrast, subject-initial declarative main clauses as e.g. (54a) are not much restricted in their occurrence. The referent of a sentence-initial subject may,
but need not be mentioned before, and may, but need not be used contrastively. Subject- and object-initial main clauses therefore differ in more respects than just word order when they are presented in isolation and the first NP is a non-deictic definite NP. This may have consequences for the strength of the order preferences: a subject-object reading for declaratives will not only be preferred on the basis of the syntactic principles addressed in Section 4, but will also be favored for discourse reasons.

5.2.2 Wh-questions

No such discourse bias for the subject-initial order is expected when the first NP is a wh-phrase, as in (58).

(58)  a. Welke dichter heeft de boeren gegroet?  
      Which poet has-SG the farmers greeted  [SO]
      
    b. Welke dichter hebben de boeren gegroet?  
      Which poet have-PL the farmers greeted  [OS]

Both subject-object (58a) and object-subject (58b) wh-questions trigger some discourse inferences. In both cases, the use of a which-N phrase presupposes that a set of N’s, poets in the example, has been mentioned before, or has been introduced into the context in a different way. When the utterance is presented in isolation, the presence of such a set is inferred. Whether the first NP is the object or subject of the clause does not make a difference for the restrictions imposed on the context.

Object- and subject-initial wh-questions therefore do not differ from each other in the way their declarative equivalents do. Both orders impose the same restrictions on the context, and these inferences can easily be drawn. Thus, in a temporarily ambiguous wh-clause, the only basis for a subject-object preference is the syntactic bias discussed in Section 4.

5.2.3 Syntactic differences

The discourse-related differences between subject-object and object-subject declarative main clauses may also be reflected in their syntactic representations. Under some syntactic approaches (e.g. Koster, 1978; Zwart 1993), object-initial declaratives such as (54b) are syntactically more complex than subject-initial declaratives as (54a). In (59) the difference is shown between an object-initial main clause declarative (b) and a neutral subject-initial declarative (a), according to the analysis proposed by Zwart (1993).6

---

6 In (59) and (60), the finite verb (V) has been moved from its base position. The details of this movement are not relevant to the present discussion, however.
Subject-object ambiguities

(59) a. [ subject V object ] [SO]
b. [ object_i ø V [ subject t_i ]] [OS]

According to this analysis, sentence-initial object NP, cf. (59b), is not part of the clause proper, but is in a left-peripheral position. It is related to the clause by means of an empty element ø. The function of this element is equivalent to the die-word in (56). Not the overt NP, but this empty element is the real object of the clause, and has been moved from the object base position. Neutral subject-initial declaratives, on the other hand, do not involve any peripheral NPs or empty elements, cf. (59a). In addition, the sentence-initial subject NP remains in its functional position in Zwart’s analysis, whereas the empty object NP moves to a higher position. Neutral subject-initial declarative main clauses may therefore be syntactically less complex than object-initial declaratives.

Subject- and object-initial wh-questions such as e.g. (58a) and (58b), do not differ to such a large extent. In Zwart’s account both subject and object wh-phrases are in the same structural position. The analysis for subject- and object-initial wh-questions is given in (60a) and (b), respectively.

(60) a. [ wh-subject_i V [ t_i object ] ] [SO]
b. [ wh-object_i V [ subject t_i ] ] [OS]

Both wh-phrases have been moved from their functional base position, represented by the trace t, to a position that is comparable to the one occupied by the empty element in (59b).7

Under this syntactic account, then, object-subject and neutral subject-object declarative main clauses differ structurally in more respects than their wh-clause equivalents do. This reflects the discourse-related differences discussed above.8

5.2.4 Predictions

Object-subject declarative main clauses require special discourse contexts and are possibly structurally more complex than subject-object declaratives. Subject- and object-initial wh-questions do not differ in these respects. The nature of the first NP may therefore affect the way in which temporarily ambiguous main clauses are processed. When the first NP is a definite NP, the subject-object order is the

---

7 Note that in this account, the subject-object preference for wh-questions must be derived from gap-filling rather than structural complexity.

8 In fact, the additional phrase structural difference between subject- and object-initial declaratives may even be motivated by discourse considerations: under certain linguistic assumptions (e.g. Chomsky, 1995), this difference can be claimed to disappear when the sentence-initial NP is contrastively stressed or referred to by a die word.
most parsimonious for several reasons: (i) a subject-object order corresponds to the base order and is therefore less costly for reasons discussed in Section 4; (ii) a subject-first interpretation involves fewer discourse inferences; and (iii), object-initial declaratives may involve the creation of additional structure, at least, under some syntactic accounts. When the first NP is a \(wh\)-phrase, a subject-first interpretation is preferred only on the basis of the structural preference for a subject-object order; the other two factors are not applicable. The subject-object bias may therefore be stronger when the first NP is a neutral definite NP than when it is a \(wh\)-phrase.\(^9\)

There is some evidence that supports this view. First, Frazier and Flores d’Arcais (1989) report that in their end-of-sentence grammaticality judgment task, object-initial \(wh\)-phrases were judged grammatical more often than object-initial declaratives. Furthermore, as opposed to the robust on-line subject-object preferences reported for declaratives in German (Hemforth, 1993), the subject-object preference for \(wh\)-questions is more controversial, at least, when case information is used to disambiguate the structure (Farke, 1994; Meng, 1995; Schlesewsky et al., to appear, see the overview in Section 3.1).

This suggests that not only abstract structural preferences play a role in ambiguity resolution but also factors tied to the nature of the first NP. This hypothesis will be tested in Experiment 1 (Chapter 3) using an on-line reading task.

5.3 The nature of the second NP

In the previous section, the potential effect of the type of the first NP were discussed. In addition, the nature of the second NP may have an effect on order preferences. In previous studies on word order ambiguities in Dutch and German, the first ambiguous NP was often followed by a non-pronominal definite NP. In this section it will be shown that pronouns, non-pronominal definite NPs and indefinite NPs differ with respect to the discourse properties of the elements they refer to. This has consequences for how these NPs are preferably syntactically encoded. Pronouns refer to salient entities that are already present in the discourse and are often encoded as subjects. This preference for a subject position is less strong for definite NPs and indefinite NPs, which need not, or may not, refer to given information. When the second NP of an ambiguous clause is a pronoun, there will be a bias towards an object-subject reading for the entire clause. Definite and indefinite second NPs, in contrast, do not introduce such a strong

\(^9\) A weaker subject-object preference for \(wh\)-clauses relative to declaratives may also be expected on the basis of frequency information. Nieuwborg (1968) reports 16 subject-initial versus 17 object-initial main clauses \(wh\)-questions, whereas the same corpus contained 1014 subject-initial and 38 object-initial main clauses starting with a non-pronominal definite NP.
object-subject bias. Therefore, a weaker subject-object preference or even a preference for the reverse order is expected when the second NP is a pronoun compared to clauses in which the second NP is a non-pronominal definite or indefinite NP.

5.3.1 NPs differ in their discourse properties

Pronouns, non-pronominal definite NPs and indefinite NPs differ with respect to the implied discourse status of the entities they refer to, even no preceding discourse is provided. Consider for instance the sentences in (61):

(61)  a. He was composing a song.
     b. The poet was composing a song.
     c. A poet was composing a song.

The use of the pronoun he in (a) implies that there is a contextually salient entity to which this pronoun can refer. This entity is salient (or: in the focus of the reader/hearer’s attention, cf. Garrod and Sanford, 1982) either because it has just been mentioned in the preceding discourse, or because it is present in the extralinguistic context (textually or situationally evoked; Prince, 1981). The definite NP the poet in (b), in contrast, does not imply the presence of a contextually salient entity. Instead, it can introduce a new entity into the discourse context. The use of a definite NP, however, does imply that its referent is in some sense familiar: the addressee of (61b) is assumed to know which particular poet is meant, or to be able to infer this from given information. Finally, the indefinite NP a poet in (c) does not refer to a given entity, but rather introduces a completely new entity into the discourse. In contrast to the definite the poet, a poet does not specify a unique poet. It is therefore not assumed that the referent is familiar to the reader.10

The difference between pronouns, full definite NPs and indefinite NPs with respect to givenness is made clearer in (62). Here, a potential referent of the NPs has already been introduced in the discourse:

(62)  William was happy.
     a. He was composing a song.
     b. The poet was composing a song.
     c. A poet was composing a song.

The pronoun in (a) can refer to the entity introduced in the first sentence. In fact, this reading is highly preferred. Although the pronoun in (a) may refer to an entity that has not been explicitly mentioned, such a reading is not initially

10 I will not discuss generic uses of indefinite NPs.
pursued. This again shows that a pronoun preferably refers to an entity that has been already been introduced.\textsuperscript{11} The definite NP \textit{the poet} in (b) may refer the person mentioned in the first sentence, but can also introduce a new (though not unfamiliar) entity into the discourse.\textsuperscript{12} The indefinite NP \textit{a poet} in (c), finally, cannot refer to \textit{William}. Instead, the NP must be interpreted as introducing a new entity into the discourse.

Briefly put, the difference between the three types of NP is that pronouns must, definite NPs may, and indefinite NPs generally may not refer to already introduced entities. Pronouns and definite NPs further differ with respect to the saliency of their referents in the context. The referent of a pronoun must be salient in the discourse, whereas the referent of an non-pronominal NP need not be. This is illustrated by an example from Sanford, Garrod, Lucas and Henderson (1983), given in (63):

\begin{itemize}
  \item (63) \begin{enumerate}
    \item John drove to London
    \item a. The car kept breaking down.
    \item b. \#It kept breaking down.
  \end{enumerate}
\end{itemize}

In (63a) the definite NP refers to an entity that has not been explicitly mentioned in the preceding sentence, but can be inferred from the use of the verb \textit{drove}. Pronouns cannot refer to such implicitly presented entities. As indicated by the hash, (b) is pragmatically odd: the dominant interpretation is the one in which the pronoun refers to \textit{London}. This shows that a pronoun is preferably related to explicitly given information.

Additional evidence comes from a study by Purkiss (1978), cited in Sanford and Garrod (1981). In this experiment, subjects were presented with short discourses such as the one below.

\begin{itemize}
  \item (64) The mother picked up \textit{the baby}.
  \item She had been ironing all afternoon.
  \item + She would not be finished for some time.
  \item + She was very tired.
  \item a. \textit{The baby} had been crying nearly all day.
  \item b. \textit{It} had been crying nearly all day.
\end{itemize}

The target sentences are given in (a) and (b). The definite NP (a) or pronoun (b) referred to an antecedent in the first sentence, \textit{the baby} in (64). Either one or three sentences (including the ones preceded by the plus signs) could intervene between the first and the target sentence. The target antecedent was not referred to

\textsuperscript{11} This tendency is so strong that a pronoun is related to an antecedent even when gender features do not allow this, cf. Osterhout and Mobley (1995).

\textsuperscript{12} I will only consider referential uses of NPs.
Subject-object ambiguities

in these intervening sentences. Instead, another entity, the mother in (64), was highlighted. This rendered the target antecedent less salient. When only one sentence intervened, target sentences containing a pronoun (it) were read marginally faster than ones containing a definite NP (the baby). However, when three sentences separated the target sentence from the antecedent, the pronoun conditions were read more slowly (at least, when the antecedent was introduced as the object of the first sentence as in (64)). This example shows first, that pronouns can only felicitously refer to salient entities; and second, that definite NPs can be used to refer to given elements even if they are not or no longer salient.

In sum, pronouns are used to refer to entities that have already given and are highly salient in the discourse. Non-pronominal definite NPs either introduce new entities, or refer to given entities. Indefinite NPs, finally, are generally used to introduce new entities into the discourse. These observations have been confirmed by several investigators (among others, for production: Karmiloff-Smith, 1980; Marslen-Wilson, Levy and Tyler, 1982; corpus studies: Prince, 1981; Ariel, 1990; comprehension: Van Dijk and Kintsch, 1983; Prince, 1982; Garrod and Sanford, 1982, 1985; Givón, 1983 (ed), 1984; Sanford and Garrod, 1985).

5.3.2 Givenness and subjecthood

These differences with respect to givenness and discourse saliency have consequences for the way pronouns, full definite and indefinite NPs are preferably syntactically encoded.

In natural texts, NPs that refer to given information are relatively more often encoded as subjects than NPs that do not. For instance, in the 56-clause text investigated in Prince (1982), 24 out of 39 NPs (62%) denoting entities that were given in the discourse were subjects. NPs that introduced new entities into the discourse only occupied the subject position in 8 out of 49 cases (16%). In addition to givenness, discourse saliency played a role. Prince (1982) made a distinction between pronominal and non-pronominal forms within the group of elements referring to discourse-old entities. As has been discussed in the previous section, pronominal forms always refer to entities that are salient in the discourse. In 13 out of the 16 cases (81%) in which a pronoun was used to refer to a given entity in the discourse, the pronominal form was the subject; in contrast, only 11 of the 23 (48%) non-pronominal forms that referred to given entities occupied the subject position. The subject to object ratio was even smaller (16 out of 41, 39%) when this class included NPs whose referents could be inferred on the basis of the preceding discourse (as e.g. in (63)). In natural texts, then, information that is given in the discourse is preferably encoded as a subject. This is especially so when this information is highly salient and referred to by a pronoun.

Why should NPs that refer to given information often be encoded as subjects? One reason is that sentences usually express statements about given
entities. Across languages, who or what the sentence is a statement about (the topic), is usually encoded in the subject position (cf. e.g. Keenan, 1976; Reinhart, 1982). This has also been shown in production studies. For instance, Bates and Devescovi (1989) presented subjects with a short scene in which, for example, a bear licked a tiger. Next, the experimenter uttered the request in (65), identifying either the bear or the tiger as the topic. In the subjects’ reply, the topic was more often encoded as the subject than as the object of the sentence: when the bear was probed, subjects always responded with sentences as in (66a); when the tiger was identified as the topic, subjects generally responded with passive sentences as in (66b).

(65) Tell me about the bear / tiger.
(66) a. The bear licked the tiger.
     b. The tiger was licked by the bear.

This shows that entities which the sentence is a statement about are generally encoded as the subject. In running text, sentences often are statements about entities that have already been introduced into the discourse: something new is said about something old. NPs that refer to given entities will thus end up in the subject position more often than NPs that introduce new entities (cf. also Yekovich, Walker and Blackman, 1979, for comprehension data).

The Prince (1982) study shows that not only givenness, but also saliency plays a role: elements that refer to salient entities in the discourse (pronouns) were even more frequently encoded as the subject. This can be accounted for along the same lines. Elements that are currently salient in the discourse will most likely be who or what the local discourse is about. Currently salient entities (referred to by a pronoun) will therefore often be the subject of one or more subsequent sentences, as e.g. in (64) (cf. Gordon, Grosz and Gilliom, 1993; Gordon and Chan, 1995).

As we have seen in the previous section, different formal types of NPs are used depending on the discourse status of their referent. The formal type of the NP may therefore have consequences for the likelihood with which the NP is encoded as the subject of the clause. Pronouns are likely to appear as the subject of the clause. Indefinite NPs generally introduce new entities and may frequently appear in a non-subject position. Non-pronominal definite NPs can either refer to given entities or introduce new, though not completely unfamiliar entities into the discourse. The frequency with which a definite NP appears as the subject will therefore be in between the subject frequency for pronouns and indefinite NPs. The numbers provided by Prince (1982) suggest that this is indeed the case. Independently of discourse status, non-pronominal definite NPs occupy the subject position in 15 out of 57 cases (26%); indefinite NPs only in 3 out of 31 cases (10%). As has been mentioned before, pronominal elements are the subject in 13
Pronouns, full definite NPs and indefinite NPs therefore form a hierarchy with relation to subjecthood: pronouns are the most likely to be encoded as subjects, and indefinites the least likely (cf. Givón’s scale of topic accessibility, Givón (1983)).

Data from Nieuwborg (1968) suggest that this hierarchy is also valid for Dutch. In Table 2.1 the number of main clause subject and direct object occurrences are given for various types of NP in the Nieuwborg corpus. The numbers are pooled over all types of predicates (intransitive, transitive, etc.).

As one can see, pronouns more frequently occupy the subject than the object position of a main clause. The relative frequency of subject use is especially striking for personal pronouns. The difference between the number of subject and object occurrences is much smaller for the definite NPs, although these, too, occur more frequently as the subject than as the object of the clauses. Indefinite NPs occur more often as the object than as the subject of main clauses. One should bear in mind that the Nieuwborg data only concern subjects and direct objects;

---

Prince does not regard the three types of NP as varying along one dimension, however, but distinguishes NPs either in terms of their (formal) definiteness (thus collapsing pronominal and non-pronominal definite NPs), or in terms of the discourse status of their referents in the text (i.e. new, old, old-plus-salient). Subjecthood was best explained by the latter factor.
other non-subject functions (indirect object, object of preposition) have not been taken into account. The relative frequency of the subject occurrences may thus be somewhat overstated.

In sum, due to their givenness properties and discourse saliency, pronouns are often encoded as the subject. This preference is less strong for definite NPs. Indefinite NPs, finally, often appear in a non-subject position.

5.3.3 Predictions

Now let us return to word order ambiguities and how the type of NP in second position may influence the resolution of this ambiguity. First, consider pronouns. If a pronoun occupies the second NP position of a temporarily ambiguous clause, the discourse-related properties bias towards a subject reading of the pronoun itself, and hence, towards an object-subject interpretation of the clause. This bias runs counter to the syntactic bias for a subject-object order already discussed. Such a bias is expected even when a preceding context is lacking. When the referent of the pronoun is not explicitly provided, as in the example in (61a), the presence of such an entity is inferred. Hence, even if the preference for the subject position is not tied to the formal properties of the NP, but is instead related to the properties of the NP referent, a subject preference will still be expected for pronominal NPs.

When the second NP is a definite NP, on the other hand, the subject bias for this NP is somewhat weaker. It may even be absent if the sentences are not presented as part of a running text: in this case, the definite NP cannot refer to a given entity, and is interpreted as introducing a new (though not unfamiliar) entity into the discourse. Although this entity is assumed to be in some sense familiar, it is not assumed that this entity is already present in the discourse (as is the case with pronominal NPs). The syntactic subject-object bias will thus receive little or no competition in the case of non-pronominal definite NPs.

Finally, when the second NP is indefinite, an object-subject interpretation is not favored at all. Instead, an indefinite second NP may even favor an object interpretation for itself and support the syntactic bias for the subject-object order. Therefore, if the nature of the second NP indeed affects ambiguity resolution, word order preferences will differ depending on which kind of NP follows the first ambiguous phrase: clauses with an indefinite second NP will show the strongest subject-object preference; clauses with a pronoun will show the weakest preference for this order, or even a preference for the reverse order.

Whether the subject-object preference can be overridden also depends on the properties of the first NP, however. For instance, if the first and the second NP both are pronouns, both NPs favor a subject reading on the basis of their discourse-related properties. The first pronoun biases a subject-object reading, the second an object-subject reading. The two competing pronominal biases are then canceled out. In this case, order preferences will appear to be determined by syntactic and other biases only.
Furthermore, as discussed in Section 5.2, if the first NP is a non-deictic definite NP, an object-subject reading may be very hard to obtain. The discourse-related properties of a definite first NP may so strongly favor a subject-object interpretation that the discourse-related properties of the second NP can hardly influence the order preference. Data from the Lamers (1996) experiment suggest that this is indeed the case. Lamers tested main clause subject-object and object-subject declaratives, using a case-marked pronoun as second NP to disambiguate the clause. Yet, the ERP data suggested that object-subject clauses (in which the pronoun was the subject) were still more difficult to process than subject-object clauses (in which the pronoun was the object).

The effects of the first and second NP may therefore interact. In the present work, the impact of the second NP will be investigated only for one type of first NP, namely a *wh*-phrase. As has been discussed in Section 5.2, subject-object and object-subject *wh*-clauses do not differ with regard to the discourse-related properties of the first NP. *Wh*-clauses are therefore more suitable than declaratives to study the effect of the second NP on order preferences.

### 5.3.4 Evidence

There is some evidence that the nature of the second NP indeed affects order preferences in *wh*-clauses. De Vincenzi (1991b) reports an effect of a definite versus indefinite second NP on the interpretation of *wh*-questions in Italian. Similarly to German and Dutch, *wh*-questions in Italian can be ambiguous between a subject-object and an object-subject reading. For instance, in (67a) either the * quale* (‘which’) NP or the second NP *il ragazzo* can be the subject of the clause.

(67) a. Quale amico ha chiamato il ragazzo?
    which friend has called the boy

b. Quale amico ha chiamato un ragazzo?
    which friend has called a boy

De Vincenzi investigated the interpretation of ambiguous *wh*-questions in Italian in an off-line questionnaire. The second NP could either be definite (66a) or indefinite (b). A general preference for a subject-object reading was found. However, the type of second NP had an effect: in *which-N* questions the *wh*-phrase was taken as the subject in 65% of the cases when the second NP was definite, against 89% when it was indefinite. This difference was significant. For *who* (*chi*) questions, the numbers were 89% versus 97% for definite and indefinite NPs, a non-significant difference. These data show that the nature of the second NP can indeed affect the order preference in the expected direction: the subject-object preference was weaker when the second NP was definite than when it was indefinite. However, the strength of the effect depended on the nature of the first NP.
Indirect evidence that the discourse status of the NPs plays a role in determining word order preferences is the effect of intonation reported by Read et al. (1980). Read et al. found that the subject-object preference for ambiguous who-questions in Dutch was strongest when the second (non-pronominal definite) NP was stressed (cf. Section 3.1.3). In general, a constituent (re)introducing new information (the focus of the sentence) is stressed (Cinque, 1995; Reinhart, 1995). The Read et al. data thus support the view that an NP introducing new information is preferably interpreted as the object of the sentence, and that this may affect the interpretation of ambiguous wh-questions.

In Chapters 4 and 5, the effect of a pronoun versus a definite second NP on the resolution of word order ambiguities in Dutch will be investigated. The pronoun condition is particularly interesting as it allows us to investigate the relative strength of the syntactic and discourse bias: if the discourse-related properties of the second NP have a stronger effect on ambiguity resolution than the syntactic bias for the subject-object order, a preference for an object-subject order is expected for clauses containing a pronoun as second NP. If the syntactic bias is strongest, however, the subject-object order will dominate.

The effects of using an indefinite NP in second position will not be further addressed in this thesis. For a comparison between clauses containing definite and indefinite NPs, see Kaan (to appear).

5.4 Summary

Although there is a general preference for a subject-object resolution of word ambiguities in Dutch, there are reasons to expect that this order preference may be influenced by the discourse-related properties of the NPs involved.

First, if the first NP is a definite NP, more restrictions are imposed on the preceding discourse when it is the object than when it is the subject of the clause. If the first NP is a wh-phrase, on the other hand, the subject- and object-initial reading are equally parsimonious in terms of discourse inferences. These discourse-related differences may also be reflected in the syntactic structure of main clause declaratives and wh-questions: object-subject declaratives may involve a more complex structure than their subject-object equivalents. This structural difference is lacking for subject-object and object-subject wh-questions. Hence, the syntactic subject-object bias will be additionally supported if the first NP is a definite NP, but not if it is a wh-phrase.

Second, word order preferences may be affected by the type of second NP. Due to their discourse-related properties, pronouns have a heavy bias towards occupying the subject position. This bias is less strong for definite NPs. The use of a pronoun as a second NP will therefore bias an object-subject reading of the clause. This may compete with the subject-object order favored on structural grounds. A smaller preference for a subject-object order, or even a preference for the reverse order is thus expected if a pronoun rather than a full definite NP is used as a second NP.
6 Hypotheses and predictions

Whether the nature of the NPs affects order preferences in Dutch, and if so, how, depends on the way the sentence processing mechanism makes use of syntactic and discourse-related information. For ease of exposition two hypotheses will be formulated, representing two rather extreme positions with respect to this issue: (i) the Syntactic Hypothesis, according to which precedence is given to syntactic biases; and (ii) the Discourse Hypothesis according to which discourse-related information mediated by the properties of the NPs is immediately made use of and competes with the syntactic bias.

6.1 The Syntactic Hypothesis

The Syntactic Hypothesis states that in ambiguity resolution precedence is given to syntactic biases, that is, the initial analysis takes place in abstraction from the specific properties of the NPs and Vs. I will take the Syntactic Hypothesis to represent an extreme view of syntax-first processing, according to which the sentence processor is initially insensitive semantic or discourse-related properties of the lexical items, even if these properties have consequences for the structural representation of the sentence. Examples of such properties are subcategorization properties of the verb (cf. Forster, 1979; Frazier, 1987a; Frazier, 1989), and, as I will assume in the following, the difference between a wh-phrase and a definite NP in Dutch. According to the extreme Syntactic Hypothesis assumed here, such lexically specific information may affect the parsing process only later.

(68) Syntactic Hypothesis

Syntactic information takes precedence; other (lexically specific, semantic, discourse, etc.) information may only have a relatively late effect on ambiguity resolution.

This predicts that across-the-board, a subject-object order is preferred for Dutch, as it is the syntactically simplest and/or most frequent reading. More specifically, this means that:

• Initially, a subject-object order should be preferred to an object-subject order.

• The strength of this preference is independent of whether the first NP is a definite NP or a wh-phrase: main clause declaratives and wh-questions will show an equally strong subject-object preference.
• The strength of this preference is independent of whether the second NP is a non-pronominal definite NP or a pronoun: *wh*-clauses containing a pronoun as second NP and *wh*-clauses containing a definite NP will both show a preference for a subject-object interpretation.

• If the discourse-related properties of the NPs have an effect at all, this effect will become apparent only relatively late.

6.2 The Discourse Hypothesis

According to the Discourse Hypothesis, no priority is given to syntactic biases. Rather, discourse-related information takes precedence if available.

(69) **Discourse Hypothesis**

If available, discourse-related information takes precedence over syntactic preferences in ambiguity resolution and does so immediately.

According to this hypothesis, word order preferences are not only syntactically driven, but are also, and even to a larger extent, determined by the nature of the NPs involved and the restrictions imposed by the discourse context.

Below only lexically mediated discourse-related information will be investigated using sentences presented in isolation. The impact of an explicitly given, preceding context on the order preferences will not be studied in the present work.

The predictions of this second hypothesis with respect to the constructions investigated here, are the following:

• There is a general subject-object preference, but

• the strength of this preference depends on the nature of the first NP: main clauses headed by a non-deictic definite NP are expected to show a stronger subject-object preference than main clauses headed by a *wh*-phrase.

• The direction of the order preference depends on the nature of the second NP: *wh*-questions containing a pronoun are predicted to show a preference for an object-subject order; whereas no such preference should be seen if a definite NP follows the *wh*-phrase.

• The discourse-related properties of an NP have an effect on the order preferences as soon as the NP is encountered.
6.3 Remarks

The reader should bear in mind that although the first hypothesis represents a syntax-first approach, and the second a more interactive, they do not correspond to the predictions made by current theories of sentence processing. The Syntactic Hypothesis represents the view that priority is assigned to syntactic information. This can either be formulated in terms of temporal priority (garden-path models) or in terms of weight of the constraints (constraint-based models). Similarly, the predictions of the Discourse Hypothesis can be captured either by interactive theories in which a larger weight is assigned to non-syntactic than to syntactic information, or by syntax-first theories that assume that the ease of syntactic reanalysis depends on the availability of non-syntactic information.

The two hypotheses are formulated for expository purposes only. The aim of the present research is to see which bias is initially strongest: the syntactically based subject-object bias (the view represented by the first hypothesis), or the lexically-mediated discourse bias (the view represented by the second). These hypotheses will be tested in the next chapters.

7 Summary

Most transitive main clauses, relative and embedded *wh*-clauses in Dutch are at least temporarily ambiguous. The subject can either precede or follow the object NP. With only a few exceptions, the experimental evidence to date shows a general preference for a subject-object order. This preference is seen even in the presence of plausibility or contextually information favoring an object-subject reading. This suggests that this bias is syntactic in nature, that is, due to generalizations or principles that abstract away from the context and the details of the lexical items used.

Accounts for this preference have mainly come from rule-based, syntax-first approaches to processing. All proposals share the assumption that the base order of Dutch is subject-object. The processing preference is then derived either by economy constraints on relating an NP to its functional or thematic base position (gap-filling approaches) or by economy constraints on structure building (structure building approaches). In addition to, or instead of, such processing strategies, the frequency of the subject-object order may play a role in the resolution of word order ambiguities.

Although the subject-object preference appears to be primarily syntactic in nature, other factors may enhance or oppose this preference. This work will investigate the potential influence of discourse-related properties of the NPs used. First, the nature of the first NP may have an effect on the order preference. A definite NP object can appear sentence-initially only in specific discourse contexts; subject-object declaratives are less restricted in their occurrence. This difference between the two orders may also be reflected in the syntactic representation of these sentences. On the other hand, subject-object and object-
subject clauses starting with a *wh*-phrase, do not differ with respect to the restrictions they impose on the discourse context. The syntactic subject-object preference may therefore be enhanced when the first NP is a (non-deictic) definite NP than when it is a *wh*-phrase.

Furthermore, the nature of the second NP may play a role in the resolution of word order ambiguities. There is a tendency to encode given, salient entities as the subject of the clause. Pronouns generally refer to such entities. If the sentence is presented in isolation, the presence of such a salient entity may be inferred. The use of a pronoun as the second NP in an ambiguous clause may therefore bias an object-subject interpretation. This may compete with the syntactic subject-object preference. Definite NPs, on the other hand, do not bias towards a subject reading to such a strong extent, especially not if the sentence is presented in isolation and the NP is interpreted as introducing a new (though not unfamiliar) entity into the discourse. The syntactically based subject-object order is therefore expected to be more prominent when the second NP is a definite NP than when the second NP is a pronoun.

Two hypotheses were formulated for expository reasons, representing two opposite views on the relation between syntactic biases and the discourse-related properties of the NPs. According to the Syntactic Hypothesis, the properties of the NPs will hardly have an effect on order preferences. If they do, this effect will only appear rather late in the process. According to the Discourse Hypothesis, the specific properties of the NPs may affect the order preferences right away. In Chapter 3, the contribution of the first NP is investigated using Dutch main clauses; in Chapter 4, the effect of the second NP is tested in embedded *wh*-questions. Chapter 5 is a corpus study providing additional data on order preferences in *wh*-questions and the type of the second NP.