RELATEDNESS OF MEANING AND HIERARCHY OF STRUCTURE

Andrej BEKES

0. INTRODUCTION

In this paper I would like to reexamine Minami's (1974) theory of levels in compound sentence in light of the concept of RELATEDNESS, a semantic concept which is possible to estimate quantitatively, as it will be shown in Chapter 2.

Minami, examining the structure of compound sentences in Japanese, arrived at four hierarchical levels of clauses, which he named A, B, C, and D respectively. Level D can only be occupied by the main clause, while some other clause on a lower level can always appear as a dependent clause embedded in a clause belonging to the same or a higher level, with D being the highest. Minami also showed that certain kinds of conjunction particles /setsuzoku-joshi/ can only connect clauses belonging to certain levels. The lowest are those operating at level A, allowing only clauses from this level, while those operating at higher levels may connect clauses from the same or lower levels. While stressing that his four levels are only a formal classification of Japanese dependent clauses, Minami also adds that these levels are involved at the making of a sentence and expressing different language functions encoded in them. These correspond roughly to what Hayashi (1960) distinguishes, from the innermost layer towards the outermost, as 'description' /byojutsu/, 'evaluation' /handan/, 'expression' /hyoshutsu/, and 'communication' /dentatsu/ .

But revealing as it is, Minami's theory has some limitations, the main being perhaps that as basically a formal classification it tells nothing about the dynamics of how clauses are combined into bigger units in an actual discourse.

I am here attempting first to objectively estasblish meaning relationship between clauses in a certain discourse (=relatedness) and then to examine the dynamics of its realisations in linguistic forms.

1. BASIC CONCEPTS

First I will introduce some basic concepts that will be used later on.

1. 1. Communication, text, discourse

Following Brown & Yule (1983, p.26) I will consider DISCOURSE as

> "the record of a dynamic process in which language was used as an instrument of communication in a context by a speaker / writer (here PRODUCER) to express meanings and achieve intentions".

Discourse defined as above is seen as a part of communicative process in which the following concepts are involved:

CONTEXT, comprising at least the PARTICIPANTS OF COMMUNICATION, i.e. the PRODUCER (speaker / writer), RECEIVER (hearer / reader), their DISCOURSE until that moment, and SETTING.

TEXT' will denote producer's hypothetical INTENDED MEANING and TEXT" receiver's INTERPRETED MEANING. The process leading from text' to DISCOURSE will be called REALISATION (encoding), and the one leading from discourse to text" INTERPRETATION (decoding). Here context with its constituents is an observable entity while concepts text' and text" are unobservable, conceptual entities. Figure 1 below shows how these concepts are interrelated.

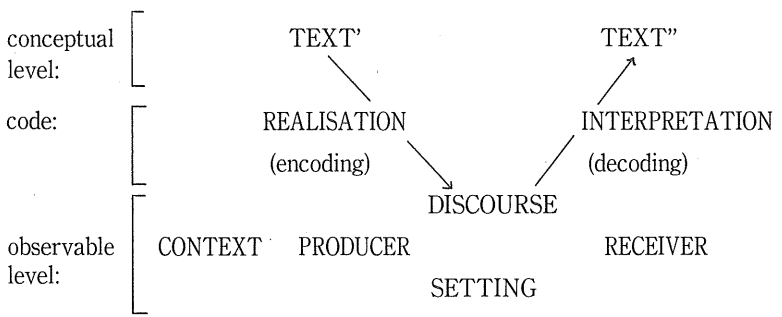


Figure 1.

The object of analysis can only be entities on observable level, while entities on conceptual level can be arrived at only indirectly through the observation of the former and are always of essentially hypothetical character.

1. 2. Realisation and motivation for sentence

As a working hypothesis about realisation of discourse I will take a model developed by Chafe (1977, 1980). Chafe hypothesizes that verbalisation, i.e. realisation as discourse goes through several stages. First, during SCHEMATISATION basically unstructured knowledge is subdivided in smaller units, CHUNKS which are then ordered. Next, suitably small chunks are subjected to FRAMING, i.e. factoring out of chunks individual entities characterised by particularity in space. Different framings of the same chunk are possible. At last, during CATEGORISATION each entity from some FRAME (a framed chunk) is encoded in language.

Chafe (1980) further hypothetises about the nature of chunks and their verbalisation and relates it to FOCI OF CONSCIOUSNESS, that are scanning the memory. The size of the smallest unit verbalised at one time, INFORMATION UNIT (i.u.) marked in discourse by intonation, pauses and syntactical characteristics / roughly a clause /, is determined by the amount of information the consciousness is able to process at a given time. When there is a piece of knowledge, that can not be dealt with within one focus of consciousness, it must be scanned by several foci. Such a piece of knowledge is called a CENTER OF INTEREST, manifested in spoken discourse by falling intonation, showing the completion of scanning. The format in which these centers of interest are expressed in discourse roughly corresponds to a sentence.

1. 3. Cohesion

Term cohesion is used differently by different authors. Here, partly following Halliday & Hasan (1976) and De Beaugrande & Dressler (1981) I will regard

cohesion as concept deeply related with interpretative aspect of communication.

This is so, because for producer concepts and relations he is trying to express are obvious, while receiver has yet to discover relations between different parts of discourse. Cohesion will be here defined so as to provide a unified view of this proccess.

1. 3. 1 Definition

COHESION will be a cover term for the following concepts and relations among them as seen from the receiver's point of view:

COHESIVE RELATION is a semantic relation between some pair of elements in discourse which is a result of successfully established relationship between a member of the pair called PRESUPPOSING and the other called PRESUPPOSED, with the help of COHESIVE MEANS.

Here, by COHESIVE MEANS are meant linguistic forms available to producer to signal with them cohesive relations, and the PRESUPPOSING ELEMENT is an element of discourse whose interpretation depends on some other element in that discourse, the PRESUPPOSED ELEMENT.

Such a definition of cohesion implies its inherent probabilistic character. Establishing cohesive relationships between different elements of discourse can never be absolutely reliable, it happens with biggger or smaller PROBABILITY, which can in turn be employed as a MEASURE OF STRENGTH of cohesive relations. Such a definition of strength of cohesive relations reveals that the difference between the three types of cohesive means, i.e. LEXICAL, SYNTACTICAL and INTERSENTENTIAL is in this respect a difference of degree and not of kind. It stems from the difference of cohesive relations that are signaled by each of them. Their probabilistic character is clear as was seen also from interpretation of time and place adverbials in compound sentences in one of my paraphrasing protocols (PROTOCOL "F › T", No.1).

1.3.2 Cohesion as means of organising discourse

From the receiver's point of view cohesion provides him with clues how to connect discrete elements of discourse into a meaningful whole. Producer, if he wants to make himself intelligible, must provide such clues. Here is an example of a variety of possible cohesive means employed by different people paraphrasing the same part of input discourse (=... otoko ga kubi o tsutte sindeiru... /a man hanging and being dead/):

(1)a ...otoko no kubitsurishitai.../man's 'hang-corpse'/ (lexical)

b Otoko ga kubi o tsutte sinde ita. /A man was hanging dead./ (syntactical)

c Otoko ga kubi o tsutta. Sosite sinda. /A man hanged himself. And then died./ (intersentential)

(PROTOCOL "F > T", No. 1)

Example above shows only a fraction of possibilities available to producer. Sytactical means are only one among the possibilities in the register that ranges between lexical at one end intersentential at the other. These three types of means form a hierarchy regarding the range within which they operate. Lexical means operate within a word, syntactical within a sentence and intersentential between sentences.

The means by which producer has to signal cohesion are at the same time the means to build up the discourse. Lexical means employ the inner organisation of a word, syntactical show how words are organised into groups and clauses and also how these into sentences. The intersentential means can show how different sentences are grouped further.

1.4 T-mode and F-mode

For this purpose I will rely on paraphrases, where input and output will be discourses produced in two distinct modes. One, T-mode will rely heavily on intersentential cohesive means and as little as possible on syntactical means. The other, F-mode will rely heavily on syntactical means. So in T-mode will

prevail short, simple sentences, while in F-mode will prevail longer, complex sentences.

This distinction will be applied to written mode of production, where usage of complex sentences is characteristic of genres like news reports etc.

1.5 Decompositions of discourse into base propositions

In order to make comparisons between input and its paraphrases and also among individual paraphrases possible discourses I propose decomposition of discourse into BASE PROPOSITIONS (b.p.-s). This will not be a conceptual representation (cf Kintsch, 1974) but an inventory of written counterparts of i.u.-s. contained in some discourse. I chose it because cognitive content of such b.p.-s seems to be more stable during paraphrasing than expression of relations among them. The latter also coincides in many cases with cohesive means, here the object of study .

To get a decomposition I used written paraphrase with output in T-mode. With a large number (here about 40) of paraphrases one gets a fair lower limit for a length of such unit, by following a procedure of extracting the shortest simple sentences from paraphrases, and comparing them among themselves and with input. Such simple sentences are considered equivalent among themselves and with corresponding part of input if they have equivalent cognitive content. In terms of Chafe's (1977) model this would correspond to equivalence at the level of framing. Roughly then, a b.p. will be a subset of all simple sentences from paraphrases, equivalent among themselves and with some part of input, and all such b.p.-s will form the set of b.p.-s of a discourse (relative to a certain paraphrase). By such a procedure is assured that decomposition is relatively independent of analyst and that the size of b.p.-s is not arbitrary but expresses an average agreement among native speakers, what would in such a context still be a sensible lower limit of a unit for a meaningful transfer of information. As such, b.p.-s can be thought of as an abstraction of a written counterpart of Chafe's (1980) idea units. When used in actual discourse, any actual manifestation of a b.p. will be as a shorthand also called

a b.p.

2. RELATEDNESS OF B.P.-S IN DISCOURSE

In this chapter I would like to introduce RELATEDNESS between b.p.-s of a discourse as a quantitatively measurable semantic concept, and try to analyse discourse structure in terms of it.

2.1 Paraphrase experiment

2.1.1 Purpose

Minami's hierarchy of conjunctive particles and the three-level hierarchy of cohesive means in Ch. 1 were hierarchies in terms of combinatory possibilities, thus representing a kind of system of brackets which while signaling relations between different parts, also bracket parts of discourse in what seem to be more or less related blocks. To verify this in the case of b.p.-s it is necessary to establish the concept of relatedness among b.p.-s independently and then see how discourse is structured in the light of this concept and what is the connection between it and cohesive means. The former will be done here and the latter in Ch. 3.

2.1.2 Method

For this purpose I made a paraphrase experiment (= PROTOCOL "T > F" No.2) where in the input b.p.-s are clearly distinguishable and the output will yield sufficiently homogeneous set of discourses. With input in T-mode and its sentences actualised b.p.-s, and with the target output in F-mode this requirement can be met. To assure higher reliability of data, the experimental task was set in a classroom as a quasi communicative activity. The task was to rewrite the T-mode input, based on a news article, as a news article for some newspaper. Thus the experimental task did not require mechanic responses but creative work. The relation between input and output is shown on Figure 2 below.

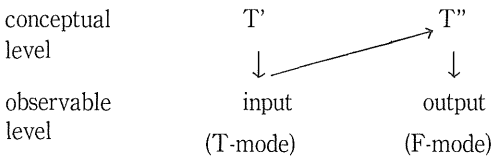


Figure 2

I hypothesize that during the task first, the text world model T" corresponding to T-mode input was reached. This next served as departing point for the production of F-mode paraphrase. At the same time the written form of input enabled testees to reuse the original categorisations in the output. Thus in majority of paraphrases input b.p.-s still retained coding comparable to that in input. Experiment analysed here was performed with 46 first year high school pupils aged between 15 and 16. The input used is given in App. 1.

2.2 Interpretation of data

2.2.1 Relatedness

Direct observation of conceptual level is impossible. The question here is how to determine meaning relatedness among b.p.-s from input. The first step in this direction is to ask how strongly these b.p.-s are related. As an implication of Chafe's hypothesis about the nature of format called sentence (see 1.2), it is possible to suppose that the fact that a pair of input b.p.-s cooccurs in the same sentence in some paraphrase can be thought of as an act of a testee, interpreting such a pair of b.p.-s to be more related in meaning within the context of input discourse than some other pair not coocurring. The bigger the proportion of testees who thus relate a certain pair, the bigger can be considerd the strength of this relatedness. This in turn can be taken as an observable measure for the strength of relatedness among b.p.-s in the input discourse and will be called RELATEDNESS. These proportions for each pair of b.p.-s can be assembled into a matrix, which will be called RELATEDNESS MATRIX for b.p.-s in a certain discourse. Example of such a ma-

trix for the experiment analysed here (PROTOCOL "T › F" No.2) is given in Appendix 2.

2.2.2 Cluster analysis of relatedness matrix

Relatedness matrix reflects the way how strongly different b.p.-s from input are thought of to be related with each other. But it is clear that such relatedness directly relates a much wider field of b.p.-s than a theory of immediate constituents could predict. For it seems that also b.p.-s that are not directly constituents are interpreted to be related. Further insight into relatedness among b.p.-s over the whole input discourse can be achieved by cluster analysis of their relatedness matrix. This is a discovery procedure which helps discover affinities that would remain hidden, if data remained presented in the matrix form. For clustering will be used hierarchical single method (cf Anderberg, 1973 Ch. 6.). This method groups together those entities from the similarity matrix that are most closely associated by the similarity measure, here provided by relatedness matrix itself. In the present case, input's b.p.-s will be grouped according to the strongest relatedness among them. Clustering diagram of the relatedness matrix (App.2) by this algorithm is given in App. 3.

2.2.3 Clustering analysis results

Clustering diagram (App.3, with s = 0.13 p = 0.90) shows how b.p.-s in input are grouped together according to the strength of their relatedness. Here the strength of link between two clusters is the strength of relatedness between their most strongly related members, forming the link via which the two clusters are linked into a bigger one.

Consider now the meaning content of these clusters. Obviously, at each level of clustering the contents of b.p.-s show high consistency. Indeed at 0.20 level shown by square brackets: [] , and 0.60 level - shown by round brackets: (), clusters and their subclusters are like this (numbers representing b.p.-s / App. 1/):

[1-9]: discovery ; (1-3): suicide, (4,5): discovery, (6): report, (7-9): investigation.

[10-17]: family's narration; (10, 11): accident, (12): aftermath, (13-16): O.'s action, (17): narrating.

[18-25]: police view; (18,19): facts, (20-24): O.'s assumed action, (25): guessing.

It is possible to assume that clusters reflect the way, small chunks of knowledge realised in b.p.-s are groupped into bigger chunks in the hypothetical, though directly unobservable T' (see Fig.2). But since this reflection was achieved through testees' interpretation of input it belongs to T". This hypothesis will be used in the next chapter.

3. REALISATION OF RELATEDNESS IN DISCOURSE

3.1 Organisation of b.p.-s in discourse

In the paraphrase output there can be discerned several patterns how b.p.-s were organised in discourse. This different patterning would in Chafe's (1977) terms correspond to schematisation, and also shows that paraphrasing was not just a mechanical excercise. The patterns were:

a) Organisation along time axis which seems to be the easiest principle to organise any narrative of events. In present experiment there was one such case.

b) Organisation around the bearers of different actions or events. This pattern was more common and looks like a list of who did what. E.g. paraphrase T-41 (PROTOCOL "T > F" No.2) regrouped b.p.-s concerning the victim like this: ...(10,11), (12,18,19), (17,13-16), ... Compared to pattern a) this pattern generally shows a higher degree of integration on b.p.-s into sentences.

c) Preserved schematisation of input, with order of b.p.-s preserved at least as far as the intermediate level of clustering (~ 0.50). This pattern is dominant, with two possible reasons why. One could be the principle of least effort and the other a choice of an efficient ordering for transmission of news.

Judgeing from the task, the latter seems to be more probable, because input order is based on the original news text input is based on. In this pattern paraphrasing activity more than with schematisation seems to be concerned with encoding of meaning relations between b.p.-s in F-mode, judgeing from the variety of forms observed in the output.

3.2 Signaling relatedness hierarchy with syntactic means

In this section I would like to examine, how relatedness hierarchy among b.p.-s was signaled by syntactic means, comparing realisations among output discourses. For this purpose it is necessary for the material to be homogeneous. That is the ordering of b.p.-s has to be similar enough for the comparison to be possible. With such a sample it can be considered, that individual discourses in it are eqivalent as paraphrases down to the level of framing. The only difference would be in how the meaning relations between different b.p.-s were encoded, which is what I would like to investigate. Thus pattern c) from the previous section seems to fit this purpose best.

3.2.1 Relatedness hierarchy in direct relation to syntax of output.

First I would like to see whether relatedness of b.p.-s has any connection with the meaning relations between such b.p.-s being directly encoded by syntactic means in the output. I will consider separately b.p.-s with frequency of cooccurrence higher than 0.50 and separately those with lower frequencies. The first group will be further subdivided in three levels, corresponding to three significant levels of clustering (see p.9 and App. 3). The three groups are divided with frequencies as this:

H: 0.87 ~ 1.00 ; M: 0.74 ~ 0.87 ; L: 0.50 ~ 0.74

The next step will be to find how the relations between b.p.-s were treated syntactically for b.p.-s in each group over the whole population (=group c)). The results are shown in the Table 1 , App.4. Here too, numbers correspond to b.p.-3 (see p.25) and brackets to cluster boundaries. So, for example (7, (8, 9)) shows that b.p.7 and cluster (8,9) form a cluster with their frequency of

cooccurrence falling within the subgroup M. On the right side of the Table 1 are first given for each group, i.e. H, M, L, observed proportions of how many of related pairs of b.p.-s (or clusters and b.p.-s) were connected directly by syntactical means by some individual. Right to it is the frequency of such observed proportions over the whole sample. In group H each pair of propositions is connected directly, while in M and L are counted connections between any members of subclusters.

From the Table 1 is apparent that the trend toward direct syntactical connecting between b.p.-s is most clearly expressed in terminal, highly related clusters consisting of single b.p.-s. But the lower the relatedness (from H towards L) the more are frequences dispersed among lower proportions as well.

And now about the b.p.-s with frequencies of cooccurrence lower than 0.35. For this case the results are shown in the Table 2 (App.4). Conventions are the same as before, except that round brackets show subclusters and square ones clusters. In this table the proportions show, how many of weakly related clusters of b.p.-s among the total had the fact explicitly marked with a sentence boundary (= period). On their right are given the frequencies for each proportion, over the whole sample. From the Table 2 it seems that there is no clear trend observed in the use of period to mark the boundaries among the weakly related clusters.

Thus, judgeing from both Table 1 and 2 it seems that the high relatedness between b.p.-s as expressed in the frequency of their cooccurrences within the same sentence is more likely to be explicitly encoded in syntactic relations among the b.p,-s concerned. Contrary to this, there is no such trend discernible with the explicit encoding of weak relatedness. From this follows that signaling explicitly the relatedness has priority over showing unrelatedness (by for example sentence boundaries).

3.2.2 Relatednesss and cohesive means

From the results of previous section it is possible to conclude that the degree of explicitness in marking relatedness is discourse-relative. This is why

such marking was less and less uniform for lower degrees of relatedness between b.p.-s. Such relativity can be seen in light of the concept of a register of cohesive means (p.6). Each testee chose the means, according to his ability and judgement, from the spectrum available to him.

Here I would like to see whether the selection of particular means is connected with the degree of relatedness, and if it is, then how it is. For this purpose I will consider sequences of b.p.-s or clusters of b.p.-s, where the middle element in the sequence will be more strongly related with one element (=H) and less strongly with the other element (=L).

I will call signals of meaning relations between b.p.-s (conjunction particles, head and modifying clause etc), CONNECTIVES. A connective revealing more about relationship between the units it connects, will be called more EXPLICIT. Obviously, sentence boundary + adjacency of sentences is the least explicit among the connectives I am concerned with here.

The results are presented in Tables 3 and 4. Table 3 is organised so that the frequency of connectives on its right is seen in terms of connectives on its left. Connectives on the left begin with the least explicit, adjacency of sentences, with plain gerund (= ren'yokei) and 'te' -gerund (= -te setsuzoku) folowing in the order of explicitness. This is also the order corresponding to Minami's levels, D, B, and B.

This order can be obtained also in another way, by looking for high and low frequences of connectives on the right, cooccuring with those on the left. Beginning with adjacency, connectives on the left are ordered so that their counterparts on the right that have appeared on the left or have relatively higher frequencies come in first, followed by those that have not yet appeared and have systematically low frequencies. It can be observed that with the hierarchy of connectives that emerged in the table's left, for high relatedness encoded in weaker connectives, i.e. , those closer to Minami's level D, accompanying lower relatedness tends to be encoded in connnectives on the same level or level closer to level D.

In Table 4 connectives on the right will be given with their frequencies in

terms of connnectives on the left, ordered from bottom up, beginning with embedded relative clause as the supposedly most explicit connective and having other following in decreasing order of explicitness, which also coincides with the order of Minami's levels. It can be observed that connectives on the right belonging to lower levels exhibit systematicaly low frequencies. Or in other words, it can be concluded, that lower relatedness encoded in a connective from some level, is very probably followed by corresponding higher relatedness being encoded in connectives from the same or a level closer to level A on Minami's hierarchy.

3.3 Conclusion

From Tables 3 and 4 it was seen that Minami's levels seem to be in close connection with regularities concerning frequencies with which connectives were chosen, and also with the degree of relatedness between b.p.-s or their clusters. Actually there seems to be a weak monotony preserving to some extent the hier archy of relatedness between b.p.-s when their relations were encoded as the hierarchy of connectives. As a basis for the hierarchy of relatedness served the clustering diagram (App. 3) and for the hierarchy of connectives, Minami's hierarchy of levels.

Thus hierarchy of levels proposed by Minami is not just a classification, as he originally intended it to be, but is revealing more about the basic nature how syntax works on discourse level. It was seen, that the way how to organise b.p.-s into discourse is neither absolute nor uniform, but that nonetheless there seem to be some global principles governing it.

First, that encoding of meaning relations between b.p.-s is positive: high relatedness is high on priority to be encoded first with lower degrees coming later (Table 1 and 2, App. 4).

Second, encoding of high priority relations may begin anywhere on the scale of cohesive means, with the adjacency of sentences /sentence boundary/ being at the lowest end and merging into one lexical item at the highest end.

Third, this choice determines the register of cohesive means to be used wi-

thin a particular sentence (though the observations show the register to be rather uniform throughout the discouse). The register available for less related b.p.-s is the part between the level of the first chosen item and the bottom of the scale. Usage of other connectives is conditioned reccursively at each point of the previous choice, starting with the first, that of the highest degree of relatedness.

As one level of possible cohesive means such a role of syntax also leads to furher speculation about the organisation of intersentential cohesive means, a question which remains open to further investigation.

REFERENCES

BROWN, G. & YULE, G. (1983) *Discourse analysis*, Cambridge UP.

CHAFE, W.L. (1977) 'Creativity in verbalisation and its implications for the nature of stored knowledge', in FREEDLE, R.O. (ed.) (1977) *Discourse production and comprehension*, Ablex, Norwood, N.J.

---(1980) 'The deployment of consciousness', in CHAFE, W.L. (ed) (1980) *The pear stories*, Ablex, Norwood, N.J.

DE BEAUGRANDE, R. & DRESSLER, W. U. (1981) *Introduction to text linguistics*, Longman, London.

GIVÓN, T. (1979) *On understanding grammar*, AP, N. Y.

HALLIDAY, M. A. K. & HASAN, R. (1976) *Cohesion in English*, Longman, London.

HAYASHI, S. (1960) *Kihon bunkei no kenkyu*, Meiji tosho, Tokyo

KINTSCH, W. (1974) *The representation of meaning in memory*, Lawrence Erlbaum, Hillsdale, N. J.

MINAMI, F. (1974) *Gendainihongo no kozo*, Taishukan, Tokyo.

Appendix 1.

Input b. p. -s

(1) Sakujitsu no yugata no koto desu. /it happened last night/
(2) Kanagawa-ken X-shi Y-yama no chufuku de otoko ga kubi o tsutte imasita. /in K. pref. on Y. mountain there was a man hanging/
(3) Sono otoko wa sinde imasita. /that man was dead/
(4) Aru hito ga yamaimohori ni dekakemasita. /a man went to dig 'yamaimo' roots' ?/
(5) Sono hito ga sono shitai o mitsukemasita. /that man found the corpse/
(6) Sono koto o Z-keisatsusho ni todokemashita. /he informed Z police station about that./
(7) Keisatsu wa sore o shirabemashita. /the police investigated that/
(8) Shitai wa Aichi-ken W-shi no kaishain O. san deshita. /the corpse was O. from W. city in A. pref./
(9) Sore wa sebiro no neemu kara wakarimashita. /this was found from the name on the suit/
(10) Kyonen no kugatsu chonan ga W. -shinai de kotsujiko o okosimasita. /in sept. last year the elder son caused a traffic accident in W. city/
(11) Chonan wa aru onna no hito ni nikagetsu no jusho o owasemashita. /the elder sone made some woman suffer two month long heavy wounds./
(12) Onna no hito wa kyonen no kure ni shibo shimashita. /the woman died last year end/
(13) O. san wa sore o shirimashita. /O. learned about that/
(14) Kare wa sono mama kaisha wa sotai shimashita. /he immediately left the office/
(15) Ie ni mo kaerimasendeshita. /he didn't even return home/
(16) Yukue fumei ni narimashita. /he became missing/
(17) Kazoku ga soo hanashimashita. /family told this/
(18) Shikashi, kono onna no hito no shibo gen'in wa shinfuzen deshita.

／actually the cause of the woman's death was heart trouble／

(19) Jiko to wa chokusetsu kankei ga arimasendeshita. ／it had no connnection with the accident／

(20) Sorenanoni, O. san wa ko omoimashita. ／though O. thought like this／

(21) Jiko ga gen'in da, to. ／that the accident was the reason／

(22) Onna no hito ga sore de shibo shita, to. ／that the woman died because of it／

(23) O. san wa sekinin o kanjita. ／O. felt responsible／

(24) Jisatsu shimashita. ／he killed himself／

(25) Keisatsu de wa ijo no yoni miteimasu. ／the above is the police view／

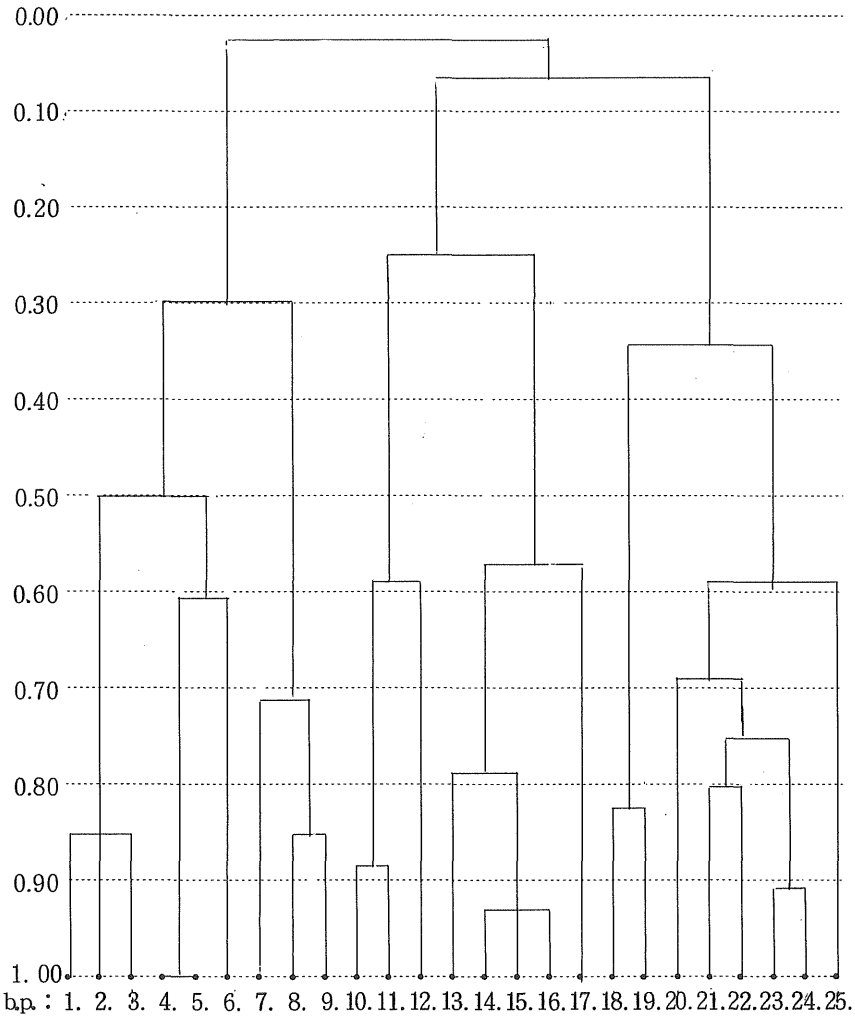
APPENDIX 2

2. Relatedness matrix

b.p	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
1	.87	.80	.41	.41	.24	.02	.02	.02																
2		.87	.48	.50	.26	.02	.02	.02																
3			.39	.41	.22	.04	.04	.04																
4				1.00	.59	.11	.04	.04																
5					.61	.09	.04	.04																
6						.30	.02	.02																
7							.76	.67	.11	.07	.04	.02	.02	.02	.02									
8								.87	.02	.02														
9									.02	.02														
10										.89	.59	.15	.09	.09	.09	.20	.02		.02	.02		.02	.02	
11											.57	.13	.09	.09	.09	.20	.02		.02	.02		.02	.02	
12												.24	.15	.15	.15	.20	.11	.07	.02	.02	.02	.02		
13													.78	.76	.78	.47	.02	.02		.02	.02	.02		
14														.91	.93	.57	.04	.02		.04	.02	.04		
15															.93	.55	.04	.02		.04	.02	.04		
16																.57	.04	.02		.04	.02	.04		
17																	.07	.07	.02	.04	.04	.04	.02	
18																		.83	.26	.30	.28	.28	.24	.17
19																			.22	.33	.30	.28	.26	.20
20																				.65	.61	.67	.63	.37
21																					.80	.76	.70	.48
22																						.67	.63	.46
23																							.91	.54
24																								.59

Appendix 3.

Clustering diagram ($s = 0.13$ $p > 0.90$)



Appendix 4.

"High level of clustering and individual usage of syntactic means to connect the clusters"

Cluster Level	Clusters of B.P.-s	individual proportion	frequency
H 87–100	(2,3);(4,5);(8,9);(10,11); (14,15);(15,16)	6/6 5/6 4/6	0.90 0.07 0.03
M 74–87	(7,(8,9));(13,(14,16)); (18,19);(21,22)	4/4 3/4 2/4	0.71 0.26 0.03
L 50–74	((4,5),6);((10,11),12); ((14-16),17);((23,24),25)	4/4 3/4 2/4 1/4	0.45 0.23 0.26 0.06

TABLE 1

"Low level of clustering and positive marking of sentence boundaries"

Cluster Level	Clusters of B.P.-s	individual proportion	frequency
0.30 and less	[(2,3)o(4,5)] [(4-6)o(7-9)]; [(10-12)o(13-16)];[(18,19)o (20-22)]	4/4 3/4 2/4 1/4	0.27 0.33 0.27 0.13

TABLE 2

Cooccurrence frequency	H → L		frequency
Cohesive Means	o	o	10/12
	o	,NP′	2/12
	ø	o	3/17
	ø	ø	11/17
	ø	te	2/17
	ø	,NP′	1/17
	te	ø	12/13
	te	te	1/13

TABLE 3*

Cooccurrence frequency	L → H		frequency
Cohesive Means	ø	ga	3/96
	ø	ø	57/96
	ø	te	16/96
	ø	,NP′	7/96
	ø	node	1/96
	ø	suru	2/96
	te	ø	1/9
	te	te	8/9
	,NP′	o	1/38
	,NP′	ø	1/38
	,NP′	,NP′	12/38
	,NP′	to	24/38

TABLE 4*

* o: sentence boundary; ø: plain gerund
-te: te-gerund; ,NP′: relative clause

意味の有縁性と構造の hierarchy

アンドレイ・ベケシュ

本稿では南（1974）のレベル説を有縁性というテキストに属する概念の立場から再考する。有縁性はディスコースの base proposition（＝b.p.）という作業的に得た、ほぼ節に相当する単位の間の、意味的関連性を数量化したものである。

ディスコースを組立てる時、有縁性はb.p.の間のシンタクス的関係に反映されている。有縁性の強いb.p.の順で、その関係が優先的にシンタクス的関係でマークされる。更に、有縁性の hierarchy は南のレベルの hierarchy に写像される。従って、南説は文中の従属節の分類だけではなく、テキストの意味的構造に動機付けられている。