A scaling approach to the study of syntactic relations

WILLEM J. M. LEVELT
Groningen University

Since my first acquaintance with linguistics I have been intrigued by phrase markers. Irrespective of whether phrase markers represent surface or deep syntactic relations, they are highly abstract representations of cognitive facts.

The study of cognitive structures, of course, belongs to the domain of psychology. It is therefore not surprising to find that psychologists also have devised ways of describing cognitive relations by abstract representations. Historically, however, the trend in psychology has been towards *spatial* representations of data. Mainly during the last decade various methods of scaling have been developed that essentially map people's judgments of relations between objects onto a spatial configuration of points. The points, then, represent the objects, and the metric distances between the points represent the strengths of the cognitive relation between the objects, as judged by the subjects. Generally speaking: the stronger the intuited relation or similarity between two objects, the closer their points are in the spatial configuration. There are numerous examples to be found in the literature. We have seen spatial representations of objects as diverse as colors, meanings, morse codes, musical intervals, phonemes, and many others. Apart from the fact that such representations are useful summaries of data that are otherwise difficult to survey, they are often more than that. In many cases spatial characteristics of the obtained configurations allow for rather direct psychological interpretations. The practice of labeling spatial dimensions or facets is a first step in building psychological theories. Moreover, if theories are already available, they can sometimes be directly checked against the experimentally obtained configurations. For instance, distinctive features of vowels may turn up as dimensions in a space obtained from judgments of vowel similarities.

Turning back to phrase markers, I can now explain why they appeared so fascinating to me. This was mainly because of two reasons.

Hierarchical representation of word relatedness data

Firstly, contrary to my experience with psychological scaling, phrase markers are *non-spatial* representations of cognitive relations. They may be metric representations, but there is no *dimensional structure* to them. Several questions suddenly crop up. Apart from syntactic structure, are there any other domains of cognitive psychology where hierarchical representations are preferable to spatial ones? For instance, what about word meanings? Are meaning relations not often determined by semantic hierarchies like "physical object" – "living" – "animate" – "human" – "male", etc.? This question is irrelevant, however, for my present topic. More important is a second question, namely whether, as in the case of spatial configurations, algorithms can be devised that will map data from relatedness judgments onto *trees*. Let me discuss this in slightly more detail.

It is quite conceivable to have subjects judge the strengths of the syntactic relations holding between words of a sentence. Take the sentence *the too expensive food was tasteless*. If you ask subjects which pair of words is more related in this sentence, *the* and *too* or *the* and *food*, nobody will hesitate to choose *the* and *food* as the more related pair. In experiments of this sort (see LEVELT, 1967, 1969a, b) it is striking how strong such relatedness intuitions are. Subjects are very consistent in their judgments, and there is little disagreement between subjects. It became more and more evident that judgments of syntactic word relations can be of great importance for the study of syntactic structure. In fact such data are not fundamentally different from judgments of grammaticality, that are intensively used by linguists as evidence for their linguistic descriptions.

Imagine now that a subject has judged the strengths of all word pair relations from a sentence. Of course, any method can be used for such judgments: triadic comparisons, 7-point rating scales, magnitude estimation, etc. In all cases, however, we end up with a relatedness matrix. This is a symmetric matrix with as many rows and columns as there are words in the sentence. Each number in the matrix expresses the intuited degree of relation between the corresponding row and column words. The question, then, is: can such a matrix be mapped on a tree?

Since 1965 various algorithms have been devised to do just this. I will not go into this issue deeply. A general review of such algorithms can be found in a paper by LANCE and WILLIAMS (1966/1967). Suffice it to say that in most of my experiments and also in the data to be presented here, I used JOHNSON's hierarchical clustering technique (1967), and preferably his connectedness method, which is especially fit for syntactic data. Details about this technique are given in the introduction to this chapter.

Grammars are theories of cognitive structure

My second source of fascination is shared by many psychologists who became interested in linguistic matters. In the study of language behavior, the psychologist finds himself in the exceptional situation of being confronted with detailed theories of cognitive structure. These theories, moreover, are based on extensive empirical evidence, which is not psychological evidence in the usual limited sense of "psychological".

It is not easy to find comparable situations in other domains of psychology. A possible instance is criminology. The study of criminal behavior necessarily implies the study of people's conception of law. What aspects of juridical legislation or custom are universal properties of human cognition? Are they generative in any sense? In this respect jurisprudence may relate to criminology as linguistics to psycholinguistics.

The natural reaction of the psychologist to a situation like this is to test the so-called "psychological reality" of the cognitive theories he is confronted with. Strictly speaking, this is a meaningless reaction. Any theory of cognition is "psychologically real" to the degree in which it can account for the data at hand. The data underlying linguistic theories are of various kinds, such as intuitions about the grammaticality of sentences, the meaning relations between sentences, segmentations into words, etc. In a slightly different sense, however, the psychological reality question is an important issue. If we take it as a question about the relevance of such theories for the explanation of other kinds of psychological data, a meaningful area of research emerges. For instance, it is a non-trivial empirical question what should be the place of a linguistic theory in, say, a theory of speech perception. In this case the data are very different from the facts a grammar is designed for. They are data like reaction times, errors of perception, etc. Still linguistic theory may have – and in fact does have! – relevance for the explanation of such data.

Word relatedness data and linguistic theory

Now, turning back to our word relatedness judgments, a similar question can be asked. We have instructed our subjects to perform a certain task: they are provided with a sentence and are required to judge the degrees of relatedness existing between the words of that sentence. The empirical issue here pertains to the role of the linguistic structure of that sentence in the judgmental behavior of our subjects. Prima facie, this role should be quite direct and obvious, much more so than for instance in a theory of speech perception. The data namely, that we want to explain, are not much different from the empirical evidence on which linguistic theories are based, being intuitions about within-sentence relations. As far as I can see they differ in only two respects from the usual linguistic data: first, they relate to a quantitative, not a qualitative state of affairs. They are judgments of *degrees* of relatedness, not of *kinds* of relations: they are not statements about the character of word relations, i.e. "this is a subject-object relation", but only about the strength of such relations, i.e. "these words are strongly related in this sentence". As a consequence of this, they are – at least – ordinal judgments, yielding a rank-order of relationships. This is the second linguistically unusual aspect of these data. But otherwise we are straightforwardly testing linguistic intuitions of native speakers, and should therefore expect a close connection with linguistic theory.

Word relatedness judgments reflect deep syntactic structure

For the present discussion I want to elucidate one hypothesis about this connection between relatedness judgments and linguistic structure. The hypothesis is that *word relatedness judgments are mainly determined by underlying structural relations of the sentence.* In a moment I will specify this hypothesis slightly more, but let us first look at some initial results. They concern a kernel sentence, i.e. a sentence where the difference between deep and surface structure is minimal. The sentence was *the boy has lost a dollar.* Twenty four native speakers of English judged the degrees of relatedness for all pairs of words from this sentence by a triadic comparisons procedure (details of procedure and data in LEVELT, 1969b). The resulting relatedness matrix was analyzed by Johnson's hierarchical clustering program, and the result is given in Fig. 1.

I took this as a gratifying result. The obtained hierarchy is in good correspondance with the usual type of phrase structure linguists assign to such a sentence. Or, to express this finding in a more precise way: for our subjects *the degree of relatedness is a function of the place in the P-marker of the smallest* *phrase including both words: the higher the dominating node, the lower the relatedness.* For the present example this can be taken as a *rule of performance.* At other places (see references) I have given several other examples confirming this same general statement.

Presently, however, I want to show that it is deep phrase structure that matters, rather than surface structure. This can be done by analyzing cases, were there are deletions in the surface structure, relative to the deep structure. Consider the following sentence: *Carla takes the book and goes to school.* This sentence is the coordination of *Carla takes the book* and *Carla goes to school*, respectively. One *Carla*, however, has been deleted in the surface structure. The remaining *Carla*, therefore, can be considered as the "trace"

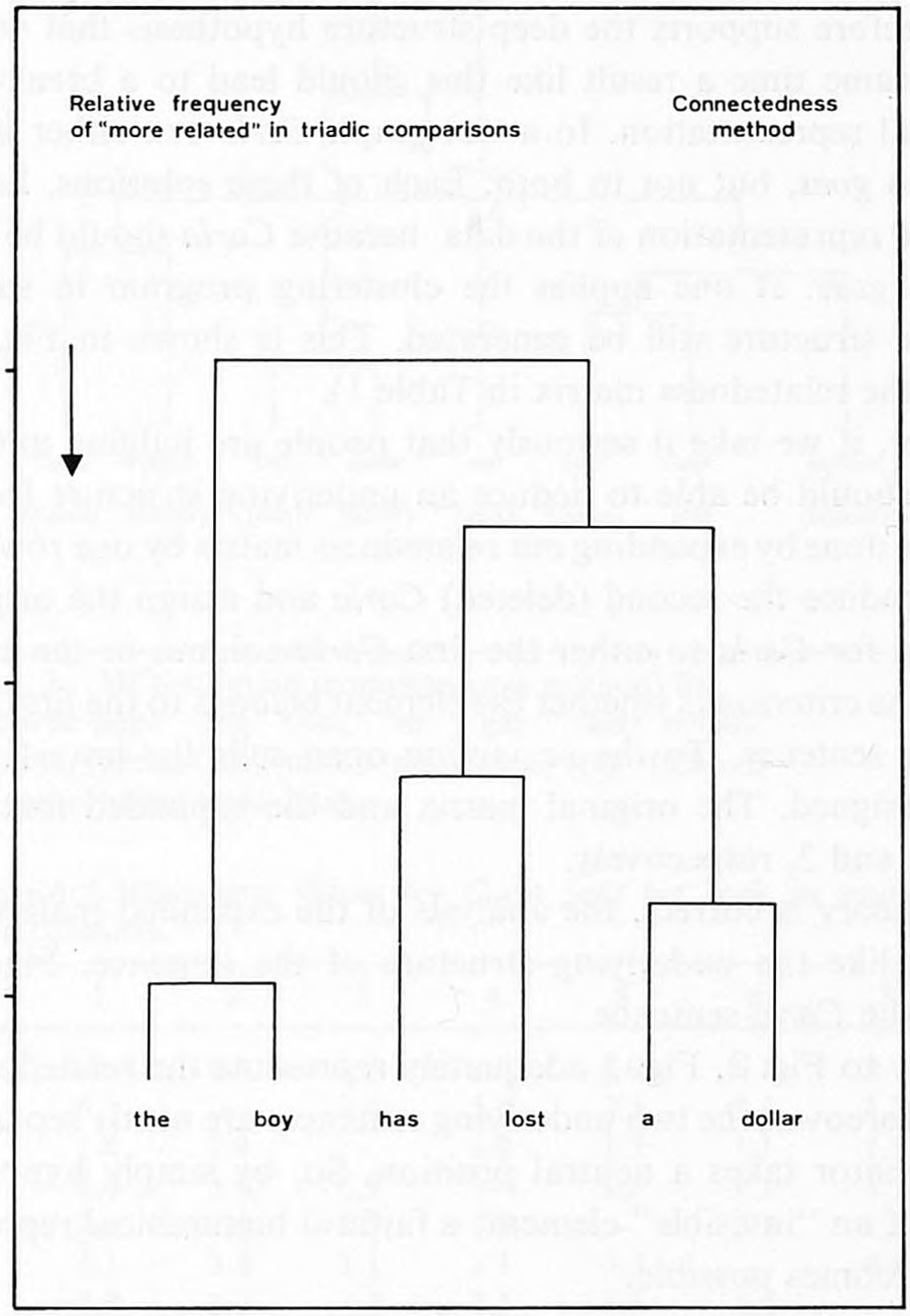

Fig. 1. HCS-solution (connectedness method) for "The boy has lost a dollar" (triadic comparisons data).

of two underlying *Carla's*. If a subject is requested to perform our judgment task, and he is only expressing surface relations in his relatedness judgments, we would expect him to estimate the *Carla – takes* relation as quite strong and the *Carla – goes* relation as relatively weak. On the other hand, if he is taking deep relations into account, he will discern the double role of *Carla* and judge these two pairs as about equally related.

We performed this experiment in Dutch. The word order is the same as in English (*Carla pakt het boek en gaat naar school*). Sixteen subjects judged all word pairs from this sentence on 7-point relatedness scales. The average scale value for the pair *Carla – takes* was 6.3, for *Carla* and *goes* it was 6.1. The tiny difference between these numbers is, of course, not significant. This result therefore supports the deep structure hypothesis that we are testing.

At the same time a result like this should lead to a break-down of the hierarchical representation. In a tree-graph *Carla* can either be coupled to *takes* or to *goes*, but not to both. Each of these solutions, however, is an inadequate representation of the data because *Carla* should be close to both *takes* and *goes*. If one applies the clustering program in such cases an inadequate structure will be generated. This is shown in Fig. 2 (which is based on the relatedness matrix in Table 1).

However, if we take it seriously that people are judging underlying relations, we should be able to deduce an underlying structure from our data. This can be done by expanding our relatedness matrix by one row and column. We re-introduce the second (deleted) *Carla* and assign the original relatedness values for *Carla* to either the first *Carla*-column or the second *Carla*-column. The criterion is whether the element belongs to the first or the second underlying sentence. To the remaining open cells the lowest possible cell-value is assigned. The original matrix and the expanded matrix are given in Table 1 and 2, respectively.

If our theory is correct, the analysis of the expanded matrix should give something like the underlying structure of the sentence. Fig. 3 gives the result for the *Carla*-sentence.

Contrary to Fig. 2, Fig. 3 adequately represents the relatedness values of Table 1. Moreover, the two underlying sentences are neatly separated out and the coordinator takes a neutral position. So, by simply hypothesizing the existence of an "invisible" element, a faithful hierarchical representation of the data becomes possible.

Let us investigate one more case, the sentence *John eats apples and Peter pears*. Again, coordination has led to the deletion of an element, in this case the main verb of the underlying phrase *Peter eats pears*. Eight subjects judged

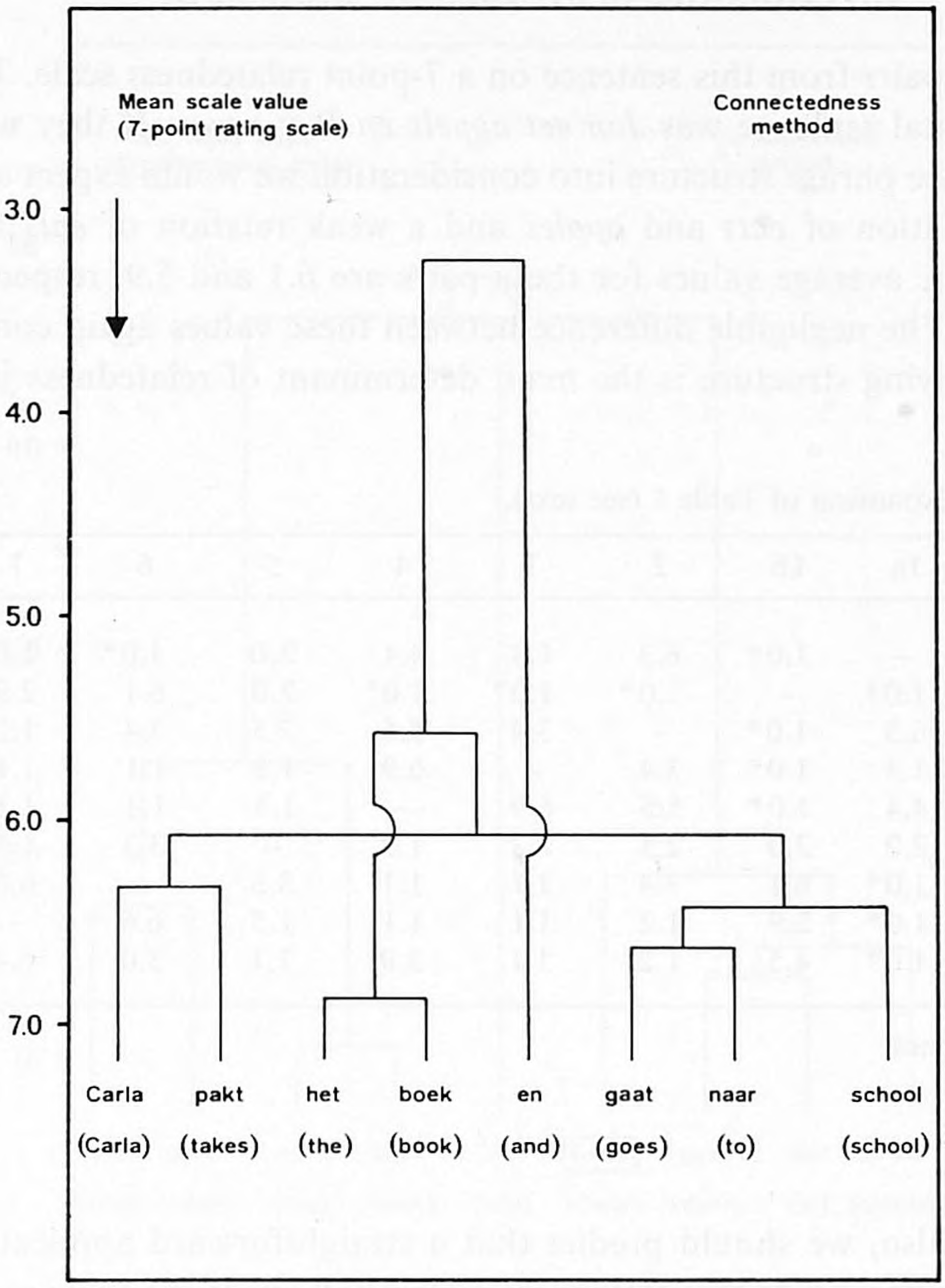

Fig. 2. HCS-solution (connectedness method) for "Carla pakt het boek en gaat naar school" (Carla) (takes) (the) (book) (and) (goes) (to) (school) (7-point rating scale data).

Table 1. Mean word relatedness values for *Carla pakt het boek en gaat naar school* (7-point scale, 16 subjects).

	1	2	3	4	5	6	7	8
1 Carla	–	6.3	1.4	4.4	2.0	6.1	2.9	4.5
2 takes	6.3	–	3.4	5.6	2.5	3.4	1.2	1.2
3 the	1.4	3.4	–	6.9	1.3	1.1	1.1	1.1
4 book	4.4	5.6	6.9	–	1.3	1.1	1.1	2.9
5 and	2.0	2.5	1.3	1.3	–	3.3	1.5	2.1
6 goes	6.1	3.4	1.1	1.1	3.3	–	6.6	5.0
7 to	2.9	1.2	1.1	1.1	1.5	6.6	–	6.4
8 school	4.5	1.2	1.1	2.9	2.1	5.0	6.4	–

the word pairs from this sentence on a 7-point relatedness scale. The Dutch experimental sentence was *Jan eet appels en Piet peren.* If they would only take surface phrase structure into consideration we would expect a relatively strong relation of *eats* and *apples* and a weak relation of *eats* and *pears*. In fact, the average values for these pairs are 6.1 and 5.9, respectively (see Table 3). The negligible difference between these values again confirms that the underlying structure is the main determinant of relatedness judgments.

Table 2. Expansion of Table 1 (see text).

	1a	1b	2	3	4	5	6	7	8
1a Carla$_1$	–	1.0*	6.3	1.4	4.4	2.0	1.0*	1.0*	1.0*
1b Carla$_2$	1.0*	–	1.0*	1.0*	1.0*	2.0	6.1	2.9	4.5
2 takes	6.3	1.0*	–	3.4	5.6	2.5	3.4	1.2	1.2
3 the	1.4	1.0*	3.4	–	6.9	1.3	1.1	1.1	1.1
4 book	4.4	1.0*	5.6	6.9	–	1.3	1.1	1.1	2.9
5 and	2.0	2.0	2.5	1.3	1.3	–	3.3	1.5	2.1
6 goes	1.0*	6.1	3.4	1.1	1.1	3.3	–	6.6	5.0
7 to	1.0*	2.9	1.2	1.1	1.1	1.5	6.6	–	6.4
8 school	1.0*	4.5	1.2	1.1	2.9	2.1	5.0	6.4	–

* added values.

Again, also, we should predict that a straightforward application of the clustering program to the relatedness matrix will result in an incomprehensive structure. This is shown in Fig. 4. It not only doesn't accurately reflect the data, but it also violates our dearest thoughts about the structure of this sentence.

If we, however, expand the relatedness matrix by an extra element *eats*, just in the same way as for the *Carla*-sentence, a more faithful (underlying) structure should arise. The result of the clustering analysis is given in Fig. 5. Here, too, we see a neat separation of the two underlying sentences, and a neutral position of the coordinator, and again the data (Table 3) are reflected more accurately in this figure.

The present sentence, moreover, gives us the possibility to check this analysis directly. We can make the missing element visible; the deletion, namely, is optional. In undeleted form the sentence is *John eats apples and Peter eats pears*, which is grammatical too. This sentence has the same underlying structure as the former one, but there is no deletion. On the basis of

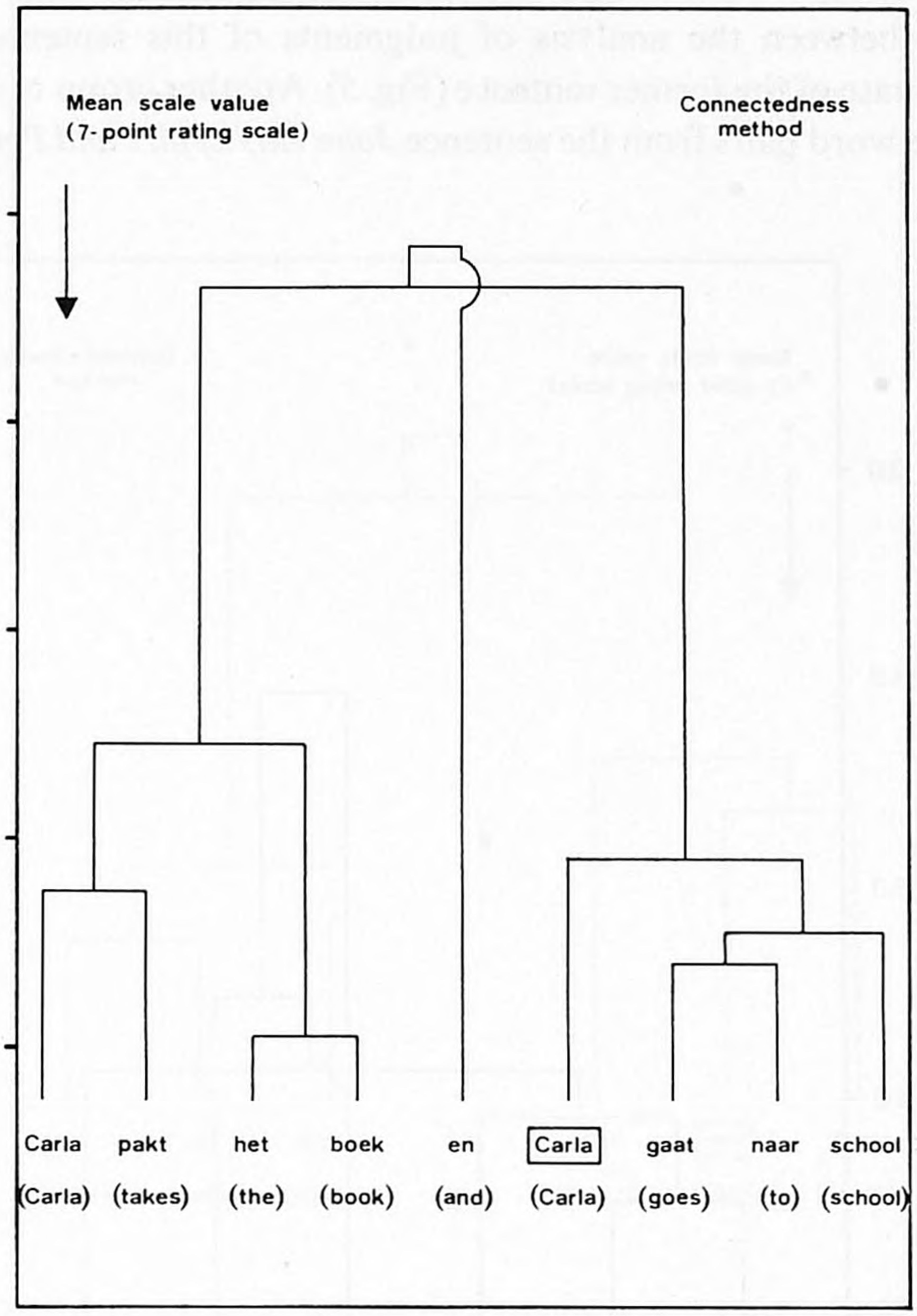

Fig. 3. HCS-solution (connectedness method) for expanded matrix of "Carla pakt het boek en $\boxed{\text{Carla}}$ gaat naar school" (Carla) (takes) (the) (book) (and) $\boxed{\text{Carla}}$ (goes) (to) (school) ($\boxed{\text{Carla}}$ added).

Table 3. Mean word relatedness values for *Jan eet appels en Piet peren* (7-point scale, 8 subjects).

	1	2	3	4	5	6
1 John	–	6.5	5.2	1.4	4.4	1.9
2 eats	6.5	–	6.1	1.9	5.6	5.9
3 apples	5.2	6.1	–	1.4	1.3	4.1
4 and	1.4	1.9	1.4	–	4.1	3.0
5 Peter	4.4	5.6	1.3	4.1	–	5.5
6 pears	1.9	5.9	4.1	3.0	5.5	–

our underlying structure hypothesis, we would thus expect a close correspondence between the analysis of judgments of this sentence and of the expanded case of the former sentence (Fig. 5). Another group of eight subjects judged the word pairs from the sentence *John eats apples and Peter eats pears*,

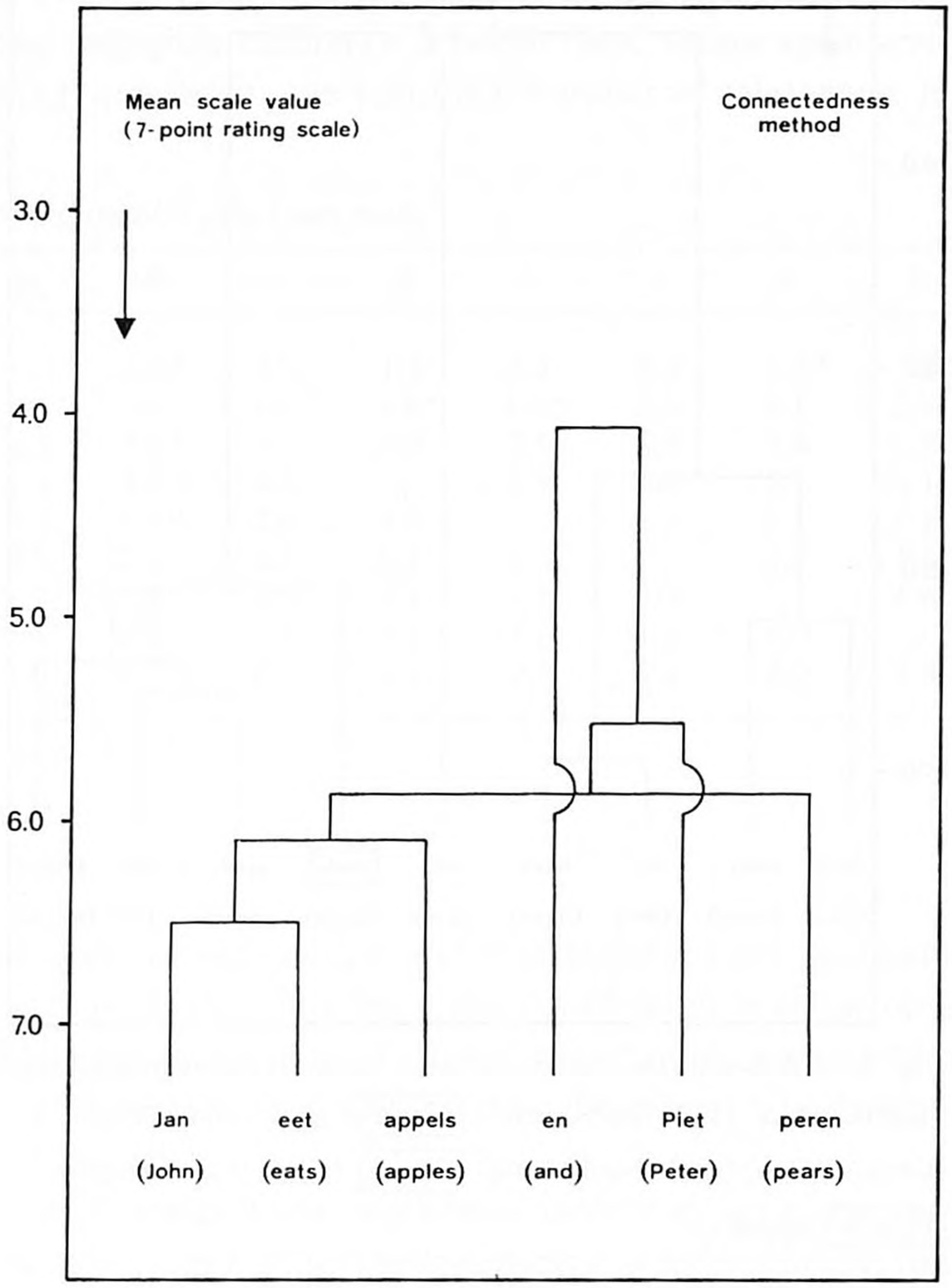

Fig. 4. HCS-solution (connectedness method) for
"Jan eet appels en Piet peren"
(John) (eats) (apples) (and) (Peter) (pears)
(7-point rating scale data).

and we applied the clustering program. The results are given in Table 4 and Fig. 6.

Apart from one small detail (*Peter eats*) the topology of this tree graph is

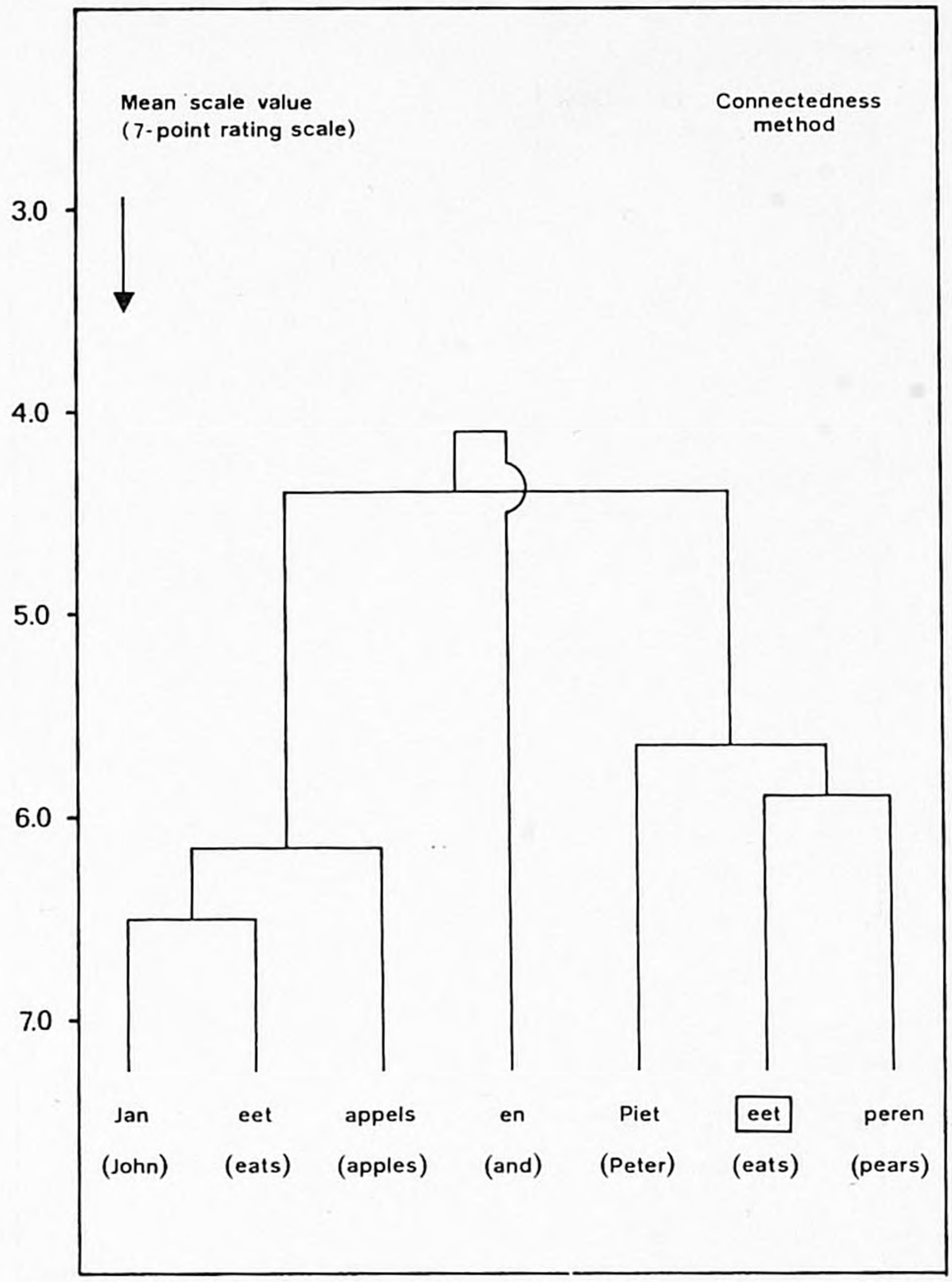

Fig. 5. HCS-solution (connectedness method) for expanded matrix of "Jan eet appels en Piet $\boxed{\text{eet}}$ peren" (John) (eats) (apples) (and) (Peter) $\boxed{\text{eats}}$ (pears) ($\boxed{\text{eats}}$ added).

Table 4. Mean word relatedness values for *Jan eet appels en Piet eet peren* (7-point scale, 8 subjects).

	1	2	3	4	5	6	7
1 John	–	6.9	6.0	3.0	4.5	1.1	1.3
2 eats	6.9	–	6.2	2.0	1.3	4.0	1.3
3 apples	6.0	6.2	–	2.1	1.3	1.3	4.1
4 and	3.0	2.0	2.1	–	2.4	1.9	2.0
5 Peter	4.5	1.3	1.3	2.4	–	7.0	5.8
6 eats	1.1	4.0	1.3	1.9	7.0	–	5.6
7 pears	1.3	1.3	4.1	2.0	5.8	5.6	–

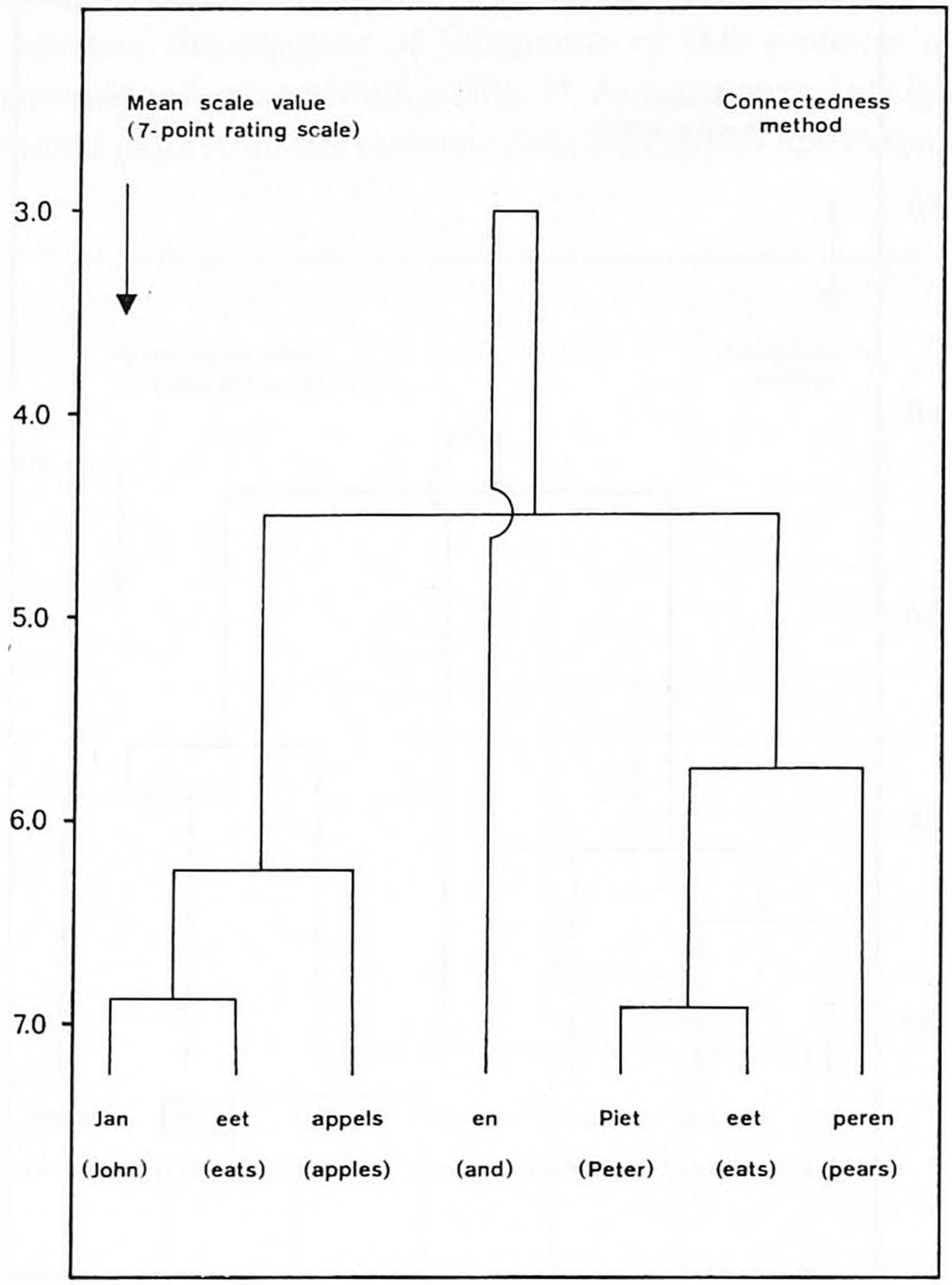

Fig. 6. HCS-solution (connectedness method) for
"Jan eet appels en Piet eet peren"
(John) (eats) (apples) (and) (Peter) (eats) (pears)
(7-point rating scale data).

identical to the former one. This is a further justification of our expanded matrix analysis.

Conclusion

In conclusion, therefore, we found that people are quite able to make judgments on the relatedness of words in a sentence. It seems that their judgments are mainly based on the underlying structure of the sentence: the intuited relation between two words is a function of the place in the P-marker of the smallest phrase including both words. This does not exclude the possibility that there are other determinants of intuited relatedness as well. In fact there are: words that are close in surface structure tend to have inflated relatedness values. Moreover, it is quite probable that certain word pairs show stronger or looser relations *independent* of the syntactic structure of the test sentence: semantic similarity may be a factor, but also the degree to which the words can form a little meaningful phrase which is not a phrase in the sentence[1]. Further study of such factors is necessary.

One could think of various applications of word relatedness scaling. One example is to be found in Flores d'Arcais' paper in this volume: certain aspects of the understanding of comparative sentences are predictable from word relatedness scaling data. Apart from behavioral applications, however, one could also think of linguistic applications. It may be a sensitive method to study the structural position of the main verb in the sentence. Middle verbs (like "eat"), for instance, tend to be more related to the subject than to the object of the sentence, whereas the full transitives (like "hit") more frequently cluster with the object of the sentence[2]. Various case-relations can be analyzed in this way. It must be added, however, that more has to be known about the "face validity" of this scaling instrument, i.e. the correspondence between the syntactic distinctions produced by this method and the generally accepted linguistic facts.

[1] I am grateful to Dr. Riegel for discussing this point at the Conference.

[2] This suggestion was made at the Conference by Dr. Bever.