The interpretation of different types of data

Corrien Blom
Free University of Amsterdam

1. Introduction

Since Koster (1975) it has generally been assumed that Modern Dutch is an OV language (leaving aside theoretical approaches such as those addressed in Kayne 1994). The OV deep structure order is reflected in the Modern Dutch subordinate clause, as in (1a). In main clauses, the finite verb is moved to second position (Verb Second movement), leaving a trace. This is shown in (1b).

(1) a. dat Jan Marie kust
    that John Mary kisses
    ‘that John kisses Mary’

b. Jan kust Marie t
   ‘John kisses Mary’

Contrary to what is the case for Modern Dutch (MoD), there is not much agreement regarding the word order of Middle Dutch (MiD) (Van den Berg 1980, Burridge 1993, Gerritsen 1978, 1987, De Meersman 1980, Neeleman and Weerman 1992). Although no one has hypothesized a strict VO grammar for MiD, it has often been reported that MiD has more VO structures than MoD. Whether these VO structures reflect underlying or surface patterns, however, is mostly left implicit. Moreover, it is not clear in what frequency VO order appears, nor whether it occurs under specific conditions and if so, what these conditions might be.

The purpose of this study is to find answers to these questions. First, the relevant literature is considered (2), after which the results from a corpus study are presented (3) and discussed (4). We will see that MiD is basically an OV language and that VO order involves marked structures, related to highly specific conditions.
2. Evidence from the literature

Three important questions that need to be answered are the following:

1. Does MiD really contain VO tokens that are unacceptable in MoD?²
2. If so, what do these VO tokens look like, in what kinds of contexts and in what frequency do they appear?
3. What do these VO tokens tell us about the underlying word order of MiD?

In traditional generative literature it is generally assumed that MiD is an OV language, but that it shows VO characteristics, especially in poetry, that are absent in MoD (Van den Berg 1980). It is not clear, however, in what frequency and under what conditions these VO characteristics appear.

Neeleman and Weerman (1992) (henceforth N&W) claim that direct objects appear both pre- and postverbally in MiD. They explain this by (a modification of) Case Theory.³ One of their examples is the italicized clause in (2), in which the direct object of a subordinate clause appears postverbally. This word order is not acceptable in normal MoD language use. However, a look at the context of this clause, given in (2), shows that it belongs to a rhyme text and that the object is in a special rhyming position.

(2) Omtewetendaeraf niemaren to learn there from news
    Si hadden onverre gevaren they had not far driven
    Dat si ontmooten ene ioncrouwe when they met a lady
    op een part drivende groete rouve on a horse driving very sad
    (Roman of Lancelot, l. 1858)

Almost all of N&W’s data involve such rhyming words, which means that these word orders can be explained by extraposition because of rhyme demands. Rhyme texts contain many unusual word orders and in this respect, MiD does not seem to differ from MoD. Rhyme data, then, need to be used with caution in word order studies; these analyses require (additional) prose data (Los 2002).⁴

Evidence from other theoretical approaches points at an OV preference varying from 66% (Gerritsen 1978, 1987) to 90% (De Meersman 1980), but here no explicit remarks regarding underlying vs. surface word order are made.

Several factors possibly related to VO order are mentioned in the literature, but there is not much agreement regarding their influence. For example, whereas De Meersman reports influence of the factors NP complexity, clause type, verbal cluster complexity, verbal cluster type, semantic category of the object NP, and the occurrence of administrative formulas, Gerritsen explicitly reports not finding any influence of such factors.

According to Burridge (1993), VO order is triggered by focality, which translates as a tendency of focal elements to appear postverbally in MiD. Burridge claims
that MiD word order is determined by the pragmatic function of the constituents, but syntax and pragmatics are better dealt with separately.

Few examples of MiD VO structures are given in the literature, and therefore it is unclear what these structures look like. In Gerritsen (1987) we see that this may lead to problems: from her few data it appears that she considers MiD data showing Verb Projection Raising (VPR) (omdat ze wilden naar huis gaan ‘because they wanted to go home’) as VO evidence. VPR, however, is still a common syntactic process in Modern Flemish (Haegeman 1992). Other scholars appear to have included accounts in their data, which even in MoD often have the structure of (3).

(3) dat zij aan Henk zullen verkopen 2 koeien
dat zij aan Piet zullen verkopen 5 koeien
dat zij aan Jan zullen verkopen 3 koeien

‘that they will sell to Henk: 2 cows’, etc.

All direct objects in (3) are positioned postverbally in order to facilitate the summing up, and this does not tell us anything about the underlying word order. An additional difficulty is presented by the different views on what constitutes underlying order. For some scholars, ‘unmarked’ (and therefore frequent) order equals underlying order, while for others the underlying order does not need to be frequent, or even attested; Kayne (1994) is a case in point. Frequency indicates that a particular structure is robustly attested in a particular period, while historical syntacticians do well to approach very infrequent ones with caution, as they can be the result of imperfect copying and scribal slips. Attested structures should always be viewed in their context. For more discussion, see Denison (1993:4, 27–28) and Fischer et al. (2000:209–210).

To conclude we can say that the literature on MiD word order phenomena reports that OV order is most frequent and also that this order probably represents the underlying word order. However, VO order does occur. Although some factors related to the occurrence of VO order are mentioned, there is very little agreement, both regarding the frequency of VO order and the influence of various VO triggers. With respect to the different types of data, it appears that rhyming lines and data from administrative texts, such as accounts, should be treated with caution when studying word order phenomena.

The questions remain whether MiD does indeed contain VO tokens that are unacceptable in MoD, in what frequency they appear, and what these VO tokens exactly look like. To find answers to these questions, I carried out a small corpus study, of which I will report the results in the next section.
3. A corpus study of Middle Dutch word order

3.1 Introduction

Poetry was excluded from the corpus to minimize possible effects of rhyme (Los 2002, but see Fischer et al. 2002: 31–32 for reasons not to exclude verse entirely).

To study possible genre-specific effects, I selected MiD texts belonging to three different genres: official, religious, and narrative texts. The official texts are legal agreements, all from the Corpus Gysseling, the religious texts are fragments of psalms and tractates, and the narrative texts are parts of prose novels. All texts were taken from the CD-rom Middelnederlands.

I then looked at clauses with direct objects and unambiguousOV or VO order, i.e. main clauses with more than one verb and subordinate clauses. I excluded clauses with sentential or pronominal direct objects. The reason for doing this is that sentential objects always follow the verb in MiD, whereas pronominal objects always precede the verb (De Meersman 1980).

I classified these clauses according to their value with respect to six parameters (based on the literature), namely clause type (main, relative, or adverbial clause), verbal cluster complexity (simplex or complex), official formula (present or not), semantic category of the object NP (denoting amount/quantity or not), verbal cluster semantics (direct objects of the verbs of naming heten and noemen ‘to be called/to call’ or not), and NP complexity (simplex or complex).

With Vanden Berg (1980), we hypothesize that MiD is an OV language. In addition, we expect that if VO order occurs were OV is expected, this will be triggered by specific conditions, as a consequence of extraposition rules that are no longer present in Dutch and are related to the six parameters above.

The corpus consists of 900 lines (about 300 lines per text type) and contains 213 direct objects. 149 (70%) direct objects show OV order and 64 (30%) show VO order. I will now discuss per text type the different patterns I found in detail.

3.2 Official texts

The numbers of OV and VO tokens in official texts are given in Table 1.

At first sight, Table 1 seems to question our hypothesis that MiD is an OV

<table>
<thead>
<tr>
<th>Position of O</th>
<th>Numbers</th>
</tr>
</thead>
<tbody>
<tr>
<td>O-total</td>
<td>67</td>
</tr>
<tr>
<td>OV</td>
<td>36 (54%)</td>
</tr>
<tr>
<td>VO</td>
<td>31 (46%)</td>
</tr>
</tbody>
</table>
language: we have almost as many VO as OV tokens. I next checked whether the six
above-mentioned parameters play a role in these VO constructions.

There was no relationship between word order and clause type or verbal cluster
complexity. However, in 12 of the 31 VO constructions the object is part of an
official formula and/or denotes an amount or quantity, as in (4–5).10

(4) dat daht & conuent van Ninuue hebben ghecoht terwert ene hofst met allen
gelegen te Bochoute bouen hare hof ane de strate.
'that the abbot and the convent of Ninove legitimately bought a farmstead with
everything belonging to it, situated in Boechoute above their own land near the
road.' (0015–33)

(5) Dat si sullen gheuen van elken ghemete iii s. jarlich te Sainte martinsmesse.
'That they will give for every 'gemet' (measure of land) 3 sol a year at Saint
Martin's Day.' (0017–2)

Constructions like (4)–(5) are typical of legal documents. The direct object gives
information about the object of the agreement, which is the amount of money,
quantity of land, or property that is bought, sold, or delivered. This crucial information
is always presented in a fixed form with the focal object in postverbal position (De
Meersman 1980, Burridge 1993) and can be characterized as a focus construction.

In four VO tokens the direct object is the object of a naming verb, as in (6).

(6) ende dat lant es gheeten gillis roden hofstede.
‘and the land is called the farmstead of Gillis Roden.’ (0051–10)

In constructions like (6) there is an implicit colon (:) behind
gheten 'called',
followed by the object, which is in fact a predicative NP rather than a direct object
NP (see also footnote 6). Conform the findings of Burridge (1993) and De Meers-
man (1980), these objects always follow the verb. Constructions with naming verbs
can be characterized as another type of focus construction.

Next, we look at the parameter of NP complexity. As a working hypothesis I
define complex direct objects as direct objects containing a relative clause (but the
object itself is not a clause). An example is given in (7).

(7) dat si segghen souden die beste wareit, diese kinden van dien dat wouters recht
ware te duffle.
‘that they would say the best truth they knew with respect to the question what
Wouter's rights in Duffel would be.' (0120–37)

In clauses such as (7) the direct object is complex and therefore cannot be posi-
tioned before the verb as a whole. It is worth noting that in MoD direct objects such
as in (8) are not completely positioned preverbally either. MoD, however, opts for
a different solution to lighten the complexity burden: the direct object is separated
in such a way that the NP is preverbal, but the relative clause belonging to it is
extraposed to the postverbal position (Geerts et al. 1997: 1246).
(8) a. Ik heb eindelijk dat boek gelezen dat jij zo geweldig vindt.
    b. Ik heb eindelijk dat boek dat jij zo geweldig vindt gelezen

‘I finally read that book that you think is so fantastic’

According to Geerts et al. the order in (8b) is possible, but the order in (8a) is certainly preferred for ease of processing.¹¹

The official texts contain 10 VO constructions with a complex direct object. It must be noted that these are not cases of ‘ordinary’ heavy NP shift, since not every heavy NP is extrapolated, but only those NPs containing a relative clause.

These parameters, then, turned out to play a role in 26 of the 31 VO constructions. This left five VO constructions. To make sure that the six parameters are really related to VO order, the OV tokens had to be checked, too.

The official texts contain 36 (54%) OV tokens. Most of the direct objects in these constructions are simple NPs (33), with or without a modifier that has no clausal status, as the one in (9).

(9) To all the ones that will see this letter.

The example in (9) seems to represent another ‘formula’, since exactly this pattern (i.e. a direct object consisting of demonstrative pronoun and a noun referring to the legal agreement itself) is found 17 times at the beginning and at the end of the legal documents. All these constructions show OV order.

From a comparison between the two types of official formulas (examples (4)–(5) vs. (9)) it can be concluded that there is a correlation between the information status of the direct object in these formulas and its position: whereas ‘new’ objects in formulas are in postverbal position (4)–(5), ‘old’ objects are preverbal (9). The pronominal modifier in constructions such as (9) signals the old information status of the direct object, which explains why these kinds of direct objects are never found in postverbal position. Only administrative formulas with a focal direct object, then, show VO order.

The three remaining OV constructions contain an object consisting of two co-ordinated simple NPs, linked by the MiD conjunction markers ende ‘and’ or noch ‘nor’. An example is given in (10).

(10) That my lord Berthout could neither establish shooting rights, nor marksmen rights on Wouter’s men and servants. (0120–210/16)

To sum up we can say that the direct objects in OV constructions are clearly different from those in VO constructions. The OV objects are simplex, i.e. they consist of a simple NP, of two co-ordinated simple NPs, or of a NP with an
adjectival modifier (without clausal status). If these direct objects are part of fixed formulas, they unambiguously give old information, signalled by a demonstrative pronoun. The VO objects, on the other hand, are often complex in the sense that they contain a relative clause. Other typical VO objects are objects in fixed focus constructions as exemplified in (4)–(6).

Although the official corpus contains more VO tokens than expected, these VO tokens occur under specific conditions. As the presence of administrative formulas turned out to be such a significant VO trigger, we needed to look at texts where they are absent.

3.3 Religious texts

The numbers of OV and VO tokens in the religious texts are given in Table 2.

<table>
<thead>
<tr>
<th>Text number</th>
<th>I</th>
<th>II</th>
<th>III</th>
<th>IV</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>O-total</td>
<td>11</td>
<td>33</td>
<td>11</td>
<td>16</td>
<td>71</td>
</tr>
<tr>
<td>OV</td>
<td>6 (35%)</td>
<td>25 (76%)</td>
<td>8 (73%)</td>
<td>14 (87%)</td>
<td>53 (75%)</td>
</tr>
<tr>
<td>VO</td>
<td>5 (45%)</td>
<td>8 (24%)</td>
<td>3 (27%)</td>
<td>2 (13%)</td>
<td>18 (25%)</td>
</tr>
</tbody>
</table>

There is no relationship between word order and clause type, verbal cluster complexity, administrative formulas, and NPs denoting amounts or quantities. There are, however, two constructions in which the direct object is the (focal) object of a verb of naming, as in (6). The only remaining parameter possibly related to VO order is NP complexity, where complex direct objects are (provisionally) defined as those direct objects containing a relative clause (cf. (7)). This kind of NP complexity is found in five VO tokens.

A striking property of the religious texts concerns the occurrence of syntactic and semantic parallelisms, in which the direct objects are generally postverbal. An example is given in (11).

(11)    Ende ghelijckerwijs dat de coninc Saul *scoerde* den mantel des propheten Samuels, alsoe pyen si hen te scoerne einicheit kersten gheloeufs, ende alle warachtighe leerlinghe, ende doechsam leven.
        'And likewise that King Saul tore apart prophet Samuel’s coat, in that way they force him to tear up the harmony of the Christian faith, and all truths of the Christian religion, and the virtuous life.' (1-12)

(11) shows a parallelism between the tearing apart of Samuel’s coat by King Saul and the tearing apart of the harmony of the Christian faith by others. The direct object is the parallel part of the construction and thus clearly has focus. Constructions
such as (11) can be seen as a type of focus construction typically found in the elevated style of religious texts (which is characterized especially by frequent repetition and style mechanisms such as parallelisms) and are present in seven VO tokens. Moreover, these focus constructions, which occur in MiD translations of (vulgar) Latin psalms, are probably influenced by their Latin source texts. As Denison (1993:4) points out, translations of Latin texts should be treated with caution in word order studies.

The four remaining VO constructions cannot be accounted for by any of the parameters discussed so far, although they may have been influenced by the Latin Vorlage too. They may well be due to other focus constructions in this text type that I have so far not yet identified.

A look at the OV constructions in religious texts again reveals important differences between the VO and the OV objects. Most direct objects in OV constructions consist of a simple NP (47). There are a few co-ordinated simple NPs (4) and some NPs with a simple adjectival modifier (2).

In sum, about \( \frac{1}{4} \) of the direct objects in the religious texts show VO order. These objects are complex or appear in genre-specific focus constructions. As the occurrence of VO order seems to be related to genre-specific constructions, possibly influenced by the Latin source texts, religious texts do not provide a reliable basis for the study of MiD word order.

3.4 Narrative texts

Both official and religious texts show genre-specific focus constructions (formulas) related to VO order and therefore we would expect the OV/VO ratio in narrative texts to be different. Table 3 shows the results of the narrative texts.

20% of the direct object constructions show VO order. No influence of the parameters clause type, verbal cluster complexity, genre-specific formulas, and NPs denoting amounts or quantities was found. In three VO constructions the direct object is the complement of a naming verb and in eight VO constructions the direct object is complex (i.e. consists of a NP and a relative clause). The four remaining VO constructions are unaccounted for, but these too may be due to specific focus constructions that I have not yet identified.

<table>
<thead>
<tr>
<th>Text number</th>
<th>I</th>
<th>II</th>
<th>III</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>O-total</td>
<td>13</td>
<td>31</td>
<td>31</td>
<td>75</td>
</tr>
<tr>
<td>OV</td>
<td>10 (77%)</td>
<td>29 (94%)</td>
<td>21 (68%)</td>
<td>60 (80%)</td>
</tr>
<tr>
<td>VO</td>
<td>3 (23%)</td>
<td>2 (6%)</td>
<td>10 (32%)</td>
<td>15 (20%)</td>
</tr>
</tbody>
</table>
As for the OV tokens, fifty direct objects consist of a simple NP, five of two co-
ordinated NPs and a further five of a NP and a simple adjectival modifier.

It is clear that in narrative texts too, the direct objects in OV constructions are
very different from those in VO constructions, mostly in terms of direct object
complexity. Section 4 will focus on these differences.

4. Discussion and conclusions

Let us see what the corpus study has revealed with respect to the three questions
asked at the beginning of Section 2.

1. **Does MiD really contain VO tokens that are unacceptable in MoD?**
The corpus results show that this is indeed the case, but that VO order only appears
under highly specific conditions.

2. **If so, what do these VO tokens look like, in what kind of contexts and with what
frequency do they appear?**
We found evidence of two specific conditions triggering VO order.

   First, complex direct objects containing a relative clause always appear post-
verbally, whereas other possible types of complex direct objects, such as those
consisting of two simple NPs (12 tokens) or of a simple NP and a simple adjectival
modifier (24 tokens) never do so. Our definition of complexity, then, seems to work
for the data.

   The second condition is related to focus. Direct objects of naming verbs such
as *heten* and *noemen* are postverbal in all texts, because they encode new, focal
information. Similarly, direct objects in genre-specific formulas in official and
religious texts that are clearly focal always appear postverbally. In official texts, these
focal direct objects encode the item that is at the heart of the legal agreement. In
religious texts, they appear as the ‘parallel’, and thus the focal, parts of parallelisms
and similar constructions.13

   With respect to the frequency of VO order we find striking differences between
the three genres. Whereas official texts show almost 50% VO order, this percentage
is much lower in religious (25%) and narrative texts (20%). The high VO frequency
in official texts seems to be related to the occurrence of many formulas, character-
ized as genre-specific focus constructions.

3. **What do these VO tokens tell us about the underlying word order of MiD?**
On the basis of the different types of direct objects that appear in OV and VO
constructions, we conclude that MiD is underlingly OV, as is MoD.14 The data
show that OV order is the default order (involving constructions with simple and
non-focal direct objects), whereas VO order is the marked order (reserved for only
two specific cases: direct objects containing a relative clause and direct objects unambiguously bearing focus). Support for this claim comes from frequency numbers: in all text types OV order is more frequent than VO order.

Obviously, then, there has been no change in the deep structure VP syntax of Dutch; both MiD and MoD have a grammar (in a strict sense) that base-generates OV constructions. What has changed is the surface positioning of complex and focal direct objects. This change results in two essential differences between MiD and MoD: the types of direct objects occurring postverbally and, as a consequence, the frequency of VO tokens.

Focal direct objects appearing in specific focus constructions are extraposed in MiD, but are assumed not to do so in MoD. However, data such as in (3) and (11) seem to be acceptable in MoD accounts and religious texts too, but these genres are usually excluded from syntactic research. In choosing our MoD data, then, we seem to be more selective than in choosing our MiD data.

To determine exactly what has changed with respect to focus, one needs to know what kinds of focus marking devices are available in MiD and in MoD and why the use of these devices has changed between the two language stages. This would require a thorough study of spoken MoD, as written MiD was probably much closer to the spoken language than written MoD (a written standard did not develop until the 17th century). For the same reason, it should also be investigated how direct objects containing a relative clause behave in spoken MoD. We leave this matter for future research.

Future research should also include variables such as the localization in time and place. The three text types studied here are from different subperiods of MiD (the official texts being older than the religious and the narrative texts), because the various genres are not evenly represented in the extant documents from the MiD subperiods. This "data-gap" makes it difficult to present a diachronic account of the evolution of style in the various text types. More MiD data should be studied to work out the genre-specific conditions and the more general MiD facts and tendencies. What has become clear from this study, however, is that the basic word order of Dutch has remained stable through time and that MiD VO order is only found under two, highly specific, conditions.

Notes

* I would like to thank Geert Booij, Ansvan Kemenade, Bettelou Los, and Fred Weerman for their valuable comments and discussion. Of course, all errors are mine.

1. In the examples the direct objects are underlined and the verbs are printed in italics.

2. VO tokens that are unacceptable in MoD are VO patterns in subordinate clauses and vVO patterns in main clauses (Jan heeft ge<underline>kust Marie instead of Jan heeft Marie <underline>gekust 'John has Marie kissed').
SVO-main clauses with only one verb cannot be taken as evidence for an underlying OV order (with V2) or an underlying VO order, since both underlying orders can account for this surface pattern. From here on I will use the terms ‘VO tokens’, ‘VO order’, etc. to describe just those (v)V(O) patterns, as these patterns cannot be derived from an OV deep structure as Koster (1975) assumes for MoD.

3. N&W’s (1992) case-marking system does not involve the concept of ‘underlying order’.

4. It is for reasons like this that syntactically enriched corpora generally exclude verse texts (e.g. The Penn-Helsinki Corpus of Parsed Texts).

5. It should be noted that MiD texts are largely Flemish texts. Flanders was a wealthy region in the Middle Ages, which explains why so many texts were produced in that area.

6. Interpunction marks are rarely found in MiD texts, but usually added in translations.

7. Unfortunately, not all three genres are attested in all MiD subperiods; the official texts are older than most religious and all narrative texts (early narrative texts are in rhyme).

8. See also footnote 2.

9. We do not expect to be able to account for every single VO token as this is a corpus of performance data, but it should be possible to formulate a number of broad tendencies.

10. The numbers in the MiD examples refer to the number of the text in the relevant subcorpus (official, religious, or narrative) and to the line number in that text.

11. Split direct objects such as in (8a) also occur in MiD, but not very often: I only found two tokens in the narrative texts. These ‘split’ direct objects are counted as OV tokens.

12. Note that the MoD gloss of this VO clause is still quite acceptable in an elevated style.

13. Support for the claim that focal direct objects are postverbal in MiD and non-focal ones are not is provided by the fact that in another type of official formula, which contains an unambiguously non-focal direct object, the object always appears preverbally (ex. 9).

14. This is also claimed in Van den Berg (1980), but the present study reveals additional evidence for this claim, and, more importantly, answers the questions regarding the frequency and the structure of the VO constructions found in MiD.

References


