

Subject–Object ambiguities in spoken and written Dutch*

Frank Jansen
UiL-OTS, Utrecht University

1. Introduction

The order of constituents in declarative clauses in written Dutch is not only dependent on their grammatical function, but also on other factors such as information status (the Left-Right principle), the internal complexity of the constituent (discussed in Haeseryn et al. 1997: 1238 and 1245 respectively) and the influence of the semantics of earlier constituents on the interpretation of later ones (Linear Modification (Jansen 2002)). Beside these factors, Jansen and Wijnands (2004) mention a putative factor that is operative on the communicative level: the avoidance of syntactic ambiguity. The purpose of this contribution is to evaluate the relevance of this factor.

Dutch declarative main clauses have a XVS structure. The sentence initial position (X) has to be filled by a constituent. This constituent may have the syntactic function of direct (or indirect) object (henceforth: O), the subject itself (S) or nearly any other syntactical function. In (1) for example S occupies the first position and O occupies the position directly after the verb. And (2) is an example of the inverted order:¹

- (1) Maar ik wil dat toch niet (MJH-16)
but I want that yet not
'But I do not want it yet'
- (2) Die heb ik nog niet gedaan (MJH-16)
this have I yet not done
'I have not done this yet'

In both sentences the alternative order is also possible:

- (1) a. Maar dat wil ik toch niet
(2) a. Ik heb die nog niet gedaan

Therefore, the position before or directly after the tensed verb in these NVN-sentences in itself gives no clue about the syntactic function. However, it is clear in the case of (1) and (2) which NP is S, and by implication which one is O, because of the nominative case of the personal pronoun *ik* 'I'.

Apart from the morphological case there are other means for disambiguating clauses: agreement (3) and the combination of the selection restrictions of the verb with the semantic characteristics of the NP's (4):

- (3) Fred Snijder zouden ze opzij schuiven voor mij (VJH-8)
Fred Snijder would they aside brush for me
'They would brush aside Fred Snijder for me'
(4) Dat heeft Janet helemaal opgevolgd (VJH-4)
that has Janet entirely observed
'Janet has observed that entirely'

Ze 'they/them' in (3) is functionally ambiguous, but the verbal inflection of the verb *zouden* 'would' is plural, which makes it impossible to let it agree with the singular name *Fred Snijder*. This leaves the other plural *ze* 'they' as the only candidate for the syntactic subject of the clause. In (4) it is a selection restriction of the verb *opvolgen* 'observe' that enables the hearer to select *Janet* as the correct subject, for he knows that if one of the NP's of the agentive verb *opvolgen* is animate and the other is inanimate, the animate NP is the subject and the inanimate NP the object. Sentences such as (2), (3) and (4) have something in common: the hearer has to wait until he has processed *ik*, *ze* and *opgevolgd* respectively before he is able to parse the sentence. Therefore these sentences are called temporarily ambiguous.

As morphological case is obsolete in Dutch outside the realm of personal pronouns, it cannot account for the disambiguation of NVN-sentences with nominal groups as S and O. The same applies to sentences where S and O have the same number, or where the selection restrictions of the verb give no clue which NP is the subject. In those circumstances the reader has to revert to other resources to disambiguate the sentence. For example in:

- (5) Een duivenliefhebber zal een kat niet doodmaken (Jansen ML-9)
a pigeon fancier will a cat not kill
'Pigeon fanciers will not kill a cat'

- (6) Een speciaal meisje hield die (sc the teacher) erop na (JansenML-9)
 a special girl held this particle particle
 ‘This teacher held a special girl on the sly’

In (5) the knowledge of the world of the reader will make him choose the meaning of a pigeon fancier killing a cat, instead of the other way around. In (6) it is clear from the context (a story from transgressions in the school life of the informant) that it is the teacher who held the girl on the sly, instead of the other way around. This type of sentence is called globally ambiguous.

How are those ambiguous clauses distributed in speech and writing? In Section 2 I will present an argument why we expect more communicative problems with (especially globally) ambiguous sentences in writing than in speech, which leads to the hypothesis that they will occur more frequently in speech. In Section 3 the corpora and the details of the method are discussed. In Section 4 the results are presented, and in Section 5 the conclusions.

2. Cognitive economy for speakers and writers

Since the introduction of the cooperative principle and its accompanying maxims (Grice 1975) it is taken for granted that we all have the basic want to use our language effectively and efficiently. The effectivity of language use pertains to the result of the communication process: the sender wants the receiver to interpret the message as the sender intended it. Efficiency has to do with the communication process itself, which should cost the language user as little time and effort as possible.

How those basic requirements are met depends on the particularities of the communication situation. An important aspect of this situation is the mode of communication: speech or writing. I assume that speakers and writers differ in the way they unconsciously make an appraisal of the most desirable message form, because a ‘principle of cognitive economy’ works differently in the two modes. In Troost, Jansen & Sanders (ms) we tested this idea by comparing the position of adverbial clauses in two writer–reader situations (chat boxes and public documents). We found that writers who are under severe time pressure (as in a chat-box) apply a rule of thumb like ‘Put adverbial clauses at the end of the sentence’. When writers have more time for composition, they apply a set of more subtle rules to achieve more thematic coherence.

When this idea of cognitive economy is applied to ambiguous sentences in speech, we expect that it is economical for the speaker to get his meaning

across by saying as few words as possible and to deliver them as fast as possible. If the speaker is too economical (by using too few words), he is not effective anymore. The hearers will immediately ask for clarification and our speaker will have to rephrase his message or elaborate it by giving extra details. A speaker will strive for efficiency by organizing his sentences according to the left-right principle, which enables them to start with phrases referring to entities already known to him, because they have an active status in his mind (Levelt 1989: Chapter 7). It is questionable whether speakers have other formulation options than these, as the speech production process is very fast, and an unprepared speaker is only able to plan just a little ahead.

If this line of reasoning makes sense, we expect all categories of ambiguous sentences to occur in speech, even the globally ambiguous ones, which are the most threatening for communication success.

In a classical writer–reader situation the cognitive economy principle predicts that it is economical for the writer to facilitate his readers to get the intended meaning right away. The professional writer knows that he will not be around to help his readers to understand his text. Therefore he will try, by refraining from globally ambiguous sentences, to minimize the risk that his readers reconstruct a wrong mental representation. Beside this want to be effective, he wishes his text to be read efficiently. To achieve this goal he will maximize the use of constituent orders that are expected by his readers. In other words he will prefer SO-orders, as these are far more frequent than the OS-order in written texts. Nieuwborg (1968: 298) presents the following counts based on literary texts from the Netherlands and Flanders: SVO 342 (90%) versus OVS 38 (10%).

SVO-sentences occur not only more frequently, they are also easier to understand. Lamers (2001) did a comprehension study on temporarily ambiguous main clauses. She found that readers rely heavily on morphological information, in particular in the case of the NP's (*hij/hem*) and on semantic information, in particular the combination of the animacy of the NP's with the selection restrictions of the verb. Furthermore, Lamers reports that readers — at least in some sentence types — first assume an SO-order and reanalyse the sentence as an OS structure if they are forced to do so by the case of the inverted subject.

Therefore professional writers will try to avoid syntactically ambiguous OS-sentences. And as writers, they are also in a better position than speakers to succeed in this, for two reasons. Firstly, the production of written text is considerably slower than speaking, which enables the writer to monitor his

formulation in this respect. Secondly the writer can reorder an OS-sentence in the prefinal version during the editing phase (Dirksen 1990).

I have analysed NVN-sentences in a corpus of speech and one of written texts, to find out if these hypotheses hold water.

3. Two corpora

3.1 Material

I used two speech corpora that are somewhat older but that I know rather well from recording or collating them. The countings of the first phase are based on five informal interviews used for the frequency dictionary of spoken Dutch (De Jong 1979).² Those data were supplemented by another 25 informal interviews of this corpus and by the informal parts of 13 interviews I recorded myself for my dissertation (Jansen 1981). These speech corpora comprise about 110,000 words in all.

The corpus of written Dutch consisted of the corpus of 18 journalistic articles (sports reports, columns, background stories) that was used by Jansen & Wijnands (2004). This corpus was complemented by a corpus of recent sports reports in *De Volkskrant* en *NRC-Handelsblad*.³ This corpus written language comprises 34,000 words in all.

3.2 Method

In the first phase (see 4.1) I analysed declarative main clauses with an 'NVN'-structure: the first three constituents had to be an NP (nominal group, noun or pronoun), followed by the tensed verb and another NP. The NP's had to meet two conditions. The first condition was that the NP's could occupy each other's position. In other words, a variant of the sentence with the first NP in the position after the verb, and the second NP on the sentence initial position, had to be acceptable as well. The consequence of this condition was that NVN-sentences in the speech corpus had to be excluded if the second NP was the oblique case of an unaccented personal pronoun (*me, je, 'm, 'r*), because those pronouns cannot occupy the sentence initial position.

The second condition for admission in the selection was that the syntactic function of the NP's was subject, direct object, or indirect object. Subject and object clauses were excluded for two reasons. Their text function is different from NP's. In the journalistic texts in our corpus, the subject and object clauses

were primarily used for citations. Furthermore the distribution pattern of dependent clauses is different from that of NP's (Hawkins 1994; Jansen 2003).

In this first phase I selected more than 100 NVN-sentences (123 in the written corpus and 103 in the speech corpus), and categorized them by the way they could be disambiguated. There were four categories: two categories of formal disambiguation, viz. case (see for example (1, 2)) and agreement (3). One semantic category: selection restrictions (4); and one pragmatic category: context (5, 6). When there was more than one formal or semantic means to disambiguate a sentence, I categorized the sentence in all categories. Therefore the row totals in the Tables 1 and 2 are greater than the number of sentences investigated.

In the second phase of the investigation (4.2) I focussed on formal ambiguous NVN sentences. Therefore the NVN-sentences in the semantic and pragmatic categories of the first phase were complemented with another 142 sentences of those categories in the written corpus and 168 in the speech corpus.

4. Results

4.1 All NVN-sentences

How frequent are the various NVN-sentences in the two corpora? The answer is in Table 1.

Before we analyse the results displayed in Table 1, a remark is in place about the striking difference in the frequencies of SO- and OS-orders in the two modi: while the SO order is predominant in the written corpus (142 SO

Table 1. Frequencies (and percentages) of disambiguation categories of SO and OS NVN sentences in a text and a speech corpus.

		disambiguated by				
		case	agreement	semantics	context	total
S-O						
	writing	26 (18)	32 (23)	66 (47)	18 (13)	142
	speech	26 (34)	13 (17)	36 (47)	2 (3)	77
	total	52 (24)	45 (21)	102 (47)	20 (9)	219
O-S						
	writing	10 (26)	10 (26)	18 (46)	1 (3)	39
	speech	45 (39)	11 (10)	57 (50)	2 (2)	115
	total	55 (36)	21 (14)	75 (49)	3 (2)	154
	total	107 (29)	66 (18)	177 (48)	23 (6)	373

(78%) versus 49 OS (22%)), it is a minority in the speech corpus (77 SO (40%) versus 115 OS (60%)). This last proportion is partly an artefact of the selection method that excluded all NVN-sentences with unaccented object pronouns in second position (see Section 3.2). In order to assess the influence of this artefact, I reanalysed the same texts of the speech corpus with incorporation of the unaccented oblique pronouns, with the following results: SO 136 (54%) — OS 115 (46%). So, although the balance is shifted somewhat towards the SO order, we still find significantly more OS- structures in speech ($\chi^2 = 18.89$, $p < .0001$), which has to be attributed to the fact that the Left-Right-principle works more rigorously in speech than in writing.

The frequencies in Table 1 demonstrate how important disambiguation by the semantic means of selection restrictions is: it accounts for nearly half of the cases in both corpora. This result is in accordance with the corpus study of word order in relative clauses by Mak, Vonk and Schrieffers (2002) who found that only 10% (29/286) of the relative clauses in their corpus could not be disambiguated by selection restrictions on the animacy of S and O.

The next category in frequency is case (29%), which is more important in speech (71/192 = 37%) than in writing (36/181 = 20%).⁴ This is not surprising when one realises that personal pronouns are far more frequent in spoken language than in written language. Disambiguation by agreement occurs less frequently (18%).

Disambiguation by context is the least frequent category: 6%. However, the results in Table 1 do not support the hypothesis that professional writers avoid global ambiguity entirely. On the contrary, there seem to be more globally ambiguous sentences in writing (8%) than in speech (2.5%). As this last result may be attributed to the relative scarcity of cases, I decided to expand this part of the corpus in the second phase of the investigation.

4.2 Formal ambiguous NVN-sentences

In order to obtain a greater understanding of the characteristics of formal ambiguous NVN-sentences, we focussed on the semantic and pragmatic disambiguation categories. The results are presented in Table 2.

Table 2 shows that the semantic cases, which I consider as exemplary for the temporarily ambiguous sentences, is three times as frequent (317 cases) as the pragmatic (globally ambiguous) category (97 cases). Furthermore the semantic and context cases together demonstrate that there seems no avoidance of ambiguity in speech, not for SO-sentences (83) nor for OS-sentences (130).

Table 2. Frequencies (and percentages) of disambiguation categories of formal ambiguous SO and OS NVN sentences in a text and a speech corpus.

	semantic	context	total
S-O			
writing	114 (64)	64 (36)	178
speech	66 (80)	17 (20)	83
total	180 (69)	81 (31)	261
O-S			
writing	20 (87)	3 (13)	23
speech	117 (90)	13 (10)	130
total	137 (90)	16 (10)	153
total	317 (77)	97 (23)	414

As for the written language, Table 2 gives no indication that ambiguity in general is avoided. However, the distribution of SO and OS-cases is skewed: in both the semantic and in the context category there are many SO-cases and few OS-cases, compared with the proportions in the speech corpus, a difference that is statistically significant ($\chi^2 = 109.15$, $p < .0001$). When we compare the degree of skewedness in the semantic and the context categories, we see that the distribution in the context category is significantly more skewed than the distribution in the semantic category ($\chi^2 = 4.81$, $p < .05$). This result is in accordance with the results of the corpus study of OS-structures in relative clauses by Mak, Vonk and Schrievers (2002), who found in their corpus of 286 cases only 8 examples (= 3%) of OS-structures with S and O either animate or inanimate.

I consider this difference in skewedness as evidence for the hypothesis that professional writers try to avoid temporary ambiguous OS sentences, and try even harder to avoid globally ambiguous OS-sentences. However, there is no indication at all for the plausibility of the idea that writers avoid ambiguous SO sentences in general.

4.3 Qualitative analysis of the context OS-sentences in writing

When quantitative research offers results as reported in 4.2, and we notice a few cases in one particular category, it is insightful to analyse those cases more qualitatively. Therefore I will discuss the globally ambiguous OS-sentences in writing now.

This type of sentences in writing turns out to be different from the context OS-cases in speech, which are relatively straightforward, like (7).

- (7) (Het is niet modern als ze Tante zeggen.) Pappa en Mamma zeggen ze
 (It is not modern when they say aunt.) dad and mum say they
 tegenwoordig ook niet meer (VJL-11)
 nowadays also not more
 ‘They do not say dad and mum anymore nowadays’
- (8) Zoveel emoties als de belevenissen van Oranje
 so many emotions as the adventures of Orange (the national team)
 in 2004 zullen de duels van volgend jaar vermoedelijk niet oproepen.
 in 2004 will the duels of next year supposedly not evoke
 ‘The matches of next year will not evoke so much emotion as the
 adventures of Orange in 2004.’ (*Volkskrant* 18–11–2004)

Sentence (8) is a rather satisfying example of ambiguity, because duels can evoke emotions and emotions can evoke duels. It is not an impeccable example, however, because *emoties oproepen* ‘evoke emotions’ can be considered as a collocation and *duels oproepen* not.

The two other cases are more problematic, for different reasons. In (9) the reader has to do some additional interpretation work before he has construed the alternative SO-interpretation:

- (9) Want Zwolle krijgt de rechtgeaarde Go-Ahead-fan niet over zijn lippen
 for Zwolle gets the right minded Go-Ahead-fan not across his lips
 ‘The rightminded Go-Ahead fan never could bring himself to say Zwolle’
 (*Volkskrant* 29–11–2004)

The reader has to interpret the town name *Zwolle* as shorthand expression for the collective of inhabitants of this town and *de rechtgeaarde Go-Ahead-fan* ‘the right minded Go-Ahead-fan’ or as a de dicto expression or as short hand for the name of this fan. Case (10) is somewhat dubious because the meanings of the two readings are very similar:

- (10) Onderdeel in die voorbereiding zal een trip naar Italië moeten vormen
 part of this preparation will a trip to Italy have to constitute
 ‘A trip to Italy will have to be a part of these preparations’ (*Volkskrant*
 15–11–2004)

The verb *vormen* ‘constitute’ is used more like a copula in (10), not unlike *constitute* in the English expression ‘to constitute an exception’, where *constitute* can also be replaced by *be*. By consequence (10) a is a good paraphrase of (10):

- (10) a. Onderdeel in die voorbereiding zal een trip naar Italië moeten zijn

Another argument for a copula analysis in (10) is the absence of the article *the* in *Onderdeel in die voorbereiding* ‘part of this preparation’ which contributes to the copula interpretation of *vormen*.⁵ Now it is an aspect of copula-constructions that the thematic roles of subject and object are less different than in sentences with a transitive verb. Therefore the syntactic ambiguity of (10) has no consequences for the role of agens versus goal, as is the case in ambiguous sentences with transitive verbs. In other words, the ambiguity of (10) is rather harmless from a communicative point of view.

5. Conclusion

The results of the quantitative and qualitative analyses of ambiguous sentences in writing and speech give partial support to the hypotheses following from the cognitive economy principle, put forward in Section 2. On the one hand, there is no indication that all categories of ambiguous sentences are avoided. On the other hand there is evidence that professional writers try to avoid temporarily ambiguous OS-sentences, and try even harder to avoid globally ambiguous OS-sentences. As these professional writers instead seem to choose the SO alternative structures that are ambiguous as well, we may conclude that these writers presume that SO-ambiguity is communicatively harmless, because their readers rely on a ‘the first NP is the subject’-strategy. Finally, we have to ask ourselves whether there is a relation between this way of ambiguity avoidance and the Left-Right principle. We have seen that the preference for OVS-structures in speech may be attributed to the unqualified application of this Left-Right principle. Now I may conclude that the avoidance of ambiguous OS-structures has a blocking effect on the operation of the Left-Right principle in writing.

Notes

* I thank the listeners to my presentation on the Taaldag for their questions and suggestions, Pim Mak, Ted Sanders and an anonymous reviewer for their comments on an earlier version of this paper, and Ester Sorm for some additional counting operations.

1. The examples in this section are taken from the speech corpora used in this investigation (for details see Section 3, and note 2).

2. The following interviews were used for the first phase of the investigation (M = male, V = female, H is higher educated, L is lower educated, J = young and O = old): MJL-16, MOH-7, MOH-13, VJH-6 and VJH-8. In the second phase were added: MJL-8, MJL-3, MJL-13,

MJH-8, MJH-9, MJH-9, MJH-11, MJH-13, MOL-3, MOH-12, MOL-19, MOL-21, VJH-4, VJH-11, VJL-9, VJL-11, VJL-15, VJL-16, VOL-9, VOL-10, VOL-12, VOL-14, VOH-1, VOH-2, VOH-9, VOH-10, VOH-11.

3. In the editions of both newspapers of 28–10, 1–11, 15–11, 18–11, 22–11, 24–11, 25–11, 26–11, 29–11, 2–12, 3–12, 4–12 in 2004.
4. If sentences with the unaccented oblique pronouns in the speech corpora are incorporated, the proportion is hardly different: 40% (instead of 37%).
5. I thank one of the listeners on the TIN-day for this observation.

References

- Dirksen, A. (1990) *Monitoring processes in language production*. Diss. Utrecht University.
- Grice, H.P. (1975) Logic and conversation. In P. Cole and J.L. Morgan, eds., *Syntax and semantics 3. Speech acts*. Academic Press, New York, p. 41–58.
- Haeseryn, W. et al. (1997²) *Algemene Nederlandse spraakkunst*. Wolters-Noordhoff, Groningen.
- Jansen, F. (1981) *Syntaktische konstrukties in gesproken taal*. Huis op de drie grachten, Amsterdam.
- Jansen, F. (2002) 'Preposed adverbial phrases in Dutch texts. Evidence for the Left Right Principle or for Linear Modification'. In H. Broekhuis and P. Fikkert, eds., *Linguistics in the Netherlands 2002*. John Benjamins, Amsterdam/Philadelphia, p. 97–106.
- Jansen, F. (2003) 'Sentence processing theories and the position of complex constituents in Dutch texts'. In L. Cornips and P. Fikkert, eds., *Linguistics in the Netherlands*. John Benjamins, Amsterdam/Philadelphia, p. 83–92.
- Jansen, F. and R. Wijnands (2004) 'Doorkruisingen van het Links-Rechtsprincipe'. *Neerlandistiek* 2004.
- Jong, E.D. de (red.) (1979) *Spreektaal. Woordfrequenties in gesproken Nederlands*. Bohn, Scheltema & Holkema, Utrecht.
- Lamers, M. (2001) *Sentence processing using syntactic, semantic and thematic information*. Groningen dissertations in linguistics 33.
- Levelt, W.J.M. (1989) *Speaking: From Intention to Articulation*. The MIT Press, Cambridge/London.
- Mak, W.M., W. Vonk, and H. Schriefers (2002) 'The influence of Animacy on Relatives Clause Processing'. *Journal of Memory and Language* 47, 50–68.
- Nieuwborg, E. (1968) *De distributie van het onderwerp en het lijdend voorwerp*. Plantijn, Antwerpen.
- Troost, C., F. Jansen, & T. Sanders (ms) 'De plaats van temporele en causaal complexe zinnen: evidentie voor cognitieve economie bij tekstproductie?'

