University of Groningen

Starting a sentence in Dutch

Bouma, G.J.

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

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

Publication date:
2008

Link to publication in University of Groningen/UMCG research database

Citation for published version (APA):
Bouma, G. J. (2008). *Starting a sentence in Dutch: A corpus study of subject- and object-fronting.* s.n.

Chapter 1

Introduction

A commercial that was broadcast on Dutch television in the past featured a song with the following lines:

(1)   a.   Koning, keizer, admiraal,
                King   emperor admiral
        b.   Popla kennen ze   allemaal!
              Popla know   they all
              'It doesn't matter whether they're king, emperor or admiral, they're all familiar with Popla.'

The direct object of (1b), *Popla*, a brand of toilet paper, directly precedes the finite verb. This position has traditionally been referred to as the *Vorfeld*.

    Dutch is a verb-second, verb-final language: In a declarative main clause the finite verb has to occur in second position, and any non-finite verbs are placed towards the end of the sentence. Otherwise, Dutch allows for a moderate amount of word order variation. One of the liberties a speaker of Dutch has is the choice of a Vorfeld occupant. For instance, the writers of the aforementioned commercial could have used (1b′) to express the proposition that everybody is familiar with their brand of toilet paper.

(1)   b′. Ze   kennen Popla allemaal!
            they know   Popla all
            'They're all familiar with Popla.'

In (1b′), it is the subject *ze* that sits in the Vorfeld. The central research question in this dissertation is what determines the choice for a Vorfeld occupant.

    There are relatively few constraints on which constituent may occupy the Vorfeld. Sentences (1b) and (1b′) only demonstrate two of the many possibilities. However, I will

not be concerned with which Vorfeld occupants are *grammatically* possible. My primary interest in this dissertation is which syntactic, discourse semantic, and communicative factors influence which constituent a speaker puts in the Vorfeld. To get more insight into this issue, I will study Vorfeld occupation by subjects, direct objects and indirect objects in naturally occurring spoken Dutch. The range of investigated constructions is thus as illustrated in (2):

(2) a. Ik heb jou dat verteld.
         I   have you that told
     b. Jou heb ik dat verteld.
         you   I   that
     c. Dat heb ik jou verteld.
         that   I   you
         'I told you that.'

In (2a), the subject occupies the Vorfeld, in (2b), the indirect object does, and in (2c), the direct object. I will use a combination of theoretical modeling and corpus investigation to pin down some of the factors that make a speaker choose (2a), (2b), or (2c).

One type of influence on Vorfeld occupation that we may find comes from known word order tendencies. These tendencies include the tendency to realize subjects earlier on in the sentence than objects, pronouns earlier on than full NPs, and definite or given material earlier on than indefinite or new material. These tendencies have been extensively studied and are fairly well established for the Dutch, and also German, postverbal domain – that is, the domain to the right of the finite verb. The part that these tendencies play in selecting a Vorfeld occupant is not well studied, however. Word order variation with respect to the Vorfeld is less restricted than postverbal word order variation, and as a result harder to investigate. Let us take the tendency to realize subjects early in a sentence as an example. This can not be a categorical constraint on Vorfeld occupation since (2b) and (2c) are grammatical. In (2b) the indirect object is realized before the subject and in (2c) the direct object is realized before the subject. However, intuitively, the subject initial sentence (2a) is the least marked of the three. We might hypothesize that the tendency to realize subjects early on in the sentence affects Vorfeld occupation, and that the unmarked status of (2a) is a result of this. One of the hypotheses to be investigated in this dissertation is the hypothesis that tendencies such as those mentioned above are *global tendencies*. If they are global tendencies, they are not restricted to the postverbal domain, but influence Vorfeld occupation, too.

The hypothesized global word order tendencies need not be the only determinants in the choice of a Vorfeld occupant. Compare the alternative answers to the question in (3A). Capitals indicate main stress.

(3) A Wie heb je het hof gemaakt?
        'Who did you court?'
     B Ik heb Grace KELLY het hof gemaakt.
        I   have Grace Kelly   courted
     B′ Grace KELLY heb ik het hof gemaakt.
        Grace Kelly   I
        'I courted Grace Kelly.'

Intuitively, the difference between (3B) and (3B′) is that *Grace Kelly* receives extra attention when it is put in the Vorfeld: Grace Kelly is presented as an extra newsworthy or unexpected object of courtship. Apparently, the Vorfeld, as a left-peripheral position, has the special function of hosting important material. We shall see that the effects of the nature of the Vorfeld as a position for important material – informal as this concept may be – can be observed separately from the global word order tendencies illustrated above.

   The effects that global word order tendencies and the Vorfeld as a position for important material have on the choice of a Vorfeld occupant have in common that a constituent property is linked to Vorfeld occupation. In the case of the global word order tendencies, we connect being a subject, being a pronoun, or being definite to appearing in the Vorfeld. In the case of the identification as the Vorfeld as a place for important material, we connect being newsworthy, highlighted, or unexpected to Vorfeld occupation. The general topic of this dissertation is what determines Vorfeld occupation. In light of this general topic, the link between constituent properties and Vorfeld occupation raises the first subquestion to be investigated in this dissertation: How do properties of a constituent influence the chance that it is selected as Vorfeld occupant? A large scale corpus study of spoken Dutch is at the heart of answering the first subquestion.

Let us return to the hypothetical task of a speaker of Dutch selecting a Vorfeld occupant. After answering the first subquestion, we know more about what makes a constituent an attractive Vorfeld occupant for the speaker. However, this ignores the purpose of producing an utterance, which is that some intended meaning is transmitted to a hearer. It may be that selecting a certain Vorfeld occupant, even if it is attractive to the speaker, jeopardizes this goal. Consider the example in (4).

(4) Fitz zag Ella
     Fitz saw Ella
     'Fitz saw Ella.' *Or, but dispreferred:* 'Ella saw Fitz.'

The sentence in (4) can in principle be an SVO sentence or an OVS sentence, depending on whether the utterer of (4) selected the subject or the object as the Vorfeld occupant. However, presented like this, the SVO interpretation is strongest, and the OVS interpretation is suppressed. On its own, this is not surprising since there is nothing about *Fitz* or

Ella to tell us which is the subject and which the object. The global tendency to have subjects early in the sentence makes SVO the strongest reading. So, a hearer interpreting (4) will be inclined to interpret the sentence as meaning that Fitz did the seeing, and Ella was seen. This observation about hearer behaviour may have consequences for the freedom of word order variation of the speaker. If the speaker intends to communicate that Ella did the seeing and Fitz was seen, the word order in (4) is a poor choice: The intended meaning would be the one that is dispreferred by a hearer. The chances of successful communication would be better if the speaker puts the subject in the Vorfeld.

The second half of the dissertation approaches the question of what determines Vorfeld occupation from the communication perspective illustrated above. The subquestion to be answered in this part of the dissertation is: How does the chance of communicative success influence the choice of a Vorfeld occupant? Building on the results from the first half of the thesis, I will show in the second half of the thesis that we can formulate a theoretical model of the interaction between speaker and hearer preferences, and that the effect of hearer preferences on speaker choices can be observed in the corpus of spoken Dutch as a statistical tendency in the use of non-canonical word order.

The dissertation is structured as follows. **Chapter 2: Preverbal Behaviour** introduces the topological model of the clause, used in traditional German and Dutch grammars. Formal and functional properties of the left-periphery of the clause are discussed, especially in relation to the Vorfeld. On the basis of this discussion, the range of constructions that will be investigated in the rest of the dissertation is delimited. I will also formulate expectations for the corpus results on the basis of existing knowledge about word order variation in other languages and in other domains in the Dutch clause. The factors to be investigated in the corpus are grammatical function, definiteness and grammatical complexity.

In **Chapter 3: Methods, Techniques & Material**, I will lay out the tools that are used in the corpus investigations, and describe the spoken Dutch corpus *Corpus Gesproken Nederlands* (CGN), that is used throughout the dissertation. A definition of the Vorfeld will be given in terms of the CGN, so that Vorfeld material can automatically be retrieved. The chapter also provides an informal introduction in *logistic regression*, one of the statistical modeling techniques that will be used in the analysis of the corpus data.

Chapter 4: A Corpus Study of the Vorfeld is the first of two empirical chapters in this dissertation. For each of the factors grammatical function, definiteness, and grammatical complexity it is investigated whether they contribute in the choice for a Vorfeld occupant. We will see that not all factors are of the same nature, and that not all factors influence Vorfeld occupation directly. Some of the factors are directly tied to the Vorfeld, some influence word order on a global level, and some only indirectly affect Vorfeld occupation by targeting other positions in the sentence.

Chapter 5: Word Order Freezing is a theoretical chapter. In this chapter, I will extend a formalization of the interaction between speaker's and hearer's preferences using so-called *bidirectional Optimality Theory*. A wide range of data on word order freezing can be elegantly captured in the resulting model. I will also contrast the bidirectional approach with other Optimality-theoretic accounts of word order freezing, and show that bidirectionality is required for a satisfactory account.

The shorter chapter that follows, **Chapter 6: A Corpus Investigation into Word Order Freezing**, is the second empirical chapter of the dissertation. In this chapter, I investigate whether we can observe the predictions that are made by the bidirectional model developed in the previous chapter as quantitative trends in a corpus. We will see that this is the case. The corpus investigations in this chapter thus provide evidence for word order freezing in spoken Dutch discourse. They also provide support for a bidirectional model of word order, and further our understanding of Vorfeld occupation as a whole.

Finally, I will summarize the findings on the determinants of word order in **Chapter 7: Conclusions**. Some of the many directions for further research will be elaborated upon, and I will speculate on some of the theoretical consequences of the findings in this dissertation.