DOCUMENT RESUME

ED 137 735 CS 003 334

AUTHOR Coke, Esther U.; Koether, Mary E.
TITLE A Search For Prose Features That Influence the Perception of Topical Organization.
PUB DATE Apr 77
NOTE 46p.; Paper presented at the Annual Meeting of the American Educational Reading Association (New York, New York, April 1977) ; Figures 1-4 may be marginally legible due to print quality of the original

EDRS PRICE MF-$0.83 HC-$2.06 Plus Postage.
DESCRIPTORS Cluster Analysis; Higher Education; *Organization; Perception; Printing; *Prose; *Reading Processes; *Reading Research; Secondary Education

ABSTRACT

This study sought to develop a method for describing readers' perceptions of the topical organization of prose and to specify text features that determine perception of organization. A hierarchical clustering analysis was applied to high-school and college readers' judgments of the topical relatedness of sentences in two prose passages. This produced semantically similar clusters of sentences that were reliable indicators of topical organization. In searching for text features related to topicality, it was found that the perception of topicality was weakly related to the lexical similarity of sentences and that both the spatial relationship between sentences and typographical conventions influenced the perception of topicality. (Author/AA)

AERA 1977
Session 12.04

A Search For Prose Features That Influence the Perception of Topical Organization

Esther U. Coke and Mary E. Koether

Bell Laboratories

The writer of a reasonably long piece of expository prose usually discusses a number of topics in presenting the general subject of the passage. Part of the reader's understanding of what the writer is trying to say involves perceiving these topics. At present, little is known about the stimulus attributes of texts that readers use in detecting the topics presented by the writer. The purposes of the present study were (a) to develop a method for describing the topical organization perceived by readers and (b) to look for text features that determine the reader's perception of topical organization.

Method of Describing Topical Organization

The method for describing topical organization is based on readers' judgments of the topical relatedness of sentences, rather than on a content analysis of the meaning of sentences. Readers mapped topics onto text sentences by identifying sentences that were topically similar. Then an hierarchical clustering analysis of the matrix of reader responses was used to determine clusters of topically identical sentences. The details of the method are as follows.

1. Mapping of Topics onto Sentences.

A. Reading. A reader was given a passage and told to read the passage carefully. The reader was told that he would be asked to summarize the topics discussed in the passage after he had finished reading it.

B. Topic generation. After reading the passage once, the reader was told to review the passage and write down from five to nine topics. A topic was defined for the reader as a general description of something the writer discussed. The reader was also given an example of a topic. To encourage the reader to describe topics clearly, he was told that another person would try to locate the topics in the passage. The reader wrote down his topics on separate 3x5 cards which were collected by the experimenter. The reader could consult the passage while generating topics.

C. Topic identification. The reader identified the sentences that discussed each of the topics he generated. The reader was given his own topic cards, one at a time, along with the passage and was told to underline all the sentences that had to do with that topic. The reader's topic cards were given to him in a haphazard order obtained by shuffling the cards. Each time the reader was given a card he was also given a fresh copy of the passage. This allowed the reader to assign a sentence to more than one topic.

2. Data analysis.

The readers' responses were entered into a matrix representing all possible sentence pairs in a passage. A cell entry in this response matrix is the number of readers who underlined both sentences of a pair to indicate that they belonged to the same topic. The following two measures of topical similarity were then derived from this response matrix:

(1) Frequency measure of similarity. This measure is the number of readers who indicated that two sentences were topically related. This measure of similarity is simply the cell entry in the response matrix for a sentence pair.

(2) Probability measure of similarity. This measure is the cell entry in the response matrix weighted by the number of readers who assigned the two sentences to topics. This measure is the geometric mean (GM) of two contingent probabilities:

$$GM = P([n,m]/m) * P([n,m]/n)$$

where

P([n,m]/m)	is the probability that sentences n and m are assigned to the same topic given that sentence m is assigned to a topic;
P([n,m]/n)	is the probability that n and m are assigned to the same topic given that sentence n is assigned to a topic;
[n,m]	is the number of readers assigning n and m to the same topic;
n	is the number of readers assigning sentence n to topics;
m	is the number of readers assigning sentence m to topics.

The two similarity matrices derived from a response matrix differ from each other in that the level of similarity between two sentences for the frequency measure is partly determined by the number of subjects who assigned the sentences to topics. This is not the case for the probability measure.

3. Hierarchical Clustering Analysis.

An hierarchical clustering analysis was used to determine the topical organization of a passage. This analysis finds the sentences that are closest together at each level of topical similarity in a similarity matrix. The analysis results in a tree structure. At the top of the tree are clusters of sentences that are highly similar to each other. Clusters at lower levels of the tree represent decreasing levels of similarity between sentences. The tree structure is hierarchical. Sentences that form a cluster towards the top of the tree are all members of the same cluster at lower levels of the tree. Also, a sentence cannot be a member of more than one cluster.

Often data that are being analyzed are not truly hierarchically organized. Johnson (1967) has developed algorithms for retrieving underlying hierarchical structures. Johnson's maximum method was chosen for the present analysis of topical organization. The maximum method is an iterative procedure for joining clusters together until only one cluster remains. This method only forms a new cluster when all the similarities between cluster sentences are at or above a minimum value.

This minimum value defines a cluster level of a tree. All clusters forming at a cluster level have the same minimum similarity value.

Any similarity matrix can be analyzed with the maximum method. Therefore, a scheme is needed for evaluating the goodness-of-fit of the tree structure to the data being analyzed. Halff (1975) has suggested a graphical evaluation technique. Halff proposed that an index of goodness-of-fit be calculated for each cluster level of a tree. This index is the proportion of unclustered sentence pairings that should have been clustered at a given cluster level. This index would be zero at all cluster levels if the data were truly hierarchically structured. Therefore, goodness-of-fit can be evaluated by looking at the deviation of the index from zero at each cluster level. Halff also proposed that a value of the index can be calculated that represents the worst fit for each cluster level. This is the expected value of the index if sentences were clustered together at random.

Halff's evaluation procedure was used in the present study to decide how well the readers' perceptions of topical relatedness were described by the hierarchical clustering analysis. Using this procedure, it was also possible to decide which of the two measures of similarity produced the better tree structure for a passage.

Application of the Clustering Technique

The cluste ing technique was used to determine the the topical organization of two prose passages. Two tree structures were derived for each passage. One was based on the frequency measure of similarity and the other, on the probability measure of similarity.

Method

Material. Two passages were taken from a publication containing technical articles of general interest. One passage (GOP) discussed gopher damage to telephone cables; the other passage (NOD) discussed pay telephone service. Each passage was approximately 2000 words long. There were 97 sentences in GOP divided into 24 paragraphs; there were 99 sentences in NOD also divided into 24 paragraphs. Both passages were fairly difficult (Flesch Reading Ease Score was 50.7 and 50.8 for GOP and NOD, respectively).

Subjects. Judgments of topical relatedness were made by two groups of readers.

(1) Twenty-eight paid, volunteer high school students. Each student only saw one of the passages.

(2) Thirty-two paid, volunteer college undergraduates. These readers also served in Experiment 1, to be reported. They read the published version of one passage and an altered version of the other passage. Only judgments for the published

versions of a passage were used. Sixteen students generated topics for GOP and 16 generated topics for NOD.

Results

The response matrices produced by the two groups of readers were compared with each other. The Pearson product-moment correlation between the matrices indicates that the two groups of readers were in general agreement about the topical relatedness of the sentences (r= .82 and .93 for GOP and NOD, respectively). This result suggests that the method for describing topical relatedness yields reliable results.

The response matrices of the two groups were combined into a single matrix for each passage. These combined matrices were used to derive the frequency and probability measures of inter-sentence similarity. Each similarity matrix was then subject to an hierarchical clustering analysis.

The tree structures representing the topical organization of a passage are shown in Fig. 1-4. Figures 1 and 3 are the tree structures for GOP and NOD derived from the probability measures of similarity. Figures 2 and 4 are the tree structures

Insert Fig. 1, 2, 3 and 4 about here.

for GOP and NOD derived from the frequency measure of similarity. The cluster levels and their similarity values are shown in the left-hand column of each figure. Sentence numbers are listed along the top of the figure. The formation of a new sentence cluster from an old cluster is indicated by a horizontal dashed line connecting the new sentences entering a cluster with the vertical line coming down from the old cluster.

In order to demonstrate how these tree structures relate to the content of a passage, Tables 1 and 2 give verbal descriptions of the content of the clusters for GOP and NOD, respectively.

Insert Tables 1 and 2 about here.

A brief summary of each sentence in a passage is shown. The sentences are organized as they are in the tree structures derived from the probability measures (Fig. 1 and 3). Sentence numbers are shown in the left margin of the page. The lines connecting sentences correspond to clusterings at different levels of the tree. The letter labels correspond to the letters labeling each major branch of the trees in Fig. 1 and 3. The sentence numbers and brackets in the right margins of the two tables represent the tree structures derived from the frequency measures of similarity (Fig. 2 and 4). This information was included in the tables to help in comparing the two tree structures for a single passage.

The two different measures of similarity produced slightly different tree structures for each passage. Clusters differed primarily in the clustering levels at which specific sentences joined clusters, rather than in which sentences formed clusters. This is illustrated by the letter labels on the figures. Each major branch of the probability-measure trees for GOP and NOD has been assigned a letter. These letters were then used to label corresponding parts of the frequency-measure trees. For GOP, four of the five major branches in Fig. 1 correspond to major branches in Fig. 2 One major branch (E) in Fig. 1 split into two branches in Fig. 2. For NOD, four of the six major branches in Fig. 3 correspond to major branches in Fig. 4. Parts of two branches (C and D) in Fig. 3 form a single branch in Fig. 4.

Evaluation of Goodness-of-fit. Halff's goodness-of-fit indices were calculated for all trees and used to determine which measure of similarity produced the better tree structure for a passage. The indices and the estimates of the worst fit are plotted in Fig. 5-8. These plots correspond to the

Insert Fig. 5-8 about here.

tree structures in Fig. 1-4. Halff's index remains reasonably near zero for all tree structures except at low levels of similarity. This finding suggests that the topical organization retrieved by the clustering technique describes the similarity data reasonably well.

Specific comparisons between similarity measures for both passages indicate that the tree structure derived from a probability measure matrix provides a better description of the readers' perception of the topical organization of a passage than the tree structure derived from a frequency measure matrix.

For both GOP and NOD, the deviations of the goodness-of-fit indices from zero appear to be less for the probability measure data (Fig. 5 and 7) than for the frequency measure data (Fig. 6 and 8).

Discussion

The hierarchical clustering analysis generated clusters of semantically related sentences. These clusters appeared to be related to the subjects discussed in the passages. In addition, the clustering analysis produced tree structures that described the similarity data for sentences reasonably well. These findings indicate that the clustering method is a valid method for describing the readers' perception of topical organization. Thus, clusters of topically similar sentences can be retrieved without evaluating the verbal labels that readers use to describe topics.

Exploration of Text Features that Determine Topical Organization

Experiment 1

The topical organizations for GOP and NOD (Fig. 1 and 3) were compared with clusters of sentences that had been formed from sentences that shared content words. This comparison was part of an effort to find text features that would predict topical organization. The co-occurrence of content words in sentences was analyzed because it was thought that writers might tend to use the same words when discussing a topic. If so, sentences that were similar in their use of content words should be topically related.

The comparison revealed little relationship between clusters based on topical similarity and clusters based on lexical similarity. One reason for this lack of correspondence was that topically similar sentences tended to be near each other in the text while lexically similar sentences could be located anywhere in the text. It was hypothesized that spatial proximity might have an influence on the reader's perception of topical relatedness. In that case, lexically similar sentences would be perceived as topically similar if these sentences were near each other. To test this hypothesis, the GOP and NOD passages were rewritten. A few lexically similar sentences that were far apart in the original texts were moved so that they were adjacent to each other in the rewritten texts. Each pair was chosen so that the sentences of a pair were lexically highly similar and topically very dissimilar. It was predicted that the sentences would become topically similar in the rewritten passages.

Method

Material. Both the original and rewritten versions of the GOP and NOD texts were used. The texts were rewritten so that nine sentences were moved in each passage. A moved sentence was placed next to the sentence with which it was lexically similar, but topically dissimilar. Topical dissimilarity was insured by picking sentence pairs that came together in a cluster at low levels of the tree representing the topical organization of a passage. Lexical similarity was insured by picking sentences that clustered together at high levels of the tree representing the lexical organization of a passage. The lexical organization of a passage was determined from a measure of lexical similarity. This measure was based on both the number of content words shared by two sentences and the number of content words in the two sentences that co-occurred in other sentences. The index of commonality (Deese, 1965) was used to determine the degree of lexical similarity.

When a sentence was moved, all anaphoral words were replaced with their referents. In additon, the tense of the moved sentence's verb was changed if it conflicted with the verbs in the sentences adjacent to the moved sentence.

Procedure. The versions of a passage seen by a reader during the topic generation and topic identification phases of the mapping procedure were varied. Some readers saw the same version of the text in both phases (OO - original version of text during both phases; RR - rewritten version during both phases). Other readers saw one version of the text while generating topics and the other version while identifying topic sentences (OR - original version during generation and rewritten version during identification; RO - rewritten version during generation and original version during identification).

The design for the experiment is shown in Table 3.

Insert Table 3 about here.

Each reader mapped topics onto sentences for both passages. After mapping topics for one passage, the reader was given a short rest before mapping the topics for the other passage.

Subjects. Sixty-four paid, volunteer college undergraduates were assigned at random to one of the conditions shown in Table 3.

Results

The mean similarity measures for the nine lexically similar sentences are shown in Table 4. To determine whether

Insert Table 4 about here.

these sentences became topically similar when the sentences were adjacent to each other, the means for conditions OR, RR and RO were compared with the means for condition OO. A t-test for correlated means was used.

The lexically similar sentences became topically similar when moved next to each other during the topic identification phase of the mapping procedure. For both passages, topical similarity was significantly greater when the sentences were adjacent ($t = 8.91$ and 9.76, $df = 8$, $p < .05$ for the OR and RR conditions of GOP; $t = 5.82$ and 5.84, $df = 8$, $p < .05$ for the OR and RR conditions of NOD).

The topical similarity between pairs did not increase when readers saw the rewritten texts only during topic generation (condition RO). For GOP, sentences actually became less topically similar ($t = 2.96$, $df = 8$, $p < .05$) while for

NOD, there was virtually no change in topical similarity ($t < 1$).

Discussion

The finding that use of the rewritten text during topic generation had little effect on topical similarity is not surprising. During the generation phase of the mapping procedure, readers were only concerned with summarizing the subjects discussed by the writer. Apparently, the displacement of a very few sentences (less than 10%) did not disrupt the reader's ability to understand what the writer discussed.

It was during topic identification that sentence proximity had its effect on topical similarity. During this phase of the mapping procedure, readers have to make judgments about each sentence. In keeping with the findings of Dooling and Lachman (1971) among others, the results of this study show that the topical context of a sentence played an important role in the reader's interpretation of the sentence's meaning.

Experiment 2

While spatial proximity and the perception of topical relatedness are strongly related, this is not invariably the case. Detailed examination of the tree structures for both passages indicated that the first sentence of a new paragraph and the last sentence of the previous paragraph were often in different clusters. This was especially so for clusters with high levels of similarity. These sentences at paragraph boundaries seemed to be perceived as topically less similar to each other than sentences within a paragraph.

This observation suggested that the perception of topical organization was influenced by the typographical convention of dividing a passage into paragraphs. This experiment evaluated the influence of paragraph structure on topical organization by comparing passages with and without paragraph indentation. Readers appear to use paragraph indentation as one cue that a new topic has begun. It was hypothesized that sentences at paragraph boundaries (i.e., the first sentence of a paragraph and the last sentence of the previous paragraph) would be perceived as more topically similar when the indentation cue to topic change was removed.

Method

Material. The original, indented GOP and NOD passages and versions of these passages without paragraph indentation were used. Each original passage was divided into 24 paragraphs. Thus, 24 of the 97 sentences in GOP were indented in the original version, and 24 of the 99 sentences in NOD were indented in the original version.

Procedure. Readers mapped topics to sentences for both the GOP and NOD texts. Half of the readers saw the GOP text first while half saw the NOD text first. After mapping topics for one passage, the reader was given a short rest before mapping the topics for the other passage.

Readers used the indented versions of the texts during the topic generation phase of the mapping procedure. The unindented version was used during the topic identification phase. This was done because we wanted readers to generate topics under optimal conditions for perceiving topic boundaries. In this way we could better assess the effect of removing a cue to these boundaries on the perception of topical relatedness. Also, the topical similarity data from the first study reported in this paper were used to evaluate changes in similarity due to the removal of indentation. Therefore, it was desirable for readers in this experiment to generate topics under the same conditions as readers in the first study.

Subjects. Eighteen paid, volunteer high school students participated in the experiment.

Results

The response matrix produced by the readers in the present experiment was compared with the response matrix for the 30 readers in the first study. The Pearson product-moment correlations indicated that there was general agreement between the two groups of readers (r = .90 and .80 for GOP and NOD, respectively). Also, the tree structures resulting from a clustering analysis of the probability measure matrices for this experiment closely resembled the tree structures for the first study shown in Fig. 1 and 3.

The finding of little difference between the topical organization perceived by the two groups could be expected. Readers in both groups generated topics under identical conditions, and removal of indentation during topic identification only affected 24 sentences. The general agreement of the groups confirms the conclusion that the present technique for studying topical organization produced reliable results.

The effect of removing paragraph indentation on the similarity of sentences at paragraph boundaries is shown in Table 5. This table shows the mean probability measure of

---

Insert Table 5 about here.

---

similarity (a) for the adjacent sentences at paragraph boundaries (i.e., the sentences at the end of one paragraph and the beginning of the next) and (b) for the adjacent sentences within paragraph boundaries.

When readers used the indented versions of texts, adjacent sentences at paragraph boundaries were less similar than adjacent sentences within paragraphs ($t = 11.06$, $df = 94$, $p < .05$ for GOP; $t = 12.10$, $df = 96$, $p < .05$ for NOD). When paragraph indentation was removed, readers still perceived the adjacent sentences at paragraph boundaries to be less similar than adjacent sentences within paragraphs ($t = 4.63$, $df = 94$, $p < .05$ for GOP; $t = 6.77$, $df = 96$, $p < .05$ for NOD). However, there was a tendency for paragraph-boundary sentences to be perceived as more similar when paragraph indentation was removed than when indentation was present. This increase in similarity was statistically significant for both passages when evaluated with t-tests for correlated means ($t = 2.58$ and 2.78 for GOP and NOD, $df = 22$, $p < .05$). This increase in similarity for the sentences at paragraph boundaries contrasts with a decrease in the similarity between adjacent sentences within paragraphs.

Discussion

The results of this experiment suggest that paragraph indentation is only one cue that readers use in determining topical boundaries. Even when paragraph indentation was removed, readers still found adjacent sentences at paragraph boundaries less similar than adjacent sentences within paragraphs. This result is consistent with the finding of Koen et al (1969) that readers can identify the first sentence of a paragraph with greater-than-chance accuracy when paragraph indentation has been removed. The results of the present study suggest that paragraph indentation acts to enhance the effects of the syntactic and semantic cues to topical boundaries.

General Discussion

The study presented a method for determining the reader's perception of the topical organization of a passage. The hierarchical clustering analysis produces clusters of semantically related sentences that seem to reflect the subjects discussed by the writer. In addition, the method yields reproducible results.

This clustering method of retrieving topical organization was devised to help in the search for text features that readers use to understand what they read. The results of Experiment 1 showed that the lexical similarity of sentences may lead readers to perceive that two sentences are topically similar if the sentences are close together. This finding suggests that the spatial relationship of sentences is an important determinant of the perception of topicality. Perhaps one of the reading strategies used by readers is to assume that spatially adjacent sentences belong to the same topic unless the writer signals the reader that a new topic has begun. Experiment 2 presented evidence that one such signal may be the typographical convention of paragraph indentation.

The hierarchical clustering method provides a reliable way of characterizing the organization perceived in a passage. Having a metric to describe organization can be valuable to both writers and researchers. This clustering method provides a way to map topical organization on texts and to detect the sequencing of topics in a passage. Using this method it should be possible to evaluate how alternative ways of organizing passages affect readers from both a practical and theoretical point of view.

References

Deese, J. The structure of associations in language and thought. Baltimore: The Johns Hopkins Press, 1965. Pp. 51 - 52.

Dooling, D.J. & Lachman, R. Effects of comprehension on the retention of prose. Journal of Experimental Psychology, 1971, 88, 216-222.

Halff, H.M. Graphical evaluation of hierarchical clustering schemes (Tech. Rep. 1). Urbana, Ill.: University of Illinois, Laboratory for Cognitive Studies in Education, October, 1975.

Johnson, S.C. Hierarchical clustering schemes. Psychometrika, 1967, 32, 241-254.

Koen, F., Becker, A. & Young, R. The psychological reality of the paragraph. Journal of Verbal Learning and Verbal Behavior, 1969, 8, 49-53.

Table 1

Verbal description of the content of the sentences of the clusters for the clustering analysis of GOP. Sentences in the same cluster are bracketed together with straight lines. A diamond ◊ encloses the number of each sentence that is the first sentence of a paragraph.

Cluster	Sentences	Description	Sentences	Cluster
A	1	introduction to general subject.	1	A
A	2,3	introdurtion of gopher and how he destroys cables.	2,3	A
A	4,5	serious problem, much evidence.	5 4	A
A	6,7	first study: time,place,gophers.	6,7	A
A	8,9	digression about first study; dead,live.	8,9	A
A	10,12	protection used prior to first study; other materials also tried in study.	10,11,12	A
A	11	main conclusion of study 1.		A
A	13	things have changed since first study.	13	A
A	14,15	new evaluation planned for 1966 with objectives described.	14,15	A
A	16,17	new evaluation. who,where,when,what.	16,17	A
B	18,19	interesting facts emerged from intense study of Geomys.	18,19	B
B	20,21	description of gopher: length,weight,habitat, and size of appetite.	20,21	B
B	22	geographical location.	22	B
B	23,24	problem of incisor growth for eating.	23,24	B
B	25,26	may lead to death.	25,26	B
B			27 30	B
B	28	ineffectiveness of chemical repellents because they depend on taste.	28	B
B	27,30	location inside mouth means can chew anything; sharpness of biting surface.		B
B	29,31	hardness of surface; large biting force results in enormous pressure.	29,31	B

C

40	ten specimens of each design.	40
42	enumeration of types and thichnesses of cable armor in samples.	39,41,42,43
41	diameter of cables and wires in sample.	
39,43	general description of samples of cables,wires and protective material including rodent repellents.	
44	armor and shield types used with wires.	44
45,46	description of buried wire and their shield and armor.	45,46
32,33,34	design of cages:who,material,details of design for experiment.	32,33,34,35
35	how test materials mounted in cage.	
36,37	how long samples left in cage;no need for incentive to chew.	36,37
38	hunger not a factor in chewing.	38
47,48	rating scheme for evaluating damage in five categories.	47,48

C

	56,57	general statement that test clarified that no damage over a certain diameter.	56	
	60	specification of diameter.	57,60	
	58,59	maximum jaw opening and how large cables can be be attacked.	58,59	
	61	field reports support minimum cable diameter for no damage.	61,62	
	62,63	field report of size and location of largest cable damage.	63	
	64,65	construction of cable sheath unimportant; discussion of steel armored cable.	64,66	
	66	results - no need to solder steel armored cable.		
D	49,51,52	one plastic offered protection; plastic to prevent corrosion did not work. thicknesses of lead and aluminum that did not work.	49,51,52	D
	50	all other plastics,even those with repellents, failed.	50	
	54,55	description of most of other metals that were effective	54,55	
	53	statement that most metals were effective.	53	
	67,68	all armored cable offered protection but corrosion of armor if penetration of plastic cover because armor then less resistant.	65,67	
	69	corrosion makes normal thickness of armor not always	68,69	
			74	
	78,79	importance of type of internal structure of wire; air space is especially bad.	78,79	
	80	one wire not damaged had no air space in core.	80	
	77	degree of overlap of windings had no effect.	77	
	74	in general, wires did not perform well.		
	75,76	odd result for two wires;description of difference in winding.	75,76	
	81	conclusion about wire construction based on results having to do with solid core.	81	

70	old theory about duration of need for cable protection.	70 72,73
71,72	ground compacted, but damage reported after 6 years.	71
73	protection needed for life of cable, not just 5 years.	
82	best armor for buried wire is stainless steel.	82,83
83,84	disadvantage of stainless steel is that it attracts lightning; limited use to service instead of rural.	84
85,86	reason for service wire resistance to lightning as opposed to rural wire; recommendation for type of rural wire.	85,86
87	why this type is best for resisting lightning damage.	87
88	Denver study only the first phase of studies.	88
89	reason for field test.	89
91,92	reasons for choice of field test area and the nature of field test.	91,92
90,93	place of field test;use of test results.	90,93
94	significant observations made to date.	94
95,96	plastics vs. metals.	95,96
97	prospects.	97

E

E

E

Table 2

Verbal description of the content of the sentences of the clusters for the clustering analysis of NOD. Sentences in the same cluster are bracketed together with straight lines. A diamond ◊ encloses the number of each sentence that is the first sentence of a paragraph.

Cluster	Sentences	Description	Sentences	Cluster
A	◊1◊,2	historical introduction - coin telephone service.	◊1◊,2,3	A
	3	accomplishments.		
	◊4◊,5	review of coin telephone service;coin service improvement program defined.	◊4◊,5	
	7,8	pre-pay (deposit) calling since 1920's.	7,8	
	9	exception to pre-pay calling	9	
	10,11	description of post-pay operation and its disadvantages.	10,11	
B	◊6◊,◊12◊	dial-tone-first a major feature of improvement; features defined.	◊6◊,◊12◊	B
			13	
	14	without dial tone don't know if phone is working before deposit of money.		
	13	risk of loss-of-money avoided with dial-tone-first.		
			16	
	16,17	type of calls with dial-tone-first made without deposit.	17	
	◊15◊,18	can dial without depositing for certain calls; including calls requiring operator assistance.	◊15◊,18	
	19	procedure for regular local call.	19	
	20,21	details of procedure for correcting deposit errors.	20,21	
			14	
	◊22◊	improvement program offers better voice transmission.		
	23,24	explanation of how and why.	◊22◊,23,24	

	36	modifications not the only problem.		
			39	
	39,40	holding time increases due to time to deposit coin after dial tone provided;also need to redial.	40	
	42	coin calls handled by ESS similarly take more time.	42	
			36	
	37,38	s me additional equipment needed in central office - example of holding time of crossbar.	37,38	
	41,43	more senders and registers;more coin-control circuits because of holding time in ESS.	41,43	
C	30	transfer of panel office service to ESS or crossbar.	30	
	28,29	equipment designs available to modify crossbar. offices;step-by-step offices in 1970.	29	C
			28,31	
	31	four key components of crossbar modification: line circuits,		
	32,33	senders, markers.	32,33,35	
	34,35	circuits;ESS conversion via program.	34	
	44,45	changes in cord switchboards.	44,45	
	46,47	conversion of post-pay offices.	46,47	
	48	compromises to minimize costs at coin telephones	48 50	
	49,50	main modification to multi-slot for dial-tone-first; none for new universal single-slot.	49	C
	51,52	modifications of multi- and single-slot described; same,different.	51,52	
			25,26 27	D

	25	ideally applicable to all central offices, switchboards and stations.	25	
	26,27	but changes over-all substantial;therefore there are compromises.	26,27	
	53	changes in signalling between telephones and central office necessary.		
	54,55	how signalling operates currently.	54,55	
			53	
	56,57	how signalling will operate with dial-tone-first.	56,57	
D	58 59	description of complication in central office because must verify initial deposit if not required.	58 59	D
	60,61	central offices differ in when and way to decide to complete or deny call;but same check for coins.	60	
			61	
	62	universal coin telephone possible with any central office.	62	
	63 64	addition of +48-volt battery and use in central office.	63 64	
	65	when battery connected with cord switch.	65 66	
	66	this precludes Touch-Tone.		
	67,68	in future, fewer cord switchboards, thus limit on Touch-Tone will lessen.	67,68	

	69	trials of dial-tone-first.	69
	70	first trial - where - Hartford.	70
	71,72	conditions of Hartford trial.	71,72
	73	purpose of Hartford trial.	73
	82,83	standard developed as result of Hartford trial and tried in Danville.	
	84	purpose of Danville trial.	
	85,86	conditions of Danville trial.	
	80	method of operating in Greenwich trial.	80
E	77	Greenwich Village trial.	77,78,79,81
	78,79	conditions of Greenwich trial.	
	81	reaction to Greenwich trial.	
			84,85,86
	87	Danville trial verified feasibility of overall plan and standard method.	
	88	transition to dial-tone as smoothly in switchboard as in TSP.	
	74	feasibility of service confirmed in Hartford trial and good customer reaction.	74,76
	75,76	specifics of results - positive and negative.	75
			91
	89,90,91	overall results discussed.	87,89,90
			82,83
	92	specific effect of one result-reporting of trouble.	88,92
	93	trial results lead to nationwide introduction of dial-tone-first.	93,94
	94,95	where being introduced.	95
F	96	all locations are crossbar offices.	96
	97,98	cost of conversion and schedule	97,98
	99	future	99

Table 3

Design of Experiment 1 showing the passage versions used during topic generation and identification and the order in which readers saw the two passages.

			First Text			Second Text	
			Mapping Procedure Phases			Mapping Procedure Phases	
Condition	N	Text	Reading and Topic Generation	Topic Identification	Text	Reading and Topic Generation	Topic Identification
1	8	GOP	Original	Original	NOD	Rewritten	Original
2	8	GOP	Original	Rewritten	NOD	Rewritten	Rewritten
3	8	GOP	Rewritten	Original	NOD	Original	Original
4	8	GOP	Rewritten	Rewritten	NOD	Original	Rewritten
5	8	NOD	Original	Original	GOP	Rewritten	Original
6	8	NOD	Original	Rewritten	GOP	Rewritten	Rewritten
7	8	NOD	Rewritten	Original	GOP	Original	Original
8	8	NOD	Rewritten	Rewritten	GOP	Original	Rewritten

Table 4

The mean topical similarity between sentences as a function of spatial proximity during the topic generation and identification phases of the mapping procedure.

Condition	Topic Generation (Test Version Used During The Mapping Procedure)	Topic Identification (Test Version Used During The Mapping Procedure)	Passages	
			GOP	NOD
OO	Original	Original	.28	.31
RO	Rewritten	Original	.11	.34
OR	Original	Rewritten	.86	.96
RR	Rewritten	Rewritten	.90	.95

Table 5

Mean probability measure of similarity between adjacent sentences at paragraph boundaries and adjacent sentences within paragraphs as a function of the presence of paragraph indentation

Version of Text Used During Topic Identification	Adjacent Sentences At Paragraph Boundaries	Adjacent Sentences Within Paragraphs
GOP:		
Indented	.60	.93
Unindented	.68	.84
NOD:		
Indented	.59	.93
Unindented	.67	.88

Figure Captions

Fig. 1 The topical organization of GOP resulting from an hierarchical clustering analysis of the similarity matrix based on the probability measure of similarity.

Fig. 2 The topical organization of GOP resulting from an hierarchical clustering analysis of the similarity matrix based on the frequency measure of similarity.

Fig. 3 The topical organization of NOD resulting from an hierarchical clustering analysis of the similarity matrix based on the probability measure of similarity.

Fig. 4 The topical organization of NOD resulting from an hierarchical clustering analysis of the similarity matrix based on the frequency measure of similarity.

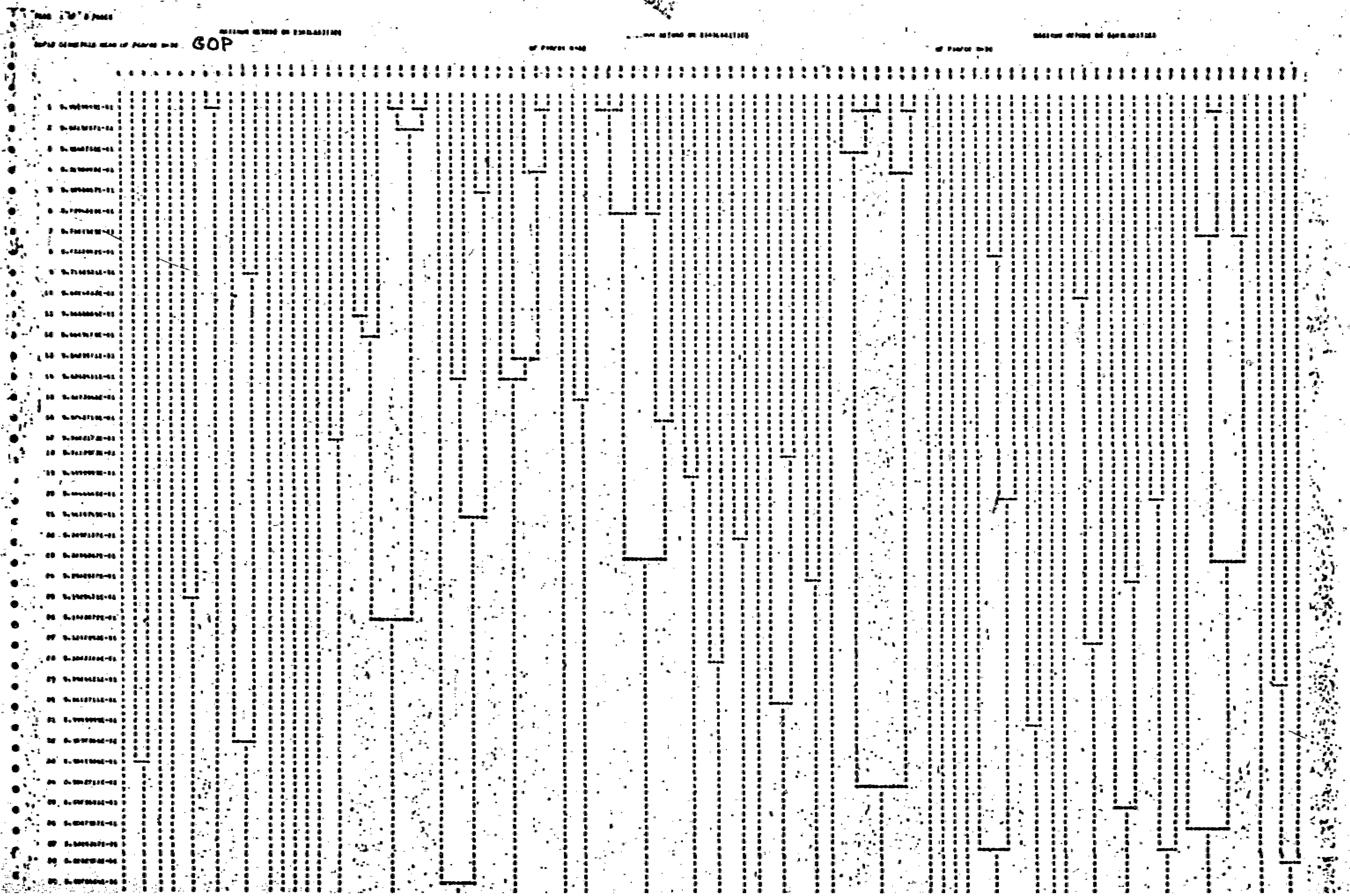
Fig. 5 Graphical evaluation of the goodness-of-fit for the GOP tree structure (Fig. 1) derived from the probability measure of similarity.

Fig. 6 Graphical evaluation of the goodness-of-fit for the GOP tree structure (Fig. 2) derived from the frequency measure of similarity.

Fig. 7 Graphical evaluation of the goodness-of-fit for the NOD tree structure (Fig. 3) derived from the probability measure of similarity.

Fig. 8 Graphical evaluation of the goodness-of-fit for the NOD tree structure (Fig. 4) derived from the frequency measure of similarity.

GOP



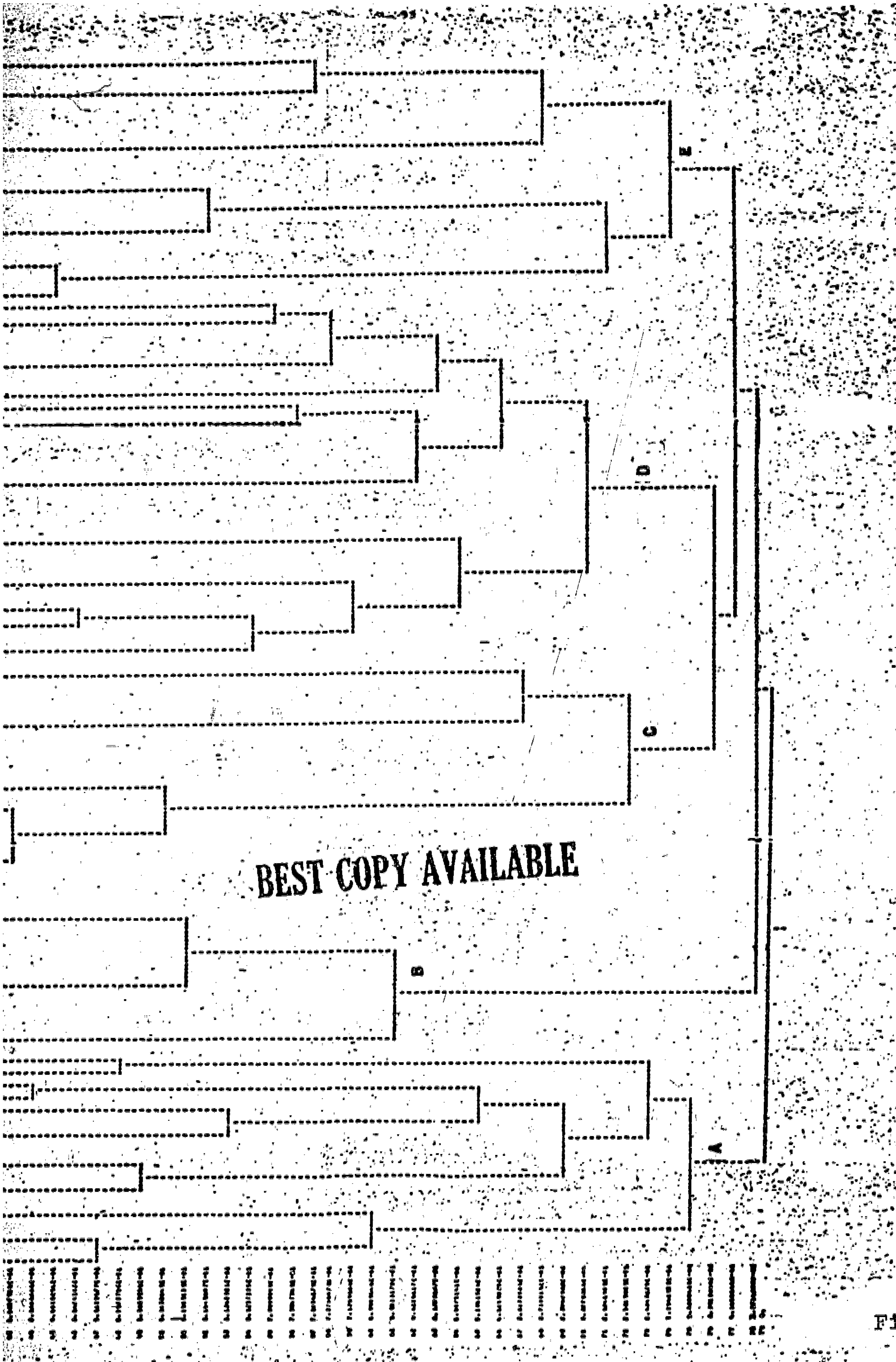
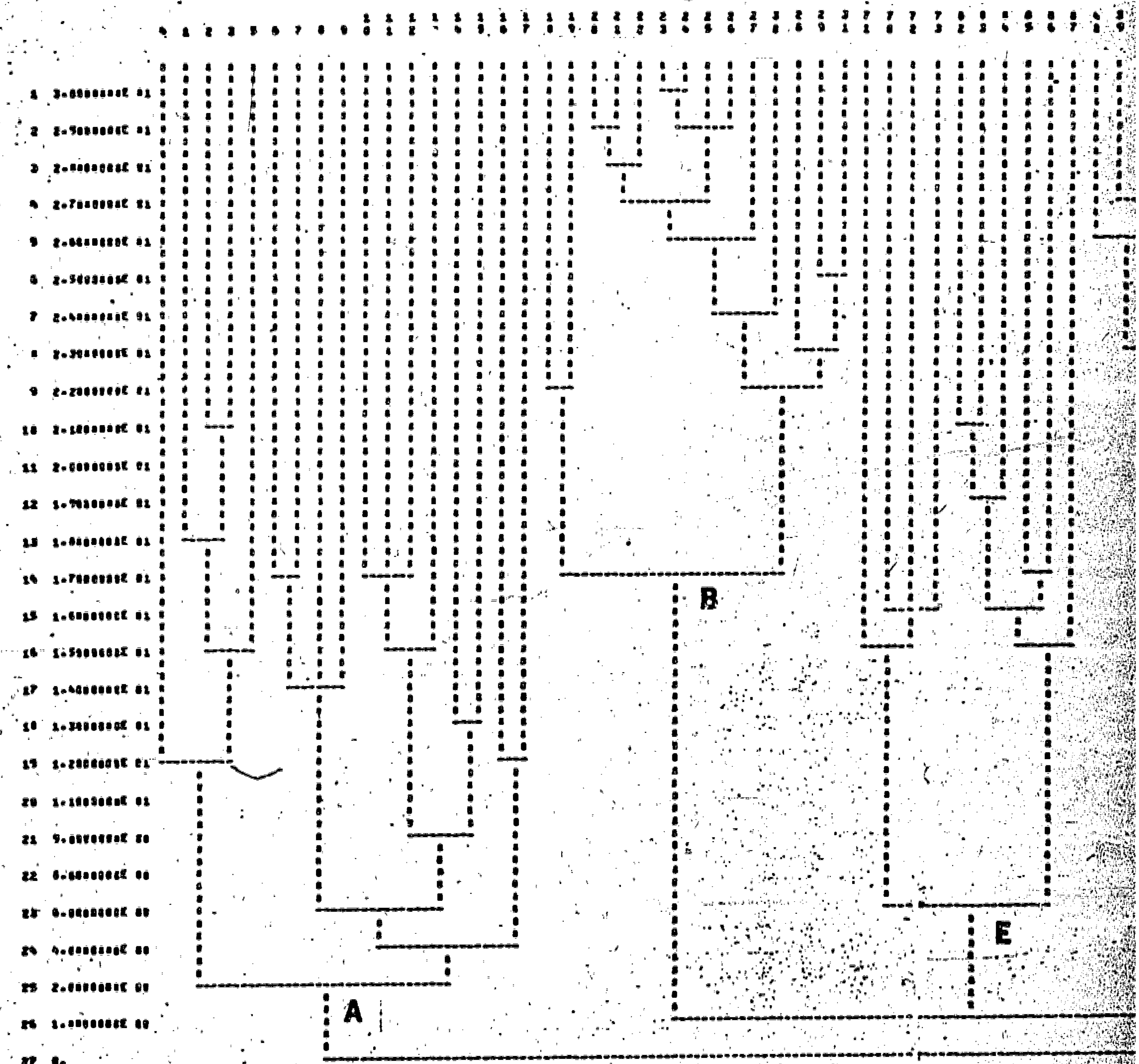


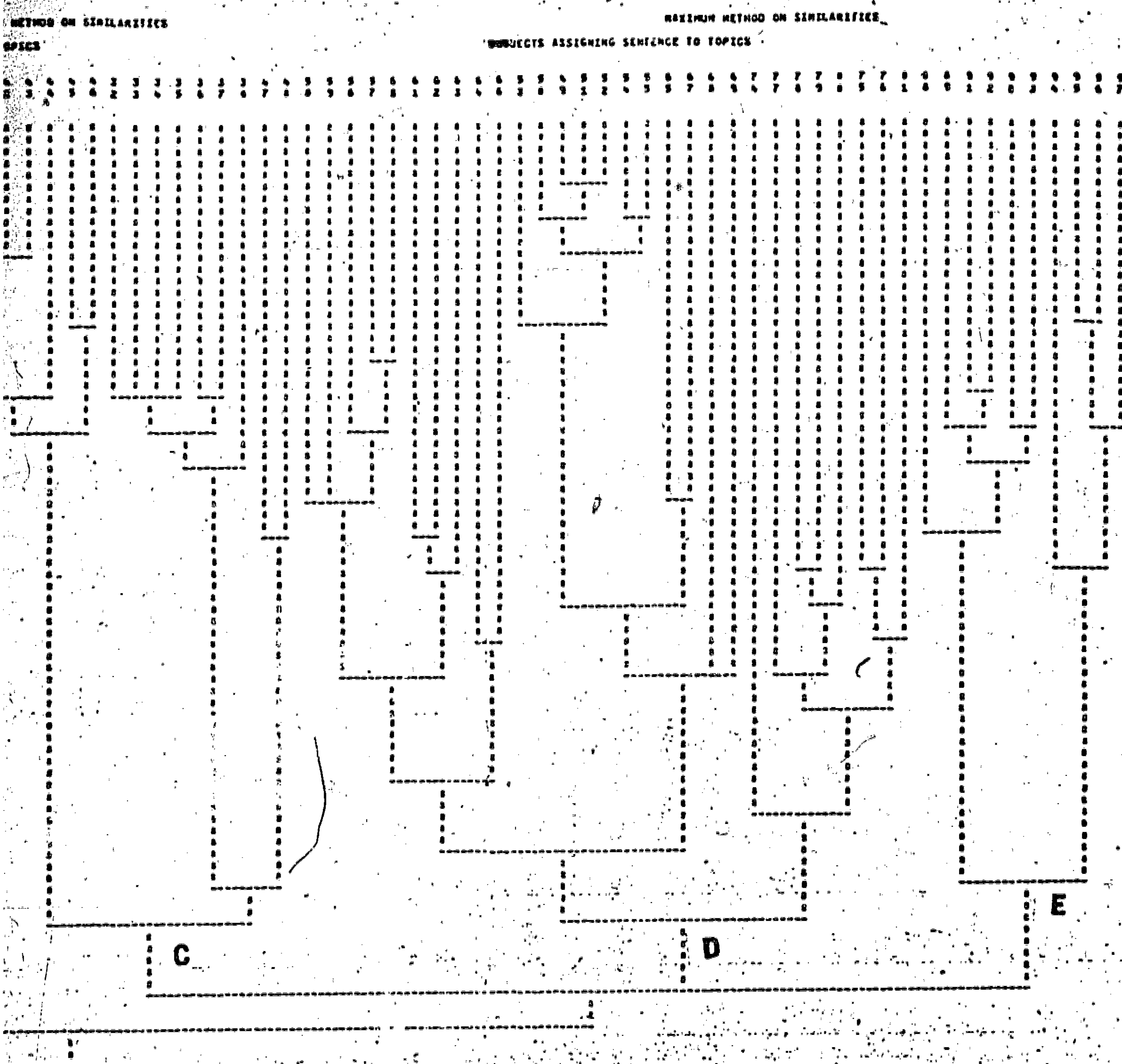
Fig. 1

GOP

MAXIMUM METHOD ON SIMILARITIES

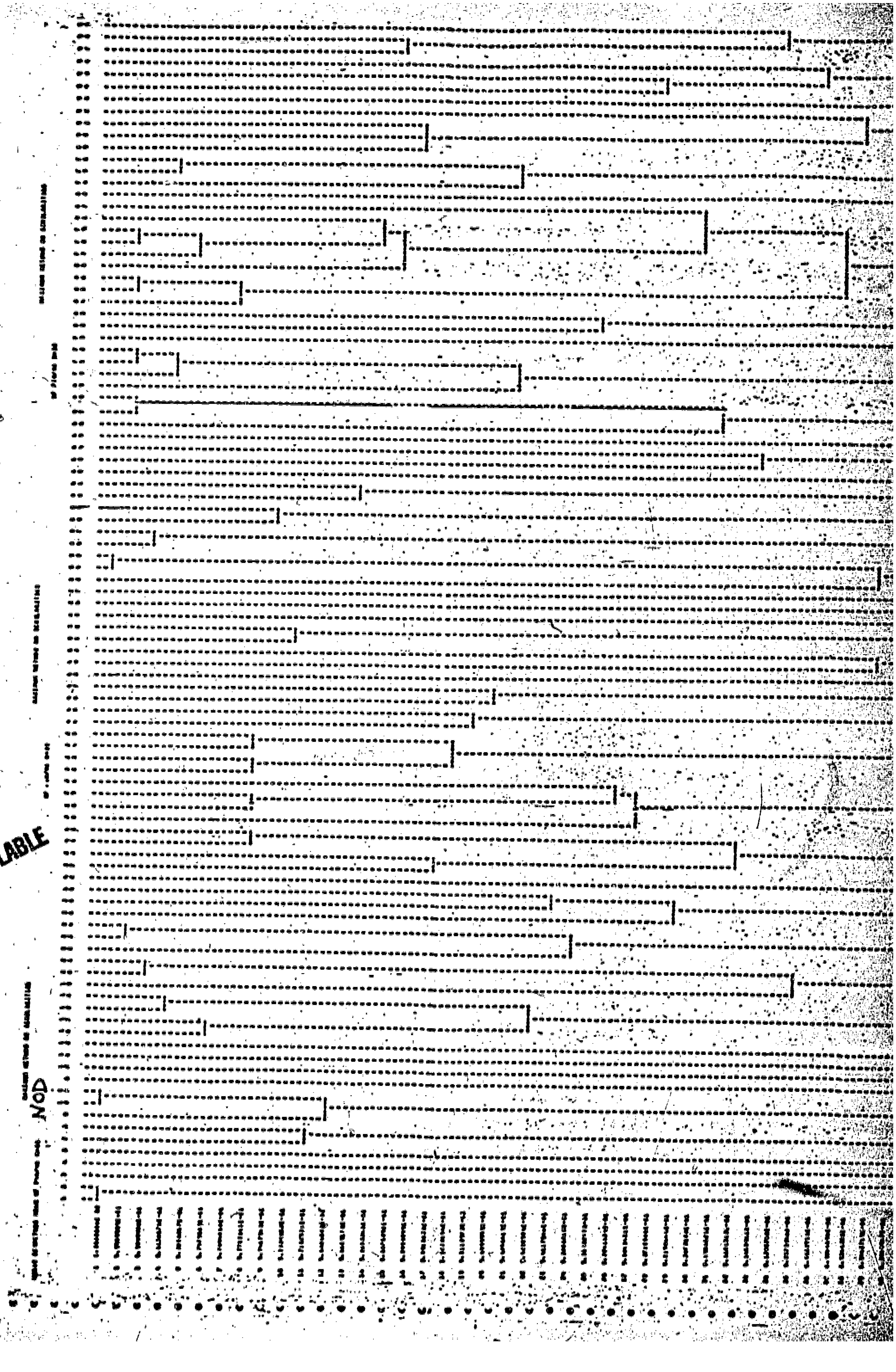
TOPIC 1 AND 2 GOP - SUBJECTS ASSIGNING SENTENCE TO TOPICS N=30





BEST COPY AVAILABLE

Fig. 2



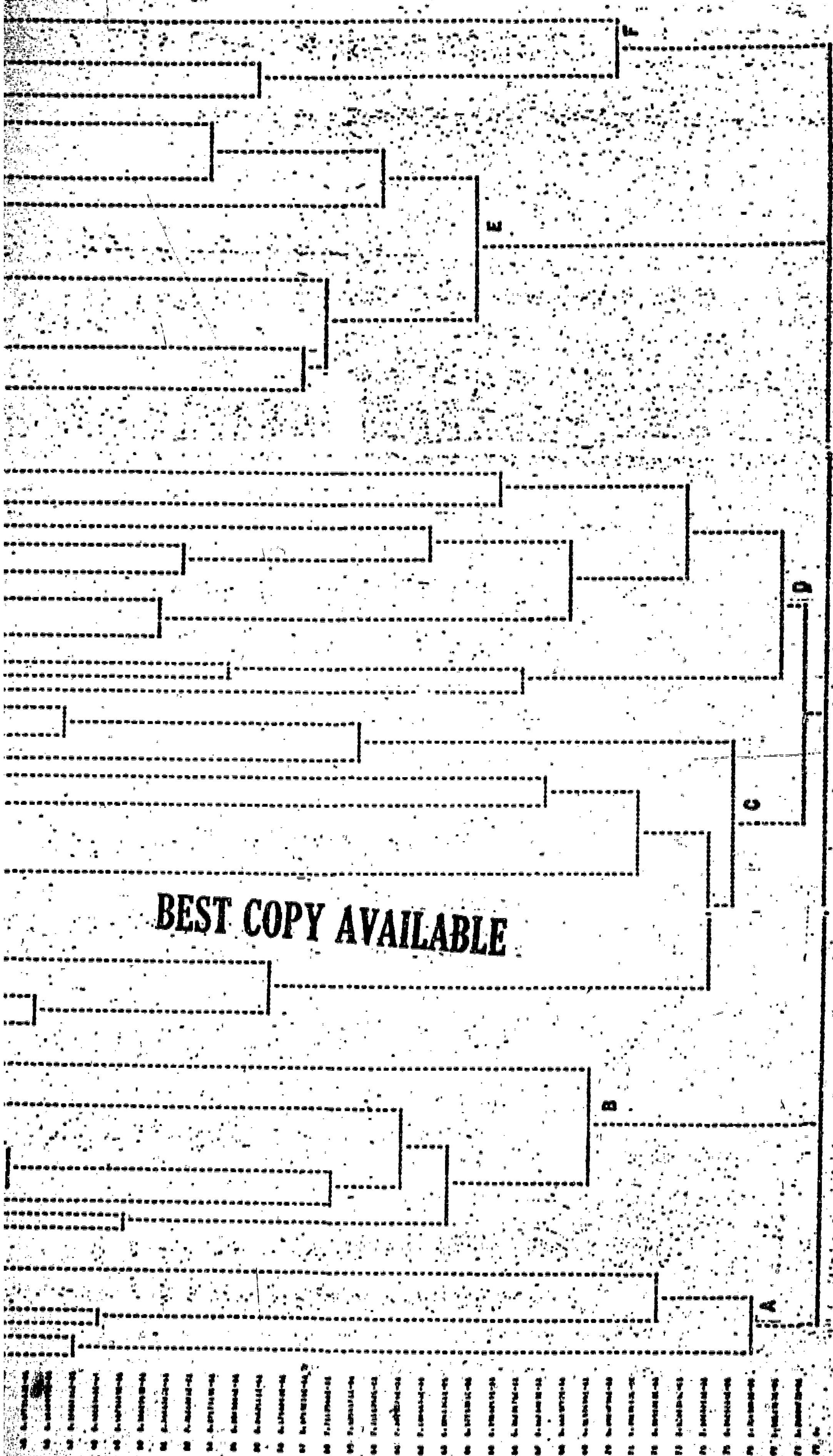
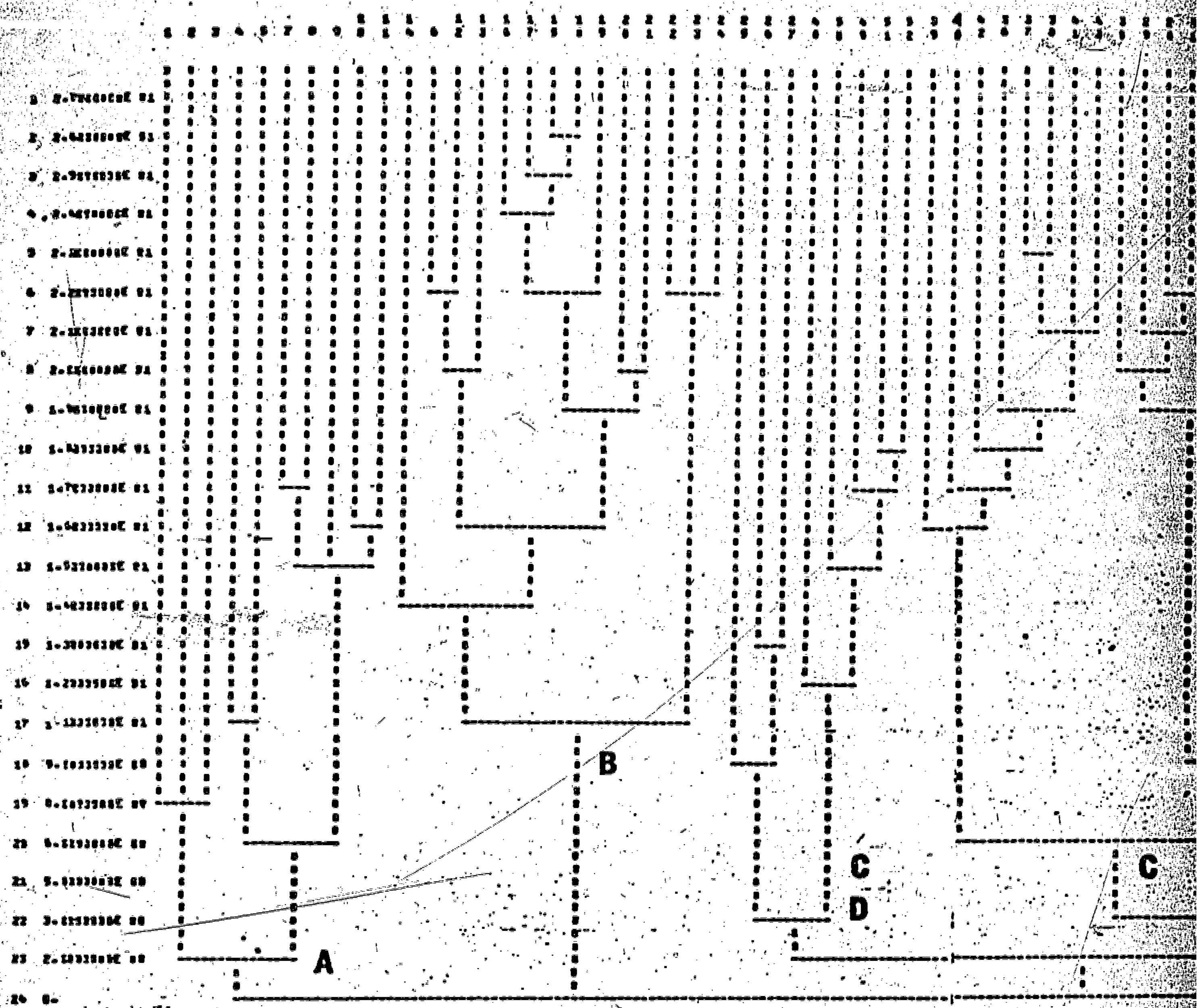


Fig. 3

PAGE 1 OF 3 PAGES NOD

MAXIMUM METHOD ON SIMILARITIES

PIC 1 AND 2 NOD - SUBJECTS ASSIGNING SENTENCE TO TOPICS NOD



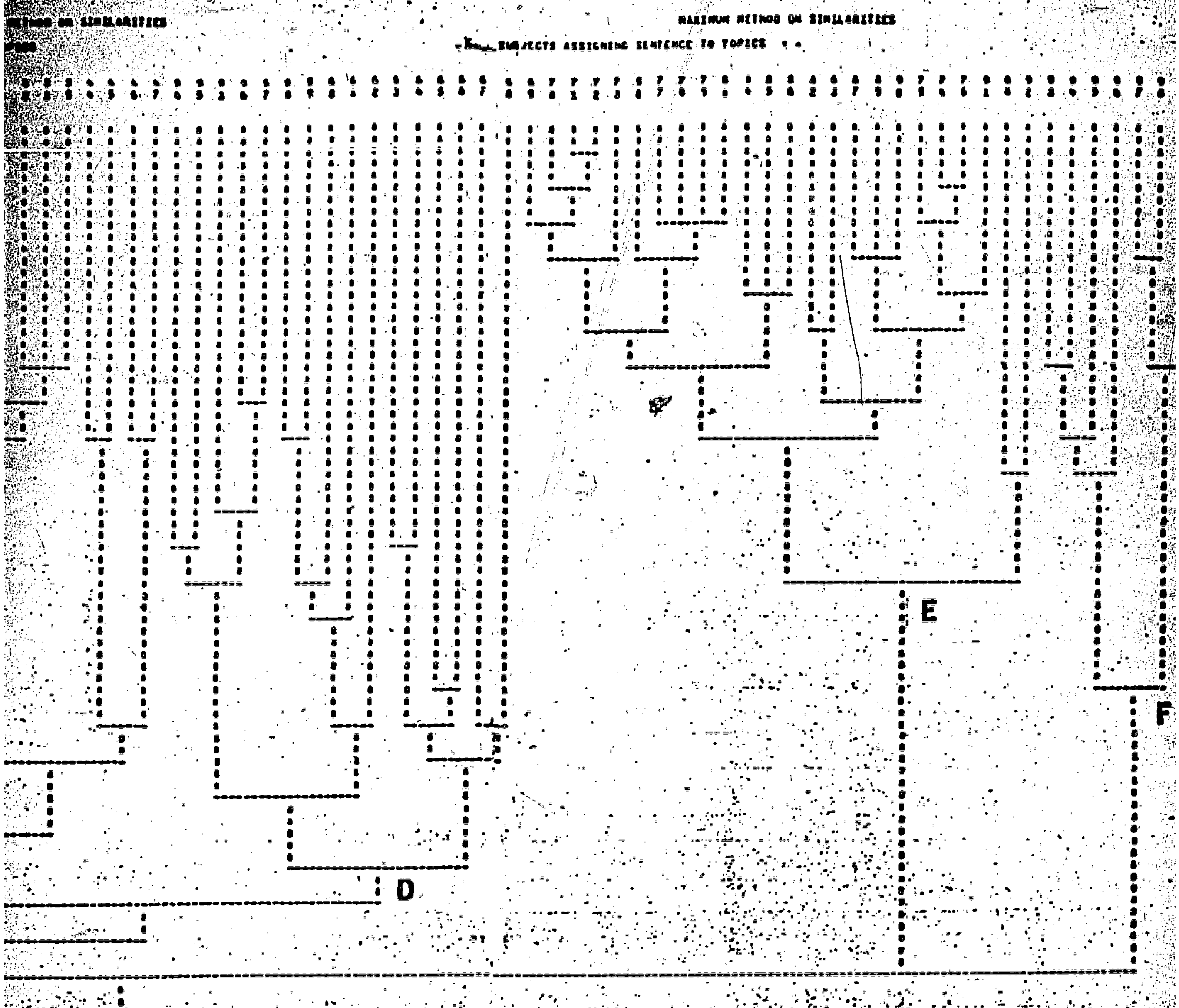
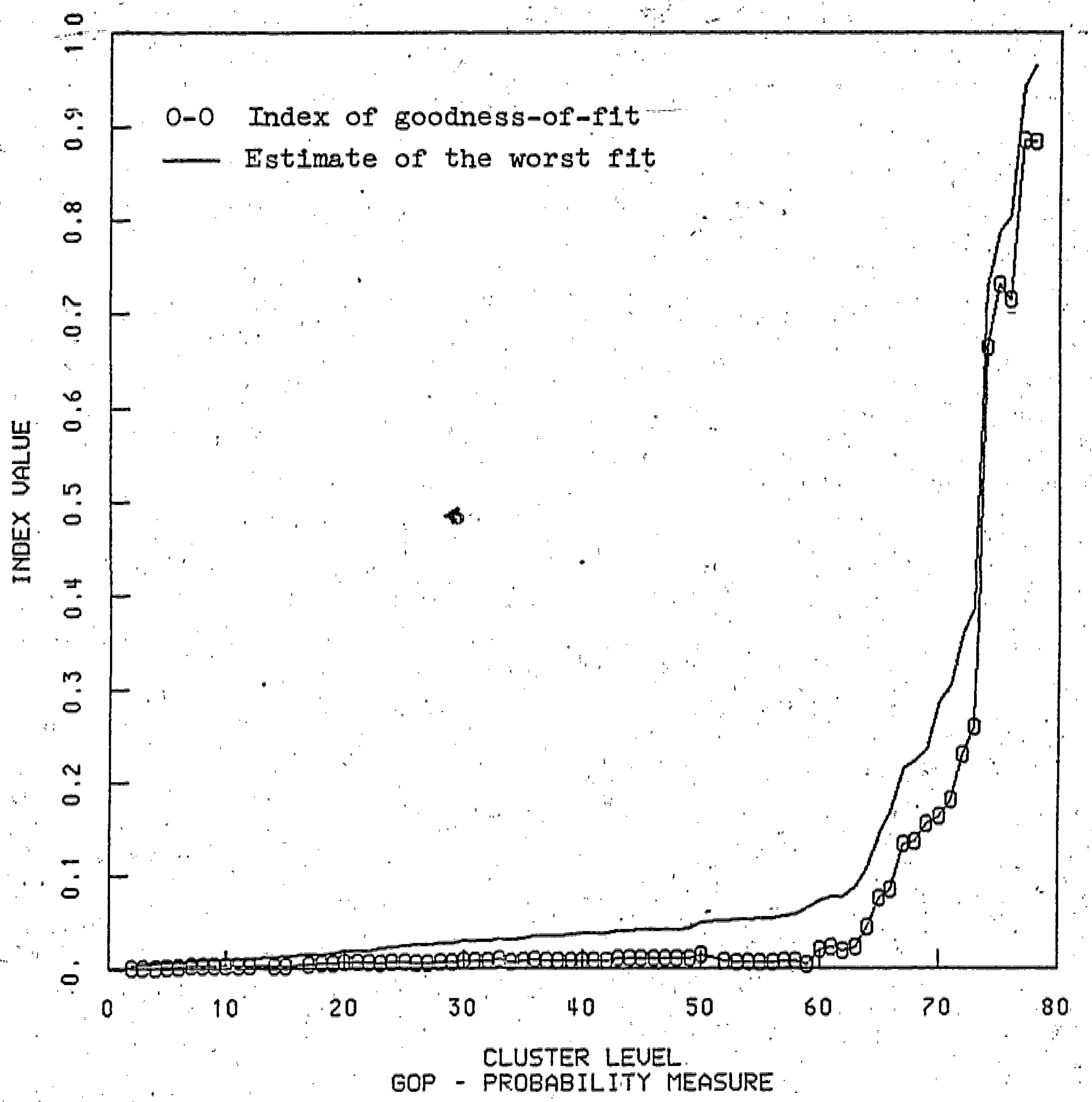

MAXIMUM METHOD ON SIMILARITIES
SUBJECTS ASSIGNING SENTENCE TO TOPICS
D
E
F

Fig. 4



O-O Index of goodness-of-fit
—— Estimate of the worst fit
INDEX VALUE
0. 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1.0
0 10 20 30 40 50 60 70 80
CLUSTER LEVEL
GOP - PROBABILITY MEASURE

Fig. 5

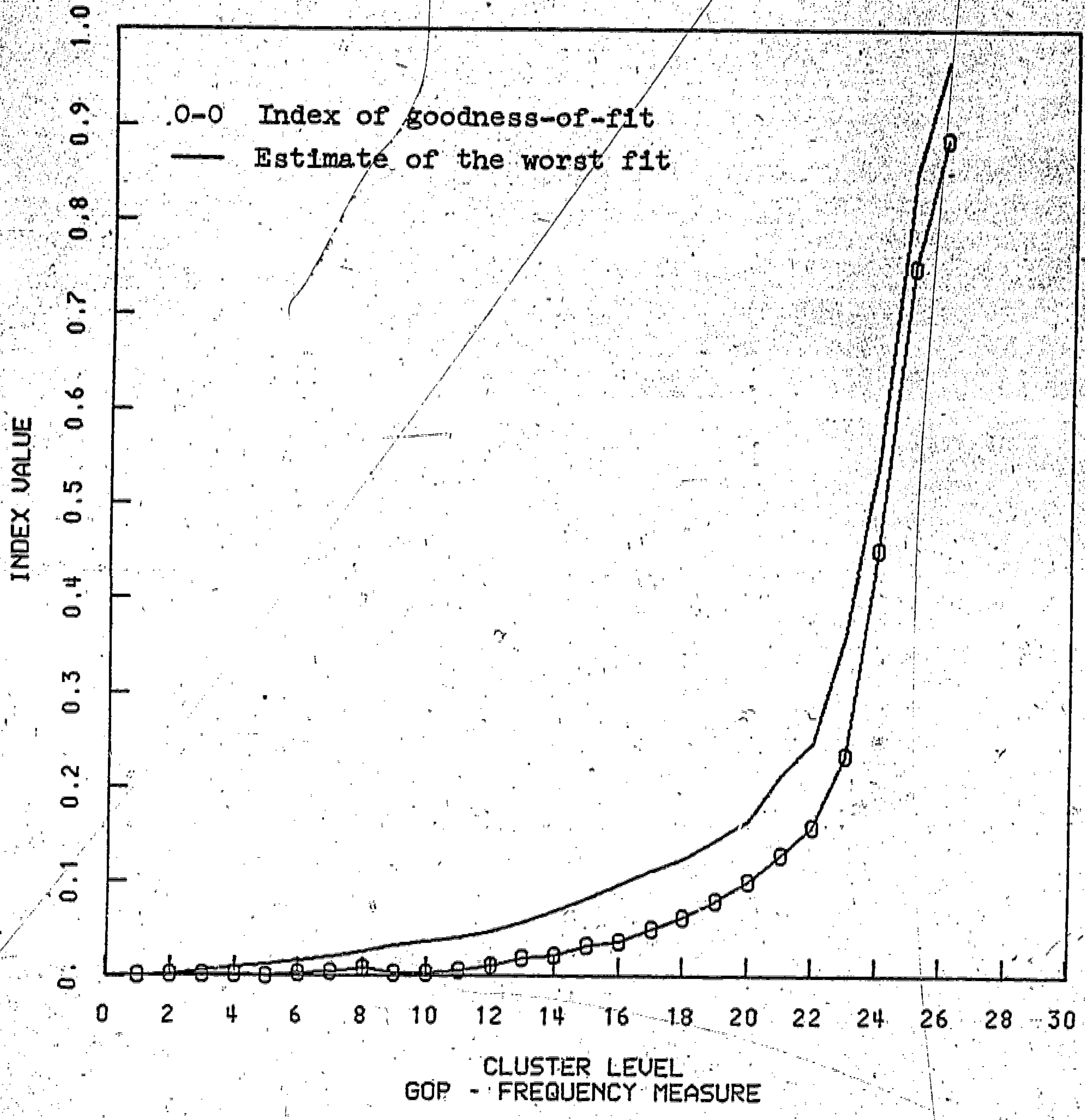
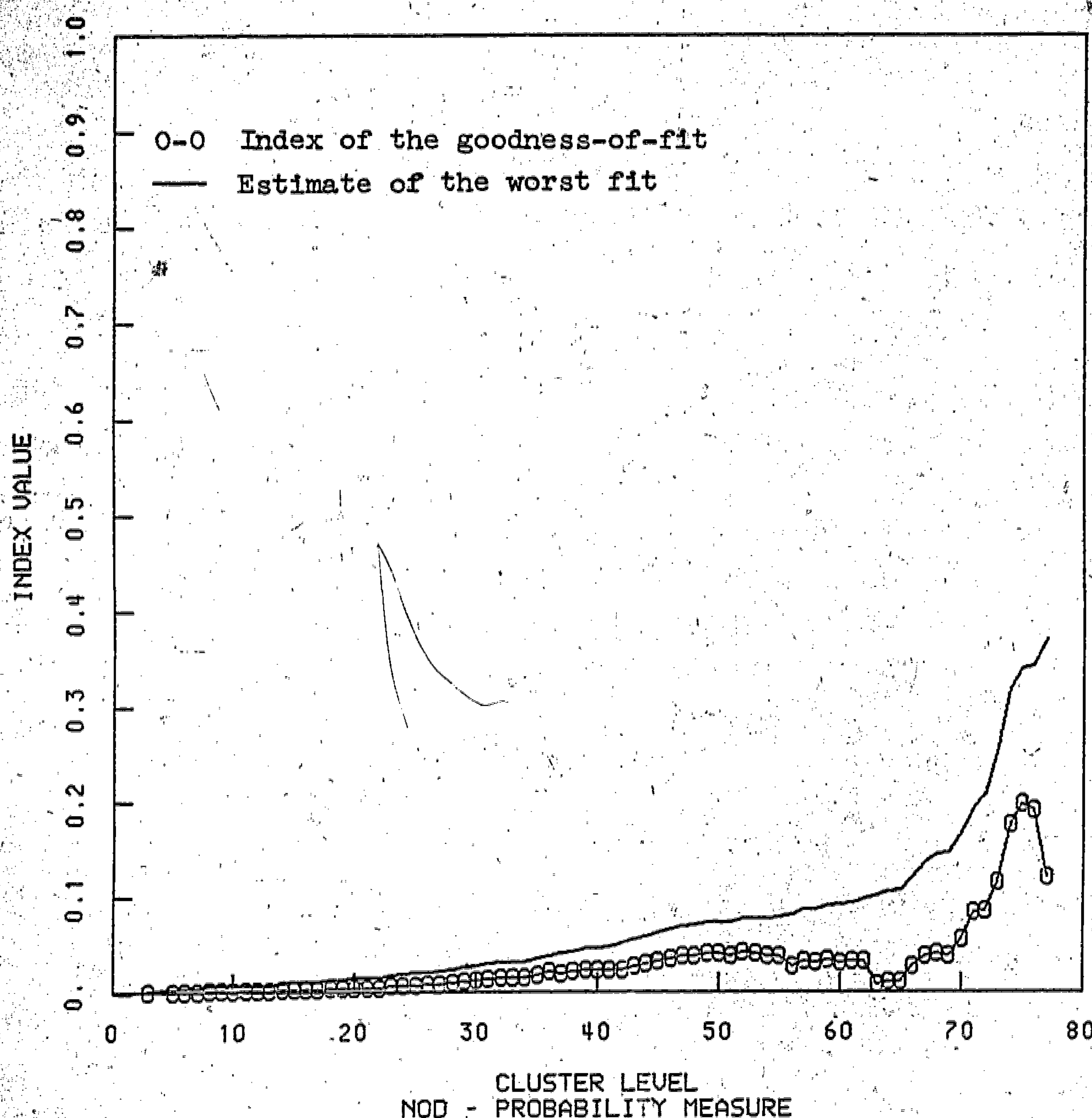

O-O Index of goodness-of-fit
Estimate of the worst fit
INDEX VALUE
0. 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1.0
0 2 4 6 8 10 12 14 16 18 20 22 24 26 28 30
CLUSTER LEVEL
GOP - FREQUENCY MEASURE

Fig. 6



O-O Index of the goodness-of-fit
Estimate of the worst fit
INDEX VALUE
0.
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0
0
10
20
30
40
50
60
70
80
CLUSTER LEVEL
NOD - PROBABILITY MEASURE

Fig. 7

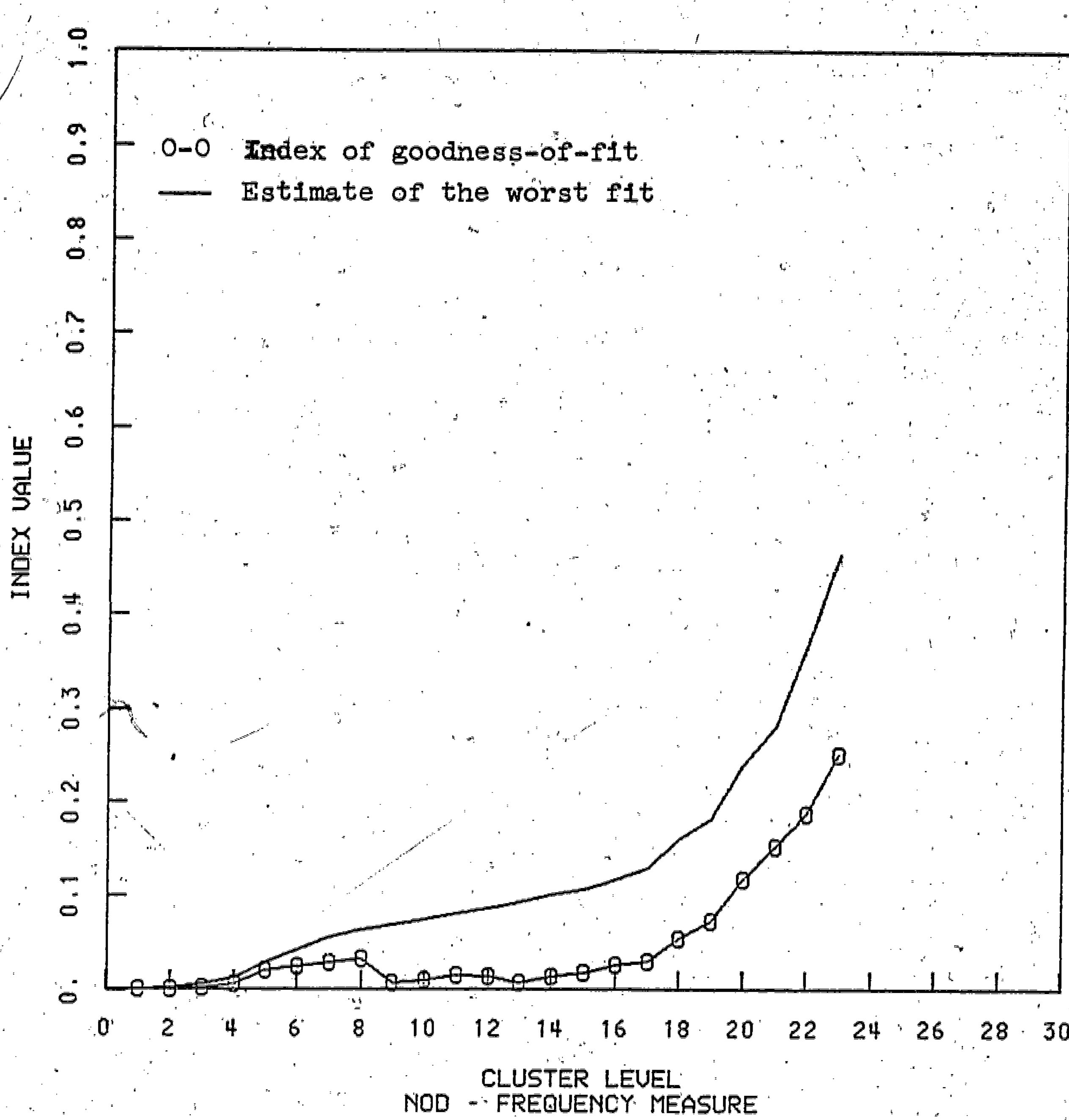

INDEX VALUE
0. 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1.0
O-O Index of goodness-of-fit
—— Estimate of the worst fit
0 2 4 6 8 10 12 14 16 18 20 22 24 26 28 30
CLUSTER LEVEL
NOD - FREQUENCY MEASURE

Fig. 8