Crossed and nested dependencies in German and Dutch: A psycholinguistic study

EMMON BACH,† COLIN BROWN,* and WILLIAM MARSLEN-WILSON*‡
*Max-Planck Institut für Psycholinguistik, Nijmegen, The Netherlands, †University of Massachusetts, Amherst, USA, and ‡MRC Applied Psychology Unit, Cambridge, UK

Received March 1986; accepted in revised form September 1986

Abstract—The clause-final verbal clusters in Standard Dutch and German differ strikingly in the kinds of dependencies they normally permit between verbs and their arguments, with Dutch preferring crossed dependencies and German nested. This study investigates the consequences of these differences for the psycholinguistic processing complexity of sentences containing either crossed or nested dependencies. German and Dutch subjects performed two tasks — ratings of comprehensibility and a test of successful comprehension — on matched sets of sentences which varied in complexity from a simple sentence to one containing three levels of embedding. The results show no difference between Dutch and German for sentences within the normal range (up to one level of embedding), but with a significant preference emerging for the Dutch crossed order for the more complex strings. We argue that this rules out the push-down stack as the universal basis for the human parsing mechanism.

INTRODUCTION

The variations in word order found in the West Germanic languages have been the focus of a great deal of attention in the linguistic literature, from the point of view of both synchronic and diachronic theory. Especially interesting have been the contrasts in the position of verbal elements (Den Besten and Edmondson, 1983), as illustrated in the following three sentences from Dutch, German, and English:

1. De mannen hebben Hans de paarden leren voeren.
The men have Hans the horses teach feed.

2. Die Männer haben Hans die Pferde füttern lehren.
The men have Hans the horses feed teach.

3. The men taught Hans to feed the horses.

For someone with even a limited competence in English and either of the other two languages, the patterns in Dutch and German seem in some sense to be more difficult to process and produce than their English counterparts. This point becomes clearer still when we add an additional clause to each of the three strings:


5. Johanna hat die Männer Hans die Pferde füttern lehren helfen.

6. Joanna helped the men teach Hans to feed the horses.

Sentences like (4) and (5) are rejected outright by native speakers of Dutch and German as unacceptable. In contrast, a sentence like (6) is only marginally more complex than (3) to a speaker of English, and is readily comprehensible and acceptable.

Address correspondence to William Marslen-Wilson, MPI für Psycholinguistik, Wundtlaan 1, 6525 XD Nijmegen, The Netherlands.
It is reasonable to assume that the strength of these contrasts reflects the different degrees of memory and processing loads imposed by the different kinds of structure. The English structures allow each verb to collect up its arguments word-by-word as the sentence is heard, with a straightforward right-branching dependency between successive clauses. But in the Dutch and German constructions illustrated in (4) and (5), as many as four NP's must be stored in memory awaiting assignment to their respective verbs in the appropriate structural configurations. In Dutch and German, furthermore, there is the contrast of the radical differences in the ordering of the sentence-final verb clusters, and the consequences of this for assignment of structure to the incoming string.

In German constructions of this sort, such as (2) and (5), the dependencies between the clause-final verbs and their arguments are fully nested — as the illustration in Fig. 1 shows. The Dutch constructions, in contrast, lead to crossed serial dependencies between verbs and their arguments — as Fig. 1 also illustrates. It is on the linguistic and the psycholinguistic consequences of this distinction that we will focus here.

Evers (1975) was the first generative grammarian to study these structures in detail, and to point out their importance for general theories of syntax. In doing this, he also raised the question of the comparative difficulty of the structures for speakers of the two languages, and claimed that the Dutch pattern was in some sense easier for the human language processor to deal with. A similar claim has also been made for Dutch vis-à-vis Frisian by Hoeksema (1981) — with Frisian following German in its nesting of dependencies. Evers also drew the rather strong inference that the fact that Dutch is easier (if true) shows that the human language processor does not use a push-down stack, since such a mechanism cannot deal with crossed dependencies. The verb 'lernen' for example, in Fig. 1, could not have access to its arguments ('De mannen', 'Hans') because they would be blocked by a subsequent entry to the stack ('de paarden').

However, it is also possible to imagine theories — for example, those that do incorporate some form of stack mechanism — which would predict that Dutch should be more difficult, or at least less preferred than a language like German, with its orderly nesting of dependencies. And certainly, because of the productivity of discontinuous and crossing dependencies in Dutch, several writers have spoken of the Dutch pattern as being more 'marked' (e.g., Ades and Steedman, 1982; Bresnan et al., 1982).

At present, however, all of these discussions of the relative ease or difficulty of Dutch and German constructions are based solely on anecdotal evidence. There has been no systematic investigation of the extent to which speakers of Dutch and German do or do not experience different degrees of 'difficulty' in processing utterances like (1) and (2). Our goal here is to rectify this, and to provide a solid empirical basis for subsequent psycholinguistic and linguistic research into verb-clusters in West Germanic.

To do this we have adopted a dual strategy, evaluating the psycholinguistic complexity of these structures both for sentences within the normal range, such as (1) and (2), and for sentences which are clearly not normally acceptable, such as (4) and (5), and (7) and (8) below, which introduce a third level of embedding:

(7) Karel heeft Jeanine de mannen Hans de paarden zien helpen leren voeren.
(8) Karl hat Johanna die Männer Hans die Pferde füttern lehren helfen sehen.
(9) Charles saw Joanna help the men teach Hans to feed the horses.

Given that sentences like (1) and (2) do exist and are used in the language, it is possible that they will not, in fact, differ in effective complexity for their native speakers. But by observing the behaviour of the system when we push it beyond its normal limits, as in the double embeddings and beyond, we can hope to obtain evidence from the way it breaks down for whether there is, in fact, a preferred order for the human parsing mechanism of arranging grammatical elements in dependent sequences.

METHOD

Materials
Test sentences. The primary test materials consisted of 18 matched sets of three sentences, in Dutch and German, that varied in complexity from single clause sentences to sentences with two levels of embedding (our Level 3). A sample set is illustrated below for both languages (with English gloss):

Level 1: De lerares heeft de knikkers opgeruimd.
Die Lehrerin hat die Murmeln aufgeräumt.
The teacher collected up the marbles.

Level 2: Jantje heeft de lerares de knikkers helpen opruimen.
Wolfgang hat der Lehrerin die Murmeln aufräumen helfen.
Wolfgang helped the teacher collect up the marbles.

Level 3: Aad heeft Jantje de lerares de knikkers laten helpen opruimen.
Arnim hat Wolfgang der Lehrerin die Murmeln aufräumen helfen lassen.
Arnim let Wolfgang help the teacher collect up the marbles.

It was not possible to extent these triplets to a third level of embedding, because of problems in finding semantically plausible extensions for many of them. We therefore constructed a separate set of nine sentences at a third level of embedding. An example of one of these is given below.

Level 4: Ingrid heeft Lotte de bewoners de blinde het eten horen leren helpen koken.
Ingrid hat Lotte die Bewohner dem Blinden das Essen kochen helfen lernen hören.
Ingrid heard Lotte teach the residents to help the blind man to cook the food.

In constructing these materials, it became apparent that there was considerable disagreement among our German informants as to the correct form of the final verb in the Level 2, 3, and 4 sentences. The informants — and the standard German texts (e.g., Dudenhoff, 1973) — did not agree on whether the final verb should be in infinitive form (e.g., ‘lehren’) or in past participle form (e.g., ‘gelehrt’). Since we had no way of resolving this disagreement, we were forced to construct two different versions of the German materials. In one version, we used only the infinitive form (as in the example set given above). The second version, which was otherwise identical, used only the past participle, except when the final verb was ‘lassen’. This can only occur in the infinitive in these constructions. Thus, for the example set given above, the final verbs would take the form ‘geholfen’, ‘lassen’, and ‘gehört’, for Levels Two to Four, respectively.

Paraphrase sentences. Since the increase in complexity from Level 1 to Level 4 is not simply syntactic in nature, but also involves semantic and propositional complexity, it was necessary to control for this by constructing an additional set of Paraphrase sentences, which express roughly the same propositions as the Test sentences for Levels 2, 3, and 4, but which do so using right-branching constructions analogous to the English glosses given earlier. An example set, matching the Test sentences given earlier, follows below (note that these paraphrases are similar in structure to the English glosses that were used for the Test sentences):

Level 2: Jantje heeft de lerares geholpen om de knikkers op te ruimen.
Wolfgang hat der Lehrerin geholfen, die Murmeln aufzuräumen.

Level 3: Aad heeft Jantje er toe gebracht om de lerares te helpen met het opruimen van de knikkers.
Arnim hat Wolfgang dazu gebracht, der Lehrerin beim aufraumen der Murmeln zu helfen.

Level 4: Ingrid heeft gehoord hoe Lotte de bewoners leerde om de blinde te helpen het eten te koken.
Ingrid hat gehört, wie Lotte die Bewohner lehrte dem Blinden zu helfen das Essen zu kochen.

To check the semantic interpretability of the events described in these sentences — and therefore, by extension, in the Test sentences as well — a pool of judges rated the plausibility of the paraphrase sentences, as well as their ease of comprehension. Following their comments, two sets of sentences were removed from the original set of 20, and some modifications were made to the remaining 18 triplets, as well as to the nine Level Four sentences.

Fillers and practise. To obscure the regularities in the grammatical structure of the Test and Paraphrase materials, a set of 36 Filler sentences was also constructed. Twelve fillers were matched to each of the three main Levels, with each filler containing the same number of verbs and nouns as its corresponding test-level. Here are three examples from the German set:

Level 1: Der Ball wurde durch das Fenster geworfen.
It was not feasible to expand the design of the experiment to allow each suitable NP in each Test or Paraphrase sentence to be questioned. Instead, items had to be nested within question-types at each Level. The 18 Level 2 Test and Paraphrase sentences were split into two groups of nine, one group for each NP that could be questioned. Each group was then distributed over the three versions in blocks of three, so that each version contained six Level 2 Test questions (three on NP1 and three on NP2), and six Level 2 paraphrase questions (three on NP1 and three on NP2).

The eighteen Level 3 Test and Paraphrase sentences were split into three groups of six, one group for each of the three NPs. Each group of six sentences was divided over the three versions, so that each version contained six Level 3 Test sentence questions (two on NP1, two on NP2, and two on NP3), and six Level 3 Paraphrase sentence questions, also distributed over the three NPs.

Each experimental version therefore had a total of 24 questions on Test and Paraphrase sentences. Twenty-four of the Fillers were also questioned, with ten questions on the Level 1 Fillers, nine on Level 2, and five on Level 3. In addition there were questions for ten of the 15 practice sentences. These questions on the filler and practice sentences did not follow the fixed format of the Test and Paraphrase questions, but varied according to the noun phrase being addressed.

Procedure
The materials were presented to the subjects over closed-ear headphones. Immediately after hearing each sentence, the subjects were required to perform a rating task, in which they made the judgement: “How easy is this sentence to understand” (in German “Wie leicht ist dieser Satz zu verstehen”; in Dutch “Hoe makkelijk is deze zin te begrijpen?”). They recorded their judgements on a 9-point scale, where point 1 was labeled ‘leicht’ or ‘makkelijk’ (meaning ‘easy’) and point 9 was labeled ‘schwer’ or ‘moeilijk’ (‘difficult’).

In addition to the judgement task, the subjects were also required to perform a second task on two-thirds of the sentences. This was the question task as described above. For those sentences where a question was asked, the subjects were allowed 4 s to make the rating judgement. The relevant question was then read aloud by the experimenter, and the subjects were instructed to give a brief but complete answer. As soon as the subjects had finished writing down their responses, they pressed a response-button to indicate that they were ready for the next sentence.

Each test-session lasted approximately 45 min, including instructions, practice, and a short break after practice to check for problems and for proper use of the rating scale.

Subjects
Three separate groups of 30 paid subjects were run. The Dutch group was recruited from the student community in the Nijmegen area, and tested by a native speaker of Dutch. The two German groups (one for the Infinitive version and one for the Participle version) were recruited at the University of Köln, and tested by a native speaker of German.

RESULTS
Rating task
The mean comprehensibility ratings for the Test and Paraphrase conditions are given in Table 1, collapsed over items and subjects, for the two German groups and for the Dutch group.

<table>
<thead>
<tr>
<th>Table 1</th>
<th>Mean ratings of comprehensibility: Test and paraphrase</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sentence type</td>
<td>Dutch</td>
</tr>
<tr>
<td>Level 1 Test</td>
<td>1.14</td>
</tr>
<tr>
<td>Level 2 Test</td>
<td>2.34</td>
</tr>
<tr>
<td>Level 2 Para</td>
<td>2.11</td>
</tr>
<tr>
<td>Level 3 Test</td>
<td>5.42</td>
</tr>
<tr>
<td>Level 3 Para</td>
<td>4.06</td>
</tr>
<tr>
<td>Level 4 Test</td>
<td>7.66</td>
</tr>
<tr>
<td>Level 4 Para</td>
<td>5.94</td>
</tr>
</tbody>
</table>

conditions are given in Table 1, collapsed over items and subjects, for the two German groups and for the Dutch group.

The overall effects are very clear. First, the rated comprehensibility declines steeply over Test Levels, from a mean of 1.16 for Level 1 to a mean of 7.86 for Level 4. The largest jump is from Level 2 (mean of 2.58) to Level 3 (5.80). These effects are not caused simply by the increase in syntactic complexity across Levels, but also by the growing propositional complexity of the sentences. This is shown by the large increase for the Paraphrase sentences as well — with overall means of 2.16, 4.01, and 5.82, for Levels 2, 3, and 4, respectively. This pattern holds without significant variation for each of the three language groups.

The role of the syntactic variables can be seen more clearly if we subtract the ratings for the Paraphrase sentences from the ratings for the Test sentences, thereby cancelling out the effects of propositional complexity. Here again, the overall difference between Test and Paraphrase increases over levels, from 0.42 at Level 2 to 1.78 at Level 3 and 2.04 at Level 4. The largest jump is still from Levels 2 and 3. As with multiple center-embeddings in English, one level of embedding (our Level 2) seems to be quite acceptable, but two levels of embedding create problems for the listener.

The acceptability of the Level 2 Test sentences for all three language groups is
illustrated not only by the similar ratings for Level 2 Paraphrase, but also by the ratings for the Level 2 Fillers. The results for the Fillers are given in Table 2, and the means for Level 2 — both overall (mean = 2.21) and for the individual groups — confirm that there is very little difference in rated comprehensibility between any of the materials at this level.

Although all sentences types and all three language groups behave very similarly at Level 2, differences start to emerge in the transition from Level 2 to Level 3 — not only between Test and Paraphrase, but also between the Dutch and the German groups. Figure 2 plots the Test-Paraphrase differences for the three language groups for Levels 2 to 4. Although all the groups show a large increase across levels, the rate of increase is smaller for the Dutch group than for any of the other two groups. The German group shows a relatively less preferred than the Dutch group throughout (F(2, 17) = 7.99, p = 0.012). More significantly, the German/Participle group, which does not differ from the Dutch group at Level 2, does differ at Level 3 (F(2, 17) = 3.705, p = 0.07).

Comparing the two German groups, there is a tendency for the Inactive materials to be rated as slightly less comprehensible at all three levels. This tendency reaches significance only at Level 4 (F(1, 8) = 8.270, p = 0.02).

Question task
The question task here has two functions. First, to test for successful comprehension of the Test and Paraphrase materials, and, second, to provide a check on the rating results. Do the perceived difficulties in comprehensibility reflect real difficulties in comprehension, and is, in particular, the apparent advantage of Dutch maintained in this different task?

To evaluate these issues, we scored the subjects' answers according to a strict definition of what counted as correct comprehension. Given a question like “Wastat Hans?” (“What did Hans do”), we scored as correct an answer which made it clear that the subject NP in question (“Hans”) had been successfully attached to its predicate verb phrase. Given that what Hans was doing was to feed the horses (“Die Pferde füttern”), we gave a score of two points to each answer which mentioned both ‘Pferde’ and ‘füttern’, and one point if only the correct verb or only the correct object was given. The subjects were not penalized for paraphrases — such as “fed the pony” — which retained the sense of the original.

The overall results for the three language groups are given in Table 3, collapsing across items and subjects. The data were also entered into a three-way anova on Level (2 and 3), Sentence Type (Test and Paraphrase), and Language (Dutch, German/Infinitive, German/Participle), with subjects as the random factor, and yielding significant main effects for all three main variables.

The overall pattern closely parallels what we saw for the rating scores. There is a big drop in performance going from Level 2 to 3, with the drop for the Test sentences (from 1.66 overall to 0.95) being somewhat larger than the drop for the Paraphrase sentences (from 1.81 to 1.36). This is reflected in a strong interaction between Level and Sentence Type, with F(1, 87) = 21.65 (p < 0.000). At Level 2 there are only marginal differences between Test and Paraphrase (0.15 overall), but this increases to a large difference (0.41) at Level 3.

Looking at the three language groups, they do not differ significantly at Level 2. But at Level 3 we again find a tendency for Dutch performance to be less severely affected by the additional level of embedding. This is reflected in an interaction between Language and Level (F(2, 87) = 3.84, p = 0.025). In going from Levels 2 to 3, the overall Dutch scores fall by an average of 0.46, whereas the Infinitive and Participle groups fall by 0.65 and 0.62 respectively. Looking at the Test sentences alone, the Dutch scores fall by 0.51, while the scores for the two German groups fall by 0.84 and 0.79. This difference in the size of the effects for Test and Paraphrase is reflected in a further interaction between Language and Sentence Type (F(2, 87) = 3.309, p = 0.045).

Paralleling the results for the rating scores, the differential pattern for the three language groups emerges most clearly when we look at the Test/Paraphrase differences. These are plotted for each Level for the three groups in Fig. 3. When the effects of propositional complexity are controlled for in this way, we see no
Figure 3. Mean differences in error-rate between Test and Paraphrase sentences in the question task, plotted for Levels 2 and 3. Means are collapsed across items and subjects within the three language groups (Dutch, German/Infinitive, German/Participles).

Table 4. Comprehension scores for Level 3: by NP

<table>
<thead>
<tr>
<th>Language</th>
<th>NP1</th>
<th>NP2</th>
<th>NP3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dutch:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Test</td>
<td>1.17</td>
<td>0.83</td>
<td>1.50</td>
</tr>
<tr>
<td>Para</td>
<td>1.28</td>
<td>1.42</td>
<td>1.50</td>
</tr>
<tr>
<td>German/Inf:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Test</td>
<td>0.88</td>
<td>0.67</td>
<td>1.12</td>
</tr>
<tr>
<td>Para</td>
<td>1.20</td>
<td>1.58</td>
<td>1.53</td>
</tr>
<tr>
<td>German/Part.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Test</td>
<td>1.02</td>
<td>0.38</td>
<td>0.97</td>
</tr>
<tr>
<td>Para</td>
<td>1.33</td>
<td>1.07</td>
<td>1.33</td>
</tr>
</tbody>
</table>

Table 5. Level 3: Differences in test/paraphrase, error-rate

<table>
<thead>
<tr>
<th>Language</th>
<th>NP1</th>
<th>NP2</th>
<th>NP3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dutch:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.11</td>
<td>0.59</td>
<td>0.60</td>
</tr>
<tr>
<td>German/Inf:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.32</td>
<td>0.91</td>
<td>0.41</td>
</tr>
<tr>
<td>German/Part.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.31</td>
<td>0.67</td>
<td>0.36</td>
</tr>
</tbody>
</table>

differences between the groups at Level 2, but with the relative error rate increasing much more steeply at Level Three for the two German groups. Looking specifically at the less disrupted of the two German groups, a two-way ANOVA comparing Dutch and German/Participle scores across Levels 2 and 3 yields a significant interaction between Language and Level ($F(1,58) = 4.036, p = 0.049$).

Tables 4 and 5 give a breakdown of the results for Level Three, according to the NP being questioned. For the Paraphrase scores (Table 4) there is relatively little difference between the three NPs, despite the serial order in which the three propositions involved were presented to the subjects. In the Test sentences, however, there are clear differences between NPs, with questions about NP2 creating the most serious problems. As the Test/Paraphrase difference scores in Table 5 make clear, this pattern holds for all three groups. Dutch nonetheless shows a small advantage both here and for NP1, with the major advantage emerging for the most deeply embedded clause, associated with NP3. The implications of this pattern for the order of parsing operations in the two languages will be discussed in more detail below.

**DISCUSSION**

We asked in the Introduction whether there were any reliable differences in the processing complexity of Dutch and German sentences containing clause-final verb clusters. The results, for Levels 3 and 4, support Evers' (1975) intuition that the crossed Dutch structures are in some sense less difficult to process than the parallel German nested constructions. This conclusion is supported not only by the subjects' comprehensibility judgements, but also by the greater error rate at Level 3 in the question answering task. What, then, are the implications of this result?

We will assume, for present purposes, that the interpretation (or semantic representation) for the Dutch and German sentences is the same. For example, using English words for our metalanguage (and with P standing for the tense operator expressed by the present perfect in the two languages), (10) and (11) are mapped onto an interpretation like (12):

(10) Henk heeft de kinderen Anneke de koeien laten zien melken.
(11) Hans hat die Kinder Anna die Kühe melken sehen lassen.
(12) P(let H, (see the children, (milk (A, the cows))))

In addition, we will also follow the widespread consensus among linguists working within transformational or related phrase-structural frameworks that the surface structures of the sentences are those originally suggested by Evers (for a differing view see, for example, Steedman, 1985). These would consist of a right-branching structure of sentence or verb-phrase 'remnants' for both languages, and a verb-cluster constituent for the two languages differing only in the direction of branching — left for German and right for Dutch (cf., Bresnan et al., 1982). Given these assumptions, the obvious processing problem posed by such sentences is that the hearer does not know how to integrate the series of NP's into an interpretation until she hears the verb-cluster, when she can start to hook up the NP's as arguments to their governing verbs.

The first point is that there is no difference in processing complexity between Dutch and German at Level 2 — either when compared with each other, or with their accompanying paraphrases. This seems to rule out the possibility that the
human interpretative process is universally dependent on something like a push- 
down store mechanism, since such a device will be unable to process sentences 
containing crossed dependencies, even at Level 2. The advantage that emerges for 
Dutch for the Level Three structures further confirms Evers' claim that Dutch 
constitutes evidence against simple push-down store accounts of psychological 
parsing.

The ease in processing of the Level 2 Dutch sentences can also be compared with 
what is known about crossing dependencies in long-distance filler-gap constructions 
in languages like English. The facility with which listeners handle Level 2 sentences 
in Dutch contrasts markedly with the reactions of English listeners to crossing 
dependencies in examples like (13), in contrast to the nested dependencies in (14).

(13) Which sonatas are these violins easy to play on D1?

(14) Which violins are these sonatas easy to play on D1?

The difference between genuine long-distance crossing dependencies in the English 
and the Dutch examples can be taken as evidence against any theory of linguistic 
structures and interpretative mechanisms which treat all dependencies by a single 
sort of process — for example, a one-pass left-to-right processor that makes use of a 
single stack.

The results for Dutch and German are also inconsistent with a processing model 
which demands that a full syntactic structure for a sentence be computed before 
interpretation is carried out. Such a model should predict no differences between 
the two languages, since in both the hearer must wait to the end of the entire 
verbcluster before being able to match up arguments and verbs.

This is in any case something of a straw man. It seems much more likely that 
listeners compute partial interpretations on-line as they hear an utterance 
(Marslen-Wilson and Tyler, 1980). From this perspective, it is clear that the crossed 
and nested orderings differ markedly in the kinds of partial interpretations they 
allow the listener to construct. Referring to the semantic structure of (13), we can 
see that, in the Dutch case, the matrix of higher verbs can begin to be built up 
immediately the verbcluster begins, and without the risk of creating structures that 
may have to be re-analysed in the light of later information. The German structures 
have the advantage that the innermost proposition can be integrated in a context-
free way, but it is not possible to figure out what to do with this structure until the 
higher verbs come in one by one.

It is perhaps in these terms that we can explain the distribution of error scores for 
the Level 3 comprehension questions, broken down according to NP in Tables 4 
and 5. The ordering NP1 to NP3 goes from highest to most deeply embedded 
clause. Dutch shows an advantage throughout, but the effect is twice as large for 
NP3, the most deeply embedded, as it is for the other two. Even though — 
continuing with the examples in (10) and (11) — the clause “Anna milk the cows” is 
the first one that the German listener can construct, and the last one that the Dutch 
listener can connect up, the Dutch listeners show a 50% higher success rate in 
answering the question “What did Anna do”.

This suggests that the most important variable in successful parsing and interpretation is not simply when information becomes available, but also what you can do 
with that information when you get it. Knowing that Anna milked the cows is of 
relatively little use to the German listener, since he has no higher structure into 
which to integrate this information at the time that he receives it. In the Level 2 
cases, this is apparently not a problem, since the higher structure can be 
completely integrated when the next verb is heard. But for Level 3, where the 
next verb also leads to a structure which must be held in intermediate store, there is 
poor recall not only of the NP3 complex, but also very poor performance indeed on 
NP2. The German Participle group, for example, scored less than 20% correct in 
questions about NP2.

The Dutch listeners also suffer some degree of overload at Level 3, as 
demonstrated by the large drop in their performance here as well. But the fact that 
they do consistently better than the German listeners — and that they show no 
decrement at all, relative to the Paraphrase control, for NP3 — testifies to the 
avantages of being able to build the higher structure first. Even though they are 
also losing material, what they do retain is clearly being integrated into a more 
coherent and long-lasting structure than seems to be possible with the German 
nested order.

We conclude with a number of questions that need to be taken up in further 
research of this sort. First — and directly relevant to the points we have just been 
discussing — is a difference between the two languages which was deliberately 
factored out in our test materials: German has a much richer system of case 
marker than Dutch, where case is marked only for the small set of personal 
pronouns (similar to the English situation). What role does this difference play in 
determining the way speakers process sentences in the two languages?

Secondly, why is there such a sharp difference between Levels 2 and 3? One 
possibility we can consider here is that the Level 3 examples are, strictly speaking, 
grammatical for some speakers. This would be the case, for example, if the 
formation of complex verbs was a lexical and non-recursive process. If this was a 
process that applied to Level 2 complexes but not to Levels 3 or 4, then it may be 
the case that these sequences force listeners to fall back on some form of non-
linguistic strategy for constructing an interpretation of these stimuli. But if so, we 
will then need to explain why Dutch speakers are more successful at this than 
German ones.

It is also possible that the process of complex-verb formation is lexical in German 
and syntactic in the narrow sense in Dutch. Note that the stress pattern of the 
German clusters conforms to the general pattern for compounding in the language 
— i.e., left branching structures with primary stress on the leftmost member of the 
compound. The Dutch clusters, in contrast, are right-stressed and right-branching. 
This goes against the general pattern of word-compounding in Dutch, which is 
basically the same as in German.

Finally, the most positive moral we can draw from this research is that it is 
possible to make meaningful comparisons across languages. Our results suggest 
that it will be worthwhile to pursue this sort of comparison in experiments that are 
designed to tease out finer-grained aspects of the on-line processes involved. But 
any further research will need to proceed by investigating much more specific 
hyotheses both about the linguistic structures involved and about the organization 
of the processor.

Acknowledgements
We thank Herbert Schriefers and Veronica Ehrlich for help with our German test
materials, Stef van Halen for help with the Dutch materials and for scoring the question-answering responses, Regina Stralke for running the German subjects, Josef Bayer and Lyn Frazier for comments on the manuscript, and Pienie Zwitserlood for advice and help throughout.

NOTES

1. In fact, according to Duden (1973, para. 276, p. 122), the usage of participle and infinitive forms is tied to specific verbs, so that verbs like *lassen* and *sehen* usually appear in the infinitive in such constructions, while *helfen* and *hören* appear with either, and *lehren* and *lernen* primarily appear in past participle form. Our informants could only agree on the infinitive preference for *lassen*, and this is what is reflected in the materials.

2. Because of the way questions were distributed over items, it was not possible to use items as the random factor in these analyses in the same way as in the analyses of the comprehensibility ratings.

REFERENCES


