

The scaling of syntactic relatedness: A new method in psycholinguistic research

W. J. M. LEVELT, *University of Groningen, The Netherlands*

Subjects were requested to judge the degrees of syntactic relation for all pairs of words from a sentence. If it is hypothesized that the judgments are determined mainly by the deep phrase structure of the sentence, it is possible to map the relatedness data on comprehensive tree graphs. Various examples are given.

The purpose of this paper is to present a scaling method that may be of some use for the psychology of grammar. The method is a judgment procedure, called *judgment of syntactic relatedness*. It can be supplemented by several types of data analysis. In this paper, we will limit ourselves to one: *hierarchical clustering analysis*.

The type of judgment under concern can be illustrated by the following example. Given the sentence *The too expensive food was tasteless*, one can ask Ss which pair of words is more related, *the* and *too* or *the* and *food*. Nobody hesitates to choose the latter pair. People have rather explicit and consistent opinions on such matters. The introspective evidence involved is not essentially different from "grammaticality," "paraphrase," and other intuitive notions on which modern grammars are based. It is also apparent from the example that syntactic structure is a determinant of such judgments. The present paper describes some exploratory experiments leading to the formulation of two hypotheses with respect to the relation between syntactic structure and relatedness judgments (for further details, see Levelt, 1967, 1969a, b).

The main purpose of hierarchical clustering analyses of word relatedness data is to obtain faithful geometric representations that allow a direct assessment of grammatical factors underlying the judgments. It should be kept in mind that such representations are not hypothesis-testing devices. They are merely given to visualize certain aspects of the data that are of grammatical interest.

EXPERIMENT 1 Method

The Ss were 24 Harvard and Radcliffe students (no language students). They were presented with the typewritten sentence, *the boy has lost a dollar*. They were instructed to judge the degrees of

relatedness between the words of the sentence. This was done by triadic comparisons: Ss were provided with a set of drawn triangles. Each triangle had three different words from the sentence at its corners. Judgments were made by first selecting the pair that seemed to be most strongly related; it was marked by a "+" along the corresponding side. Then, the least related pair was marked with a "-." The complete method of triads was used. Each S had a different random order of triads (for other details, see Levelt, 1967).

Results

For each word pair, we computed the relative frequency over triads and Ss with which it was judged as more related than other pairs. These numbers are given in Table 1. It is apparent from Table 1 that the relatedness data are intimately related to the phrase structure of the sentence; words within the same constituent have stronger mutual relations than words from different constituents. One way to visualize this state of affairs is by analyzing the data by a clustering method, mapping the matrix on a tree graph. We used S. Johnson's hierarchical clustering scheme, connectedness method (Johnson, 1967). An extensive justification of applying this method to syntactic data is given in Levelt (1967). The result is given in Fig. 1.

Discussion

The gratifying correspondence between this tree and the linguistic phrase structure of the sentence allows us to formulate a first hypothesis about the relation between syntactic structure and relatedness judgments: *The degree of relatedness is inversely related to the number of phrase boundaries separating one word from the other*. For further confirmation of this hypothesis, we must refer to our former

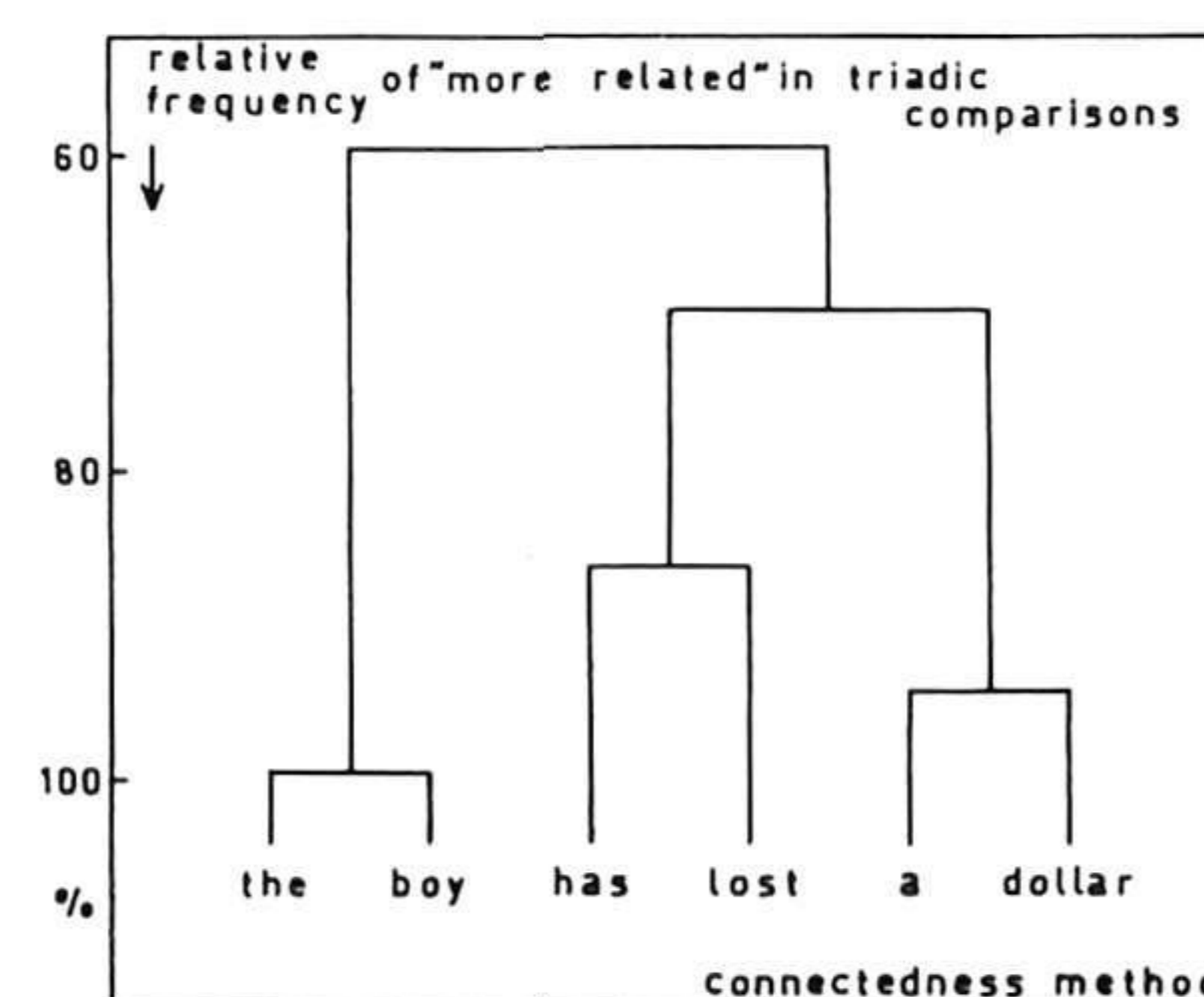


Fig. 1. Tree graph for *the boy has lost a dollar*.

reports. Here we want to proceed a little further by considering a case of coordination.

EXPERIMENT 2

Method

Eight Groningen University students (no language students) were instructed to make relatedness judgments on the sentence, *Jan eet₁ appels en Piet eet₂ peren*. (English: *John eats apples and Peter eats pears*.) Ss were presented with the typewritten sentence as well as with a list of all pairs of words from the sentence. Each pair was accompanied by a 7-point relatedness scale. The word order in the pairs and the order of pairs were randomized for each S. Ss judged by encircling 1 of the 7 points of each scale.

Results

The average scale value (7 for "strongly related") for each word pair is given in the right upper half of Table 2.

Analysis and Discussion

The result of a clustering analysis is given in Fig. 2. Again, a satisfactory structure is obtained. The two sentences that are coordinated turn up as independent structures with identical topologies. The coordinator (*and*) is least related to any other word. Subject-verb clusterings often occur for so-called middle verbs (verbs with an optional object). This is not surprising from the standpoint of a case grammar.

The sentence is also interesting because the second *eats* is deletable: *John eats apples and Peter pears*. How will this sentence be judged, by surface relations or by underlying grammatical relations? In the latter case, we should find that the relation *eats, apples* will be about equal to the relation *eats, pears*. If this happens, a tree representation will break down, because the hierarchy can only represent one of these relations at a time.

EXPERIMENT 3

Method

A new group of eight Ss judged the sentence *Jan eet appels en Piet peren* (*John eats apples and Peter pears*), as in Experiment 2.

Results and Discussion

Average scale values are given in Table 2, lower half. A clustering analysis resulted in a confused structure. It can be summarized

Table 1
Relatedness Values for the boy has lost a dollar

Words	1	2	3	4	5	6
1 the	--	.99	.43	.29	.19	.16
2 boy		--	.63	.65	.16	.31
3 has			--	.86	.31	.40
4 lost				--	.42	.70
5 a					--	.94
6 dollar						--

