

Subject first or topic first

A corpus study of the position of bare
and definite plural subjects in Dutch

Jorrig Vogels

Subject first or topic first

A corpus study of the position of bare
and definite plural subjects in Dutch

Jorrig Vogels

Optimal Communication
Department of Linguistics
Radboud University Nijmegen
P.O. Box 9103
6500 HD Nijmegen
The Netherlands
www.ru.nl/pionier
JorrigVogels@student.ru.nl

Subject first or topic first

A corpus study of the position of bare
and definite plural subjects in Dutch

Master's thesis

General Linguistics

Faculty of Arts

Radboud University Nijmegen

September 2009

Jorrig Vogels

0408751

First supervisor: Geertje van Bergen

Second supervisor: Helen de Hoop

Acknowledgements

It was a struggle. An educational, interesting and from time to time even fun struggle, but a struggle, up till coming up with a nice title. When I had completed my Bachelor's thesis *The placement of bare plural subjects in Dutch* almost a year after I started it, many questions remained about bare plural subjects in natural language, since I had only looked at tightly restricted experimental data. Therefore, a follow-up corpus study seemed like a logical next step for my Master's thesis. I would also save time, because I would not have to dig into a completely new topic, and no experiments had to be set up. But never again will I dare to think that conducting a corpus study is less of an effort than conducting a production experiment. Admittedly, although it certainly was a lot of work, my own perfectionism and stubbornness in wanting to reinvent all wheels myself largely contributed to the fact that it took me seventeen months to complete this thesis. It is for sure, however, that it would have taken a lot longer without the tremendous help I got from Geertje van Bergen, for which I am very grateful. She taught me how to work with the Corpus Gesproken Nederlands (Corpus of Spoken Dutch, CGN) and with the statistical software program R, she helped me a great deal with the collection and annotation of the data and with the conduction of the statistical analyses, and also in the writing process her advice was of great support. Even after she heard the news that her father was seriously ill, she continued to help me successfully round off this thesis, which I appreciate very much. I wish Geertje and her family every strength.

I would like to thank Helen de Hoop for sketching the first outlines of my thesis and for reading and assessing the final version in the little spare time she has. I am also glad that she hired me as a student assistant, which gave me a welcome change in activities besides working on

my thesis. I would also like to thank Monique Lamers, without whom I could never have written this thesis in the first place. I am further grateful to Steven Westelaken for being the second annotator of the corpus data and to Roeland van Hout for giving his approval of the statistical techniques. Finally, I would like to thank my friends, relatives and colleagues both within and outside the OC group for their support and useful comments, with a special mention of Peia Prawiro-Atmodjo for cheering me up when I was down, and of Margot van den Berg for being a pleasant roommate and for assuring me that an MA thesis is not the most important aspect of a scientific career.

Table of contents

Acknowledgements	vii
Table of contents	ix
1 Introduction	1
2 The placement of bare and definite plural subjects in Dutch	7
2.1 Introduction	7
2.2 Subjects, topics and prominence	9
2.3 Indefinites	12
2.4 Bare plurals	15
2.5 Definites	20
2.6 Conflicting preferences	22
2.6.1 Prominence scales	22
2.6.2 Existential, generic and “third” readings of bare plural subjects	23
2.6.3 The role of the predicate	27
2.6.4 Definite plural subjects	33
2.7 Conclusion	35
3 Definiteness as a cue for word order: a production experiment	37
3.1 Introduction	37
3.2 A drag-and-drop production experiment	38
3.2.1 Methods	38
3.2.2 Results	40
3.2.3 Discussion	41
3.3 Conclusion	43
4 Prominence and Topicality	45
4.1 Introduction	45
4.2 Topics and prominence revisited	46
4.3 Topicalization of other constituents than the subject	49
4.4 Other factors contributing to prominence	51

4.5 Conclusion	54
5 There is more than definiteness: a corpus study	57
5.1 Introduction	57
5.2 The factors	59
5.2.1 Sentences with bare plural subjects	60
5.2.2 Sentences with definite plural subjects	70
5.3. Methods	73
5.3.1 Materials	73
5.3.2. Procedure	76
5.4 Sentences with a bare plural subject	78
5.4.1 Data exploration per factor	78
5.4.2 Multifactorial analysis	86
5.4.3 Discussion	89
5.5 Sentences with a definite plural subject	92
5.5.1 Data exploration	92
5.5.2 Multifactorial analysis	99
5.5.3 Discussion	100
5.6 Conclusion	103
6 Dutch word order variation: an analysis	105
6.1 Introduction	105
6.2 Bare plural subjects versus definite plural subjects	106
6.2.1 Topicalization of the PP	108
6.2.2 Properties of the subject	109
6.2.3 More conflicting preferences	115
6.3 Context	121
6.4 Conclusion	124
7 Conclusion	125
References	129

Introduction

An important tendency in the majority of the languages of the world is the preference to start a sentence with the subject. Other constituents, such as direct objects, do not generally take up the first position in the sentence. Cross-linguistically, SVO and SOV are by far the most frequent canonical word order patterns (e.g. Tomlin, 1986; Comrie, 1989; Dryer, 2008). Starting a sentence with the subject can therefore be considered a universal syntactic preference. Still, many cases can be found in languages in which the subject does not occupy the sentence-initial position. English, for example, is a language with a rather strict word order: a sentence almost always starts with the subject. In some constructions, however, the subject can appear in a position behind the finite verb, as is illustrated in (1) below.

- (1) a. There is *something fishy* going on here.
b. Down the stairs came *John*.

Sentence (1a) is an example of an existential construction, in which the standard subject position is occupied by an expletive *there*, while the actual subject – usually an indefinite – appears postverbally. In sentence (1b), locative inversion has taken place. This is a stylistic operation, in which a locative or directional PP can switch places with the subject. Note that in these two constructions, the postverbal position of the subject is not obligatory: the sentences in (1) are still perfectly grammatical when the subject is in its standard, sentence-initial position, as in (1)′.

- (1)′ a. *Something fishy* is going on here.
b. *John* came down the stairs.

Besides the strong preference to start a sentence with the subject, there is also a general tendency in languages for given or old information, if present, to be expressed early in a sentence and to precede new information. This tendency is seen as iconic: what comes first in a sentence is older than what comes later (Haiman, 1983; De Swart and De Hoop, 2000). In some languages, word order is much less based on the grammatical function of the constituents, such as subject and object, but rather on the information structural content of the sentence. Such languages are called *topic prominent languages*. Hungarian, for example, has a free word order in terms of grammatical function: all theoretically possible configurations of S, V and O can be found in this language (É. Kiss, 2002). Instead of grammatical function, topic-focus structure is used for the organization of information in the sentence. There are special syntactic slots for topic and focus in Hungarian, to which constituents from the main sentence can be moved. This gives the following configuration: [[T][F][S⁰]] (De Swart and De Hoop, 2000). In the topic position (T), given information can be expressed. Material in the focus position (F) provides new information and receives main stress. Some examples are given in (2) below (from É. Kiss, 1987).

- (2) a. [T e][F e] *Szereti János Máriát.*
 loves John Mary.ACC
 ‘John loves Mary.’
- b. [T *János*][F e] *szereti Máriát.*
 ‘As for John, he loves Mary.’
- c. [T e][F *Máriát*] *szereti János.*
 ‘It is Mary that John loves.’
- d. [T *János*] [F *Máriát*] *szereti.*
 ‘As for John, it is Mary that he loves.’

Alternatively, topicality can be defined in terms of aboutness: a topic is what the sentence is about. In Sign Language of the Netherlands (NGT), for example, a sentence typically starts with the topic, which indicates what the sentence is about (Crasborn, Van der Kooij, Ros and De Hoop, 2009). The material that follows the topic is the *comment*, which provides additional information about the topic. Topics in

NGT may be either subjects or objects (*argumental topics*), as in (3a), or adjuncts expressing time or space (*spatiotemporal topics*), as in (3b).

- (3) a. [_T MEISJE *PT_{Left}*], *PT_{Left}* BOEK WEGGOOIEN *PT_{Left}*
 girl there/that she book throw.away she
 ‘That girl, she threw away the book.’
- b. *PT_{Right}* [_T AFGELOPEN DINS DAG], #HEMA GEBOUW
 he last Tuesday HEMA building
PT_{Left} *PT_{Right}* BINNENGAAN HORLOGE VINDEN *PT_{Left}*
 there he enter watch find there
 ‘Last Tuesday he found a watch in the HEMA.’

Although in English information structure can sometimes be expressed by word order variation as well, as in *At six o'clock, Jane left* as opposed to *Jane left at six o'clock*, it is usually marked by intonation, e.g. by stressing the focused constituent (*JANE left at six o'clock*; *Jane LEFT at six o'clock*; *Jane left at SIX o'clock*). Dutch, a language that, like English, has a preference to start a sentence with the subject, also makes frequent use of intonation to encode information structure. However, in contrast to English, Dutch has a special syntactic position at the beginning of the sentence to which topics can move, similar to Hungarian. As a result, word order in Dutch is freer than in English, and word order variation is frequently used to signal differences in information structure, as illustrated in (4).¹

- (4) a. [_T Piet] draagt een hoed.
 Pete wears a hat
 ‘Pete is wearing a hat.’
- b. [_T Deze hoed] draagt Piet graag.
 this hat wears Pete gladly
 ‘It is this hat that Pete likes to wear.’

¹ De Swart and De Hoop (2000) note that the relation between information structure and syntactic structure in Dutch is in fact still weaker than that between information structure and intonation.

Thus, whereas English word order is largely dependent on grammatical function (start a sentence with the subject), and languages such as Hungarian and NGT have their sentence configuration based on information structure, Dutch can be considered a mixed language, in which both grammatical function and information structure are important determinants of word order. On the one hand, there is a strong syntactic preference to start a sentence with the subject. On the other hand, the sentence-initial position is also a topic position, resulting in a preference to start a sentence with the topic. As we will see, the subject and the topic of a sentence will coincide in many cases: subjects are often also topics, and topics are often made subject. However, there are types of noun phrases (NPs) that can be subjects, but are not usually topics. These are bare nominals, NPs that are not preceded by an article or any other determiner.

In this thesis, I will show that bare plurals in Dutch cause a conflict in word order preferences when they are subjects. Using experimental data, it will be argued that this conflict is responsible for the word order variation found with bare plural subjects, and that it can account for the semantic difference that correlates with this syntactic variation. I will relate the word order preferences and their correlation with the different meanings of bare plural subjects to topicality. In addition, grammatical function and topicality will be taken together in the notion of *prominence*. I will discuss several semantic and syntactic factors that determine the prominence of a constituent, and present Dutch corpus data that show that prominence influences the position of bare and definite plural subjects in Dutch. It will appear that subjects with properties characteristic for topics will prefer the preverbal position. Less prototypical topic characteristics will create a preference for a postverbal position. Also for definite subjects, these topic properties can cause word order variation.

In Chapter 2, I will introduce the notions of prominence and topicality, and I will show how they can influence word order in Dutch. Their relevance in the placement of bare and definite plural subjects will be explained. Next, I will present some experimental data to show that in out-of-the-blue sentences two conflicting preferences can predict the word order variation with bare plural subjects in Chapter 3. In Chapter 4 I will return to the notions of prominence and topicality and

show that more factors than the difference between bare and definite subjects play a role in word order variation. The effects of some of these factors will be tested in a corpus study presented in Chapter 5. The results of the corpus study will be the input for an analysis of Dutch word order variation in terms of conflicting preferences in Chapter 6. I will round off this thesis with the conclusions in Chapter 7.

The placement of bare and definite plural subjects in Dutch

2.1 Introduction

Dutch is a verb second language, which means that the finite verb in main clauses is always in the second position of a sentence. Consequently, only one constituent can precede the finite verb. This preverbal, sentence-initial position is often taken up by the subject, but it can in principle be filled by any type of constituent. This makes Dutch a language with a rather free word order. When the preverbal constituent is not the subject, the subject automatically ends up in a postverbal position. Thus, in contrast to English, where postverbal subjects are restricted to certain constructions, Dutch readily allows for subjects that are not in sentence-initial position.

This thesis is about the question what determines the position of Dutch subjects. Under which circumstances does a subject find itself in a postverbal position, and when is it preferred in sentence-initial position? I will not be concerned with the exact syntactic position of a subject. For example, the finite verb in Dutch main clauses is often assumed to be in the head of the CP. A postverbal subject in Dutch could then be either in SpecIP or in SpecVP. It can also originate in some position other than where it is found in the surface structure. I will not go into these discussions. Instead, I will investigate subjects that appear to the left of the finite verb versus subjects that appear to the right of the verb, and seek for properties and preferences that cause a subject to be in the one or the other position. I will also restrict the investigation to main clauses, since these are the only sentence types that exhibit the verb second rule, and consequently allow for variation in subject posi-

tion relative to the verb. In subordinate clauses, for example, the finite verb is always sentence-final, and postverbal subjects are impossible.

In this thesis, I will concentrate on two types of subjects: bare plural subjects and definite plural subjects. Bare plural subjects in Dutch are known to have different readings associated with their position. Consider the following Dutch sentence pair:

- (5) a. *Op straat spelen kinderen.*
 on street play children
 'Children are playing in the street.'
 b. *Kinderen spelen op straat.*
 children play on street
 'Children play in the street.'

The sentences in (5) are constructed from the same constituents: a bare plural subject *kinderen* 'children', a verb *spelen* 'play' and a prepositional phrase (PP) *op straat* 'in the street'. Only the word order differs. In (5a) the subject *kinderen* follows the verb and the PP *op straat* precedes it, while in (5b) it is the other way around. Both word orders are perfectly natural. Interestingly, the difference in word order also involves a different interpretation of the subject. In (5a), *kinderen* 'children' is preferably read as a statement about some unspecified group of children, while in (5b) it is understood as a general statement about children.

The same sentences with definite subjects do not exhibit such a meaning difference. Compare (6):

- (6) a. *Op straat spelen de kinderen.*
 on street play the children
 'In the street, the children are playing.'
 b. *De kinderen spelen op straat.*
 The children play on street
 'The children are playing in the street.'

Both (6a) and (6b) are utterances about a specific, probably familiar group of children. Clearly, the word order of (6b) is the more natural of the two in a neutral context. The word order of (6a) appears somewhat

more marked, perhaps suggesting that the children and/or the street are contrasted with other individuals on a different location (e.g. the adults in the living room). In contrast to the sentences in (5), the denotation of the NP *de kinderen* 'the children' does not change with the different word orders. Apparently, bare plural subjects in Dutch can occur in different positions typically combined with a change in meaning, while definite subjects do not show such a meaning change and seem to prefer the preverbal position.

In the following sections, I will discuss properties of bare plural subjects and definite plural subjects that might account for their variation in sentence position found in Dutch. In Section 2.2 I will introduce the notions of topicality and prominence. In Sections 2.3 to 2.5, I will elaborate on the role of definiteness in determining prominence and topicality. I will show in Section 2.6 that with bare plural subjects conflicting preferences cause variation in word order, while this is not the case for definite plural subjects. This variation in word order with bare plural subjects is accompanied by a change in meaning that can not only be attributed to the nature of the predicate.

2.2 Subjects, topics and prominence

Two important preferences in the placement of Dutch subjects can be identified. First, the universal syntactic preference to start a sentence with the subject is predicted to influence Dutch word order. Let us call the preference to place a subject at the sentence-initial position the *Subject First* preference (Vogels, 2008; Vogels and Lamers, 2008). Second, for reasons of iconicity given/old information and information that indicates what the sentence is about tend to be expressed early in the sentence. Because topics typically convey such information, there will also be a preference to start a sentence with the topic. I will call this the *Topic First* preference. Trivially, since material at the beginning of a sentence is uttered first, the sentence-initial position can be considered a *prominent* position: it directly connects to the preceding discourse. Thus, both subjects and topics prefer a prominent position in the sentence. The reason for this preference is that subjects and topics are typically very prominent constituents as well. Subjects are said to be *syntactically prominent* (e.g. Levelt, 1989; Levin and Rappaport Hovav, 2005):

they are highest on the hierarchy of grammatical functions. Topics, on the other hand, are *semantically* and *discourse prominent*: they typically possess properties that make it highly salient, either semantically or in discourse. The prominence of a constituent can be understood in terms of *accessibility*, the ease with which particular information is accessed in memory (Ariel, 1990). For instance, information is accessible when it has been mentioned in the immediately preceding discourse, when it is salient in the physical or linguistic context, or when it is available through world knowledge. Highly accessible information will be more active and can be retrieved with less effort than information with a low accessibility. Consequently, a constituent conveying highly accessible information will be prominent within a linguistic utterance.

Some constituents can be considered inherently more accessible than others. According to the Accessibility Hierarchy of Keenan and Comrie (1977), subjects are highly accessible. Topics also “constitute the most salient entities more often than not” in discourse (Ariel, 1990: 22/3). A topic conveys what the sentence is about, and frequently contains given information. It is therefore highly accessible. Their high accessibility makes subjects and topics very prominent information in a sentence. As a result, they are good candidates to occupy a prominent position, i.e. the sentence-initial position. Thus, we could formulate the following general relation between syntactic/semantic/discourse properties and syntactic position: ‘prominent constituents prefer a prominent position’.

Since both subjects and topics are usually prominent, they are expected to compete for the sentence-initial position. Often, however, there will be no competition, because the two frequently coincide. In many cases, the subject of a sentence will also be the topic, because a subject is often what the sentence is about (although other constituents may very well be topics as well). Similarly, topics frequently get the grammatical function of subject. Brunetti (2009) proposes that the reason why subject and topic coincide so often, is that both notions are typically associated with agentivity. The thematic agent role often corresponds to the grammatical subject, and being an agent is also a prototypical property of a topic, because people like to talk about agents. A prototypical agent is volitional, sentient and individuated, for instance

(Hopper and Thompson, 1980; Brunetti, 2009). These are all properties that contribute to prominence.

Another important property of topics is that they are often definite (e.g. Givón, 1979). Topics are typically referential and frequently convey old information, because sentences are usually about some familiar entity, either discourse-familiar or present in the real world. When topics are old information, they relate to the common ground of the speaker and the hearer. It is easier for the hearer when it is first stated what the coming sentence is about, before any new information about that topic is given (Skopeteas and Verhoeven, 2009), and when the beginning of the sentence refers back to something present in the immediately preceding context. Thus, information should preferably be given iconically, in linear order: what is older precedes what is newer (Haiman, 1983). Familiarity and reference to the common ground of the speaker and the hearer are often expressed by definite NPs. Thus, since definite NPs tend to be referential and to refer to old information, they can be considered intrinsically topical. Put differently: definiteness is a property that contributes to a constituent's semantic/discourse prominence, which is an important determinant of what will be the topic of a sentence.

To sum up, we have seen that subjects and topics frequently coincide, because they are both prominent: subjects are syntactically prominent, while topics are semantically/discourse prominent. An important contributor to semantic/discourse prominence is definiteness: definite NPs are typically referential and often refer to old information. Since prominent constituents prefer a prominent position, both subjects and topics are frequently found in sentence-initial position. Consequently, the Subject First and the Topic First preference will go hand in hand in many cases, since the subject and the topic of a sentence are often the same constituent. However, not all subjects are equally good topics. When a subject is indefinite, for example, it has properties that are less prototypical for topics. In this case, the Subject First and the Topic First preference are expected to be in conflict. In the following sections, I will zoom in on definiteness as a factor determining topicality. I will start with indefinites in Section 2.3, continuing with a special type of indefinite NP, the bare plural, in Section 2.4, which I will compare with its definite counterpart in Section 2.5.

2.3 Indefinites

In contrast to definite NPs, indefinite NPs are likely to contain new information (e.g. Vallduví, 1990). Firstly, indefinite NPs are typically used when no unique reference is made. They do not refer to particular individuals. In truth-conditional semantic terms, a non-unique reading of an indefinite NP states that the intersection between a set A and a set B is not empty. Thus, the NP merely describes the presence or existence of one or more individuals. Secondly, indefinite NPs can also express that the referent is not familiar. Familiarity means that the referent is part of the common ground of the speaker and the hearer: it contains information that is familiar to both speech partners. This could be because the referent has been introduced earlier in the discourse, because the referent is present in the physical context, or because it belongs to the shared (world) knowledge of speaker and hearer (Ariel, 1990). Consider the examples in (7) to (9):

- (7) Linguistic context: *“Chocolate and Co. has been struck deeply by the crisis.”*
- a. The company will fire hundreds of its workers.
 - b. #A company will fire hundreds of its workers. (*probably not Chocolate and Co.*)
- (8) Physical context: *Jenny walks by.*
- a. The poor girl lost her job.
 - b. #A poor girl lost her job. (*probably not Jenny*)
- (9) World knowledge:
- a. The queen will come and visit the affected families.
 - b. #A queen will come and visit the affected families. (*probably some foreign queen*)

The definite descriptions in the a-sentences are readily interpreted as referring to the familiar referent present in the immediate context or in shared world knowledge, while such an interpretation is hardly possible when an indefinite NP is used. Thus, the referent of a non-familiar

indefinite is not part of the common ground of the speaker and the hearer.

An indefinite NP that is both non-unique and non-familiar is called a quantificational or weak indefinite (Milsark, 1979; De Hoop, 1992). Weak indefinites are also called *non-specific indefinite* or *existential* NPs. Because a non-specific indefinite is non-unique and non-familiar, it cannot be referential (cf. Farkas and De Swart, 2006). Non-specific indefinites do not refer back to anything in the preceding discourse or in the non-linguistic context. Consequently, they are not accessible from memory at all. This means that they will not constitute very prominent information in a sentence. As a result, indefinites are not easily interpreted as what the sentence is about. Consider (10), from Reinhart (1982):

- (10) a. There is a fly in my tea.
 b. *As for a fly, it is in my tea.

Since the indefinite subject ‘a fly’ in (10a) most probably expresses non-referential, new information, it cannot be used in the ‘as for’-construction like in (10b), which explicitly presents the subject as being ‘what the sentence is about’. From this, it follows that non-specific indefinites are not good topics.

In addition to their non-specific weak reading, however, indefinite NPs can also get strong, referential readings (Milsark, 1977; De Hoop, 1992). For instance, an indefinite NP can have unique reference, i.e. it can refer to a particular individual. In this case, we speak of a *specific indefinite* NP. Consider the Dutch sentence pair in (11), from Hogeweg and De Hoop (to appear).

- (11) a. *Er gleet twee keer een meisje uit.*
 there slipped two time a girl out
 ‘Twice, a girl slipped.’
 b. *Een meisje gleet twee keer uit.*
 a girl slipped two time out
 ‘A girl slipped twice.’

In (11a), the indefinite subject *een meisje* 'a girl' is a non-specific indefinite: the sentence is preferably interpreted as two different girls slipping. The indefinite does thus not refer to a specific individual. In (11b), by contrast, the same indefinite subject is interpreted specifically: in this sentence it is the same girl that slips twice.

Another possible strong reading of indefinite NPs is a generic reading, in which the NP refers to some generality, as in (12) (Farkas and De Swart, 2006).²

(12) A dog is dangerous when it is hungry.

In (12), the indefinite subject 'a dog' refers to a generalization over all dogs. An indefinite with a strong, referential reading is more accessible in discourse than a non-specific indefinite, since there is a fixed referent or group of referents. As a consequence, specific and generic indefinites can be considered more prominent and therefore better topics than non-specific indefinites.

Indefinite NPs thus have low discourse prominence on one of their readings, the non-specific one. On the two other readings, the specific and the generic reading, an indefinite NP has a somewhat higher discourse prominence and consequently is a better topic. There are however indefinite NPs that lack the specific reading and can therefore be considered worse topics than other indefinites: bare plurals. In the next section, I will discuss some properties of bare plurals, and show that they are bad topics. After that, I will turn to a type of NP that is generally a good topic: the definite plural. I will show that in Dutch bare plurals cause a conflict between preferences when they are subjects, whereas definite plurals do not.

² A further possible strong reading for indefinites would be a partitive reading (De Hoop, 1992). However, this reading is usually found with NPs preceded by quantifiers like *enkele* 'some' or *veel* 'many', or by numerals. Since I am not concerned here with such NPs, I will not discuss this reading.

2.4 Bare plurals

Bare plurals are plural indefinites that are not preceded by an article or any other determiner. Bare plurals could be analyzed as the ordinary plural form of their singular indefinite counterparts with the indefinite article. In this case one would expect them to behave like singular indefinites and show the same contrast in weak and strong readings. However, there are several semantic properties of bare plurals that led scholars (e.g. Carlson, 1977; Dayal, 2003) to assume that bare plurals are not simply the plural of indefinite singulars. First of all, singular indefinites can take either narrow scope or wide scope with respect to operators such as quantifiers or negation, while bare plurals normally only take narrow scope. As a result, specific readings are excluded for bare plurals. Consider the Dutch sentences in (13).

- (13) a. *Een hond blaft.*
 a dog barks
 ‘A dog barks/is barking.’
 b. *Honden blaffen.*
 dogs bark
 ‘Dogs bark.’

The singular indefinite *een hond* ‘a dog’ in (13a) can receive either a specific or a generic interpretation, but for the bare plural *honden* ‘dogs’ in (13b) only the generic reading is available.³ The sentence cannot mean that there is a particular group of dogs that is barking.

Secondly, as opposed to singular indefinites, bare plurals allow for distributive readings. Consider the examples in (14).

- (14) a. *Jan was een uur lang konijnen aan het vermoorden.*
 John was an hour long rabbits on the murder
 ‘John was killing rabbits for an hour.’

³ In the right context, a contrastive existential reading is also possible, see section 2.5.2. A specific reading is however still excluded.

- b. *Jan was een uur lang een konijn aan het vermoorden.*
 John was an hour long a rabbit on the murder
 'John was killing a rabbit for an hour.'

The sentence in (14a) is straightforwardly interpreted as John killing several rabbits subsequently over the course of an hour. The sentence in (14b) on the other hand only has the odd reading that it took John one hour to kill one single rabbit.

Furthermore, bare plurals allow for kind readings, while singular indefinites do not. For example, (15a) is okay, while (15b) is semantically impossible.

- (15) a. *Reuzenalken zijn uitgestorven*
 'Great auks are extinct'
 b. **Een reuzenalk is uitgestorven*
 'A great auk is extinct'

In addition, there is a class of bare nouns that has these same properties as bare plurals, namely mass nouns, such as *zand* 'sand' and *water* 'water' (Krifka 2004; Chierchia 1998). Like bare plurals, mass nouns cannot receive a specific interpretation (e.g. Dayal, 2003), and allow for kind readings. Compare for example (16a) with (16b) and (16c) and compare (17a) with (17b) and (17c).

- (16) a. *Barack is op zoek naar een schilderij.*
 'Barack is looking for a painting.'
 b. *Barack is op zoek naar schilderijen.*
 'Barack is looking for paintings.'
 c. *Barack is op zoek naar kunst.*
 'Barack is looking for art.'
- (17) a. **Een waterplas is enorm schaars geworden*
 'A pond has become extremely scarce.'
 b. *Waterplassen zijn enorm schaars geworden.*
 'Ponds have become extremely scarce.'
 c. *Water is enorm schaars geworden.*
 'Water has become extremely scarce.'

The sentence in (16a) can either mean that Barack is looking for a particular painting or that he is just looking for some painting with no specific piece of work in mind. The bare plural *schilderijen* ‘paintings’ in (16b) can only get this latter, non-specific reading. The same holds for the mass noun *kunst* ‘art’ in (16c): the sentence cannot mean that Barack is looking for specific works of art. In (17a), the singular indefinite *een waterplas* ‘a pond’ cannot get the kind reading evoked by the kind level predicate *schaars worden* ‘to become scarce’. For both the bare plural *waterplassen* ‘ponds’ in (17b) and the mass noun *water* in (17c) this reading is available. Relying on these observations, it seems that bare plurals are closer to mass nouns than to singular indefinites.

Based on such differences between singular indefinites and bare plurals as presented above, Carlson (1977) analyzes English bare plurals as essentially referring to kinds. He observes that overtly kind denoting NPs such as ‘this kind of animal’ exhibit the same semantic properties that distinguish bare plurals from singular indefinites. He then argues that bare plurals are names of kinds, and that kinds should be considered individuals, with the peculiarity that kinds can be distributed in space. Kinds do not have quantificational force by themselves and always take narrow scope.

Still, besides kind readings bare plurals can also have existentially or universally (generically) quantified interpretations. These quantificational readings, Carlson argues, are triggered by the predicate. Predicates, he says, can be divided into predicates that are true of separate ‘events’ or *stages* that make up an individual, and predicates that are true of an individual as a whole, i.e. that characterise an individual. The former predicates Carlson calls *stage level predicates*; the latter *individual level predicates*. When a bare plural is the subject of a stage level predicate, it will get an existential reading, such as in (18a). As the subject of an individual level predicate it will receive a generic interpretation, as shown in (18b). Only with kind level predicates such as ‘to be extinct’ will bare plural subjects retain their original kind denotation, as in (18c).

- (18) a. Dogs are running (*stage level predicate*)
 b. Dogs are intelligent (*individual level predicate*)

c. Dogs are extinct (*kind level predicate*)

More recently, Chierchia (1998) takes a new approach to Carlson's view, including ideas from type shifting. Within type theory, (bare) noun phrases are considered to be predicates of type $\langle e, t \rangle$ (Partee, 1987; Chierchia, 1998; Van der Does and De Hoop, 1998; Krifka, 2004). Consequently, bare NPs as such cannot be used as arguments of a predicate. Determiner phrases (DPs), on the other hand, are suitable as arguments. They can be of type $\langle e \rangle$ (referential) or of type $\langle \langle e, t \rangle, t \rangle$ (a generalized quantifier) (Partee, 1987; Chierchia, 1998; Van der Does and De Hoop, 1998). One way to use a (bare) NP as an argument is to make it a DP, by adding a determiner (e.g. an (in)definite article or a quantifier). However, given that bare nominals can be subjects or objects in English and in Dutch, determinerless NPs do occur as arguments. Using bare nominals as arguments will result in a type mismatch, because bare nominals are of type $\langle e, t \rangle$, while arguments need to be of type $\langle e \rangle$ or $\langle \langle e, t \rangle, t \rangle$ (Chierchia 1998). Instead of assuming an empty D projection on every argumental bare NP, Chierchia (1998) proposes that this mismatch can be solved through a type shift by the operator $\hat{}$ ('down'), which maps a property of type $\langle e, t \rangle$ to a kind. Since kinds are of type $\langle e \rangle$, they can be arguments of a predicate. Thus, through the $\hat{}$ operator bare nominals can be interpreted as kinds, and consequently be used as arguments.

Following this reasoning, Chierchia (1998) agrees with Carlson that argumental bare plurals should be analyzed unambiguously as kinds. Kinds in his view can be represented as individual concepts, which are derived from properties (predicates) by the $\hat{}$ operator. Kinds can thus be regarded as nominalizations of predicates. From the kind reading, all other interpretations of bare nominals can be derived. Mass interpretations, for example, can be derived from kinds by the operator \cup 'up', and can be regarded as predicativizations of kinds (i.e. they constitute group properties). Existential readings are derived from kind readings by the \cup operator in addition to the application of a type shifting mechanism Chierchia calls *Derived Kind Predication* (DKP), which allows existential quantification over instances of kinds. Generic readings can be derived from kinds by the addition of a Gen operator. This operator quantifies over events or situations, and functions similarly to

quantificational adverbs such as ‘usually’ (cf. Farkas and De Swart, 2006).

Thus, in the neo-Carlsonian approach of Chierchia (1998), bare plurals are basically kinds, and type shifts are needed to get the existential and generic readings. It could be assumed that the type shift to an existential interpretation in Dutch occurs when the subject is in postverbal position. For strong readings in postverbal and sentence-initial position to arise, the bare plural subjects should be of a referential type. However, a type shift to a definite or specific NP is blocked, because the existence of a definite and indefinite article with these particular readings (which act as overt type shifters) prevents a bare NP to take on such an interpretation. What is possible is a type shift to a generalized quantifier (GQ; type $\langle\langle e, t \rangle, t \rangle$, Chierchia, 1998; Van der Does and De Hoop, 1998) with the Gen operator, which acts similarly to overt type shifters like adverbs such as *always* and *usually*. With this type shift, a bare plural subject takes on a generic interpretation.

Covert (i.e. not morphologically visible) type shifting can be considered a costly operation in terms of language processing (production or comprehension), because NPs are preferably interpreted in their basic type (Partee, 1987). In addition, there seems to be a rule that prohibits covert type shifting when the operation can be done overtly (Chierchia, 1998). Therefore, Chierchia (1998) argues that covert type shifting is a ‘last resort’ mechanism, which is only used when no overt type shifter is present. For example, in a language that has a definite article, covert use of the ι operator is blocked. A language that lacks a definite article, on the other hand, may freely use ι for covert type shifting operations on bare nominals.

If bare plurals are basically (nominalizations of) properties or predicates, they are non-referential by definition. They can be type shifted, but they will never have quantificational force by themselves: they are either bound by existential closure, or indirectly bound by a Gen operator that quantifies over situations. In neither case will they be able to refer to specific individuals. This makes bare plurals not very prominent in discourse. They are even less prominent than singular indefinites, because the latter have inherent quantificational force due to the presence of the indefinite article. Consequently, singular indefinites can either take narrow scope and receive a non-specific reading or

take wide scope and be interpreted specifically, whereas bare plurals always take narrowest scope (Chierchia, 1998).

Although a generic reading, which is considered a strong, referential reading, is available for bare plurals, the non-specific, existential interpretation is the one that occurs most frequently. Therefore, it can be said to be the unmarked reading of a bare plural. Consequently, bare plurals are prototypically non-prominent in terms of discourse accessibility. Since topics are prototypically discourse prominent, bare plurals should be considered bad topics.

2.5 Definites

In contrast to bare NPs, definite NPs are usually good topics. A singular definite NP prototypically refers to one uniquely identifiable, specific individual, object or concept. In truth-conditional semantics, the uniqueness property of singular definites is defined by stating that the intersection of a set A and a set B contains exactly one element, while $(A - B)$ is empty. For example, consider the Dutch sentence in (19).

- (19) *De hond blaft.*
'The dog barks.'

This sentence is true when the intersection of the set of individuals that are dogs with the set of individuals that bark contains exactly one element, while the set of individuals that are dogs but do not bark is empty.

By using a definite NP the speaker can also indicate that s/he assumes that the hearer is familiar with the identity of the referent. Definite NPs with unique reference to a familiar entity will generally be highly accessible in discourse. Being definite is therefore a property that contributes to the discourse prominence of an NP. Consequently, definites are prototypically prominent NPs, which makes them good candidates for the topic of a sentence.

Thus, the unmarked interpretation for a definite NP is a referential interpretation: most frequently, it refers to a unique set of entities, and is also discourse familiar. Alternatively, however, singular definites can also be used for kind reference. In this case, although a sense of

uniqueness is present (reference is made to a clear-cut concept, the kind, cf. Chierchia (1998)), the referent need not be familiar. For example, the sentence in (19) is ambiguous between a referential reading and a kind reading. In the first, the sentence means ‘there is a familiar dog that barks’, while in the second it means ‘it is a characteristic of the dog species that its members bark’. Non-familiar definites such as kinds are less referential and can therefore be considered less prominent than other definites.

For plural definite NPs, on the other hand, a kind reading is not possible. Consequently, plural definites will most likely have a referential denotation. Consider the examples in (20).

- (20) a. *De reuzenalk is uitgestorven.*
 ‘The great auk is extinct.’
 b. *?De reuzenalken zijn uitgestorven.*
 ‘The great auks are extinct.’

Because of the lack of a kind reading, a definite plural NP can be said to be more categorically prominent than a singular definite NP. In addition, referential definites are highly preferred over other types of NP when indicating what the sentence is about, even when it is a non-uniquely referring definite (Epstein, 2002). For these reasons, definite plurals can be considered good topics.

To recapitulate, I have discussed some semantic properties of indefinite and definite NPs in Dutch. I have shown that bare plurals do not have a very high accessibility in discourse, because of their lack of inherent quantificational force. Definite plurals, on the other hand, are prototypically referential and are therefore highly accessible. Because of this, definite plurals can be considered good topics, while bare plurals are less good topics. Now, as we have seen, prominent NPs prefer a prominent position (i.e. the sentence-initial position), while non-prominent NPs do not prefer a prominent position (and will end up in a less prominent position, such as a postverbal position). Both subjects and topics are prominent and therefore prefer a prominent position. Since subjects are often topical, this will go hand in hand in many cases. But what happens when a subject is not a good topic, for instance a

bare plural? In the next section, I will show that the conflict between preferences caused by the occurrence of non-topical subjects results in variation in word order in Dutch.

2.6 Conflicting preferences

2.6.1 Prominence scales

In the preceding discussion, I have specified definiteness as a factor that affects the topicality of an NP: referential definites are generally good topics, while weak indefinites are bad topics. The degree of definiteness can be given in a definiteness scale, such as in (21) (Aissen, 2003).

- (21) *Definiteness scale:*
 Personal pronoun > Proper name > Definite NP > Indefinite specific NP > Non-specific NP

NPs more to the left of the definiteness scale are more prominent, and therefore better topics, than those to the right. Weak, non-specific NPs are on the far right end of the scale, while personal pronouns and proper names are even more prominent than definite NPs. Since prominent constituents typically combine with a prominent position, NPs high on the definiteness scale are likely to occur in a prominent position, while NPs low on the definiteness scale are not.

We have also seen that subjects are the most syntactically prominent constituents. This can again be depicted in a prominence scale, as shown in (22) (Skopeteas and Verhoeven, 2009).

- (22) *Scale of grammatical functions:*
 Subject > Object > Oblique complement > Adjunct

Subjects are more prominent than objects, which in turn are more prominent than oblique complements and adjuncts. Consequently, subjects are preferred in a prominent position, whereas this is less so for adjuncts. A prominent position in terms of discourse salience is the preverbal, sentence-initial position. A postverbal position is less promi-

ment. This can also be represented by a prominence scale, as in (23) (cf. Van Bergen, 2006; Yang and Van Bergen, 2007).

- (23) *Word order scale:*
Preverbal > Postverbal

Thus, the preferences associated with definiteness and grammatical function could be described more generally as a tendency that says: the higher an NP is on a prominence scale, the more likely it is produced closer to the beginning of the sentence.

However, the preferences mentioned above can also run counter to each other: an NP can be prominent on one scale, but non-prominent on the other. This is the case with bare plural subjects. Since they are non-specific NPs, bare plurals are typically non-prominent on the definiteness scale. As subjects, however, they are syntactically prominent. Thus, two preferences are in conflict here: the part of a bare plural subject that is a subject prefers a prominent position, whereas the part that is a non-specific indefinite does not prefer a prominent position. This conflict between preferences might account for the variation found in the position of bare plural subjects in Dutch sentences. Bare plural subjects in Dutch can easily occur either at the high prominent sentence-initial position or in a low prominent postverbal position. In the former situation, the bare plural subject is in the unmarked position for a subject, but in a marked position for a bare plural. In the latter it is the other way around: the bare plural subject is in the unmarked position for a non-specific indefinite, but in a marked position for a subject. In addition, bare plural subjects receive a different interpretation in these two positions. This meaning difference will be the topic of the following subsection.

2.6.2 *Existential, generic and “third” readings of bare plural subjects*

Dutch bare plural subjects can easily occur either in the preverbal, sentence-initial position or in a postverbal position. When a bare plural subject is in postverbal position, the sentence-initial position is often filled by the expletive *er* ‘there’, like in (24) (taken from the Corpus Gesproken Nederlands (Corpus of Spoken Dutch, CGN)).

- (24) *Er verbleven toen in Jeruzalem vrome joden uit alle*
 there stayed then in Jerusalem pious Jews from all
delen van de wereld. [fn009113.44]
 parts of the world
 'At that time, pious Jews from all parts of the world stayed in
 Jerusalem.'

Alternatively, if there is a locative PP adjunct present in the sentence, it can replace the expletive (locative inversion), as in (25).

- (25) *In alle hoeven lagen soldaten ingekwartierd.*
 in all farms laid soldiers quartered
 [fv800675.29]
 'On all farms, soldiers were quartered.'

It is also possible that a locative PP occupies the sentence-initial position, and that *er* 'there' appears postverbally, as shown in (26).

- (26) *In Amerika zijn er dus zomerkampen.* [fv400175.116]
 in America are there thus summer.camps
 'So in America there are summer camps.'

The bare plural subjects in (24)-(26) refer neither to particular Jews, soldiers and summer camps, nor to Jews, soldiers and summer camps that were introduced earlier in the context: the subjects have an existential interpretation. Although they are subjects, they are not in a prominent position, but in a non-prominent, postverbal position. This position is marked for a subject, but unmarked for a bare plural. In this position bare plurals receive their unmarked existential reading. Thus, the unmarked, non-prominent position of the bare plural corresponds with its unmarked, non-prominent interpretation. Apparently, the fact that the bare plurals are at the same time syntactically prominent as subjects is of minor importance than the preference to combine the non-prominent, existential reading of the bare plurals with a non-prominent, postverbal position.

Bare plural subjects can also occur at the sentence-initial position, which is the unmarked position for a subject, as in (27).

- (27) *Vrolijke mensen kopen in de Grote Vrolijke*
 happy people buy in the Big Happy
Supermarkt. [fv801255.14]
 Supermarket
 ‘Happy people buy at the Big Happy Supermarket.’

Although the subject is now in its standard position, this position is dispreferred for a non-prominent bare plural. Hence, a weak, existential meaning for the bare plural subject *vrolijke mensen* ‘happy people’ in (27) is not likely. Indeed, indefinites in sentence-initial position favour a strong reading in Dutch (De Hoop, 1992). However, in contrast to singular indefinites, a specific meaning for the bare plural is not an option here. The subject in (27) most probably has a generic reading: it does not denote some specific happy people, or just some random happy people; the sentence is about all happy people, or at least about happy people in general.

According to Farkas and De Swart (2006), a generic reading is obtained when it is bound by the operator *Gen*, which imposes *maximal reference* on the otherwise non-referential bare plural. Maximality says that the referent comprises all individuals that satisfy its descriptive content. It corresponds to the ι (‘iota’) operator in Partee (1987) and Chierchia (1998), which is used to interpret the definite article (the dog, the dogs): ιX is the largest plurality of X if X is plural, and the only X (uniqueness) if X is singular. Thus, maximality boils down to uniqueness in case of a singular NP. Applied to the sentence in (27), the NP *vrolijke mensen* ‘happy people’ refers to all individuals that are both human and happy. Because generic bare plurals have maximal reference, they can be considered more prominent than existential bare plurals (but less prominent than definite plurals, since they are still non-familiar), as represented in (28).

- (28) *Maximality scale:*
 Generic (maximal) bare plural > Existential (non-maximal)
 bare plural

Now, generic bare plural subjects are both syntactically (being the subject) and semantically (having maximal reference) prominent. As a consequence, bare plural subjects with a generic denotation can be good topics. In fact, the subject of (27) is the topic: the sentence is about happy people. While for unmarked (existential) bare plurals a prominent, sentence-initial position is avoided, bare plurals are not bad in this position when they can be interpreted generically. Here, the for bare plurals marked, sentence-initial position is combined with a marked, strong, generic meaning.

Besides a generic reading, De Hoop (1992) distinguishes another possible interpretation for a bare plural in sentence-initial position, which she calls the “third reading”. In this reading, the bare plural has contrastive focus, while retaining an existential interpretation. Hence, when a non-specific, bare plural subject has contrastive focus, it can be in sentence-initial position. Consider the examples in (29) below ((29a) is taken from De Hoop (1992), and (29b) from the CGN; capitals indicate stress).

- (29) a. *MUIRBLOEMEN* *bloeden* *voor* *het* *lage*
 wall-flowers blossomed in.front.of the low
raam.
 window
 ‘Wall-flowers blossomed in front of the low window.’
- b. *PASPOORTNUMMERS* *moeten* *gecontroleerd* *worden* [...].
 passport.numbers must checked become
 [fn000603.236]
 ‘Passport numbers have to be checked.’

These examples can be accounted for by stating that to be in sentence-initial position, a subject has to be prominent. Besides by a costly type shift to a generic denotation, this can be achieved for weak subjects by putting focal stress on them. Choi (1999) analyzes contrastive focused constituents as having the features [+Prominent] and [+New] (as opposed to non-contrastive, completive focused constituents, which are [-Prominent] and [+New]). Topics are [+Prominent] and [-New] in Choi’s model. This means that “third reading” subjects should be seen as more

topical than postverbal weak subjects, which makes them more fit for the sentence-initial position. This analysis would imply that contrastive focus like in the above examples actually has topic-like properties. This is exactly what has been proposed for preposed focus (e.g. Vallduví, 1992): for topics as well as for preposed focus constituents a certain presupposition is needed. For topics, this presuppositional information lies in the fact that what is talked about is part of the common ground of speaker and hearer: it is accessible information. For the contrastive focus case, the presupposition constitutes the base for the contrast (e.g. all things that could be checked, in (29b)). This information is then combined with a new element: the focus part (e.g. it is emphasized that it is *passport numbers* that have to be checked). In this way, an existential bare plural can still be prominent and appear in sentence-initial position.

2.6.3 *The role of the predicate*

Whether a bare plural subject gets an existential or a generic reading has also been found to be dependent on the nature of the predicate. Carlson (1977) argued that in English individual level predicates induce generic readings on a bare plural, while stage level predicates induce existential readings. If the different positions and readings of bare plural subjects in Dutch can be accounted for by relating them to prominence and topicality, what would still be the influence of the predicate?

In English, bare plural subjects of individual level predicates can only receive a generic reading (Diesing, 1992). For example, they are bad in existential *there*-sentences with an individual level predicate. Chierchia (1998) argues that individual level predicates are inherently generic and that they always occur in the scope of a Gen operator. The assertion that bare plural subjects of individual level predicates cannot occur in existential constructions also holds for Dutch (cf. the ungrammaticality of (30)).

- (30) **Er zijn mensen intelligent.*
 there are people intelligent

De Hoop (1992), however, shows that in Dutch, it is not the case that individual level predicates require the subject to be generic, but they induce a strong reading on the subject. Thus, individual level predicates can occur in existential constructions in Dutch when the postverbal subject has a strong reading.

Stage level predicates, on the other hand, were argued by Carlson (1977) to induce an existential reading of their bare plural subjects. In English, a predicate can be made stage-level by using progressive aspect. For example, the sentence in (31b) is preferably read as a general property of children, while (31a), with the progressive, is interpreted as a description of an event (a stage), and hence the bare plural subject will be interpreted existentially.

- (31) a. Children are playing in the street
 b. Children play in the street

In Dutch, however, the progressive is far less common, and both sentences in (31) are preferably translated with imperfective aspect on the verb, as in example (5), repeated here as (32).

- (32) a. *Opstraat spelen kinderen.*
 on street play children
 'Children are playing in the street.'
 b. *Kinderen spelen op straat.*
 children play on street
 'Children play in the street.'

The predicate of the sentences in (32) is the same: *spelen* 'play'. Still, the bare plural subject in (32a) gets an existential reading, while the subject in (32b) is most likely interpreted generically. Thus, the predicate cannot account for the interpretation difference between the bare plural subjects in (32a) and (32b).

According to Diesing (1992), a stage level predicate like *spelen* 'play' can select subjects with either an existential or a generic meaning, dependent on their syntactic position. She maps the semantics of indefinites onto syntactic structure in her Mapping Hypothesis:

- (33) *Mapping Hypothesis:*
 Material from VP is mapped into the nuclear scope.
 Material from IP is mapped into the restrictive clause.

This hypothesis is based on the logical representation of sentences proposed by Kamp (1981) and Heim (1988), in which a sentence consists of a restrictor, a nuclear scope and an operator that quantifies over the restrictor. NPs interpreted in the nuclear scope of a predicate will receive a weak reading, while NPs interpreted in the restrictive clause will receive a strong reading. Consequently, the Mapping Hypothesis suggests that VP-internal bare subjects will receive a weak, existential reading, while VP-external bare subjects will receive a strong, generic reading. If the postverbal subject in (32a) is assumed to be within the VP, it is correctly predicted to receive a weak, existential reading. If the preverbal subject in (32b) is outside of the VP, it will receive a generic reading, which is also correct.

Assuming for the moment that the VP-internal position corresponds to a postverbal position and the VP-external position to the sentence-initial position, the Mapping Hypothesis also seems to make the right predictions for Dutch bare plural subjects of individual level predicates. Diesing (1992) argues that, whereas subjects of stage-level predicates can appear either VP-internally or VP-externally, subjects of individual-level predicates can appear only VP-externally. Consequently, bare plural subjects of individual-level predicates can only receive a generic reading. Consider the sentences in (34).

- (34) a. *Opschool zijn kinderen intelligent.*
 on school are children intelligent
 'At school, children are intelligent.'
 b. *Kinderen zijn op school intelligent.*
 children are on school intelligent
 'Children are intelligent at school.'

The bare plural subject *kinderen* 'children' has a generic reading in both sentences. The sentence in (34b), with the generic subject in sentence-initial position, has the most natural word order in a neutral context. The sentence in (34a) probably needs a contrastive context: it could be

followed by 'At home, on the other hand...'. Despite its postverbal position, an existential reading for the subject (34a) is not possible, as predicted by the Mapping Hypothesis. This is further shown by the ungrammaticality of (35) below.

- (35) **Op school zijn er kinderen intelligent.*
 on school are there children intelligent

However, De Hoop (1992) shows that the Mapping Hypothesis is actually wrong in stating that VP-internal subjects can only get weak readings. For example, other kinds of weak NPs can get a strong, partitive reading in VP-internal position, as in (36).

- (36) a. *Er zijn twee vliegtuigen neergeschoten.*
 there are two aircrafts shot.down
 'Two (non-specific) aircrafts have been shot down.' or 'Two (of the) aircrafts have been shot down.'
- b. *Op straat spelen twee kinderen.*
 on street play two children
 'There are two children playing in the street.' or 'Two (of the) children are playing in the street.'

Both (36a) and (36b) are ambiguous between a weak (existential, non-specific) and a strong (partitive, presuppositional) reading. Moreover, subjects can occur in VP-internal position with individual level predicates when they have a partitive reading. Compare (37) below to (35), for example.

- (37) *Op school zijn er twee kinderen intelligent.*
 on school are there two children intelligent
 'At school, two (of the) children are intelligent.'

The sentence in (37) shows that in Dutch strong readings of indefinites are possible in a VP-internal position, and that therefore individual level predicates can occur in existential constructions, as long as the postverbal subject has a presuppositional reading. In English, strong readings are not possible in a VP-internal position. Because individual

level predicates can only take subjects with a strong reading, English existential constructions cannot combine with individual level predicates. In VP-external position, subjects can get only strong readings in Dutch, while they can be either weak or strong in English. In sum, the only influence of the predicate on bare plural subjects so far is that individual level predicates induce a strong reading on the subject. However, since strong readings are possible in both VP-external and in VP-internal position in Dutch, this does not say anything about the position of the bare plural subject.

Another restriction on the possible interpretations of a bare plural subject induced by the predicate is that, at least in English, for an existential bare plural subject to be in postverbal position, the verb has to be unaccusative. Compare (38a) to (38b), for example.

- (38) a. In the bedroom were children. (*unaccusative*)
 b. *In the bedroom cried children. (*unergative*)

In (38a), the verb ‘to be’ is unaccusative, and the existential subject is fine in postverbal position. In (38b), on the other hand, the verb ‘to cry’ is unergative, which makes a postverbal existential bare plural subject ungrammatical. This difference could be explained in terms of topicality. Subjects of unaccusative verbs are not prototypical agents. Rather, they have properties that are typically associated with patients, such as being causally affected by the event (Brunetti, 2009). Since agentivity is a prototypical property of topics, subjects of unaccusative verbs might not be good topics. Following the iconicity in topic-focus structure, it is expected that subjects of unaccusative verbs can occur naturally in a postverbal position. When there is another constituent that is a good topic, this constituent may occupy the preverbal topic position, allowing the patient-like subject to remain postverbally. In addition, when the subject is indefinite, and thus has even less topic properties, the preverbal position may be occupied by an expletive ‘there’, which results in an existential construction. Looking at (38a), this analysis would suggest that the locative adjunct *in the bedroom* is a good topic, allowing the bare plural subject *children* to be in a postverbal position. I will come back to this in Chapter 4.

Contrary to unaccusative verbs, unergative verbs in English are generally disallowed in existential there-sentences, as the ungrammaticality of (39a) shows. However, unergatives in existential sentences become better when a locative PP adjunct is present. Zwart (1992) suggests that the locative adjunct makes the expletive 'there', which he analyzes as a small clause predicate, more or less locational, which turns the unergative verb into an unaccusative verb, as in (39b).

- (39) a. *There hung pictures.
 b. There hung pictures on the wall.

The presence of a locative adjunct PP in (39b) adds to the meaning of (39a) that the pictures exist on a certain location. This makes the verb *hung* unaccusative, which allows the expletive 'there' to occupy the standard subject position. De Hoop (1992) argues that the subject in English existential there-constructions is inside VP, but that it cannot receive (weak or strong) nominative case in this position. Only weak objective case is available. Since subjects of unaccusative verbs are thematically patients, they can be assigned objective case. This is not the case for unergative verbs. Hence the impossibility of postverbal subjects with unergative verbs.

In Dutch, this restriction of unaccusativity in existential sentences is not present. Dutch existential *er*-sentences can occur with unaccusative as well as with unergative predicates. Zwart (1992) argues that the reason for this is that the expletive *er* 'there' in Dutch, besides a small clause predicate, can also be a meaningless element. With unaccusative verbs, the expletive is a small clause predicate. With unergative verbs this is not possible and the expletive is a meaningless element occupying the standard subject position. This last option is not available in English. De Hoop (1992), on the other hand, says that subjects of unergative verbs can be VP-internal in existential *er*-constructions because weak nominative case is licensed in the VP in Dutch. Thus, subjects of all kinds of verbs can be in the VP, as opposed to English. Strong nominative case can be licensed either in the VP or in the IP in Dutch; hence, postverbal subjects in existential constructions may either get a weak or a strong reading, while for sentence-initial subjects only a strong read-

ing is available. In neither of the two analyses does the predicate directly influence the interpretation or the position of the subject.

In sum, the nature of the predicate cannot account for the different interpretations of a bare plural subject in Dutch and for the positions associated with these interpretations. Although individual level predicates always induce a strong reading on their subject, strong readings are possible preverbally as well as postverbally. Other types of predicates can occur freely with either existential or generic bare plural subjects in Dutch. Neither can the nature of the predicate explain why VP-internal bare subjects can get weak or strong readings, while VP-external bare subjects can only get strong readings. Neither individual level predicates nor unergative verbs require a VP-external subject in Dutch. I will now turn to definite plural subjects.

2.6.4 *Definite plural subjects*

For definite plural subjects, no conflict between the Topic/Definite First and the Subject First preference resulting from the prominence scales in (21) and (22) is present. Definites are semantically prominent NPs and are therefore not dispreferred in sentence-initial position. Since both definite NPs and subjects are highly prominent, definite plural subjects are expected to be very good candidates for the sentence-initial position. Indeed, whereas bare plural subjects can appear equally natural in pre- as in postverbal position, be it with a shift in meaning, for definite plural subjects the preverbal position is clearly the unmarked position, as we have seen in (6), repeated here as (40).

- (40) a. *Op straat spelen de kinderen.*
 on street play the children
 ‘In the street, the children are playing.’
 b. *De kinderen spelen op straat.*
 The children play on street
 ‘The children are playing in the street.’

The sentence in (40b) has the unmarked word order for a definite plural subject, while (40a) is more marked. In contrast to the same examples with bare plural subjects (see (32)), there is no structural semantic dif-

ference between the two word orders in (40). There is thus no a priori reason for a definite plural subject to be placed in a non-prominent position.

That definite subjects are not easily placed in a non-prominent position is also shown by the so-called *definiteness effect*. The Definiteness effect states that the postverbal subject in existential sentences must be a non-specific (existential) indefinite NP, and that definite NPs and strong indefinite NPs are not found in this position (Milsark, 1977; Beaver, Francez and Levinson, 2005). Indeed, definite plural subjects do not generally occur as the postverbal subject in existential sentences in Dutch, as the ungrammaticality of (41) shows.

- (41) **Er spelen de kinderen op straat.*
 there play the children on street

Note, however, that the preference to put a definite subject in topic position is only a tendency; there might be other preferences at work than definiteness that influence the position of definite subjects. For instance, note that the sentence in (40a) is certainly not ungrammatical; it just needs a special context to be acceptable. If the discourse imposes a certain information structure on the sentence, for example that the location *op straat* 'in the street' is contrasted with some other location (e.g. in the garden), the locative PP has contrastive focus and is preferred in a prominent position (and probably gets heavy stress). In addition, it has frequently been noted for English that sometimes definite subjects are allowed in existential *there*-sentences, as in (42) (Lumsden, 1988).

- (42) There were the same tendencies noted in France.

Comorovski (1991), for example, argues that definite NPs can occur as the subject in existential *there*-sentences when they express new information. But also other, more semantically defined factors than information structure might influence the position of definite plural subjects. I will come back to this in Chapter 4.

In sum, while bare plural subjects can occur either in sentence-initial position or in postverbal position, resulting in different interpre-

tations, definite plural subjects seem to prefer the sentence-initial position, which can be considered their unmarked position. Although post-verbal definite subjects are possible, this is a marked position and no change in meaning is involved.

2.7 Conclusion

In this chapter, I have argued that two important preferences are at work in determining Dutch word order: the Subject First preference and the Topic First preference. These two preferences are compatible in many cases: subjects are often topics, and topics are often made subject. Bare plurals, however, can be considered bad topics, whereas they can perfectly well be subjects. Therefore, bare plural subjects cause a conflict between the two preferences: as subjects they prefer the sentence-initial position, but as bare plurals they prefer a non-prominent position. This conflict has been shown to result in variation in the position of bare plural subjects, in which a prominent position is associated with a strong, generic meaning, while a non-prominent position is associated with a weak, existential meaning. For definite plural subjects, which are generally good topics, no such variation in position and interpretation is found.

In the next chapter, I will present experimental data that support this analysis. I will show that when there is no a priori preference for a particular interpretation of a bare plural subject, free variation will occur in the position of bare plural subjects. Definite plural subjects, on the other hand, are generally produced in sentence-initial position.

Definiteness as a cue for word order: a production experiment

3.1 Introduction

In Chapter 2, two main preferences have been identified that are assumed to influence the position of a bare plural subject in Dutch. First, since subjects are syntactically prominent, they preferably occur in a prominent position, which is defined here as the sentence-initial position (Subject First preference). Second, there is a general preference to start a sentence with the topic (Topic First preference), because topics are semantically and discourse prominent. Since bare plurals are typically non-topical, they are not preferred in a very prominent position, which results in a conflict between the two preferences.

This conflict between preferences manifests itself in Dutch word order in the occurrence of variation in the position of bare plural subjects. In the unmarked reading for a bare plural, a bare plural subject is placed postverbally, violating the Subject First preference. To satisfy the Subject First preference, the bare plural subject can also be in preverbal, sentence-initial position, which requires either a type shift to a strong, generic reading or focal stress on the subject. Now, the question arises what the strength of these two preferences is. Suppose that sentences with bare plural subjects are produced without any intention to express a certain reading (existential or generic) for the bare plural. Which word order would prevail? The one in which the subject is in its standard position, or the one in which the bare plural is in its unmarked position? In other words, what is worse: a postverbal subject or a preverbal bare plural (with a concomitant generic or contrastive focus

reading)? Or are they equally bad, resulting in variation around chance level?

To answer these questions, a drag-and-drop language production experiment was set up (Vogels, 2008; Vogels and Lamers, 2008). It tested participants' word order preferences in out-of-the-blue sentences with bare and definite plural subjects. In the next section, I will discuss the production experiment and its results in more detail.

3.2 A drag-and-drop production experiment

3.2.1 Methods

3.2.1.1 Participants

Sixteen native speakers of Dutch (9 males, 7 females; mean age 22.5 years) participated in the experiment. All had normal or corrected to normal vision, no neurological disabilities and enough experience in working with a computer to complete the task.

3.2.1.2 Materials

The materials contained 28 stimuli. Each stimulus consisted of an intransitive verb, a plural animate NP and a locative PP. The NP could either be a bare plural or a definite plural, which formed the two conditions. The NP and the PP both could be placed either in front of or behind the verb. Care was taken that all stimuli were grammatical in both word orders. In the experiment, the verb was presented in the middle of a computer screen, while the NP and the PP were presented in the upper part of the screen. This was done in random order, to avoid an effect of the order of presentation. An overview of the different conditions in one set of items is given in Table 1.

Multiplied by the four conditions in Table 1, the 28 stimuli yielded $28 \times 4 = 112$ different experimental items. These items were distributed over two lists, in such a way that from each set one list contained an item with a bare plural subject in a certain order of presentation and the other an item with a definite plural subject in the reverse order of presentation. In addition, 32 filler items unrelated to the experimental items were added to each list, resulting in a total of 88 items per list.

Table 1 Overview of the two conditions (bare or definite) in two presentation orders of experiment 1 (BP = bare plural NP; DefP = definite plural NP).

Plural NPs	Order of Presentation	
	PP-NP	NP-PP
BP	in de tuin ⁴ kinderenspelen.....	kinderen in de tuinspelen.....
DefP	in de tuin de kinderenspelen.....	de kinderen in de tuinspelen.....

The two lists were divided into two blocks of 44 items, taking care that items from the same set were kept as far apart from each other as possible. To keep the participants attentive during the task, 8 so-called response items were included as well, randomly distributed over each list.

3.2.1.3 Procedure

Participants were seated in front of a computer screen in an experimental room. They were tested individually. The participants were instructed to build grammatically correct Dutch sentences on the computer screen using three given constituents. They had to drag one of the two phrases appearing on the screen to the desired position on



Figure 1 Example of an experimental item presented on the computer screen.

either side of the verb with the mouse. Each block started with the text *Let op! Het gaat nu beginnen!* ('Attention! It starts now!') ap-

⁴ in = in; de = the; tuin = garden; kinderen = children; spelen = play

pearing on the screen. Next, an asterisk (*) appeared for 800 ms in the middle of the screen. Then, at the same position a verb appeared with dotted lines on both sides to indicate the pre- and postverbal slots. After 1500 ms, the NP and the PP appeared in the upper part of the screen, positioned right above each other (see Figure 1).⁵ Participants could drag one of the phrases to one of the slots. Once one of the constituents had been dropped in one of the positions, the other position was automatically filled with the remaining constituent and the computer program continued to the next item. An asterisk appeared on the screen before the next item started. A participant had 6000 ms to complete a sentence. Only by exceeding this time limit an item could be skipped. Once the participant released the mouse button in a designated area no correction was possible.

To check whether participants were performing the task attentively, some of the filler items were followed by a word in capitals immediately after a sentence was formed (the response items). Participants had to indicate whether the word had occurred in the sentence they just built by clicking with the mouse on the *ja* 'yes' or *nee* 'no' boxes on the screen.

The stimuli were presented in white on a black screen, using the software program *Presentation* (Neurobehavioral Systems). A single experiment consisted of 2 blocks, separated by a short pause.⁶ Each experiment was introduced with a practice block of 14 items that were similar to the items in the experimental blocks. It took approximately 15 minutes to complete the experiment.

3.2.2 Results

The results of the experiment are shown in Table 2.⁷

⁵ The filler items were presented in a similar way as the experimental items.

⁶ There was no fixed time for this pause. As soon as the participant indicated that he or she was ready to go on, the next block was started.

⁷ One participant was excluded from the analysis, because this participant showed a significant overall bias. In addition, there were 5 unscorable experimental items due to exceeding the time limit. All response items but one were answered correctly, indicating that all participants paid attention to the task.

Table 2 Production of the two different word orders in the drag-and-drop experiment, for two types of subject (n=15; S=subject, V=verb, PP=prepositional phrase).

Subject	Produced word order				TOTAL	
	S-V-PP		PP-V-S			
Bare plural	240	57.4%	178	42.6%	418	100%
Definite plural	338	81.1%	79	18.9%	417	100%
TOTAL	578	69.2%	257	30.8%	835	100%

As can be seen in Table 2, for bare plural subjects the postverbal position is much more frequent (178 cases) than for the definite plural subjects (79 cases). The results also show that for both definite and bare plural subjects the sentence-initial position was the most frequent. A repeated measurement ANOVA was performed on the percentages of sentences with the subject in sentence-initial position (S-V-PP). The effect of type of subject (bare plural or definite plural) on word order was found to be significant ($F_1 = 9.374$; $p = .011$; $F_2 = 60.042$; $p < .001$).

3.2.3 Discussion

Because of the conflicting preferences due to the difference in prominence between subjects and bare plurals, it was expected that bare plural subjects would be placed more often in a postverbal position than definite plural subjects. This indeed appears to be the case: 42.6% of the bare plural subjects were produced postverbally, against 18.9% of the definite plural subjects.

Still, both definite and bare plural subjects are more frequently produced in sentence-initial position than in postverbal position. For definite plural subjects this is expected, since definites are very prominent and consequently prefer a prominent position. For bare plurals, on the other hand, a non-prominent position is expected to be preferred. Apparently, the Subject First preference is stronger than the preference to start the sentence with a good topic (i.e. not with a bare plural), at least in an artificial situation in which there is no intention to commu-

nicate a certain meaning and in which information structure does not dictate word order. It seems, then, that even though Dutch has a special position where topics can be expressed, the syntactic preference to start a sentence with the subject is stronger than the semantic/discourse preference to start a sentence with the topic. Thus, as in a strict word order language like English, the sentence-initial position is highly preferred by subjects, even if the subject is a non-prominent bare plural. It is also clear from Table 2, however, that the Topic First preference still has a considerable impact on word order in Dutch.⁸

Although the speakers' intentions were not examined in the experiment, it was assumed that a change in placement of a bare plural subject combined with a shift in meaning, i.e. when a bare plural subject was placed at the sentence-initial position, the speaker intended to make the sentence a generic statement; otherwise, an existential reading of the subject was meant (contrastive focus was not considered). When the preference to start a sentence with a good topic is violated, the bare plural subject is placed in the sentence-initial position, which is the unmarked position for a subject, but a marked position for a bare plural. Because it is in a prominent position, the meaning of the bare plural is shifted to a more prominent, generic reading. A generic interpretation makes a bare plural more topical, and hence more fit for the sentence-initial position. Thus, although bare plurals are normally preferred in a non-prominent position and a generic reading is a marked interpretation for a bare plural, bare plural subjects with a generic reading are not bad in the sentence-initial position. The reason is that here the marked, sentence-initial position combines with a marked, generic meaning.

In sum, the results of the drag-and-drop experiment support the analysis that the position of bare and definite plural subjects depends on at least two important preferences: the Subject First preference and

⁸ Recall that we have only tested the intrinsic, semantic topicality in this experiment: it was argued that definite NPs are inherently good topics, while bare plurals are not, and the experimental items were presented without any context. However, if discourse factors would be taken into account, the effect of Topic First would be expected to be even greater (see Vogels, 2008; Vogels and Lamers, 2008).

the Topic First preference. Since these preferences are in conflict with bare plural subjects, variation in word order occurs. There is much less word order variation for sentences with a definite plural subject. The Subject First preference can be considered somewhat stronger than the Topic First preference in Dutch. Because bare plural subjects have the option to escape their non-topicality by taking on a generic reading, the aversion to preverbal bare plurals is tempered. This could have contributed to the higher frequency of bare plural subjects in sentence-initial position.

The conflict also predicts that definite plural subjects always prefer the sentence-initial position, because they are good topics and there is no a priori reason why they would be put in a non-prominent, postverbal position. However, in the experiment almost 20% of the definite plural subjects is produced in postverbal position (Table 1). This is not yet accounted for in the present analysis. A possible explanation would be that even though all items in the experiment were presented out of context, participants were creating their own context for the sentences they had to produce. Thus, they could have imagined a discourse for the sentence, which could also include a discourse topic. For example, when their discourse was about the location of the event described by the verb, the locative PP would be the discourse topic, and this would consequently influence the choice for the topic in the sentence to be produced. Since contextual factors were not tested in the present experiment, this possibility cannot be checked. Alternatively, the relatively high number of postverbal definites could also be the result of participants deliberately adding random variation in the word order of their sentences. Having produced several sentences with an S-V-PP word order in a row, participants could have been inclined to use a different word order for the next item, even if the resulting word order would not really match their linguistic preferences.

3.3 Conclusion

The results of the experiment described above support the idea of the existence of different preferences working on the production of sentences with a bare or a definite plural subject, which can be in conflict and which are of different importance: the preference for a subject at

the sentence-initial position is very strong, followed by the preference to start a sentence with the topic. Both the Subject First and the Topic First preference influence the word order of sentences with bare and definite plural subjects. For definite plural subjects, the two preferences point in the same direction: a definite plural subject is preferably placed in sentence-initial position. For bare plural subjects, the preferences are in conflict: a bare plural is a bad topic and does therefore not prefer a prominent position. As a subject, on the other hand, it does prefer a prominent position. This conflict results in word order variation. The observation that bare plural subjects are more frequent in preverbal than in postverbal position suggests that the Subject First preference is the strongest of the two in Dutch. Apparently, the fact that the preverbal position is a marked position for bare plurals and that this results in a marked, generic reading of the bare plural subject is of minor importance. Thus, creating a generic sentence is less bad than not starting a sentence with the subject. The relatively high frequency of definite plural subjects in postverbal position could be due to contextual factors not included in this experiment, or could just be random variation. In the next chapter, I will reconsider the notions of prominence and topicality introduced in Chapter 2. I will argue that the drag-and-drop experiment discussed in the present chapter only shows a small part of the influence of prominence on word order, and that there are more factors than definiteness alone that contribute to prominence, and therefore to the choice of a topic.

Prominence and Topicality

4.1 Introduction

The drag-and-drop experiment discussed in Chapter 3 showed that in Dutch out-of-the-blue sentences definite plural subjects are preferably placed in sentence-initial position, while for bare plural subjects this preference is less strong. This can be explained by stating that definites are inherently good topics while non-specific indefinites are bad topics, and that topics in Dutch are preferred in sentence-initial position. The results of the experiment therefore suggest that semantic prominence plays an important role in the placement of bare versus definite plural subjects in Dutch. Even if no context is present to determine the topic-focus structure of the sentence, word order is influenced by the inherent topic characteristics of the different constituents.

However, the experiment dealt only with a restricted set of experimental items, which satisfied a number of restricting conditions. The only topic characteristic manipulated was definiteness of the subject (bare plural subjects vs. definite plural subjects). Other possible sources of variation in word order were kept constant. For example, the subject was always animate, the verb was always intransitive, active and in the present tense, and the PP was always a locative adjunct. Although definiteness is a strong determinant of topicality (both definites and topics tend to refer to old information; see also Epstein (2002)), it is likely that only taking into account the definiteness of a subject is not sufficient in determining its prominence and its topicality, and therefore its position in the sentence. Other factors may be of influence here.

At least two main points spring to mind. First, topic characteristics of other constituents than the subject should be taken into account. It is possible that such a constituent has better topic characteristics than the

subject, and consequently takes over the sentence-initial position, or has worse topic characteristics and causes the subject to be in sentence-initial position. Second, the definiteness of the subject is not the only factor that determines its prominence and therefore its ability to be a good topic. Additional factors might make a definite NP a less good topic or a bare NP a better topic. In this chapter, I will further explore these two options. First, in Section 4.2 I will reconsider the notions of prominence and topicality and give a structured overview of how I will use the terms in the remainder of this thesis, based on the analysis started in Chapter 2. Then, I will discuss topicalization of other constituents than the subject in Dutch in Section 4.3. In Section 4.4 I will show that the notion of topicality can be subdivided into a range of factors that all contribute in their own way to the semantic prominence of a constituent. I will conclude in Section 4.5.

4.2 Topics and prominence revisited

In Chapter 2, the topic of a sentence has been defined as the constituent that the sentence is about.⁹ It has also been noted that topics often contain old information. Choi (1999) developed a model in which topics are defined by two binary discourse features: [\pm New] and [\pm Prominent]. According to this model, topics are [-New] and [+Prom], i.e. they refer to old information and they are discourse prominent (which could be taken to relate to aboutness on this account). Defined this way, topicality is seen as the junction of the prominence on the one hand and given/old information on the other. However, I will not follow Choi in her definition of topicality. I will call topics only those constituents that indicate what the sentence is about. I will not consider the fact that topics often express old information as a defining feature of topics, but rather as a prototypical discourse feature of topics (following e.g. Reinhart, 1982). The reason is that topics not necessarily convey old information. Reinhart (1982) gives the example in (43).

⁹ This is different from the notion of 'discourse topic', which is the topic of a larger portion of speech/text and need not correspond to a single constituent.

- (43) ...*The public benches* that used to be west of their restaurant are gone also, it has been rumoured that *the removal of the benches* has been brought about by pressure from certain business people who want to discourage those who can't afford to get drunk in public behind iron work railings, from annoying those who can. // Of course, one of the consequences is that *the tenants of 1415 Ocean Front Walk* don't have their benches to sit on...
(Beachhead, Venice, CA, December, 79, 15)

The first part of the text in (43) is about (*the removal of*) *the public benches*. Since these benches have been introduced here, they can be considered old information when they are mentioned again in the last sentence. However, the topic of this sentence, in terms of aboutness, is *the tenants of 1415 Ocean Front Walk*, which is new information.

Similarly, definiteness will be treated as a prototypical semantic property of topics. Although there is a strong tendency for topics to be definite, this is not a necessary feature of topics, as is shown for example by the possibility for a generic bare plural to be topic, as in example (44). Neither is it the case, for that matter, that definite NPs always express old information, as is shown by the use of the definite article in (45), from Epstein (2002) (emphasis mine).

- (44) [T Gnomes] live in mushrooms.
- (45) M: Did you hear about *the fight*?
A: What fight?
M: Between Bob and Grandpa...

Both the old/new distinction and definiteness are thus factors that contribute to the topicality of a constituent, and therefore to its semantic or discourse prominence. Both factors can be presented as prominence scales. The prominence hierarchy of the old/new distinction is depicted in (46) below. Old information is more prominent, and therefore a better topic property than new information. The definiteness scale in (21) is repeated here as (47) for convenience. Since I am not

concerned with personal pronouns and proper names in this thesis, I have omitted these categories from the hierarchy.

(46) *Information structure scale* (discourse prominence):
Given/old information > New information

(47) *Definiteness scale* (semantic prominence):
Definite NP > Indefinite specific NP > Non-specific NP

This semantic/discourse prominence of a constituent is contrasted with its syntactic prominence, i.e. its place on the hierarchy of grammatical functions. This hierarchy is presented again in (48).

(48) *Scale of grammatical functions* (syntactic prominence):
Subject > Object > Oblique complement > Adjunct

Since the properties on the left side of the hierarchies in (46) and (47) (i.e. being given/old information and being definite) are highly prominent, constituents that have those properties can be considered good topics. With being a good topic I mean that a constituent has a high probability of being chosen as the topic, that is: it is likely that it constitutes what is talked about in the sentence. Because people often talk about prominent things, prominence co-determines what is going to be the topic.

As soon as a topic has been selected, this constituent is preferred in the preverbal position (Topic First preference). The reason for this is that, as we have also seen in Chapter 2, highly prominent constituents prefer a prominent position, which is taken to be the preverbal, sentence-initial position. A sentence is preferably started with what the sentence is about. The hierarchy of word order is repeated in (49).

(49) *Word order scale*:
Preverbal > Postverbal

I will refer to prominent constituents in preverbal position as topicalized constituents. Thus, topicalization is defined here as the promotion of a constituent to the topic position to the left of the finite verb accord-

ing to the Topic First preference. With this definition I follow Brunetti (2009), who says that topics are “preverbal, sentence initial arguments in sentences with a neutral, descending intonation” (Brunetti, 2009: 3).

In sum, for the remainder of this thesis I will define topics semantically as ‘what the sentence is about’. Topics are typically very prominent. What constitutes a good topic is determined by different semantic and discourse factors that contribute to the prominence of a constituent, such as being definite or being old information. A topic is consequently prototypically placed at the first position of a sentence, following the Topic First preference (a prominent constituent prefers a prominent position). A semantically/discourse prominent constituent that is promoted to the preverbal position can therefore be considered a topicalized constituent.

4.3 Topicalization of other constituents than the subject

Subjects often coincide with topics, because they are syntactically prominent. If, in addition, a subject is semantically prominent as well, for example by being definite, it is also semantically a good topic. If, on the other hand, the subject is not semantically prominent, it will be a less good topic, as we have seen with bare plural subjects. Since the prototypical topic position is the sentence-initial position, semantically prominent subjects will prefer the sentence-initial position. However, prominent constituents other than the subject may also be topicalized. Although constituents that are not subjects are necessarily less high on the syntactic prominence scale, they can still be high on a semantic prominence scale. As a result, constituents like objects or adjuncts may also occupy the sentence-initial position when they have a high semantic prominence and are consequently good topics. This does however not mean that the remaining constituents in the sentence, which are necessarily postverbal, should be non-prominent by definition. For example, definite or generic subjects can easily appear postverbally when another constituent is topicalized.

Since I am only concerned with preverbal (sentence-initial) and postverbal subjects, regardless of their exact syntactic position, we can say that because a variety of constituents can be easily topicalized and appear in sentence-initial position, subjects are more frequent in post-

verbal position in Dutch than in a strict word order language like English. If the topic is not the subject, the subject automatically ends up in a postverbal position, irrespective of its inherent prominence or the type of predicate involved. As a result, a postverbal subject is not required to be a very bad topic; it is enough that another constituent be a better topic.

In the drag-and-drop experiment discussed in Chapter 3, the only constituent present other than the subject (and the verb) was a PP adjunct. In the sentences with a postverbal subject, this adjunct appears in sentence-initial position and is therefore assumed to be topicalized.¹⁰ However, on the grammatical functions scale given in (48) above, adjuncts are presented as lowest in syntactic prominence. Thus, according to this scale, adjuncts are not preferred in a very prominent position. Indeed, Brunetti (2009) only talks about *arguments* in her definition of topics. But PP adjuncts or adverbs are often topic in Dutch, as in other topic prominent languages. This is especially the case when they refer to the spatial or temporal setting of the predicate in question. Crasborn et al. (2009) call these topics *spatio-temporal topics*. Spatio-temporal adjuncts situate the event described in the sentence in a concrete spatial or temporal location. Therefore, they are highly suitable to serve as sentence topics (cf. Givón, 1976). Thus, although adjuncts have low syntactic prominence, they can be semantically prominent when they refer to a concrete location. Concreteness might therefore be considered another good topic property. But still other factors may influence topicality as well, to which I will now turn.

¹⁰ It is difficult to distinguish between a topicalized locative PP and a PP that has undergone locative inversion: in both cases the PP is in sentence-initial position, while the subject is postverbal. When in addition an expletive *er* 'there' appears postverbally, such as in (22) above, it can be assumed that the locative PP has been topicalized, since the PP and *er* are in complementary distribution (i.e. locative inversion always involves replacement of *er*). However, I will not make this distinction here.

4.4 Other factors contributing to prominence

Up till now I have concentrated on definiteness as a determinant of topicality. Indeed, topics are often definite, because they typically tap on the common ground of speaker and hearer. However, it has also become clear that definiteness is not the only factor determining the topicality of a constituent. It has already been noted that topics are also prototypically agents and typically convey old information. In addition, the discussion in the previous section suggests that concreteness should also be considered a prototypical topic property. Givón (1976) proposes several hierarchies of properties that make a constituent a more or a less prototypical topic. Some of these hierarchies are given in (50) below:

- (50) good topic >> bad topic
 human > non-human
 definite > indefinite
 agent > dative > patient
 1st person > 2nd person > 3rd person
 possessor > possessed

The reason why the properties on the left side of the hierarchies in (50) make good topics, according to Givón, is that we, as humans, like to talk about ourselves or at least about individuals that are similar to ourselves, such as other humans or entities that are considered to have control or volition in their actions, such as agents or possessors. These are all semantic properties that contribute to the prominence of a constituent in terms of aboutness: sentences are often about things that are either humans, or have human-like properties. In the remainder of this thesis, I will mainly concentrate on such intrinsic properties of constituents that contribute to the semantic or discourse prominence of topics. Factors dependent on the (physical or linguistic) context will not be my main focus, because they are difficult to detect.

Since prominent constituents are preferably combined with a prominent position in the sentence, constituents that refer to agents, humans and/or possessors will be good candidates for the preverbal, sentence-initial position. This can also be understood in terms of iconic-

ity: since (spoken) language is mostly a linear system, there always has to be some ordering in the linguistic structure. It seems logical, then, to map natural orderings perceived in the real world onto this linear linguistic structure. For example, causal events always start out with a causer: the entity that initiates the event. It can thus be seen as iconic to also start a description of the event with the causer. Since it is often agents and humans that initiate events, they are frequently placed at the beginning of a sentence. Likewise, entities that are affected by the event come only into vision when the result of the event becomes clear, i.e. when it has already ended. Hence, such entities (causees, patients, inanimates) are placed further to the end of the sentence, in a non-prominent postverbal position.

It is clear that the semantic prominence of a constituent is determined by multiple factors, such as animacy, agentivity, definiteness and concreteness. These factors can themselves be represented as prominence hierarchies: constituents high on these hierarchies contribute to prominence more than constituents low on the hierarchies. Thus, semantic prominence can be seen as a cover term for a set of different prominence hierarchies. However, a constituent high on one of those hierarchies could at the same time be low on others. For example, a referent can be high in animacy, but low in definiteness. As a result, these two factors are in conflict: they influence semantic prominence in opposite directions. If the influence of both factors is equally strong, they would probably cancel each other out, leaving a constituent with neither high, nor low semantic prominence. However, it is likely that some factors contribute stronger to prominence than others. Hence, in case of a conflict, stronger factors might overrule the influence of weaker ones. Since semantically prominent constituents are inherently better topics than constituents that are not semantically prominent, we can also say that factors like animacy, agentivity, definiteness and concreteness contribute to topicality. Constituents with many good topic properties (human, agent, etc.) are inherently good topics, which makes them good candidates for the sentence-initial position. Constituents with few good topic properties, on the other hand, are inherently bad topics, and are consequently dispreferred in a prominent position.

Coming back now to the placement of bare and definite plural subjects in Dutch, we have seen that there is a possible conflict between a

syntactic preference (Subject First) and a semantic preference (Topic First). Now, it is clear that there might be additional mutual conflicts between different semantic preferences. In turn, these semantic preferences may together cause a conflict with a syntactic preference. In the production experiment described in Chapter 3 such additional conflicts were not likely, because all semantic factors except definiteness were kept constant. For example, when a subject was definite, it was considered a good topic in the experiment; hence, no conflict between Subject First and Topic First would occur. However, when this definite subject is at the same time inanimate, we do have a conflict: being high on definiteness makes the constituent a good topic, but being inanimate reduces semantic prominence, causing the constituent to be a less good topic. Now it will depend on the relative strength of the two factors whether the subject will eventually be a good or a bad topic. If definiteness is stronger than animacy, the effect of being inanimate might be less manifest, causing the subject to retain its strong semantic prominence. The subject will then still be a relatively good topic. If animacy is stronger than definiteness, on the other hand, the subject will overall be a bad topic, which will consequently result in a conflict between Subject First and Topic First.

For bare plural subjects, the conflict between Subject First and Topic First, resulting in word order variation as shown in the production experiment, may either be reinforced or weakened by including more semantic factors that influence prominence. For example, when a bare plural subject is inanimate, this might make the subject an even less good topic, resulting in a greater preference for a non-prominent position. When it is animate or human, on the other hand, it will be a better topic. Again, the relative strengths of the different factors will have to determine whether an animate bare plural is a good enough topic to wipe out the conflict with the Subject First preference or not. A complicating factor with bare plurals is further that a difference in position is associated with a difference in meaning: a topicalized bare plural is likely to receive a generic interpretation. For example, Cohen and Erteschik-Shir (2002) argue that topic-focus structure parallels the division in sentence structure between restrictor and nuclear scope: topics correspond to the restrictor, and focused constituents correspond to the nuclear scope. Following Diesing's Mapping Hypothesis, this would

mean that topics are in a VP-external position and receive a strong reading, while focused constituents reside within the VP and receive a weak reading. As for bare plural subjects, they are predicted to have a generic reading when they are topic, and to have an existential reading when they are in focus. If a generic interpretation is not desired, the sentence-initial position might be avoided, even if the bare plural is otherwise semantically prominent (animate, agent, etc.).

Rather than posing such a strict relation between syntactic position, topicality and semantics of bare and definite plural subjects, I would like to take an approach that takes into account different preferences on sentence production, some of which might be in conflict with each other. Since the preferences are not absolute, sometimes one preference is satisfied, while another is violated, depending on their relative strengths. Thus, there is not one factor that dominates in determining the choice for a particular word order. Rather, it is the interaction of multiple factors with variable strengths that can account for the variation in word order found in Dutch main clauses.

Summarizing, two main preferences have been identified that are assumed to influence Dutch word order: the Subject First preference and the Topic First preference. It has been made clear that what constitutes a good topic is dependent on different properties relating to semantic and discourse prominence. I have concentrated mainly on properties intrinsic to a given constituent, since these can be identified relatively easily. Contextual factors, on the other hand, are difficult to capture and were not taken into account. I have briefly illustrated that the different topic characteristics could either point in the same direction or be in conflict. These conflicts are a possible source of word order variation in Dutch.

4.5 Conclusion

In this chapter, I have taken a slightly more structured view on topicality than in Chapter 2 and 3. I have defined topics semantically as 'what the sentence is about' and syntactically as the preverbal constituent in Dutch main clauses. Since topics are often, but not necessarily old information, I consider being old information as a prototypical, rather

than a defining, property of topics. Similarly, definiteness is also considered as a prototypical topic property, but on a more semantic level.

I have argued that besides looking at the topicality of bare and definite subjects themselves, the topicality of other constituents in the sentence also needs to be taken into account. Furthermore, there are more factors than definiteness that contribute to the semantic prominence of a subject and therefore make it intrinsically topical. In the next chapter, I will present a corpus study in which a number of these topic characteristics are investigated more systematically. I will show that whereas in the production experiment discussed in Chapter 3 definiteness was shown to have an effect on word order of sentences with bare or definite subjects, while other factors were kept constant, in natural language other semantic factors than definiteness also have an effect on word order in Dutch.

There is more than definiteness: a corpus study

5.1 Introduction

In Chapter 3, I described a drag-and-drop experiment that tested the influence of the Subject First and the Topic First preference on the word order of Dutch main clauses with bare and definite plural subjects. What determined topicality was solely based on the definiteness of the subject: definite subjects were considered good topics, whereas bare subjects were less good topics. In Chapter 4, it was pointed out that although definiteness strongly correlates with topicality, other factors are likely to also contribute to topicality and therefore to have an effect on word order. Furthermore, it is not sufficient to look at the topicality of the subject alone: even though a subject is syntactically prominent, another constituent in the sentence may be a good topic. Topicalization of this constituent will cause the subject to be in a postverbal position.

In the drag-and-drop experiment only definiteness was manipulated while other sources of variation were kept constant. However, in natural language nothing is kept constant, and all kinds of factors may influence word order. To be able to better map the interaction of multiple factors that are expected to contribute to topicality, and to better predict the influence of topicality on the behaviour of bare and definite plural subjects in Dutch, one should look at natural language data. Therefore, a corpus study was conducted on the position of bare and definite plural subjects in Dutch, using the Corpus of Spoken Dutch (Corpus Gesproken Nederlands, CGN). The CGN is a large corpus of spoken language, containing both spontaneous and prepared speech and both dialogues and monologues. It consists of about 9 million

words, spoken by native speakers of Dutch in The Netherlands and Belgium. The corpus can be searched on orthography, phonology, part of speech and metalinguistic data. In addition, a selection of 1 million words has been annotated with syntactic information.

From the CGN, the same kind of sentences as were examined in the drag-and-drop experiment were selected, i.e. main clauses in which the preverbal (sentence-initial) position is either filled by a bare or a definite plural subject, or by a PP adjunct.¹¹ In this way, the position of the bare and definite plural subjects in the corpus can be compared to what was found in the experiment, and thus it should be possible to see if the results of the experiment can be corroborated. In addition, given that corpus data is unconstrained and non-manipulated, the effects of multiple factors that were not included in the experiment can be investigated simultaneously by using the right statistical methods.

In this chapter, I will give a description of the corpus study as it was conducted. In Section 2, I will present an overview of the different factors that were selected and I will discuss what the predictions are regarding their possible effects. In Section 3, I will describe how the data were selected and annotated. Then, in Section 4, the results of the corpus study will be presented for a subset of the data concerning only sentences with bare plural subjects. First, the effects on word order of each of the factors separately will be examined in a quantitative exploration, and after that, all relevant factors will be included simultaneously in a multifactorial model using a logistic regression analysis. In Section 5, the results for the subset containing sentences with a definite plural subject will be presented in a similar way. I will conclude in Section 6.

¹¹ In contrast to the experimental items, the sentences from the corpus selection were allowed to also contain other constituents, such as direct objects or adverbials, as long as it contained at least a plural subject, a verb and a PP adjunct. Although they might have an influence on word order as well, other constituents are left undiscussed.

5.2 The factors

An inventory was made of possible factors that were likely to play a role in determining topicality, and therefore word order in Dutch, and which could straightforwardly be operationalized in the corpus study. The point of departure in selecting the factors was formed by the restrictions on the items in the drag-and-drop experiment. For example, the subject was always either preceded by a definite article, or not preceded by anything, and it was always animate. The verb was always active and in the present tense, and the PP adjunct was always a locative. Table 3 gives an overview of all factors that were investigated in the present study.

Table 3 Overview of the selected factors in the corpus study

<u>Factors</u>
1 Specificity of the subject
2 Animacy of the subject
3 Presence of <i>er</i> 'there'
4 Tense
5 Aspect
6 Voice
7 Type of PP

Not all factors will, if at all, affect sentences with bare plural subjects and definite plural subjects in an equal manner. For example, the specificity of the subject has been defined differently for bare and for definite subjects. Similarly, the presence of *er* 'there' is expected to have an effect on sentences with bare subjects, but not on sentences with definite subjects. Therefore, two separate analyses were performed; one for sentences with bare plural subjects and one for sentences with definite plural subjects. Consequently, two different statistical models will be created in the multifactorial analysis. In Chapter 6, the two models will be brought together again. In the following sections, I will discuss the different factors under investigation separately for the two types of sentences.

5.2.1 Sentences with bare plural subjects

5.2.1.1 Specificity of the subject

In the production experiment, sentences with a bare plural subject were compared to equivalent sentences with a definite plural subject. Bare plural subjects included only NPs not preceded by anything. However, bare plurals can also be modified. Modified bare NPs could be considered somewhat more specific than non-modified bare NPs, because a modifier can narrow down the referent to a more specific set. In Italian, for example, bare plural subjects are normally disallowed, but they become significantly better when modified (Chierchia, 1998; Longobardi, 2001; Krifka, 2004). Consider the following examples:

- (51) a. **Hanno telefonato studenti.*
 have phoned students
 b. *Hannotelefonato studenti che volevano sapere la data dell'esame.*
 have phoned students who want find.out the date of.the.exam
 'Students who want to find out the date of the exam have phoned.'
 (from Chierchia, 1998)
- (52) a. **Elefanti possono creare grandecuriosità.*
 elephants can create great curiosity
 b. *Elefanti di colore bianco possono creare grande curiosità.*
 elephants of.the colourwhite can create great curiosity
 'White-coloured elephants may raise a lot of curiosity.'
 (from Longobardi, 2001)

While he argues that bare nominals essentially are kinds, Chierchia (1998) observes that not every bare nominal constituent can act as a kind. When bare nominals are modified, they do not always correspond to a regular established kind. *Depressed pink zebras in our kitchen*, for example, is a modified bare nominal that is not likely to constitute a

regular kind for most people. Chierchia therefore argues that such NPs are not type shifted to kinds, but to existentially quantified NPs. They would then be interpreted as ordinary weak indefinites, with the accompanying possibility of specific interpretations. This would mean that modified bare nominals would allow for referential readings, contrary to non-modified bare nominals.

We have seen that, because they are non-specific indefinites, bare plurals typically prefer a non-prominent, postverbal position. Based on the observations above, it is expected that bare NPs preceded by an adjective are positioned somewhat more towards definite NPs on the definiteness scale than purely bare NPs (see 47)). We might therefore formulate another prominence scale, as a subpart of the definiteness scale:

- (53) *Definiteness subscale:*
 Modified bare plural > Non-modified bare plural

Since definite/specific NPs are better topics than indefinite/non-specific NPs, modified bare plurals could be somewhat better topics than their non-modified counterparts. This could make it more likely to find them in sentence-initial position.

I did not include bare plural subjects in the corpus data set that were modified by a relative clause, such as in (51b), because such constituents tend to become quite long. This makes it very likely that a factor like heaviness is going to play a role in the position of the bare plural subject, which could obscure possible effects of other factors. However, I did include bare plural subjects modified by an adjective (excluding NPs preceded by a quantifier like *alle* 'all' or *sommige* 'some', since such NPs are considered definite, see Section 5.3.1). Thus, the bare plural subjects in the data set can be divided into 'true' bare NPs and bare NPs preceded by an adjective. Because bare plurals modified by an adjective are considered less non-specific than true bare plurals, I will name this factor *Specificity of the subject*.

If it is true that bare NPs modified by an adjective are better topics than non-modified bare NPs, we would predict that the two levels of this factor show a difference in their influence on the placement of bare plural subjects. Modified bare plural subjects are expected to be more

frequent in sentence-initial position than non-modified bare plural subjects.

5.2.1.2 Animacy of the subject

It has been shown in a wide range of studies (e.g. Bock and Warren, 1985; Van Nice and Dietrich, 2003; Hendriks, De Hoop and Lamers, 2005; Rosenbach, 2005; Bresnan, Cueni, Nikitina and Baayen, 2007; Lamers, 2007; Bresnan and Hay, 2008; Lamers and De Hoop, 2008; Van Bergen, to appear) that animacy has clear effects on word order. There is a universal preference in languages for animate NPs to precede inanimate NPs. Just as for definiteness, there is also a prominence scale for animacy (Aissen, 2003), which is shown in (54).

- (54) *Animacy scale:*
Human > Animate > Inanimate

Human NPs are more prominent than animate NPs, which in turn are more prominent than inanimate NPs. Humans are highly salient in discourse: people like to talk about sentient and volitional beings (Givón, 1979). Because topics are usually prominent, being human or animate can be considered an important topic property. For instance, in a sentence in which the subject is inanimate, and in which there is another constituent that is higher in animacy, this other constituent is likely to occupy the topic position, even if it is not an agent (Lamers, 2007; Lamers and De Hoop, 2008; Brunetti, 2009).

The high percentages of preverbal bare subjects found in the production experiment (see Table 2 above) could partly be due to the fact that the items included only human and animate subjects. In the corpus, subjects can also be inanimate. When a bare plural subject is inanimate, this could reinforce the tendency for non-specific NPs to take up a less prominent position. Thus, the prediction is that animacy influences the placement of bare plural subjects: it is expected that inanimate bare plural subjects are more frequent in a non-prominent, post-verbal position than human/animate bare plural subjects, which will be more frequent in sentence-initial position.

5.2.1.3 Presence of *er* 'there'

In the production experiment participants had two options: either the subject could be placed in preverbal position and the PP in postverbal position, or the PP could be placed in preverbal position and the subject in postverbal position. What was not included in the experiment was the possibility to include an expletive *er* 'there' in the sentence.

In Dutch, expletive *er* 'there' is commonly found in the sentence-initial position of presentative sentences, which causes all other constituents to be postverbal. This so-called presentative *er* is assumed to have some locative properties, since it is homophonous with locative *er*, and also semantically related in that both types of *er* can introduce the (spatio-temporal) context of the sentence (Coppen, Haeseryn and De Vriend, 2002). In addition, presentative *er* can easily be replaced by a locative PP in Dutch (locative inversion). With locative inversion, the expletive must be omitted, because the fronted PP fills the same structural position (the standard subject position). Thus, *er* and the fronted locative cannot co-occur (Zwart, 1992).

In this study, I will not be concerned with *er*-initial sentences, because only sentences in which the preverbal position is filled by the subject or by a PP adjunct are investigated. However, even if the sentence-initial position is filled by a locative PP, *er* 'there' can still be present postverbally in Dutch, as shown in (55).

- (55) *Aan de muur hingen er foto's.*
 on the wall hung there pictures
 'There hung pictures on the wall.'

Zwart (1992) argues that such sentences should not be analyzed as cases of locative inversion, since with locative inversion the fronted PP and *er* cannot co-occur. Instead, *aan de muur* 'on the wall' must be a topicalized adjunct. In addition, Zwart (1992) argues that the fronting of a temporal adjunct, such as in (56), should not be analyzed as locative inversion, and consequently that omission of the expletive is not possible when a temporal adjunct is fronted. According to him, this is because temporal adjuncts cannot take up the standard subject position. A fronted temporal adjunct should thus be analyzed as topicalization.

- (56) *Op zaterdag hingen*(er) foto's.*
 on Saturday hung there pictures
 'On Saturday, there were hanging pictures.'

As mentioned above, I will not make the distinction between locative inversion and topicalization here. Rather, I will take a more functional approach to the use of *er* 'there' in Dutch sentences with a PP adjunct, based on Grondelaers and Speelman (2007). Grondelaers and Speelman analyze presentative *er* as an inaccessibility marker: it signals a pragmatically inaccessible subject, a subject that cannot be predicted by inferencing on the preceding context. Since we have defined prominence in terms of accessibility in Chapter 2, inaccessible constituents can be considered non-prominent. Inaccessible subjects are typically indefinite and non-referential, and contain new information. In other words, they lack prototypical topic properties. As a result, they are expected to be frequent in non-prominent positions. In addition, since *er* introduces an inaccessible subject, it should precede the subject to have an effect. Because in principle nothing can precede the preverbal position in Dutch main clauses, the subject should be in a postverbal position for *er* to function.

Adjunct PPs, on the other hand, provide a means to access an otherwise inaccessible subject. They help the hearer in building predictive inferences about the subject through the lexical meaning, syntactic structure and world knowledge associated with the PP. In presentative sentences with sentence-initial *er* 'there' the occurrence of such a PP adjunct before the postverbal subject would cancel the effect of *er*, since whereas *er* marks the subject as inaccessible, the adjunct makes it appear as if the subject can be predicted. In addition, abstract PP adjuncts have less inferencing power than concrete locative adjuncts (see Section 5.2.1.7). As a consequence, the former are found more frequently before an inaccessible subject accompanied by postverbal *er*.

In sentences with an initial PP, such as (55), postverbal *er* is present more frequently as the PP is less concrete. Since abstract PPs do not have very strong predictive inferences, the subject remains inaccessible, and postverbal *er* is preferred. This might explain the ungrammaticality resulting from the omission of *er* 'there' in (56). However, omission of

the expletive in sentences with an initial temporal PP is possible in other cases, as (57), from Grondelaers and Speelman (2007), shows.

- (57) *Op die dag komen (er) ook speciale acties.*
 on that day come there also special actions
 'On that day there will also be special actions.'

When the PP is concrete (e.g. locative), its predictive inferences about the subject are blocked when it is followed by a postverbal *er*, resulting in a preference to leave out the expletive. Thus, the difference between *er*-initial sentences and sentences with postverbal *er*, according to Grondelaers and Speelman, is that in the latter *er* blocks all predictive inferences raised by the PP, while in the former *er* deletes all previous occurrences of the upcoming material (the subject) in the discourse, such that it cannot be interpreted as topical information. As said above, I will only look at occurrences of postverbal *er* 'there' in the present corpus study, since sentence-initial *er* yields a word order where both the subject and the PP are postverbal. I will leave such constructions for future research.

In sum, the main prediction for the occurrence of postverbal *er* is that the presence of *er* influences the position of bare plural subjects relative to the PP: since *er* signals an inaccessible subject, and inaccessible subjects are non-topical, subjects that are accompanied by *er* should be in a non-prominent position (postverbal, after *er*). In addition, postverbal *er* should occur especially when the initial PP is abstract.

5.2.1.4 Tense

The production experiment only included present tense verbs. However, it might be the case that the tense of the verb has an influence on the position of the bare plural subject. The present tense can be used for habitual or general statements, and therefore combine easily with a generic reading of the bare plural subject. Past tense verbs, on the other hand, tend to place the event that is described in a situational, episodic context. Thus, past tense verbs are expected to combine less easily with a generic reading of the subject. If this is true, one would expect generic readings of bare plural subjects to be less frequent with past tense verbs than with present tense verbs. Since NPs with a generic reading are

more prominent than NPs with an existential reading, less bare plural subjects in a prominent, preverbal position are expected with past tense verbs. The main prediction is then that bare plural subjects occur less frequently in sentence-initial position when the main verb is in past tense than when it is in present tense.

5.2.1.5 Aspect

Something similar as for tense can be said for aspect: the production experiments only included verbs in the (unmarked) imperfective aspect. However, it might be the case that verbs with perfective aspect trigger an episodic reading of the sentence.¹² Again, we would expect generic readings of bare plural subjects to be less frequent with perfective aspect verbs than with imperfective aspect verbs. Consequently, less bare plural subjects in preverbal position are expected with perfective aspect verbs than with imperfective aspect verbs.

In addition, however, verbs with perfective aspect might be considered more transitive than verbs with imperfective aspect. According to Hopper and Thompson (1980), perfective aspect indicates that an action is “more effectively transferred to a patient” (Hopper and Thompson, 1980: 252). As a result, subjects of perfective verbs may be considered more agentive (they have more control over the patient) than subjects of imperfective verbs. For example, Hopper and Thompson cite a number of languages in which agents receive ergative case when they are subjects of perfective verbs, while they have absolutive or nominative case when they are subjects of imperfective verbs.

It might be the case, then, that this property of aspect also has an influence on the placement of bare plural subjects in Dutch. Subjects of perfective verbs would be higher on the prominence scale of agentivity given in (58) than subjects of imperfective verbs.

- (58) *Agentivity scale*
Agent > Recipient > Patient

¹² The same would probably hold for progressive aspect, but since progressive aspect is commonly expressed by the imperfective form of the verb in Dutch, I will not consider it.

Since high agentivity is a good topic property, subjects of perfective verbs might be considered better topics than subjects of imperfective verbs. Consequently, bare plural subjects of verbs with perfective aspect might have a somewhat higher preference for the sentence-initial position than bare plural subjects of verbs with imperfective aspect. However, since the sentence-initial position triggers a generic reading of the bare plural subject and this is less compatible with an episodic (perfective) reading of the verb, it is difficult to predict the preferred position of a bare plural subject in this case.

5.2.1.6 Voice

Another factor that might have an influence on the placement of bare plural subjects is *voice*, or in this case basically the difference between active and passive sentences in the corpus data set. In a passive sentence, the direct object of a transitive verb becomes the subject, whereas the original subject is either left out or demoted to an oblique phrase. It could be the case that passivization occurs when the original direct object (the patient) is highly topical and consequently is promoted to the prominent subject position. As Givón (1979) notes:

“In general, the function of passive sentences in language is to code sentences in the context in which *the non-agent is more topical*. This automatically means that the agent is *less topical* in a passive sentence [...]” (Givón, 1979: 57)

However, as Brunetti (2009) points out, it is more likely that it is just the other way around. The function of passivization is to indicate that the agent (the original subject) is not important, i.e. has low prominence, in the sentence. Thus, the agent is demoted, and it follows that some other constituent has to become subject. Although this will often be the patient, this does not mean that this patient therefore automatically becomes more topical. It is still a patient, which is a thematic role low on the semantic prominence scale in (58) above. In fact, Brunetti shows that when a third constituent is present that is more agentive than the patient subject (e.g. an indirect object), this constituent is preferred in topic position over the subject.

Thus, because subjects of passive verbs often bear the thematic role of the patient, they could be considered inherently degraded subjects; they are low on the prominence scale given in (58) above. Subjects of passive sentences might therefore be less good topics than subjects of active sentences. If this is true, we would expect to find a higher frequency of postverbal bare plural subjects in passive sentences than in active sentences in the corpus data set. Thus, the main prediction is that bare plural subjects of passive sentences are found more frequently in postverbal position than bare plural subjects of active sentences.

5.2.1.7 Type of PP

Every item in the production experiment contained a locative PP adjunct. This allowed for locative inversion to occur with bare plural subjects, causing variation in the position of the subject. Apart from locative inversion, PPs can also be in preverbal position due to topicalization. However, I do not wish to pursue the distinction between locative inversion and topicalization any further. I will assume that locative inverted PPs are topics (they are in sentence-initial position), and therefore I will call both operations ‘topicalization’. According to the prominence hierarchy of syntactic functions in (48), repeated here as (59), adjuncts are low in syntactic prominence.

- (59) *Scale of grammatical functions (syntactic prominence):*
 Subject > Object > Oblique complement > Adjunct

Thus, according to their syntactic prominence, PP adjuncts are not easily selected as topics (see also Brunetti, 2009). Yet, there might be semantic properties that increase the prominence of PP adjuncts, and consequently make them better topics.

In Dutch, topicalization is possible for all kinds of PP, but it might be the case that some types of PP are more easily topicalized – have better topic characteristics – than others. This would mean that with PPs that are good topics, bare plural subjects would be more frequent in postverbal position, while with PPs that are bad topics, more bare plural subjects would go in preverbal position. Thus, the type of PP might be another factor that could influence the position of a bare plural subject.

In the corpus set, all sentences were required to contain one PP adjunct, such that there was a fixed constituent in every sentence, next to the subject, that could in principle be topicalized. Three types of PP are distinguished: locative PPs, temporal PPs and abstract PPs. Locative and – to a lesser extent – temporal PPs express concrete localization in space or time of the situation described in the sentence. Temporal PPs could be considered somewhat less concrete than locative PPs, because time is generally perceived as a more abstract concept than space. For example, descriptions of time are often derived from how we talk about space (e.g. ‘a *long* night’; ‘*on* Saturday’). Nevertheless, both types of PP are high on the prominence scale of concreteness given in (60) below.

- (60) *Concreteness scale:*
Concrete > Abstract

Since locative and temporal PPs are typically prominent, it is expected that they can easily be topicalized, and consequently are likely to be found in sentence-initial position. Abstract PPs, on the other hand, do not express a concrete spatio-temporal localization. Therefore, they can be considered less good topics, and predicted to prefer a non-prominent, postverbal position. If, at the same time, the subject of the sentence is bare, the subject will also typically prefer the postverbal position. A bare plural subject will then compete with the abstract PP for this position. If both preferences are assumed to be equally strong, some abstract PPs will be expected to appear sentence-initially when the subject is bare, even though they are bad topics.

In addition, inaccessible, postverbal subjects that are preceded by an abstract PP are expected to be accompanied by the inaccessibility marker *er* ‘there’, because abstract PPs are not very suitable to create predictions on the identity of the subject. Thus, in the PP-V-Subj word order with an abstract PP and a bare subject, a higher occurrence of (postverbal) *er* is predicted than in the same word order with a concrete PP.

In sum, the main prediction is that locative PPs will be the most frequent in preverbal position when there is a bare plural subject. There will also be a lot of temporal PPs in this position. Abstract PPs are less expected in preverbal position, but due to a conflict in preference with

bare plural subjects, they are still expected to occur in that position, sometimes accompanied by *er* 'there'.

5.2.2 Sentences with definite plural subjects

5.2.2.1 Specificity of the subject

In the production experiment, the definite plural subjects included only NPs preceded by a definite article. In natural language, other definite determiners than the definite article can be found, and these might behave differently with respect to topicality. Some determiners may cause an NP to be a somewhat better or worse topic than others. For example, it could be argued that determiners such as demonstrative and possessive pronouns are more explicitly referential because they allow for deictic interpretations. Strong quantifiers, like *alle* 'all', *somme* 'some' or *de meeste* 'most', on the other hand, are much more opaque in their denotation, and might therefore be considered less prototypical definites (cf. Haspelmath, 1997). Taking referentiality as a prototypical topic characteristic, it follows that subjects preceded by a demonstrative or possessive pronoun are very good topics, while subjects preceded by a strong quantifier are less good topics. Subjects preceded by a definite article are expected to be somewhere in between. We could further subdivide the prominence hierarchy for definiteness in (47) to include these different kind of definites, as in (61).

- (61) *Definiteness subscale:*
 Demonstrative/Possessive > Definite article > Strong quantifier

It has already been found in the production experiment that NPs preceded by a definite article usually occur in sentence-initial position. If it is true that different determiners have different topic properties, it is expected that they also have a different effect on word order. Subjects with the best topic characteristics, i.e. those preceded by a demonstrative or a possessive pronoun, are predicted to occur more frequently in sentence-initial position than subjects with less good topic characteristics, i.e. those preceded by a strong quantifier. Note that being a 'less good topic' here is only relative to other types of definite plural sub-

jects. Subjects preceded by a strong quantifier are still expected to be better topics than bare plurals. For example, they pattern with other definites in that they do not usually occur in existential *er*-sentences.

To be consistent in naming the factors across the two models (bare vs. definite), I will refer to the factor under consideration here as *Specificity of the subject* as well. It is predicted that subjects preceded by a strong quantifier will be the least topical and relatively more frequent in postverbal position, while subjects preceded by a demonstrative or possessive pronoun will be the most topical and the most frequent in preverbal position.

5.2.2.2 Animacy of the subject

The factor Animacy is defined in the same way for definite plural subjects as for bare plural subjects. Inanimate definite plural subjects are expected to be more frequent in postverbal position than animate definite plural subjects. However, in contrast to bare plural subjects, definite plural subjects are already typically good topics. As a result, the effect of animacy might be less apparent with definite plural subjects: when they are animate, the preference for the preverbal position will be reinforced; when they are inanimate, however, the preference for the preverbal position will still be strong because of definiteness.

5.2.2.3 Presence of *er* 'there'

Since definite subjects are usually referential, they are likely to be accessible in discourse. Therefore, it seems unlikely that a definite NP would be the subject in a presentational *er*-sentence, since *er* 'there' signals an inaccessible subject. This prediction is captured by the definiteness effect mentioned in Section 2.5.4. Some cases can be found in which definite subjects appear in existential sentences, such as in (62) below, from the CGN.

- (62) *Maar goed dus voor die tijdgevoelige technologieën*
 but good so for those time-sensitive technologies
is er de zogenaamde future and emerging technologies
 is there the so.called
 [...] [fn000056.42]

'Anyway, so for those time-sensitive technologies there is the so-called future and emerging technologies.'

However, since such cases seem to be marginal, there will be hardly any to no definite plural subjects in existential *er*-constructions in the corpus set. Therefore, this factor will not be included in the model for the definite plural subjects.

5.2.2.4 Tense

For definite plural subjects, a difference in syntactic position is not associated with a semantic difference. Neither generic, nor existential readings are possible. Therefore, the tense of the verb is not expected to trigger any interpretation differences of definite plural subjects; hence, the prediction is that the factor Tense will not have an effect on the position of definite plural subjects.

5.2.2.5 Aspect

The fact that perfective aspect triggers a certain reading of bare plural subjects does not affect the interpretation and/or the position definite plural subjects. However, the relation between aspect, transitivity and agentivity noted by Hopper and Thompson (1980) could have an effect on the position of definite plural subjects. Since verbs with perfective aspect are more transitive than verbs with imperfective aspect, their subjects can be considered more agentive. As a consequence, definite plural subjects of perfective verbs are more prototypical topics, making them even more preferable in sentence-initial position. Of course, definite subjects are already highly topical by themselves, which means that the contribution of high agentivity will not be very apparent, but if an effect of aspect is found, it is expected that definite plural subjects of perfective verbs are more frequent in sentence-initial position than definite plural subjects of imperfective verbs.

5.2.2.6 Voice

As with bare plural subjects, the difference between active and passive sentences is also expected to have an influence on the position of definite plural subjects: because subjects of passive verbs often bear the thematic role of the patient, they are less prominent on the agentivity

scale. Therefore, they are expected to be less topical and to be more frequent in a non-prominent, postverbal position. However, this effect should be less strong for definite than for bare subjects, since definite subjects are already good topics. Thus, the main prediction is that definite subjects of passive sentences are more frequent in postverbal position, but because these subjects are definite, a lot of them will also be placed in preverbal position.

5.2.2.7 Type of PP

When a PP has a concrete (spatiotemporal) denotation, it becomes a more likely candidate for topicalization. Concrete PPs, such as locative PPs, are therefore expected to be frequent in preverbal position, automatically causing the subject to be postverbal. Definite subjects, however, are also good topics. As a result, they will compete with concrete PPs for the sentence-initial position. Consequently, more postverbal concrete PPs are expected when the subject is definite than when the subject is a bad topic (i.e. a bare plural).

In contrast, when a PP has an abstract denotation, it will be less suitable for the topic position, and hence it is expected that abstract PPs will be most frequent in postverbal position. Since this does not cause a conflict with – topical – definite subjects, Subject-Verb-PP word orders will be expected to be predominant. Thus, the main prediction is that abstract PPs are very frequent in postverbal position, and that locative and, to a lesser extent, temporal PPs appear more often in preverbal position, but not as frequently as with bare plural subjects.

5.3. Methods

5.3.1 *Materials*

From the Corpus Gesproken Nederlands (Corpus Spoken Dutch, CGN), 223 sentences with a bare plural subject and 334 sentences with a definite plural subject were selected, resulting in a total of 557 sentences. The selection was restricted to only main clauses that contained a plural subject and a PP adjunct. The PP adjunct was mandatory because its presence ensures that there are always at least two constituents that can compete for the sentence-initial position. Recall that in

Dutch the preverbal position always has to be filled. If a sentence has only one argument/constituent (most probably the subject) that can fill this position, there is no choice and the constituent will be in preverbal position regardless of whether it is a good or a bad topic. As a result, nothing can be said about the effects of topicality on word order preferences when considering such sentences.

Both intransitive and transitive sentences, active, passive and reflexive constructions in all tenses were included in the set. All subjects were plural; they were either bare nominals, or NPs preceded by a modifier, a quantifier, a definite article, a demonstrative pronoun or a possessive pronoun. Subjects could be accompanied by a postverbal occurrence of the expletive *er* 'there'. PPs were locative, temporal or abstract. All selected sentences occurred in one of two different word orders: one in which the preverbal position was occupied by the subject (Subj-V-PP), and one in which it was occupied by the PP adjunct (PP-V-Subj). Word orders other than these two are possible in Dutch: next to Subject-V-PP and PP-V-Subject, X-Verb-Subject-PP and X-Verb-PP-Subject are possible, where X can be filled by some other constituent, such as *er* 'there', a direct object or an adverb. However, including these additional word orders would make the analysis quite complex, so that I will stick to the two word orders that were also used in the experiment.

A number of constructions have been left out of the selection. First of all, subordinate clauses and interrogative sentences were excluded, because these involve different word orders (verb-final and verb-initial or wh-initial, respectively). Secondly, only sentences with verbal predicates have been considered. Sentences with a subject complement were excluded, because they have a different syntactic structure than sentences with a verbal predicate. Thirdly, all bare subject NPs that were modified by a relative clause were also omitted. The reason for this is that relative clauses can be separated from the NP they modify. In addition, conjunctions of subjects were left out as well. Subjects preceded by an indefinite quantifier, such as *veel* 'many', *twee* 'two', *allerlei* 'all kinds of', *enkele* 'some' or *verschillende* 'various' have not been considered either, because quantified NPs are not bare nouns.

Furthermore, sentences containing a PP with a directional meaning have not been taken into account, because contrary to other kinds of

PP, directional PPs cannot be easily placed in sentence-initial position. Consider for example the sentence in (63).

- (63) *?Naar de rechterkunnen mensen gaan.*
 to the judge can people go
 'People can go to court.'

The reason for the awkwardness of (63) might be that directional PPs have a closer relation to the verb. They usually need a verb that expresses motion, such as *gaan* 'to go'. A directional PP that precedes the verb of motion is strange from an incremental view of language comprehension, since the direction is already given, while the manner of movement is not clear yet. Note, however, that not all directional PPs in sentence-initial position are bad in Dutch. Moreover, in English, stylistic inversion of subject and a directional PP is quite common, as illustrated in (64) (see e.g. Levin and Rappaport Hovav, 1995).

- (64) Down the hill rolled the baby carriage.

Other PPs that are strongly connected to the verb or the subject, such as fixed expressions and prepositional objects (as in (65a)), and prepositional phrases appearing as pronominal adverbs (as in (65b)) have been removed from the corpus selection.

- (65) a. *Dronkaards moeten luisteren naar klassieke muziek.*
 drunks must listen to classical music
 [fv600029.75]
 'Drunks should listen to classical music.'
- b. *Daar moeten directeuren bijvoorbeeld in sturen.*
 there must directors for.example in direct
 [fn000008.131]
 'Managers should be directing in that, for example.'

Sentences like (65a) were omitted because it is hard to move the PP to the preverbal position, and sentences like (65b) were omitted because pronominal adverbs are often discontinuous. Finally, sentences with more than one PP have been taken out of the selection if it was not clear

which of them would end up in sentence-initial position when the subject would be placed in postverbal position (cf. Grondelaers and Speelman, 2007). This was also the case when the PP occurred together with a locative expression that was not a PP, such as *daar* ‘there’.

5.3.2. Procedure

After all sentences that fell outside the restrictions described above were manually removed from the data set, the remaining 557 sentences were annotated for the seven factors discussed in Section 5.2. The annotation was done twice by different annotators, independently from each other. Cases of disagreement between the annotators were resolved through discussion with a third person. The factors were operationalized such that for each factor a sentence could be assigned a unique value. The factors could take on the following values (Table 4):

Table 4 Operationalization of the factors

Factors	Values	Cohen's κ
Specificity of the subject	Bare plural subjects: [modified] [non-modified] Definite plural subjects: [strong quantifier] [definite article] [strong referential]	.98
Animacy of the subject	[animate] [inanimate]	.96
Presence of <i>er</i> ‘there’	[+er] [-er]	1
Tense	[present] [past]	.98
Aspect	[imperfect] [perfect]	.94
Voice	[active] [passive]	.92
Type of PP	[locative] [temporal] [abstract]	.90

A few notes on the operationalization should be made. Firstly, the three levels of the factor Specificity for the definite plural subjects were defined as follows: *alle/al dezelal die* ‘all (these/those)’, *de meeste* ‘most’, *sommige* ‘some’ and *beide* ‘both’ were categorized as strong quantifiers.

The definite article was always *de* 'the' (no *het* since all subjects were plural). Demonstrative pronouns were *deze* 'these' or *die* 'those'. Possessive pronouns could be *mijn/m'n* 'my', *jouw/je* 'your', *zijn/z'n* 'his', 'its', *haar/d'r* 'her', *onze* 'our', *jullie* 'your' (pl.) and *hun* 'their'. The last two categories were taken together, because there are no clear predictions as to which of these has better topic properties. Secondly, although the animacy scale in (54) has a three-way distinction, the factor Animacy in the present study is subdivided into two levels, to ensure that each level is covered by enough data. The label Animate was given to human and animate NPs, as well as NPs referring to companies or organizations with a collective voice that are treated as if they are sentient beings. All other NPs were categorized as Inanimate. For the factor Voice, all sentences with the passive auxiliary *worden* 'to become' (or in the perfect *zijn* 'to be') or with a substituting auxiliary like *raken* 'to get' were counted as passive. Reflexive and medial constructions were counted as active.

For each sentence the word order (Subj-V-PP or PP-V-Subj) was determined. The resulting matrix of sentences with their word order and their values on each factor was then used as the input for the creation of statistical models. These models should reveal which factors have a significant effect on word order, and what the direction of this effect is.

Two models were calculated: one for sentences with a bare plural subject, and one for sentences with a definite plural subject. The two models could not be compared directly, due to the fact that some factors either did not have the same values across the models (e.g. specificity) or were predicted to have an effect in one model, but not in the other (e.g. the presence of *er* 'there'). The effects of all different factors and interactions were calculated with a logistic regression analysis using the statistical software program R.¹³

In the following sections, the effect of each factor on word order will be discussed separately for both models. In addition, a multifactorial analysis will be presented in which all factors are included simultaneously. The results for bare plural subjects will be discussed in Section 5.4; definite plural subjects will be analyzed in Section 5.5.

¹³ <http://www.r-project.org/>

5.4 Sentences with a bare plural subject

5.4.1 Data exploration per factor

In this section, the effects of the different factors discussed above on word order (Subj-V-PP or PP-V-Subj) in sentences with a bare plural subject will be presented separately for each factor. The individual effects of the factors will be tested with Chi-square (χ^2) tests. But first I will give the overall frequencies of both word orders in Table 5 below.

Table 5 Frequencies (absolute and relative) of the two different word orders (S-V-PP and PP-V-S) in sentences with a bare plural subject in the corpus data set

Condition	Produced word order				TOTAL	
	S-V-PP		PP-V-S			
Bare plural	81	36.3%	142	63.7%	223	100%

As Table 5 shows, in almost two third of the 223 sentences, the bare plural subject is in postverbal position. Let us now have a closer look at the different factors that are expected to influence the placement of bare plural subjects.

5.4.1.1 Specificity of the subject

As can be seen in Table 6, there are 56 cases in which the bare plural subject is preceded by an adjective. Although these modified bare plural subjects are still more frequent postverbally than in preverbal position, they occur relatively more often in sentence-initial position than non-modified bare plural subjects (44.6% versus 33.5%; see also Figure 2). However, this difference is not significant ($\chi^2(1) = 1.78$; $p = 0.18$).

Table 6 Word order (S-V-PP vs. PP-V-S) x Specificity of the subject (non-modified vs. modified)

Bare plurals	Produced word order				TOTAL	
	S-V-PP		PP-V-S			
Non-modified bare plural	56	33.5%	111	66.5%	167	100%
Modified bare plural	25	44.6%	31	55.4%	56	100%
TOTAL	81	36.3%	142	63.7%	223	100%

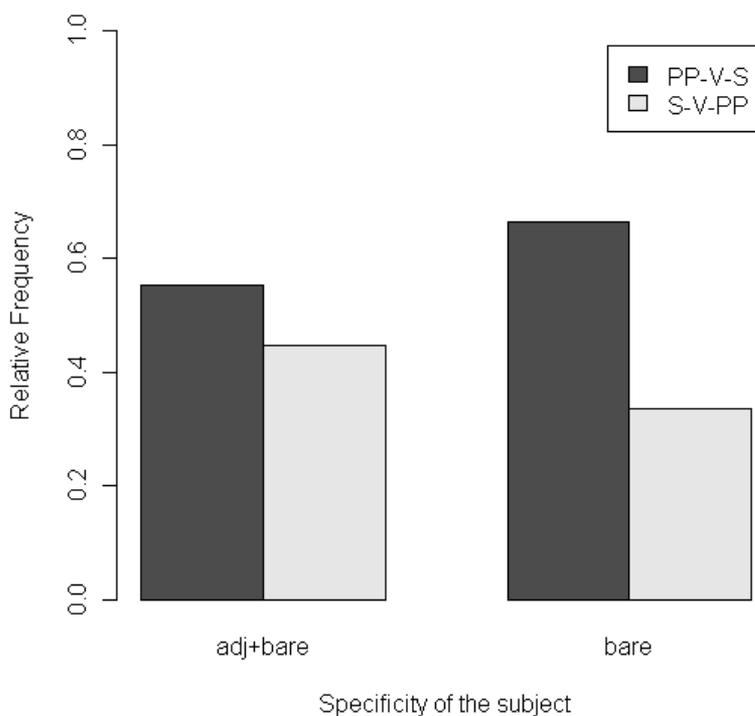


Figure 2 Word order (S-V-PP vs. PP-V-S) x Specificity of the subject (modified (*adj+bare*) vs. non-modified (*bare*))

5.4.1.2 Animacy of the subject

Table 7 and Figure 3 show that animate bare plural subjects appear about as frequently in preverbal as in postverbal position, a pattern that is close to that found for the bare plural subjects in the production experiment, which only included animate subjects (see Table 2 above). Inanimate bare subjects, on the other hand, show a clear preference for the postverbal position (72.9%). The effect of animacy on word order is found to be significant ($\chi^2(1) = 8.35$; $p < 0.01$). Note further that more than half of the bare plural subjects in the set is inanimate (118 out of 223). This is remarkable, because it is a general tendency in languages that subjects are animate (e.g. Comrie, 1989; Dahl and Fraurud, 1996; Øvrelid, 2004; Van Tiel and Lamers, 2007).

Table 7 Word order (S-V-PP vs. PP-V-S) x Animacy of the subject (animate vs. inanimate)

Bare plurals	Produced word order				TOTAL	
	S-V-PP		PP-V-S			
Animate	49	46.7%	56	53.3%	105	100%
Inanimate	32	27.1%	86	72.9%	118	100%
TOTAL	81	36.3%	142	63.7%	223	100%

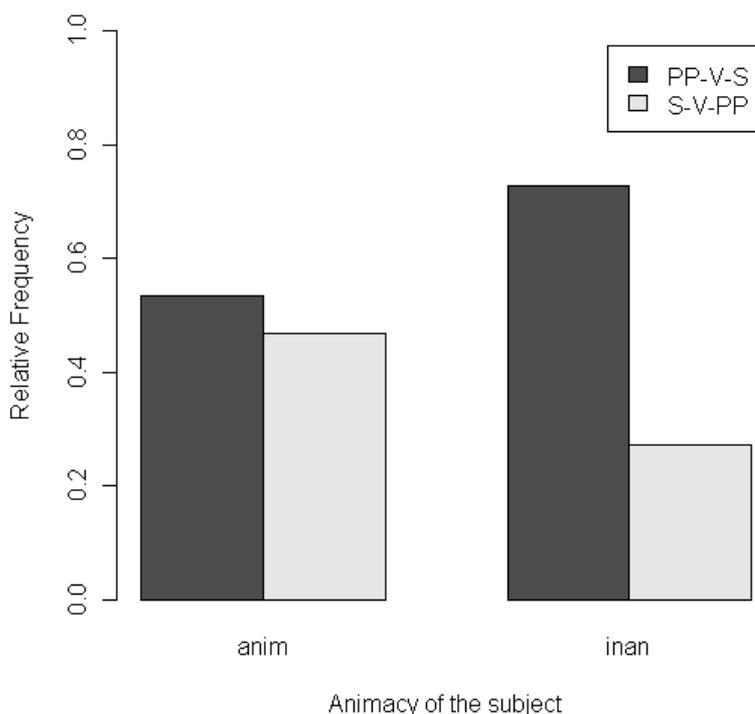


Figure 3 Word order (S-V-PP vs. PP-V-S) × Animacy of the subject (animate (*anim*) vs. inanimate (*inan*))

5.4.1.3 Presence of *er* 'there'

A postverbal expletive *er* 'there' occurred in 16 sentences. As shown in Table 8 and Figure 4 below, in all cases but one the subject is also postverbal when *er* occurs (only absolute frequencies are given because of the low numbers in some of the cells).¹⁴ This effect is significant ($\chi^2(1) = 5.41$; $p = 0.02$).

¹⁴ The figures in the upper half of Table 6 are so low, because only the S-V-PP and PP-V-S word orders are included here. In these cases, when *er* occurs, it appears in a postverbal position. A more common position for *er* in Dutch,

Table 8 Word order (S-V-PP vs. PP-V-S) x Presence of *er* 'there' (+*er* vs. -*er*)

Bare plurals	Produced word order		TOTAL
	S-V-PP	PP-V-S	
+ <i>er</i>	1	15	16
- <i>er</i>	80	127	207
TOTAL	81	142	223

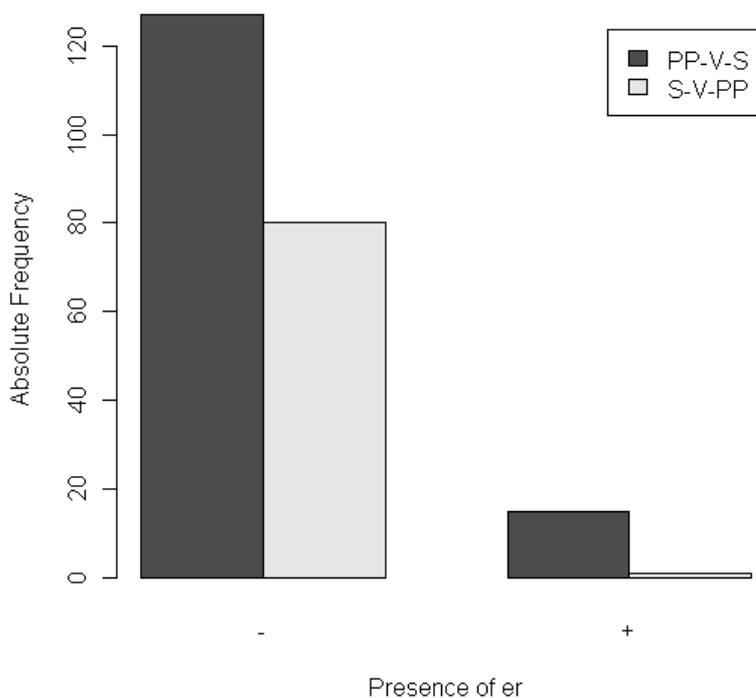


Figure 4 Word order (S-V-PP vs. PP-V-S) x Presence of *er* 'there' (*er* does not occur (-*er*) vs. *er* does occur (+*er*))

however, is the sentence-initial position, yielding a word order in which both the subject and the PP are postverbal (Er-V-S-PP or Er-V-PP-S).

5.4.1.4 Tense

As is shown in Table 9 below, present tense sentences are about twice as frequent in the set as past tense sentences. However, there seems to be no difference in the distribution over the two word orders between the two conditions. Indeed, no significant effect was found ($\chi^2(1) < 1$; $p = 0.95$).

Table 9 Word order (S-V-PP vs. PP-V-S) x Tense (present vs. past)

Bare plurals	Produced word order				TOTAL	
	S-V-PP		PP-V-S			
Present	57	36.8%	98	63.2%	155	100%
Past	24	35.3%	44	64.7%	68	100%
TOTAL	81	36.3%	142	63.7%	223	100%

5.4.1.5 Aspect

The number of verbs with (unmarked) imperfective aspect in the set is much higher than the number of verbs with perfective aspect. As can be seen in Table 10, however, there is no significant difference in the distribution over the two word orders between the two conditions ($\chi^2(1) < 1$; $p = 0.87$).

Table 10 Word order (S-V-PP vs. PP-V-S) x Aspect (imperfect vs. perfect)

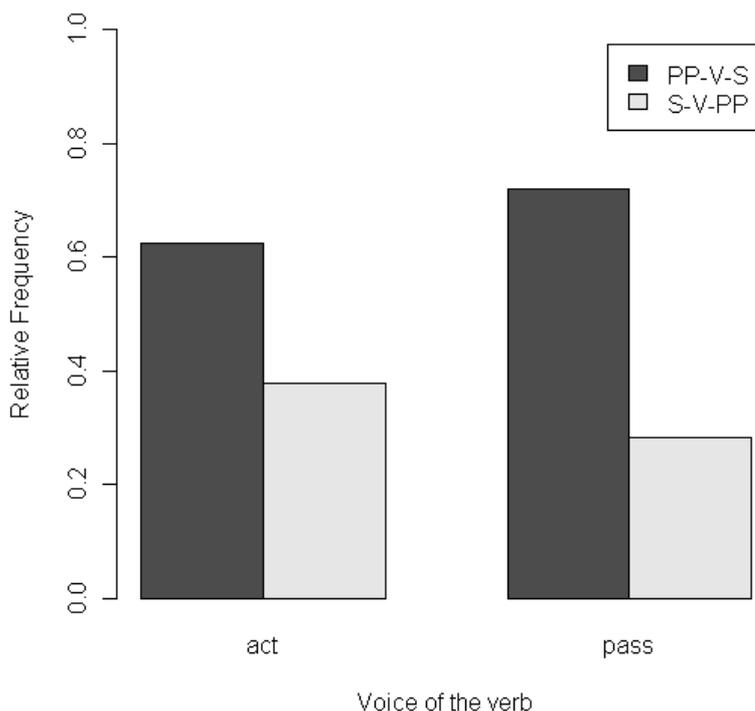
Bare plurals	Produced word order				TOTAL	
	S-V-PP		PP-V-S			
Imperfect	71	36.8%	122	63.2%	193	100%
Perfect	10	33.3%	20	66.7%	30	100%
TOTAL	81	36.3%	142	63.7%	223	100%

5.4.1.6 Voice

As can be seen in Table 11 and Figure 5, the PP-V-Subj word order is relatively more frequent for passive than for active sentences. However, this difference is not significant ($\chi^2(1) < 1$; $p = 0.40$), perhaps due to the relatively low number of passive sentences in the set.

Table 11 Word order (S-V-PP vs. PP-V-S) x Voice (active vs. passive)

Bare plurals	Produced word order					
Voice	S-V-PP		PP-V-S		TOTAL	
Active	72	37.7%	119	62.3%	191	100%
Passive	9	28.1%	23	71.8%	32	100%
TOTAL	81	36.3%	142	63.7%	223	100%

**Figure 5 Word order (S-V-PP vs. PP-V-S) x Voice (active (*act*) vs. passive (*pass*))**

5.4.1.7 Type of PP

The frequencies of the two word orders for each of the three types of PP are given in Table 12 and Figure 6.

Table 12 Word order (S-V-PP vs. PP-V-S) x Type of PP (locative, temporal and abstract)

Bare plurals	Produced word order					
	S-V-PP		PP-V-S		TOTAL	
Locative	38	30.9%	85	69.1%	123	100%
Temporal	3	12.5%	21	87.5%	24	100%
Abstract	40	52.6%	36	47.4%	76	100%
TOTAL	81	36.3%	142	63.7%	223	100%

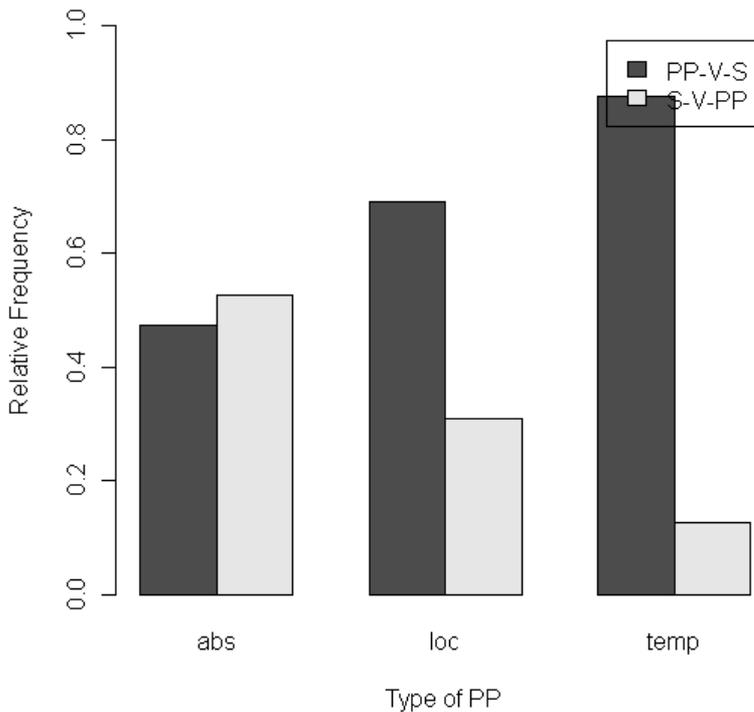


Figure 6 Word order (S-V-PP vs. PP-V-S) x Type of PP (abstract (*abs*), locative (*loc*) and temporal (*temp*))

The effect of the type of PP on word order is found to be significant ($\chi^2(2) = 16.2$; $p < 0.01$). As can be seen in Table 12 and Figure 6, there is a big difference between concrete PPs on the one hand and abstract PPs

on the other: concrete PP adjuncts (locative and temporal) seem to have a strong preference for the sentence-initial position. Abstract PP adjuncts, on the other hand, show a more equal distribution over the two word orders, appearing slightly more often in postverbal position than in sentence-initial position. There is also a notable difference between locative and temporal PPs, the latter occurring more frequently in sentence-initial position. Note, however, that the corpus data set only includes a restricted amount of temporal PPs (Table 12).

In Table 13 below all seven factors discussed in this section and their individual effects are summarized.

Table 13 Summary of the individual effects of the factors in the bare plural subjects subset

Factors	df	χ^2	p
Specificity of the subject	1	1.78	0.18
Animacy of the subject	1	8.35	< 0.01
Presence of <i>er</i> 'there'	1	5.41	0.02
Tense	1	< 1	0.95
Aspect	1	< 1	0.87
Voice	1	< 1	0.40
Type of PP	2	16.2	< 0.01

As can be seen in Table 13, three factors are found to have a significant effect ($p < 0.05$): Type of PP has the strongest effect, followed by Animacy of the subject and Presence of *er* 'there'.

5.4.2 Multifactorial analysis

In the previous subsection, the frequency distributions over both word orders (Subj-V-PP and PP-V-Subj) were presented separately for the seven factors. This gave an indication of the effects of each single factor on word order. However, in natural language, and thus in the corpus data, all factors are at work at the same time. To determine the effects of multiple factors affecting the data simultaneously, all factors have to be included in one statistical model. This was done by performing a logistic regression analysis. With a logistic regression model, the com-

bined effect of all factors can be investigated. Also, it can be determined which factors are more important than others and which factors do not have an effect at all. In addition, it can give an indication of how much variation (i.e. in word order that was found in the data) can actually be explained by these factors, and how much variation is not yet covered.

First all seven factors were included in the model. Factors that were found to be non-significant in the model were removed one by one, starting with the factor with the highest p-value (stepwise backward regression). Each time a factor was removed, an ANOVA was performed to ensure the model did not change significantly due to the removal of a factor.

A summary of the logistic regression model for bare plural subjects is given in Table 14 below. In the final model, there are four factors with a significant effect on word order: Animacy of the subject, Specificity of the subject, PP Type and Presence of *er* 'there'. For each factor one of its levels is taken as the baseline to which the effect of that factor is measured. For example, for the factor Animacy the value 'animate' is taken as the baseline. The estimate (log odds) then says that for inanimate subjects the odds of a PP-V-Subj word order are (inverse log of 1.0207 =) 2.77 times the odds of a PP-V-Subj order for animate subjects. Negative estimates point towards a higher probability of a PP-V-Subj order; positive estimates point towards a higher probability of a Subj-V-PP word order. For the factor PP Type the value 'abstract' is taken as the baseline, to which both the effect of locative PPs and that of temporal PPs are measured. In Table 14, we can see that all factors have a negative estimate, meaning that they all increase the probability of the PP-V-Subj word order. Thus, when the bare plural subject is inanimate, non-modified, when the PP is concrete and *er* 'there' is present, all this increases the odds of a PP-V-Subj word order. The presence of *er* has the strongest effect, followed by PP Type, where the difference between temporal PPs and abstract PPs (-2.1818) is larger than between locative and abstract PPs (-0.9079). Then comes Animacy with an estimate of -1.0207; Specificity has the smallest effect (-0.8121). The effects of Tense, Aspect and Voice were not significant. In addition, no interactions were found to be significant.

Table 14 Logistic regression model: bare plural subjects

Factor	Estimate	p
<i>Intercept</i>	4.5648	0.01
Animacy (inanimate)	-1.0207	< 0.01
Specificity (non-modified) ¹⁵	-0.8121	0.02
PP Type (locative)	-0.9079	< 0.01
PP Type (temporal)	-2.1818	< 0.01
Er (+er)	-2.2174	0.04
Statistics		
n	223	
Model L.R.	38.11	
d.f.	5	
P	0	
C	0.73	
D _{xy}	0.459	
R ²	0.215	
Predicted correctly	70.0%	
Baseline	63.7%	

The accuracy of the logistic regression model is given by the percentage ‘predicted correctly’. The model predicts the correct word order in 70.0% of the cases. A naïve model, that is, a model in which no factors are included and the word order choice is solely based on the most frequently occurring word order, has a prediction accuracy of 63.7% (the baseline). Thus, by including these factors the predictive power of the model increases by 6.3%.

¹⁵ The factor Specificity of the subject is included here as an ordinal variable, which means that it is treated as a scale with fixed intervals. The value of the estimate should be interpreted as follows: if the value of the factor increases with 1 (in this case, from *modified* to *non-modified*), the probability of a Subj-V-PP word order increases with the value of the estimate.

5.4.3 Discussion

The logistic regression model in Table 14 largely corroborates the effects found for the factors when considered separately, which were presented in Section 5.4.1. The factors that were found to have a significant individual effect on word order, namely Animacy of the subject, Presence of *er* 'there' and Type of PP, also come out as significant contributors to word order in the multifactorial analysis. However, one factor, Specificity of the subject, did not show a significant effect when tested separately from the other factors, but it does come out as significant in the logistic regression analysis. Apparently, only when multiple factors are included simultaneously, the effect of specificity becomes manifest. The remaining factors, Tense, Aspect and Voice do not have a significant effect on word order in either the monofactorial or the multifactorial analysis. This supports the assumption that the nature of the predicate does not influence the position of a bare plural subject.

Of the factors that were found to have a significant effect in the logistic regression analysis, the direction of the effect is always as predicted. The effects of the factors Animacy of the subject, Specificity of the subject and Presence of *er* 'there' all confirm the prediction that more semantic variables than definiteness alone have an influence on the word order of sentences with bare plural subjects: although bare plurals are generally bad topics, their prominence is affected by other factors. First of all, for sentences with an inanimate subject a higher probability of PP-V-Subj word orders is found. This means that inanimate bare plural subjects have a higher chance to be in postverbal position than animate bare plural subjects. This finding fits with the analysis that animacy contributes to the prominence of a constituent: animate NPs are more semantically prominent than inanimate NPs. Therefore they are better topics and more easily placed at the sentence-initial position. Second, non-modified bare plural subjects (i.e. bare plural subjects not preceded by an adjective) are found to have a greater probability to be in postverbal position (PP-V-Subj word order) than modified bare plural subjects. This can be explained when it is assumed that modified bare plurals are somewhat less non-specific than 'true' bare plurals. Consequently, they are less bad topics and are therefore found more frequently in sentence-initial position. Third, when a bare

plural subject is accompanied by a postverbal expletive *er* 'there', it is again more likely to be in postverbal position. This connects to the analysis of Grondelaers and Speelman (2007) that *er* 'there' is an inaccessibility marker that signals an inaccessible subject. Since something that is inaccessible is non-prominent in discourse, inaccessible subjects are not easily made topic and consequently prefer a postverbal position. The fact that bare plural subjects are already semantically non-prominent NPs is thus reinforced when they are marked as inaccessible by *er*.

The factor Type of PP, by contrast, does not influence the topic characteristics of the bare plural subject, but those of the PP adjunct. It illustrates the fact that other constituents than the subject can be topics, and that topicalization of such a constituent automatically results in a postverbal subject, regardless of the inherent topic characteristics of that subject. As was mentioned earlier, spatiotemporal adjuncts are frequently topics in Dutch, because they indicate the concrete spatiotemporal setting in which an event takes place. In this way, they can also help to build predictive inferences on a coming subject that would otherwise be inaccessible. Indeed, as is shown in the logistic regression model in Table 14, both locative and temporal PPs create a higher probability of PP-V-Subj word orders (i.e. with a PP in the topic position) with respect to abstract PPs. Thus, since locative and temporal PPs create a concrete spatiotemporal setting, they can be considered prominent constituents that are easily topicalized. Abstract PPs lack such a situationalizing function, and are therefore less frequently found in sentence-initial position. When a PP adjunct has been topicalized, the (bare plural) subject that is then in postverbal position could be an inaccessible NP that is made more accessible by way of the PP adjunct, but this need not be the case.

Nevertheless, the prediction that postverbal bare subjects are often combined with a sentence-initial locative PP is only partly borne out by the data. Indeed, when the subject is bare, a locative PP appears in sentence-initial position (and the subject in postverbal position) in over two third of the cases. However, the fact that the effect of temporal PPs is found to be stronger than the effect of locative PPs (with estimates of -2.1818 and -0.9079, respectively) is remarkable. It was predicted that locative PPs are more concrete and therefore better topics than tempo-

ral PPs, because time can be considered somewhat more abstract than space in human perception. This is not borne out by the data. Temporal PP adjuncts seem to have a much higher preference for the sentence-initial position than locative PP adjuncts. A possible explanation could be that time indications often situate an entire event within a temporal context, while a spatial location applies more to entities than to events. Thus, the scope of a temporal adjunct might be the whole clause, while a locative adjunct has only scope over, in this case, the subject. Crasborn et al. (2009) find that when a sentence contains both a spatial and a temporal topic in Sign Language of the Netherlands (NGT), the temporal topic typically precedes the spatial topic. Also in Dutch, when a sentence contains both a locative and a temporal PP, the temporal PP preferably occupies the sentence-initial position, rather than the locative PP. This is illustrated in (66).

- (66) a. *In augustu*_{TEMP} *waren er rellen in Hoek van Holland*_{LOC}.
 in August were there riots in HvH
 'In August there were riots in Hoek van Holland.'
- b. *#In Hoek van Holland*_{LOC} *waren er rellen in augustu*_{TEMP}.
 'In Hoek van Holland there were riots in August.'

Summarizing, the factors Animacy of the subject, Specificity of the subject and Presence of *er* 'there', on the one hand, influence the prominence of bare plural subjects, and significantly affect the word order of sentences with a bare plural subject: bare plural subjects that are animate and/or modified by an adjective have a higher probability to occur in preverbal position than inanimate and non-modified bare plural subjects. Bare plural subjects accompanied by the expletive *er* 'there' are more frequent in a postverbal position than subjects without *er*. PP adjuncts, on the other hand, are often topicalized, and significantly more when it is a spatiotemporal PP than when it is an abstract PP. Topicalized PPs result in a higher frequency of postverbal subjects.

5.5 Sentences with a definite plural subject

5.5.1 Data exploration

In this section, the effects of the factors that were predicted to influence word order of sentences with a definite plural subject are discussed. The overall occurrence of the two word orders (Subj-V-PP or PP-V-Subj) in sentences with a definite plural subject in the data set is given in Table 15.

Table 15 Frequencies (absolute and relative) of the two different word orders (S-V-PP and PP-V-S) for definite plural subjects in the CGN data set

Condition	Produced word order				TOTAL	
	S-V-PP		PP-V-S			
Definite plural	178	53.3%	156	46.7%	334	100%

Table 15 shows that definite plural subjects occur about as much preverbally as postverbally. The following subsections will explore the influence of the different factors included on this word order pattern.

5.5.1.1 Specificity of the Subject

Table 16 and Figure 7 show that subjects preceded by a demonstrative or a possessive pronoun are more frequent in sentence-initial position than subjects preceded by a strong quantifier (64.5% vs. 47.4%). The difference between subjects preceded by a strong quantifier and subjects preceded by a definite article is not very large, both occurring about as much preverbally as postverbally. The effect of the different degrees of specificity on word order almost reaches significance ($\chi^2(2) = 5.0$; $p = 0.08$).

Table 16 Word order (S-V-PP and PP-V-S) x Specificity of the subject (subjects preceded by a strong quantifier, a definite article or a strong referential pronoun)

Definite plurals	Produced word order					
	S-V-PP		PP-V-S		TOTAL	
Strong quantifier	9	47.4%	10	52.6%	19	100%
Definite article	120	50.2%	119	49.8%	239	100%
Strong referential	49	64.5%	27	35.5%	76	100%
TOTAL	178	53.3%	156	46.7%	334	100%

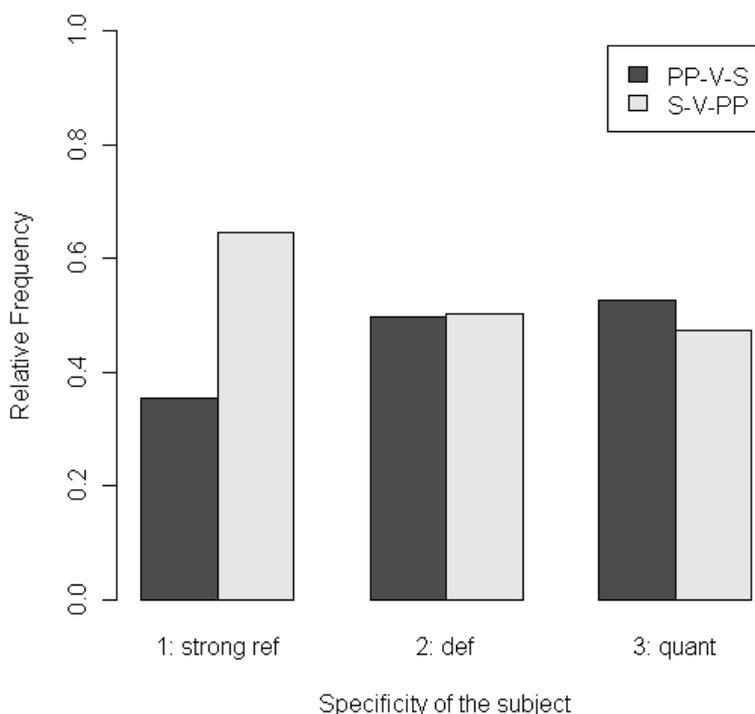


Figure 7 Word order (S-V-PP and PP-V-S) x Specificity of the subject (subjects preceded by a strong referential (*strong ref*; demonstrative or possessive), a definite article (*def*) or a strong quantifier (*quant*))

5.5.1.2 Animacy of the Subject

For definite plural subjects, no significant effect of animacy is found ($\chi^2(1) = 1.2$; $p = 0.28$, see Table 17 and Figure 8). Exactly half of the animate definite subjects appear in preverbal position, the other half in postverbal position. For the inanimate subjects, there seems to be a slight preference for the preverbal position. However, this tendency is not significant. Again, as with bare plural subjects, more than half of the total number of definite plural subjects is inanimate (170 out of 334).

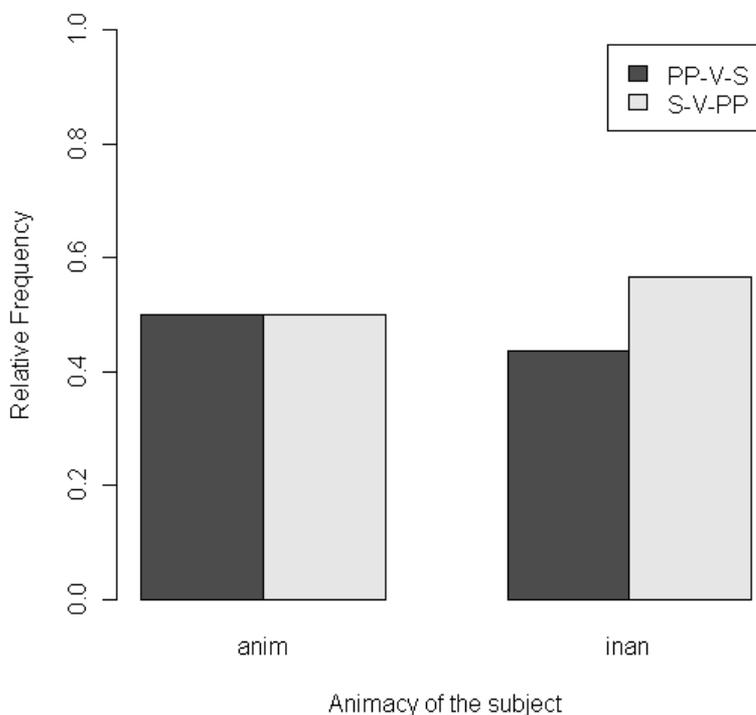


Figure 8 Word order (S-V-PP and PP-V-S) x Animacy of the subject (animate (*anim*) vs. inanimate (*inan*))

Table 17 Word order (S-V-PP and PP-V-S) x Animacy of the subject (animate vs. inanimate)

Definite plurals	Produced word order					
	S-V-PP		PP-V-S		TOTAL	
Animate	82	50.0%	82	50.0%	164	100%
Inanimate	96	56.5%	74	43.5%	170	100%
TOTAL	178	53.3%	156	46.7%	334	100%

5.5.1.3 Presence of *er* 'there'

Expletive *er* 'there' never occurs with a definite plural subject in the CGN data selection. This does of course not mean that such a construction is always ungrammatical, but one would predict that if a definite subject ever occurs in an *er*-construction, it will be not be a prototypical definite, but for example one that contains new information.

5.5.1.4 Tense

As with sentences with bare plural subjects, present tense verbs are in the majority in sentences with definite plural subjects, as shown in Table 18. The frequencies of the occurrence of present tense and past tense verbs lie very close to each other, with only a slightly higher occurrence of definite subjects of past tense verbs in sentence-initial position. There is however no significant effect on word order ($\chi^2(1) < 1$; $p = 0.41$).

Table 18 Word order (S-V-PP and PP-V-S) x Tense (present vs. past)

Definite plurals	Produced word order					
	S-V-PP		PP-V-S		TOTAL	
Present	111	51.4%	105	48.6%	216	100%
Past	67	56.8%	51	43.2%	118	100%
TOTAL	178	53.3%	156	46.7%	334	100%

5.5.1.5 Aspect

Again, there are many more sentences in the set with imperfective aspect than with perfective aspect. From Table 19 and Figure 9 there seems to be a small effect of aspect on word order, albeit in the opposite direction from what was expected: definite subjects of perfective verbs are somewhat more frequent in postverbal position, while definite sub-

jects of imperfective verbs are somewhat more frequent in preverbal position. This effect does however not reach significance ($\chi^2(1) = 2.9$; $p = 0.09$).

Table 19 Word order (S-V-PP and PP-V-S) x Aspect (imperfect vs. perfective)

Definite plurals	Produced word order				TOTAL	
	S-V-PP		PP-V-S			
Imperfect	152	55.7%	121	44.3%	273	100%
Perfective	26	42.6%	35	57.4%	61	100%
TOTAL	178	53.3%	156	46.7%	334	100%

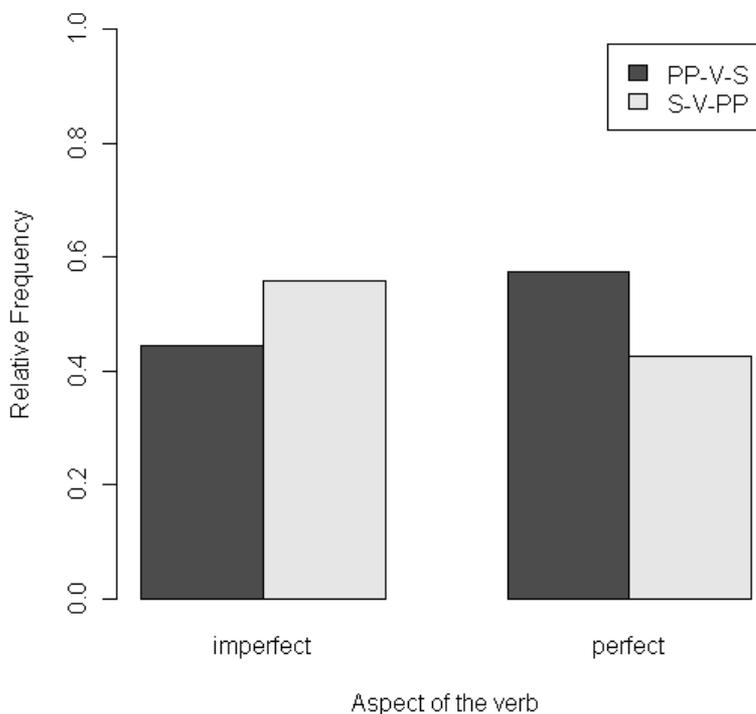


Figure 9 Word order (S-V-PP and PP-V-S) x Aspect (imperfect vs. perfect)

5.5.1.6 Voice

The relative frequencies of occurrence of both word orders are very similar across the active and the passive condition, as is shown in Table 20. Definite subjects of passive verbs occur relatively more frequently in sentence-initial position than definite subjects of active verbs. However, this is a very small difference, especially when considering the small number of passive sentences in this set, and it is far from significant ($\chi^2(1) < 1$; $p = 0.81$).

Table 20 Word order (S-V-PP and PP-V-S) x Voice (active vs. passive)

Definite plurals	Produced word order					
Voice	S-V-PP		PP-V-S		TOTAL	
Active	149	52.8%	133	47.2%	282	100%
Passive	29	55.8%	23	44.2%	52	100%
TOTAL	178	53.3%	156	46.7%	334	100%

5.5.1.7 Type of PP

The type of PP has a significant effect on word order in sentences with a definite plural subject ($\chi^2(2) = 10.3$; $p < 0.01$). A locative PP is placed in preverbal position in a little less than half of the cases (45.2%; see Table 21 and Figure 10). For abstract PPs, the preverbal position is less common (39.2%). Temporal PPs occur even more frequently in sentence-initial position than locative PPs (63.2%).

Table 21 Word order (S-V-PP and PP-V-S) x Type of PP (locative, temporal and abstract)

Definite plurals	Produced word order					
Condition	S-V-PP		PP-V-S		TOTAL	
Locative	80	54.8%	66	45.2%	146	100%
Temporal	25	36.8%	43	63.2%	68	100%
Abstract	73	60.8%	47	39.2%	120	100%
TOTAL	178	53.3%	156	46.7%	334	100%

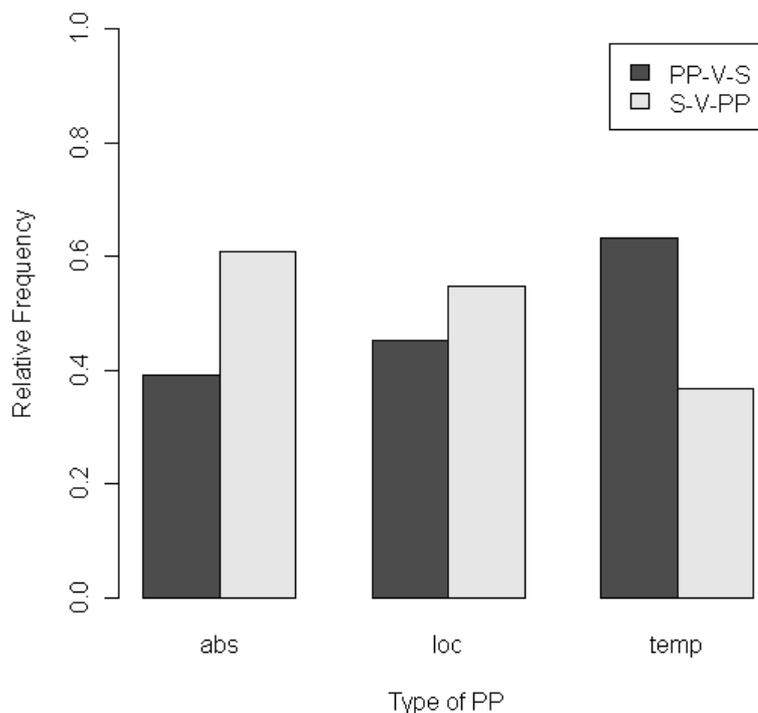


Figure 10 Word order (S-V-PP and PP-V-S) × Type of PP (abstract (*abs*), locative (*loc*) and temporal (*temp*))

All seven factors and their effects presented in this section are summarized in Table 22 below. As can be seen in Table 22, the only factor that has a significant effect on word order ($p < 0.05$) for definite plural subjects is Type of PP.

Table 22 Summary of the individual effects of the factors for definite plural subjects

Factors	df	χ^2	p
Specificity of the subject	2	5.0	0.08
Animacy of the subject	1	1.2	0.28
Presence of <i>er</i> 'there'	(n.a.)	(n.a.)	(n.a.)
Tense	1	< 1	0.41
Aspect	1	2.9	0.09
Voice	1	< 1	0.81
Type of PP	2	10.3	< 0.01

5.5.2 Multifactorial analysis

A summary of the logistic regression model for the definite plural subjects is shown in Table 23 below. The model includes three factors: Specificity of the subject, Type of PP and Aspect, of which Type of PP has the strongest effect, followed by Specificity and Aspect. As for the factor Type of PP, only temporal PPs differ significantly from abstract PPs (est. -1.0241; $p < 0.01$). The small difference between locative and abstract PPs is not significant (est. -0.3150; $p = 0.22$). The effect of Aspect is marginally significant ($p = 0.05$). The effects of the factors Animacy of the subject, Tense and Voice were found to be non-significant. The factor Presence of *er* 'there' was not included, since it was not applicable for this subset.¹⁶

As in the logistic regression model for the bare plural subjects, the estimates of all factors are negative. This means that they all increase the probability of a PP-V-Subj order. Thus, the chances for a PP-V-Subj word order increase when the subject is lower on the definiteness scale (preceded by a strong quantifier or a definite article as opposed to when it is preceded by a strong referential (demonstrative or possessive) pronoun), when the PP adjunct is concrete instead of abstract and when the verb has perfective aspect as opposed to imperfective aspect.

¹⁶ Two small interaction effects were found as well (Voice x Type of PP; Voice x Specificity), but since no meaningful interpretation could be given to these effects, the interactions were omitted here. The prediction accuracy of the model improved by this omission.

Table 23 Logistic regression model: sentences with definite plural subjects

Factor	Estimate	p
<i>Intercept</i>	1.4090	< 0.01
Specificity (1=strong referential; 2=definite article; 3=strong quantifier)) ¹⁷	-0.4512	< 0.05
Type of PP (locative)	-0.3150	0.22
Type of PP (temporal)	-1.0241	< 0.01
Aspect (perfect)	-0.5663	0.05

Statistics

n	334
Model L.R.	18.37
d.f.	4
P	0.001
C	0.628
D _{xy}	0.256
R ²	0.071
Predicted correctly	61.4%
Baseline	53.3%

Finally, a word on the accuracy of the definite plural subjects model. The model predicts 61.4% of the cases correctly. This is an improvement of 8.1% compared to the baseline frequency of 53.3%, which represents the prediction accuracy of the naïve model, i.e. choosing the most frequent word order (in this case, Subj-V-PP) all the time.

5.5.3 Discussion

When the effects of each factor on the position of definite plural subjects were examined separately in Section 5.5.1, the only factor that was found to have a significant effect was Type of PP. However, in the lo-

¹⁷ As in the bare plural subjects model, the factor Specificity of the subject has been transformed into an ordinal variable. See footnote 15.

gistic regression analysis two more factors, namely Specificity of the subject and Aspect, reach significance. These factors were already relatively close to significance in the individual Chi-square tests (with p-values of 0.08 and 0.09, respectively), but the inclusion of more factors at the same time apparently made their effects more manifest. The factors Animacy of the subject, Tense and Voice were neither significant in the monofactorial, nor in the multifactorial analysis.

The factor with the strongest effect on word order in the definite plural subjects model is Type of PP. When a sentence contains a temporal PP, the PP is significantly more frequently placed in sentence-initial position than when an abstract PP is concerned. This could be explained in terms of concreteness: temporal PPs situate an event in a certain point or period of time, making the event more concrete. Concreteness adds to the prominence of the PP, making it a better topic and hence a good candidate to start a sentence with. Since locative PPs are considered more concrete than temporal PPs, an effect of locative PPs was expected as well. However, as shown in Table 21 and Figure 10 above, locative PPs occur more often in postverbal position than in preverbal position. In fact, as can be seen in the regression model in Table 23, locative PPs do not differ significantly from abstract PPs. An explanation could be that locative PPs often only have scope over an entity (they indicate the location of the subject), while temporal PPs usually have scope over the entire event, which makes them highly preferred in sentence-initial position.

When a PP adjunct has been topicalized, the (definite plural) subject automatically ends up in a postverbal position. This does not mean that the subject is a bad topic. On the contrary, definite plural subjects are normally good topics, since they are generally referential and refer to old information. Still they can be placed in a non-prominent position when another constituent – in this case a PP adjunct – has been topicalized.

The other two factors with a significant effect on word order, Specificity and Aspect, tap on the properties of the subject. As can be seen in the regression model, the probability of a postverbal subject is higher when it is lower on the definiteness scale (preceded by a strong quantifier or a definite article as opposed to preceded by a strong referential pronoun (a demonstrative or a possessive)). Or put the other way

around: the chances of a preverbal subject become higher when the subject is higher on the definiteness scale. This fits with the predictions regarding different degrees of specificity/definiteness. NPs preceded by a definite article are generally good topics, since they are often referential, as became clear from the production experiment presented in Chapter 3, which only contained definite subjects of this type. Other types of definite NPs, however, might be either more strongly referential, and therefore even better topics (such as those preceded by a demonstrative or a possessive pronoun), or less referential, and therefore less good topics, (NPs preceded by a strong quantifier). Since topics favour the sentence-initial position, strong referential NPs were predicted to be more frequent in this position than other definites, while strongly quantified NPs would be less frequent in this position. This prediction is thus borne out by the present model.

As for the factor Aspect, it was predicted that subjects of verbs in perfective aspect are more frequent in sentence-initial position, because subjects of verbs that describe an event as closed or telic can be considered more agentive, as they have a tighter control over the event (provided they are agents in the first place). As a result, subjects of perfective verbs are more prominent in terms of agentivity and are consequently better topics; hence, they were expected to occur sentence-initially more often. However, according to the present logistic regression model, the opposite is the case: given the negative estimate of the factor Aspect, which takes imperfective aspect as its baseline, the probability of a PP-V-Subj order increases with verbs that have perfective aspect. A possible direction in which an explanation could be sought is the following: verbs with perfective aspect are more likely to be interpreted as descriptions of non-stative, episodic events (Alexiadou, 2000). For example, perfective verbs can be used in an answer to the question 'what happened?'. Consequently, subjects of verbs with perfective aspect can easily be associated with presentational focus: they provide new information about an episodic event. Since new information/focus is often associated with a postverbal position, subjects of perfective verbs could be more frequent in postverbal position. In Greek, for instance, postverbal subjects are ungrammatical with stative verbs. For a stative predicate like 'hate', a VSO word order is not possible, as in

(67a), unless it is made non-stative by adding perfective aspect to the verb, as in (67b).

- (67) a. **Misi i Maria ton Petro.*
 hates the Mary.NOM the Peter.ACC
 b. *Misise i Maria ton Petro.*
 hated.PERF.3SG the Mary.NOM the Peter.ACC
 (from Alexiadou, 2000)

This analysis somewhat resembles my comments on tense and aspect properties in Section 5.2.1.4 and 5.2.1.5, namely that they can induce an existential reading of bare plural subjects. However, no effects of either tense or aspect have been found for sentences with bare plural subjects. I will leave this for future research.

In sum, although definite plural subjects are generally good topics, they can still be placed in a non-prominent, postverbal position. The present multifactorial analysis suggests that an important factor contributing to this is the topicalization of other constituents than the subject, in this case PP adjuncts. Especially temporal PP adjuncts are often placed in sentence-initial position, probably due to their ability to set an event in a certain point of time. The position of a definite plural subject is further influenced by at least two other factors, one concerning the aspect of the verb, and the other the degree of referentiality of the subject. The animacy of the subject and the tense and voice of the verb do not have a significant effect on word order.

5.6 Conclusion

In this chapter I have presented a corpus study that was conducted to investigate word order variation in Dutch sentences with bare and definite plural subjects in natural language. Besides a clear difference in word order preferences between the two subsets (bare vs. definite), as was found in the production experiment, a number of additional factors have been found to significantly affect word order. Animacy and specificity of the subject, the type of the PP and the presence of *er* 'there' had an effect on sentences with a bare plural subject, while the specificity of the subject and the type of the PP had an impact on sen-

tences with a definite plural subject, as well as the aspect of the verb. These effects can be accounted for when they are considered the result of one or both of the constituents (subject or PP) becoming more or less prominent, leading to an increase or a decrease in its preference for the sentence-initial position. In the next chapter, I will make a comparison between the bare plural subjects model and the definite plural subjects model and combine the factors found to have an effect into an analysis relating them to the two main preferences introduced in Chapter 2: Subject First and Topic First.

Dutch word order variation: an analysis

6.1 Introduction

Recall that in Chapter 2, two main preferences have been identified that have an influence on word order in Dutch: the Subject First preference and the Topic First preference. The Subject First preference captures the fact that subjects are syntactically prominent; the Topic First preference that topics are semantically and discourse prominent. In Chapter 2 and 3, definiteness was taken as the predictor of topicality: definite NPs are good topics, while bare NPs are bad topics. Thus, when a bare NP is a subject, a conflict arises between the two preferences: the NP is syntactically prominent (subject), but semantically non-prominent (bad topic). This conflict gives rise to word order variation. Now, the corpus study presented in the previous chapter revealed some additional factors besides definiteness that can be assumed to influence topicality. These factors were however different for the subset concerning only sentences with bare plural subjects and the subset including sentences with definite plural subjects. In Section 2 of this chapter, I will compare the two subsets and answer the question why the regression models for the two subsets differ. I will also show that the different properties associated with the factors can be in conflict with each other with respect to the question whether a constituent is a good topic. Such conflicts may or

may not give rise to a conflict between the Subject First and the Topic First preference for both bare and definite plural subjects. In Section 3, I will give a brief note on the role of context and I will finish this chapter with a conclusion in Section 4.

6.2 Bare plural subjects versus definite plural subjects

The production experiment discussed in Chapter 3 showed a clear Subject First effect: 69% of all sentences (578 out of 835) were produced with the subject in sentence-initial position. This effect was found to be much less strong for bare plural subjects than for definite plural subjects. Bare plural subjects were found in sentence-initial position in 57% of the cases (240 out of 418), while definite plural subjects were produced sentence-initially in 81% of the cases (338 out of 417).

These results can be compared to the overall word order pattern found in the corpus data set. This is done in Table 24 below.

Table 24 Relative frequencies of the two different word orders (S-V-PP and PP-V-S) for both types of subject (bare plural and definite plural) in the production experiment from Vogels (2008) (left column) and in the present corpus study (right column)

	Produced word order			
	Experiment		Corpus	
Definiteness	S-V-PP	PP-V-S	S-V-PP	PP-V-S
Bare plural	57.4%	42.6%	36.3%	63.7%
Definite plural	81.1%	18.9%	53.3%	46.7%
TOTAL	69.2%	30.8%	46.5%	53.5%

Table 24 shows that there is no overall Subject First effect in the corpus data: the overall distribution between the two different word orders is around chance level (46.5% vs. 53.5%). This difference compared to the experimental results is most probably due to the fact that the data set selected from the CGN deviates from the material in the production experiment in a number of ways. The experimental material consisted of a carefully selected set, limited by fixed criteria, while the corpus data are subject to a lot of factors that could influence the position of the subject or the PP. In addition, whereas one probably would expect

to find a general Subject First preference in natural language production, the corpus data do not constitute a representative subset of Dutch natural language production in general. They constitute a representative subset of main clauses with a plural subject and a PP adjunct. For example, the obligatory presence of a PP adjunct could have increased the number of postverbal subjects, because they formed another group of candidates for the topic position. Sentences in which the subject was the only candidate for the topic position, on the other hand, were not included in the corpus set.

A clear effect of definiteness (bare vs. definite) can be seen in the corpus data: only 36% of the bare plural subjects (81 out of 223) are produced sentence-initially, as against 53% of the definite plural subjects (178 out of 334). This supports the analysis that definite NPs are more prominent, and therefore better topics, than bare NPs. Again, the differences with the production experiment can be attributed to the nature of the data set and the influence of factors not present in the experiment. It is clear from these percentages that the definiteness of the subject alone cannot account for the word order variations of both bare and definite plural subjects. If the position of a bare plural subject would be solely determined by the conflict between being subject on the one hand and being non-specific (and thus a bad topic) on the other, the number of sentence-initial bare subjects would be expected to be much higher, since the Subject First preference is quite strong in Dutch. Clearly, there are other preferences at work that are causing the high number of postverbal bare subjects. Similarly, since definite plural NPs are considered inherently good topics, no conflict would be expected between the Subject First and the Topic First preference, yielding no a priori reason for a definite plural subject to be postverbal. Still, almost half of the definite plural subjects in the corpus set are in a postverbal position. Again, other factors than definiteness might have contributed to this distribution. Now, let us have a closer look at the factors that were found to influence word order in the two regression models. Recall that these models only included factors that tap on the inherent prominence of the subject and the PP. Factors relating to the context of a sentence were not taken into account. Although context probably has a big impact on word order, it is nevertheless interesting to see that

inherent semantic properties of the constituents in a sentence can still influence word order. I will come back to this in Section 3.

As was mentioned in Chapter 4, there are at least two main options in which topicality can cause variation in the position of a subject. Either the properties of the subject itself cause that constituent to be either a good or a bad topic and subsequently contribute to its preference for a pre- or a postverbal position, or a constituent other than the subject – in this case always a PP adjunct – is topicalized and automatically causes the subject to be postverbal. In the following subsections, I will go through these two points again, now using the results of the logistic regression models to examine word order variation in the corpus data set.

6.2.1 Topicalization of the PP

Many PP adjuncts in sentence-initial position can be found in the corpus set. This is the case in 63.7% of the sentences with bare plural subjects and in 46.7% of the sentences with definite plural subjects. This could either be due to the PP being a semantically or discourse prominent constituent, which makes it a good candidate for the topic position (i.e. the preverbal position), or to the subject being a bad topic and leaving the topic position for another constituent. The semantic prominence of a PP was measured in terms of its concreteness: locative and temporal PPs were considered concrete and therefore more prominent, while abstract PPs were seen as less prominent. The fact that Type of PP came out as a very strong factor in both logistic regression models suggests that the semantic prominence of the PP adjunct contributes a great deal to the high frequency of sentence-initial PPs.

As was predicted, concrete PP adjuncts have a higher preference for the sentence-initial position than abstract PP adjuncts, resulting in a higher frequency of PP-V-Subj word orders. This can be explained by the higher semantic prominence of concrete PPs, which makes them better topics. The effect is especially strong in sentences with bare plural subjects. Here, the subject inherently is a less good topic, which already creates a preference for PP-V-Subj word orders. When a concrete PP is present at the same time, this preference is even reinforced. Yet, concrete PP adjuncts appear more frequently in postverbal position

in sentences with a definite subject (54.8% of the locative PPs; 36.8% of the temporal PPs) than in sentences with a bare subject (30.9% of the locative PPs; 12.5% of the temporal PPs). This is in line with the analysis that definite subjects compete with concrete PP adjuncts for the sentence-initial position.

Abstract PP adjuncts in both subsets have a higher preference for a postverbal position. This results in a higher frequency of Subj-V-PP word orders. The preference for the Subj-V-PP word order with an abstract PP can be explained by the fact that abstract PPs are not so easily topicalized. In contrast to locative and temporal PPs, which are much more concrete and are often used to provide the situational setting of the sentence in space or time, abstract PPs are less suited to have this function, simply because they are not concrete enough. This tendency is much stronger when the subject is definite than when it is bare. This can be explained by saying that bare subjects are bad topics as well, and thus compete with abstract PP adjuncts for the postverbal position. This conflict leads to a more equal distribution in word order.

The results of the factor PP Type further corroborate the analysis of Grondelaers and Speelman (2007). When the subject is inaccessible (i.e. bare, non-specific), a concrete PP helps making the subject more predictable for the hearer. In an incremental approach to language comprehension this concrete PP has to precede the subject to have an effect. An abstract PP does not contribute much to the predictability of the subject. Hence the higher percentage of abstract PPs in the Subj-V-PP word order. For the same reason, a higher occurrence of postverbal *er* 'there' is expected with a preverbal abstract PP than with a preverbal concrete PP. However, because of the small frequencies of occurrence of *er* in the data set with these word orders, no judgements can be made based on the present data.

6.2.2 *Properties of the subject*

Besides topicalization of the PP adjunct, inherent semantic properties of the subject itself can also cause variation in the position of the subject. First of all, the distinction between bare and definite NPs was shown to influence word order: bare plural subjects appear in postverbal position much more frequently than definite plural subjects. In Chapter 5, addi-

tional factors were tested for their influence on word order, which resulted in two logistic regression models: one for sentences with a bare plural subject and one for sentences with a definite plural subject. These models appeared to be quite different from each other: in the bare plural subjects model the animacy, specificity and accessibility (indirectly measured by the presence of *er* 'there') of the subject were found to have a significant effect on word order, while in the definite plural subjects model only the specificity of the subject was significant. A marginally significant effect of the aspect of the verb was also reported, which probably has to do with a presentational focus status of the subject. In what follows, I will go more deeply into the differences between the two models.

A factor that turned out to have a significant effect in both logistic regression models is Specificity of the subject. This factor was defined differently for bare plurals and definite plurals. For bare plurals, specificity referred to whether or not they were modified by an adjective, while for definite plurals it signified three different types of definite NPs. However, the two factors can be combined into one scale of specificity/referentiality, as represented in (68).

- (68) *Specificity scale:*
 Strong referential (demonstrative/possessive) > Definite article > Strong quantifier > Modified bare plural > Non-modified bare plural

NPs more to the left on the scale in (68) are considered more prominent than those more to the right. Consequentially, the former are better topics than the latter. This means that strong referential NPs are expected to be found most frequently in sentence-initial position, while non-modified bare plurals are expected to appear most often in a post-verbal position. That this prediction is borne out by the data can be shown when the graphs for the factor Specificity in the two subsets (Figures 2 and 7) are combined into one figure:

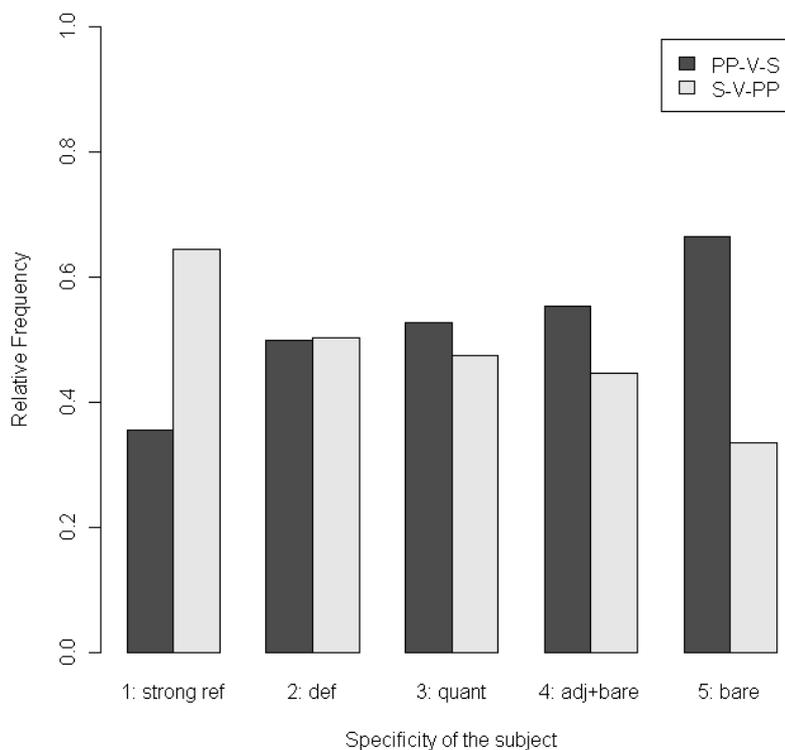


Figure 11 Word order (S-V-PP and PP-V-S) x five levels of the factor Specificity of the subject (a combination of Figure 2 and Figure 7)

Figure 11 nicely depicts the frequencies of the Subj-V-PP word order decreasing for each subsequent level of specificity (and increasing for PP-V-Subj). In this way, the factor Specificity can really be seen as a more fine-grained subdivision of definiteness, in which the two types of subject that were originally included in the drag-and-drop production experiment discussed in Chapter 3 are represented by the second category to the left in Figure 11 (NPs preceded by a definite article) and the rightmost category (NPs not preceded by anything). The distinction between bare and definite NPs should thus be seen as two points on a more or less continuous scale of definiteness/specificity/referentiality along which prominence gradually increases, and with that the chances

that an NP will be selected as a topic.¹⁸ It is therefore nice to see that not only the overall word order patterns between bare and definite plural subjects differ, but also that a more subtle division can be made within these categories, each level contributing differently to the position of the subject.

Animacy appeared to have a significant effect on the word order of sentences with a bare plural subject, but not on the word order of sentences with a definite plural subject. The direction of the effect of animacy on the word order of sentences with bare plural subjects was as predicted: animate bare plural subjects are significantly more frequent in preverbal position than inanimate bare plural subjects, which are more frequent in postverbal position. Animacy is an important contributor to the prominence of a constituent (e.g. Aissen, 2003; Lamers and De Hoop, 2008) and the property of being animate therefore increases the probability that a constituent is selected as a topic. Consequently, animate NPs are preferred in the prominent, preverbal position, whereas inanimate NPs are not. The fact that this effect is not found with definite plural subjects can be explained as follows: being definite is already a strong property to determine the position of a subject. When the subject is definite, it is a good topic, resulting in a strong preference for the sentence-initial position, regardless of whether it is animate or inanimate. Thus, although animacy is a strong contributor to prominence, its effects on word order are not visible when a definite NP is concerned. If, on the other hand, the property 'definite' is not present, the influence of animacy can become manifest. When the subject is bare, a conflict between the syntactic prominence of subjects and the semantic non-prominence of bare nominals causes variation in the choice for either the preverbal or the postverbal position. Given this situation, animacy comes into sight as a possible decisive factor for the position of the subject. An animate bare subject has a higher semantic prominence, making it a better topic, which reduces the conflict and increases the chances for the sentence-initial position. An inanimate bare subject, by contrast, might be an even worse topic, resulting in a

¹⁸ Note that the scale presented in Figure 11 is not exhaustive, since categories like indefinite NPs and pronouns are not included.

higher frequency of postverbal subjects, despite the Subject First preference.

The presence of the expletive *er* 'there' also has a significant effect on the word order of sentences with a bare plural subject. Since there were no occurrences of *er* in the subset of the data with definite plural subjects, this factor was not included there. In the bare plural subjects subset, almost all sentences in which *er* appears contain a postverbal subject. This was predicted: *er* signals an inaccessible subject, which is not suitable as a topic. In addition, *er* should precede the subject to be effective. There is however one case in which a postverbal *er* occurs in the sentence, but in which the subject is in sentence-initial position. The subject thus precedes the constituent that should have introduced it. This sentence is given in (69) below.

- (69) *Kansen liggen er nou voor je Jacco.* [fn000818.103]
 chances lie there now for you Jacco
 'Chances are there for you now, Jacco.'

It is likely that we are dealing with a different use of *er* 'there' here. It seems that in this sentence *er* does not signal a weak subject, but indicates a certain (abstract) location. It could then be considered a weak, neutral variant of *hier* 'here' and *daar* 'there' (locative *er*; Coppen, Haeseryn and De Vriend, 2002). This analysis is supported by the fact that making the subject definite does not make the sentence ungrammatical:

- (70) *De kansen liggen er nou voor je Jacco.*
 'The chances are there for you now, Jacco.'

When *er* is put in front of the subject, it probably gets a presentative reading, and a definite subject does make the sentence ungrammatical. Compare (71a,b) to (71c,d).

- (71) a. *Nou liggen er kansen voor je Jacco.*
 'Now there are chances waiting for you, Jacco.'
 b. *Er liggen nou kansen voor je Jacco.*
 'There are chances waiting for you now, Jacco.'
 c. *?Nou liggen er de kansen voor je Jacco.*

- 'Now there are the chances waiting for you, Jacco.'
- d. **Er liggen nou de kansen voor je Jacco.*
'There are the chances waiting for you now, Jacco.'

In addition, the bare plural subject *kansen* 'chances' in (69) is probably not an inaccessible subject. It is likely that in this sentence it is emphasized that it is *chances* that are there, i.e. the subject is in focus and receives heavy stress (De Hoop, 1992). Thus, the analysis that inaccessible subjects prefer a non-prominent position and are preceded rather than followed by presentative *er* 'there' still holds.

The fact that no occurrence of *er* was found with definite plural subjects supports the assumption that definite NPs are inherently more accessible than non-specific indefinite NPs (but probably still less accessible than e.g. pronouns), because they are usually referential and refer to old information. However, the effect of perfective aspect found in the definite plural subjects subset might indicate that an influence of accessibility is present in this half of the data as well. As noted in Section 5.6, the finding that definite subjects are more frequent in postverbal position when the verb has perfective aspect could relate to the association of perfective aspect with presentative focus. Perspective aspect is often used in sentences with an episodic reading, which can be used as answers to questions like 'what happened?'. Sentences that can be regarded as typical answers to the question 'what happened?' are news headlines. The corpus data set contains a number of records of news reports, which can be considered the ultimate places to find new information presented to a naïve hearer. In other words, it is expected that news headlines will contain many cases of presentational focus. An example is given in (72).

- (72) *Opde linkeroever van de Schelde zijn de*
on the left.bank of the Scheldt are the
werkzaamheden aan het Deurganckdok hervat.
works on the D.-dock resumed
[fv600906.1]
'On the left bank of the river Scheldt, the works on the Deurganck dock have been resumed.'

In (72), the definite plural subject *de werkzaamheden aan het Deurganckdok* 'the works on the Deurganck dock' is placed in postverbal position. Given that it is a definite NP, it is probably not completely new information (the works might have been mentioned in earlier broadcastings), but since this sentence is the start of the message, the subject is new in the discourse. It can therefore be considered inaccessible: the subject is not predictable from the preceding discourse. It might be the case, then, that the locative PP *op de linkeroever van de Schelde* 'on the left bank of the river Scheldt' has a function similar to locative PPs preceding inaccessible bare subjects, namely creating predictive inferences on the subject. The sentence in (72) might thus be an answer to the implicit question 'what happened (on the left bank of the river Scheldt)?'. This is supported by the use of perfective aspect on the verb (*zijn hervat* 'have been resumed'). It could be the case that just like presentative *er* 'there' perfective aspect is in a way associated with inaccessible subjects. However, since the effect of the factor Aspect is only small and no effect of Aspect has been found for bare plural subjects, I will leave this topic for future research.

Finally, neither in the regression model for the bare plural subjects, nor in that for the definite plural subjects were the factors Tense and Voice found to have a significant effect on word order. As for tense, indeed no effect was predicted for sentences with definite plural subjects. For sentences with bare plural subjects, however, the fact that the past tense elicits episodic, and therefore existential, readings on the subject does not significantly influence the choice of word order. As for voice, the theory that subjects of passive sentences are less agentive and therefore less suited to be topics is not reflected in the results, although a tendency for more postverbal bare subjects in passive sentences can be seen.

6.2.3 More conflicting preferences

The regression analyses revealed a number of factors affecting word order in sentences with a bare or definite plural subject and a PP adjunct. These effects can be accounted for when they are related to the prominence of either the subject or the PP. On the one hand, PPs are often topicalized, especially when they indicate concrete localizations

of the event in time or space: concrete PPs are typically good topics, while abstract PPs are generally bad topics. Note, however, that in the present study concreteness is only considered as a property of the PP adjunct: it is not applied to NPs (subjects). Consequently, this topic property of PPs is treated separately from properties of subjects.

On the other hand, the semantic or discourse prominence of the subject contributes to whether it is selected as a topic. In the present corpus study, semantic/discourse prominence has been measured by a number of factors, which tap on a constituent's semantic and discourse properties. First of all, definiteness has been shown to be important in determining prominence: definite NPs are better topics than bare NPs, and more fine-grained subdivisions of these two categories show a gradual increase in prominence along the definiteness/specificity scale. These differences in prominence give rise to word order variation. When a subject is definite, there already is a strong preference for the sentence-initial position, since the constituent is both syntactically and semantically prominent. Other semantic factors, like animacy, are not able to alter this. When a subject is a bare noun, however, semantic prominence is low, while syntactic prominence (being a subject) is still high. As a result, no strong overall preference for a pre- or a postverbal position is present, creating room for a semantic property like animacy to have an effect. Thus, the property of being animate makes a bare subject a better topic, increasing the chances for a Subj-V-PP word order, while being inanimate lowers the probability for a bare noun to become a topic, increasing the chances for a PP-V-Subj word order.

In addition, a discourse property like accessibility, indirectly measured by the factor Presence of *er* 'there', also has a large effect on the position of bare plural subjects: *er* signals an inaccessible subject (Grondelaers and Speelman, 2007), which is preferably placed in a non-prominent position. Note, however, that the absence of *er* does not necessarily mean that the subject is accessible. The factor only works in one direction. Even for definites, which are generally more accessible than bare nouns, discourse accessibility might have an influence on word order. Here, the factor Aspect could be a factor tapping on accessibility: perfective aspect might be associated with less accessible subjects. However, subjects of imperfective verbs are not automatically more accessible. Thus, the factor Aspect only goes in one direction as well.

Below I will briefly sketch how the results of the present study can be accounted for in terms of conflicting preferences.

As was already mentioned earlier, the different properties contributing to prominence can be in conflict with each other in determining whether something is a good topic. For example, an NP can be a good topic because it is definite, but at the same time a bad topic when it is also inanimate. Thus, before it can be established whether some constituent is a good topic and is therefore preferably placed in a prominent position, first the contribution to the prominence of that constituent by each of the semantic and discourse properties should be examined. Not all properties need to be equally strong, so that some property might overrule the influence of another. For example, it could be assumed that when a factor has an effect in the bare plural subjects model, but not in the definite plural subjects model, definiteness (i.e. whether the subject is bare or definite) is overruling this other factor. I will illustrate this with a couple of examples. First, consider (73), repeated from (26).

- (73) *In Amerika zijn er dus zomerkampen.* [fv400175.116]
 in America are there thus summer.camps
 'So in America there are summer camps.'

The bare plural subject *zomerkampen* 'summer camps' in (73) is probably not accessible: it is accompanied by the expletive *er* 'there'. It is not animate either, nor is it modified by an adjective. This NP is thus not topical according to any of these three properties. The PP adjunct *in Amerika* 'in America', on the other hand, is a locative and therefore high on the concreteness scale. As a consequence, the PP can be considered a good topic.

Since the PP *in Amerika* 'in America' is a good topic, while the subject *zomerkampen* 'summer camps' is a bad topic, the PP is more likely to be selected as the topic of the sentence, and consequently, according to the Topic First preference, preferred in sentence-initial position. However, we also have the Subject First preference: subjects are syntactically more prominent than objects or adjuncts, and are therefore preferably placed at the beginning of the sentence. Thus, the two preferences are in conflict here: the PP is the best candidate for the topic position, but

the subject is the best candidate for the sentence-initial position on syntactic grounds. The sentence in (73) has a PP-V-Subj word order, suggesting that the Topic First preference overrules the Subject First preference. Note, however, that the Subj-V-PP word order would also have been possible, as shown in (74).

- (74) *Zomerkampen zijn er dus in Amerika.*
 'So summer camps are in America.'

Because (73) and (74) involve a bare plural subject, the change in word order is accompanied by a shift in meaning: in (73), the subject *zomerkampen* 'summer camps' gets an existential reading, while the same subject in (74) probably needs some kind of presupposition, which could come from a contrast, for example. This presuppositional reading thus arises when the bare subject is placed in the topic position, since this position is not only associated with syntactic prominence, but also with semantic and discourse prominence. With a presuppositional/contrastive focus reading the bare subject is more discourse prominent than with an existential reading.

A bare plural subject can also have more prominence semantically. Consider the subject *Nederlandse moslims* 'Dutch Muslims' in (75).

- (75) *Nederlandse moslims werden op de site aangespoord*
 Dutch Muslims became on the site urged
om zich militair te trainen bij schietclubs en
 to themselves militarily to train at shooting clubs and
de landmacht. [fn001692.3]
 the land forces
 'Dutch Muslims were urged on the website to get a military training at shooting clubs and the land forces.'

Here, the subject is a modified bare noun, which makes the NP semantically more specific. It is also animate: it refers to humans. Consequently, the subject in (75) is more semantically prominent than the subject in (74). It is therefore more likely to be selected as a topic. On the other hand, the locative PP *op de site* 'on the website' in (75) is also a candidate for the topic position. Thus, there could be a conflict as to

which constituent will take the topic position. However, since the subject is also syntactically prominent while the PP adjunct is not, the preference to start the sentence with the subject might be stronger, resulting in a Subj-V-PP word order, as in (75).

Alternatively, when a bare plural subject is for instance a modified NP and inanimate at the same time, the two properties ‘animacy’ and ‘specificity’ are in conflict in determining whether the subject is a good topic. Dependent on which of the two properties is dominant, this may or may not result in a successive conflict between the Subject First and the Topic First preference. Examples are given in (76).

- (76) a. *Tijdens de enquête zijn nieuwe feiten aan 't*
 during the survey are new facts on the
 licht gekomen. [fn007226.20]
 light come
 ‘During the survey, new facts came to light.’
- b. *Mensen liepen hier rond met DAT-recorders.*
 people walked here around with DAT-recorders
 [fn000058.4]
 ‘People were walking around here with DAT-recorders.’

In (76a), the bare plural subject *nieuwe feiten* ‘new facts’ is an inanimate NP, but modified, while the subject *mensen* ‘people’ in (76b) is animate but non-modified. When the subject is selected as a good topic, no conflict between Subject First and Topic First occurs, and the subject is preferred in sentence-initial position, as might be the case in (76b). When the subject is not prominent enough to be selected as a topic, a conflict between the two preferences arises. Depending on their relative strength, the subject could be placed either in sentence-initial position, as could also have been the case in (76b), or in postverbal position, as in (76a). Table 25 below shows that the relative frequency of sentence-initial bare subjects increases when going from non-modified to modified, but even more when going from inanimate to animate, suggesting that being animate is the stronger property, as also came out of the regression analysis. When a subject is both animate and modified, the number of sentence-initial subjects exceeds the number of postverbal subjects.

Table 25 Frequencies of the two different word orders (S-V-PP and PP-V-S) for four types of bare plural subject (inanimate non-modified; inanimate modified; animate non-modified; animate modified)

Word order	Bare plural subjects							
	inan + non-mod		inan + mod		anim + non-mod		anim + mod	
S-V-PP	16	21%	16	39%	40	44%	9	60%
PP-V-S	61	79%	25	61%	50	56%	6	40%
TOTAL	77	100%	41	100%	90	100%	15	100%

As for definite subjects, they are intrinsically higher in semantic prominence than bare subjects, and are consequently predicted to be more frequently selected as sentence topics, as illustrated in (77).

- (77) *De glazen stonden op de rand.* [fn001049.23]
 the glasses stood on the edge
 'The glasses were on the edge.'

Although the subject *de glazen* 'the glasses' is inanimate, and the locative PP *op de rand* 'on the edge' could perfectly be a topic, the subject is selected as the topic because it is definite. Hence, no conflict between the Subject First preference and the Topic First preference will occur. However, definite subjects are not always good topics. An NP preceded by a strong quantifier, for example, is lower on the definiteness scale. As a consequence, a conflict between the two preferences could arise. This is illustrated in (78).

- (78) *Allee bij de Griekse tragedies uhm sterven ook*
 INTER]by the Greek tragedies HESIT die also
de meeste uhm personages [fv400165.32]
 the most HESIT characters
 'In the Greek tragedies, most characters in fact die.'

Although the PP adjunct *bij de Griekse tragedies* 'in the Greek tragedies' in (78) does not describe a concrete spatio-temporal location and is

therefore semantically not particularly prominent, the sentence has a PP-V-Subj word order. The fact that the definite subject *de meeste personages* 'most characters' is not a prototypical referential NP, and therefore less prominent than a typical definite, might have contributed to this word order.

Of course, the examples (73)-(78) above are only illustrations of the possible influence on topic selection and therefore on word order of some semantic and discourse properties inherent to NPs and PPs. It is not said that these properties should be seen as necessary or even sufficient conditions for a constituent to be chosen as topics. An example like (78) shows for instance that constituents that are neither semantically nor syntactically very prominent can be in topic position in Dutch. In addition, I have focused in this thesis on sentences with a definite or a bare plural subject and a PP adjunct. Besides PPs, there are other non-subject constituents that can fill the preverbal position in Dutch, such as direct objects, adverbs and *er* 'there'. Which properties determine their position was not investigated here. Furthermore, other types of subjects, such as pronouns and singular indefinites have not been considered yet.

Although the present study shows that semantic factors are certainly influential, probably the most important factor in determining what will be the topic of a sentence is context. This is what I will turn to in the next section.

6.3 Context

The regression models presented in Chapter 5 predict 70.0% of the word order of the sentences with bare plural subjects and 61.4% of the word order of sentences with definite plural subjects correctly. These are improvements of 6.3% and 8.1%, respectively, compared to the baseline. While this might not seem to be very good, keep in mind that the factors included constitute only a selection from the totality of factors that can possibly have an influence on word order. Most importantly, I have only examined factors that determined the inherent semantic or discourse prominence of constituents. I have not considered the influence of the context on topic selection, because it is difficult to determine and to operationalize in a study like the present one. Never-

theless, the linguistic or physical context of a sentence is probably the most important factor in predicting what will be the topic of the sentence, and therefore a strong determinant of word order variation in Dutch natural language. While what a sentence is about can partly be predicted by inherent properties of constituents (e.g. people often talk about humans and agents), people can in principle talk about anything. The intention to convey a certain message already imposes a certain information structure on the sentence to be produced. For example, if you want to make a general statement about, say, penguins, you would probably use a bare plural in topic position, even though bare plurals are normally not very good topics, since it is in this position that bare plurals tend to get a generic interpretation. Furthermore, the topic of the preceding discourse will also influence what will be the topic of the next sentence.

To test the effect of the immediate preceding context on word order of sentences with bare and definite plural subjects and a PP adjunct, a second drag-and-drop production experiment was conducted (Vogels, 2008; Vogels and Lamers, 2008). This experiment was similar to the one discussed in Chapter 3, the only difference being that each experimental item was preceded by a *wh*-question that manipulated the topic-focus structure of the sentence. The *wh*-question could address either the subject (*who*-question) or the locative PP (*where*-question) in the sentence to be produced. Participants were instructed to build a sentence such that it formed a plausible answer to the question. The constituent that was already mentioned in the question became old information in the answer and was therefore likely to become the topic, irrespective of its inherent topic properties; the constituent which the question addressed was new information and was more likely to become focused in the answer. By adding the questions, individual topic characteristics of the constituents were predicted to play only a minor role. It was expected that it would not really matter if the subject was a bare or a definite NP: if a bare subject happened to be the topic, it would go in preverbal position, in spite of it being a bad topic; if a definite subject happened to be the new information, it would go in postverbal position, despite being a good topic.

The prediction that the results would be similar for bare and definite plural subjects was indeed borne out. It was observed that partici-

pants produced their answers to the questions in either of two ways. Some participants started their answers with the old information, placing the new information in postverbal position, according to the iconicity of topic-focus structure. Most participants, however, began their sentence with the new information. Thus, when the subject was the new information (after a *who*-question), it was put more frequently in sentence-initial position than in postverbal position. Similarly, when the locative PP was the new information (after a *where*-question), it appeared more frequently in sentence-initial position than in postverbal position. The explanation for this finding was that old information is likely to be omitted, since this is a repetition of information present in the *wh*-question. The most economic way to answer a question is to provide only the focused information. In the experiment, however, both constituents – the old as well as the new information – had to be used in the answer. Thus, in most cases participants apparently started their answer to the question with the new information, while backgrounding the old information, which was automatically placed in the remaining postverbal slot. In either case, effects of definiteness (definite NPs more preverbal than bare NPs) completely disappeared in this experiment: no significant differences were found between bare and definite subjects regarding their pre- and postverbal occurrences, for both types of *wh*-questions. Apparently, the information structure of a sentence is more important in topic selection than the individual topic characteristics of the constituent in question.

The results of this experiment show that when information structure is manipulated, a semantic factor like definiteness can be overruled in selecting the topic of a sentence. The difference in word order variation between bare and definite subjects found in an experimental situation in which no context was present disappears when context is added. However, in the present thesis a corpus study was performed on natural language data, in which context is naturally present. Still, factors such as definiteness that could be defined independently from context have been shown to affect word order. It is not the case, then, that when context is present, it completely overrules other factors in determining the topic of a sentence and consequently which constituent will be promoted to the topic position. I will leave the influence of context on topics, prominence and word order in natural language for fur-

ther research. However, it is clear that properties inherent to NPs and PPs should definitely be examined when explaining word order variation in Dutch.

6.4 Conclusion

In this chapter, the results of the two logistic regression models presented in Chapter 5 have been compared. I have argued that being definite is an important property that increases the probability of a constituent to be selected as the topic of a sentence. As a consequence, other semantic properties, like being animate, will not have much influence on topic selection, since definiteness is already a strong determinant of topichood. When the property of being definite is absent, however, as is the case with bare nouns, a factor like animacy can come into sight as a decisive factor as to whether the constituent is a good or a bad topic. Other than animacy, the position of both bare and definite plural subjects is affected by their discourse accessibility as measured by the presence of *er* 'there' and the presence of perfective aspect, respectively. In addition, topicalization of (spatio-temporal) PP adjuncts also causes a high frequency of postverbal subjects in both models. Again, however, effects are less strong with definite plural subjects than with bare plural subjects.

I have further illustrated that the different properties of bare and definite plural subjects can be in conflict with each other in selecting a good topic. When such a conflict ultimately leads to not selecting the subject as the topic, a subsequent conflict will arise between the Subject First preference and the Topic First preference. Finally, it is clear that the semantic/discourse properties presented here are certainly not the only factors that determine topic selection. The intentions of the speaker and the (linguistic or physical) context of a sentence are probably for a large part responsible for the choice of the sentence topic and therefore for the position of the subject in Dutch. As additional experimental data suggested, the effect of definiteness can be overruled by the influence of the context. Still, the present study has shown that a constituent's inherent properties can have a significant effect on topic selection and word order in Dutch natural language.

Conclusion

I started out this thesis by distinguishing two types of languages with respect to how word order is determined. On the one hand, there are languages, like English, that have a rather strict word order according to syntactic function: a sentence almost always starts with the subject. On the other hand, in topic prominent languages, like Hungarian and NGT, word order is largely determined by topic-focus structure: a sentence starts with the topic, which may or may not correspond to the subject. I have presented Dutch as a language that is in between these two types. Like English, Dutch has a strong preference to start a sentence with the subject. At the same time, however, the preverbal position in Dutch can be seen as a position to which topics can move. This results in two potentially conflicting preferences: the Subject First preference and the Topic First preference. Since subjects and topics often coincide, the two preferences will go hand in hand in many cases. However, when a subject is not a good topic, the preferences can be in conflict.

In this thesis, I have examined one type of NP that is normally not a very good topic, but can perfectly well be a subject: the bare plural. In contrast to its definite plural counterpart, a bare plural is low on the prominence scale of definiteness. I have discussed experimental data that show that this difference in prominence can indeed influence Dutch word order: definite plural subjects are strongly preferred in sentence-initial position, while bare plural subjects occur about as fre-

quently in preverbal and in postverbal position. This can be explained by the conflict between the Subject First and the Topic First preference: as a subject, a bare plural subject prefers the preverbal position, but as a bad topic it does not prefer this position. This conflict yields variation in word order.

Besides definiteness, more factors can be identified that determine whether something is a good or a bad topic, because they contribute to the semantic or discourse prominence of a constituent. The corpus study presented in this thesis examined the effects of a number of those factors on word order in natural language. The effects appeared to differ substantially between sentences with bare plural subjects and definite plural subjects. Most strikingly, the animacy of the subject has a significant effect on word order when the subject is bare, but not when it is definite. I argued that the reason for this is that being definite is already a strong property in making a constituent a good topic. Topics are often referential and refer to old information. Although animacy is an important contributor to the semantic prominence of a constituent, it is not as strong as definiteness in selecting a topic. When the property of being definite is not present, however, as is the case with bare nouns, animacy can become decisive in whether something is a good or a bad topic. Animate bare plural subjects are more likely to be topics and are therefore more frequently found in sentence-initial position than inanimate bare plural subjects.

In addition, the degree of specificity/referentiality of the subject affects the position of both bare and definite plural subjects. The more specific, the higher the probability that a subject is chosen as a topic and placed in sentence-initial position. Discourse accessibility as measured by the presence of the expletive *er* 'there' and perhaps by the presence of perfective aspect also affects word order: inaccessible subjects are not usually topics, and are thus not preferred in the preverbal position. Finally, PP adjuncts are often topicalized in Dutch, especially when they situate the event described by the sentence in a concrete spatio-temporal location. All these effects seem to be less strong for sentences with definite plural subjects than for sentences with bare plural subjects, supporting the analysis that definiteness is a strong factor in determining topicality.

The different properties contributing to the selection of a good topic can be in conflict with each other: some properties of a particular subject can be prototypical for topics, while others make it a worse topic. Depending on the relative strengths of the different properties the subject will or will not be eventually selected as a good topic. In the former case the Subject First and the Topic First preference will probably go hand in hand, since the subject and the topic will be likely to be the same constituent. In the latter case, it is likely that a subsequent conflict between the two preferences will arise, since the subject will not be the preferred topic. As a consequence, there will be variation in word order: in some cases, the Subject First preference will overrule the Topic First preference, resulting in a bad topic at the sentence-initial position. In the case of bare plural subjects, this will often be accompanied by a shift to a more prominent generic interpretation. In other cases, the Topic First preference will overrule the Subject First preference, resulting in a postverbal subject.

In this thesis, I have shown that although the context and the intentions of the speaker probably have the biggest influence on what a sentence is about, factors inherent to a constituent that can be determined independently from context have an impact on topic selection and on word order in Dutch. For bare plurals, the absence of the strong topic property of definiteness leaves room for other properties, such as animacy, to have an influence.

References

- Aissen, Judith (2003). "Differential Object Marking: Iconicity vs. Economy". *Natural Language and Linguistic Theory* 21: 435-483.
- Alexiadou, Artemis (2000). "Some remarks on word order and information structure in Romance and Greek". *ZASPIL* 20: 119-136.
- Ariel, Mira (1990). *Accessing Noun-Phrase Antecedents*. London/New York: Routledge.
- Beaver, David, Itamar Francez and Dmitri Levinson (2005). "Bad Subject: (Non)-canonicity and NP Distribution in Existentials". In: *Proceedings of Semantics and Linguistic Theory XV*, Effi Georgala and Jonathan Howell (eds.). Ithaca, New York: CLC Publications.
- van Bergen, Geertje (2006). To *ba* or not to *ba*. Differential object marking in Chinese. MA thesis, Radboud University Nijmegen.
- van Bergen, Geertje (to appear). "Placing objects before subjects in Dutch. A comparison between word order variation in experimental and natural sentence production". To appear in *Linguistics in the Netherlands* 26.
- Bock, J. Kathryn and Richard K. Warren (1985). "Conceptual accessibility and syntactic structure in sentence formulation". *Cognition* 21:147-67.
- Bresnan, Joan, Anna Cueni, Tatiana Nikitina and R. Harald Baayen (2007). "Predicting the dative alternation". In: *Cognitive foundations of interpretation*, Gerlof Bouma, Irene Krämer and Joost Zwarts (eds.), 69-94. Amsterdam: Royal Netherlands Academy of Science.
- Bresnan, Joan and Jennifer Hay (2008). "Gradient grammar: An effect of animacy on the syntax of *give* in New Zealand and American English". *Lingua* 118, 245-259.
- Brunetti, Lisa (2009). "On the semantic and contextual factors that determine topic selection in Italian and Spanish". In *Topics Cross-linguistically*, Geertje van Bergen and Helen de Hoop (eds.), *The Linguistic Review* Vol. 26:2-3, 173-176.

- Carlson, Greg N. (1977). A unified analysis of the English bare plural. *Linguistics and philosophy*, 1.413-56.
- Chierchia, Gennaro (1998). "Reference to kinds across languages". *Natural Language Semantics* 6: 339-405. Kluwer Academic Publishers.
- Choi, Hye-Won. (1999). *Optimizing structure in context: scrambling and information structure*. Dissertation, Stanford University. Stanford: CSLI Publications.
- Cohen, Ariel and Nomi Erteschik-Shir (2002). "Topic, focus, and the interpretation of bare plurals". *Natural Language Semantics* J. 10: 125-165.
- Comorovski, Ileana (1991). "Partitives and the definiteness effect". *Proceedings of WCCFL 10*, 91-102.
- Comrie, Bernard (1989). *Language universals and linguistic typology*. 2nd edition. Amsterdam/Philadelphia: Blackwell.
- Coppen, Peter-Arno, Walter Haeseryn and Folkert De Vriend (2002). *De elektronische ANS*. Available on <http://www.let.ru.nl/ans/e-ans/>. Accessed on 07-24-2009.
- Crasborn, Onno, Els van der Kooij, Johan Ros and Helen de Hoop (2009). "Topic agreement in NGT (Sign Language of the Netherlands)". In *Topics Cross-linguistically*, Geertje van Bergen and Helen de Hoop (eds.), *The Linguistic Review* Vol. 26:2-3, 173-176.
- Dahl, Östen and Kari Fraurud (1996). "Animacy in grammar and discourse." In: *Reference and referent accessibility*, Thorstein Fretheim and Jeanette K. Gundel (eds.), 47-64. Amsterdam: John Benjamins.
- Dayal, Veneeta (2003) "Bare Nominals: Non-specific and Contrastive Readings under Scrambling". In: *Word Order and Scrambling*, Simin Karimi (ed.), 67-90. Blackwell Publishers.
- Diesing, Molly (1992). *Indefinites*. The MIT Press.
- van der Does, Jaap and Helen de Hoop (1998). "Type shifting and scrambled definites". *Journal of Semantics* 15(4): 393-417.
- Dryer, Matthew S. 2008. "Order of Subjects, Objects and Verbs". In: *The World Atlas of Language Structures Online*. Martin Haspelmath, Matthew S. Dryer, David Gil and Bernard Comrie (eds.) Munich: Max Planck Digital Library, chapter 39. Available online at <http://wals.info/feature/39>. Accessed on 07-15-2009.

- Epstein, Richard L. (2002). "The definite article, accessibility, and the construction of discourse referents". In: *Cognitive Linguistics* 12(4): 333-378.
- Farkas, Donka and Henriëtte de Swart (2006). "Article choice in plural generics". *Lingua* 117, 1657-1676.
- Givón, Talmy (1976). "Topic, pronoun and grammatical agreement". In: *Subject and topic*, Charles N. Li (ed.), 149-188. New York/San Francisco/London: Academic Press.
- Givón, Talmy (1979). *On Understanding Grammar*. Perspectives in Neurolinguistics and Psycholinguistics. New York/San Francisco/London: Academic Press.
- Grondelaers, Stefan and Dirk Speelman (2007). "A variationist account of constituent ordering in presentative sentences in Belgian Dutch". *Journal of Corpus Linguistics and Linguistic Theory* 3(2), 161-193.
- Haiman, John (1983). "Iconic and economic motivation". *Language* 59: 781-819.
- Haspelmath, Martin (1997). *Indefinite Pronouns: A Typological Perspective*. Oxford: University Press.
- Heim, Irene (1988). *The Semantics of Definite and Indefinite NP's*. New York: Garland.
- Hendriks, Petra, Helen de Hoop and Monique Lamers (2005). "Asymmetries in language use reveal asymmetries in the grammar". *Proceedings of the Fifteenth Amsterdam Colloquium*, 113-118.
- Hogeweg, Lotte and Helen de Hoop (to appear). Children and transitivity: the subject-object asymmetry in a natural setting.
- de Hoop, Helen (1992). *Case Configuration and Noun Phrase Interpretation*. Dissertation, University of Groningen.
- Hopper, Paul J. and Sandra A. Thompson (1980). "Transitivity in Grammar and Discourse". *Language* 56(2), 251-299.
- Kamp, Hans (1981). "A Theory of Truth and Discourse Representation". In: *Formal Methods in the Study of Language*. Jeroen Groenendijk, Theo Janssen and Martin Stokhof (eds.). Mathematical Centre, Amsterdam.
- Keenan, Edward L. and Bernard Comrie (1977). "Noun Phrase Accessibility and Universal Grammar". *Linguistic Inquiry* 8(1), 63-99.
- Kiss, Katalin É. (1987). *Configurationality in Hungarian*. Dordrecht: Reidel.

- Kiss, Katalin É. (2002). *The Syntax of Hungarian*. Cambridge Syntax Guides. Cambridge: University Press.
- Krifka, Manfred (2004). "Bare NPs: Kind-referring, Indefinites, Both, or Neither?" *Proceedings of Semantics and Linguistic Theory (SALT) XIII*, University of Washington, Seattle. Edited R. B. Young & Y. Zhou, CLC Publications, Cornell, 2004.
- Lamers, Monique (2007). "Verb type, animacy and definiteness in grammatical function disambiguation". *Linguistics in the Netherlands* 24. 125-137.
- Lamers, Monique and Helen de Hoop (2008). Animacy and object fronting in Dutch. Unpublished manuscript.
- Levelt, Willem (1989). *Speaking. From intention to articulation*. Cambridge, MA: The MIT Press.
- Levin, Beth and Malka Rappaport Hovav (1995). *Unaccusativity: At the Syntax-Lexical Semantics Interface*. Linguistic Inquiry Monograph 26, Cambridge, MA.: MIT Press.
- Levin, Beth and Malka Rappaport Hovav (2005). *Argument Realization*, Research Surveys in Linguistics Series. Cambridge: Cambridge University Press.
- Longobardi, Giuseppe (2001). "How comparative is semantics? A unified parametric theory of bare nouns and proper names". *Natural Language Semantics* 9: 335-369.
- Lumsden, Michael (1988). *Existential Sentences: Their Structure and Meaning*. London: Croom Helm.
- Milsark, Gary L. (1979). *Existential Sentences in English*. New York/London: Garland Publishing.
- Øvrelid, Lilja (2004). "Disambiguation of syntactic functions in Norwegian: modeling variation in word order interpretations conditioned by animacy and definiteness". *Proceedings of the 20th Scandinavian Conference of Linguistics*, Fred Karlsson (ed.). Helsinki: University of Helsinki.
- Partee, Barbara H. (1987). "Noun phrase interpretation and type-shifting principles". In: *Studies in Discourse Representation Theory and the Theory of Generalized Quantifiers*, GRASS 8, Jeroen Groenendijk, Dick de Jongh, and Martin Stokhof (eds), 115-143. Dordrecht: Foris.
- Reinhart, Tanya (1982). "Pragmatics and Linguistics: An Analysis of Sentence Topics". *Philosophica* 27(1), 53-94.

- Rosenbach, Annette 2005. "Animacy versus weight as determinants of grammatical variation in English". *Language* 81, 613-644.
- Skopeteas, Stavros and Elisabeth Verhoeven (2009). "The interaction between topicalization and structural constraints: evidence from Yucatec Maya". In *Topics Cross-linguistically*, Geertje van Bergen and Helen de Hoop (eds.), *The Linguistic Review* Vol. 26:2-3, 173-176.
- de Swart, Henriëtte and Helen de Hoop (2000). "Topic and focus". In: *The First Glot International State-of-the-Article Book*, Lisa Cheng and Rint Sybesma (eds.), 105-130. Berlin/New York: Mouton de Gruyter.
- van Tiel, Bob and Monique Lamers (2007). "Animacy in verschillende teksttypes". *Tabu* 36: 19-38.
- Tomlin, Russell S. (1986). *Basic word order: Functional principles*. London: Croom Helm
University of Massachusetts, Amherst. New York: Garland.
- Vallduví, Enric (1990). The informational component. Dissertation, University of Pennsylvania.
- Van Nice, Kathy Y. and Rainer Dietrich (2003). "Task sensitivity of animacy effects: evidence from German picture descriptions". *Linguistics* 41(5): 825-849.
- Vogels, Jorrig (2008). The placement of bare plural subjects in Dutch. Bachelor thesis, Radboud University Nijmegen.
- Vogels, Jorrig and Monique Lamers (2008). "The placement of bare plural subjects in Dutch". *Linguistics in the Netherlands* 25, 169-180.
- Yang, Ning and Geertje van Bergen (2007). "Scrambled objects and case marking in Mandarin Chinese" *Lingua* 117, 1617-1635.
- Zwart, Jan-Wouter (1992). "Dutch Expletives and Small Clause Predicate Raising". In: *Proceedings of the North East Linguistic Society* 22, Kimberley Broderick (ed.), 477-491.