

THE STRUCTURE OF TEXTS: A PRELIMINARY REPORT

PETER H. FRIES and JULIE FAIR

Central Michigan University, Mt. Pleasant

Learning to form coherent groups of sentences relating to one topic (that is — learning to form texts) is one of the most important tasks an individual faces as he learns his first language, for without this ability he would be virtually unable to communicate. Children begin to control this aspect of language from an early age when they learn the first rudiments of intonation and stress location. The difference between *John walked home last night* and *John walked home last night* lies in the relation the two sentences have to previous context. This signal is learned quite early. All normal adult native speakers, therefore, are able to form coherent groups of sentences on a given topic. It is abundantly clear to the composition teacher, however, that not all normal adult native speakers control this aspect of language equally well in written language. Some, indeed, are quite poor at writing coherent sequences of sentences.

This paper asks the question — what do good writers do to show how the sentences they write form a coherent structure? What signals do they use? In order to see the signals in their clearest and most obvious form we have begun our analysis with a body of texts which have been analyzed and presented by Francis Christensen as examples of good paragraphs. These paragraphs have been chosen by Mr. Christensen because they are clear examples of various kinds of paragraph structures. His analysis consists of outlining the sentences of each paragraph. (See Chart A).

It should be mentioned that an outline is essentially an expression of levels of relevance, or, in other words, coordination and subordination. If two points are at the same level of relevance, they are coordinate, while if two points are *related* but at *different* levels of relevance, one is subordinate to the

Chart A

- 1 *The other* [mode of thought] is the scientific method.
- 1..... 2 *It* subjects the conclusions of reason to the arbitrament of hard fact to build an increasing body of tested knowledge.
- 2..... 2 *It* refuses to ask questions that cannot be answered, and rejects such answers as cannot be provided except by Revelation.
- 3..... 2 *It* discovers the relatedness of all things in the universe — of the motion of the moon to the influence of the earth and sun, of the nature of the organism to its environment, of human civilization to the conditions under which it is made.
- 4..... 2 *It* introduces history into everything.
- 5..... 3 *Stars and scenery* have their history, alike with plant species or human institutions,
- 6..... *and nothing* is intelligible without some knowledge of its past.
- 7..... 4 *As Whitehead has said*, each event is the reflection or effect of every other event, past as well as present.
- 8..... 2 *It* rejects dualism.
- 9..... 3 *The supernatural* is in part the region of the natural that has not yet been understood, in part an invention of human fantasy, in part the unknowable.
- 10..... 3 *Body and soul* are not separate
- 11..... *and Man* is that portion of the universal world-stuff that has evolved until it is capable of rational and purposeful values.
- 12..... 4 *His place in the universe* is to continue that evolution and to realize those values.¹

other. Chart A reproduces almost exactly Christensen's analysis of one paragraph. We have made two changes. (1) We have numbered the sequence sentences for ease of reference — the leftmost number. (2) We have moved the *and* of the clauses six and eleven from the preceding line.

Before we go further in our presentation, we should say here that we do not believe that all well written paragraphs are outlinable. The ability to outline a paragraph merely shows that the paragraph has a simple structure. If we cannot outline a paragraph, the structure is complex. This complexity may arise through faulty writing but that is not necessarily true. The question we have asked about these paragraphs is — what signals in the language used tell Mr Christensen that his outline is the correct one?

M. A. K. Halliday has isolated the signals speakers use to form texts: they are the information system, the thematic system and the four types of cohesive relations: referential, substitutive-elliptical, conjunctive, and lexical. Since the primary signal of the information system is the location of sentence stress (*John walked home* vs. *John walked home*) and since our writing system does not indicate stress, the information system is extremely

¹ Julian Huxley, "Man in the modern world", (Mentor), pp. 146-47.

difficult to trace in written work. As a result, the information system has been ignored in this analysis.

The theme of a clause is the first clause level constituent of that clause. "The theme specifies the point of departure of a clause as message" (Halliday 1967:212). The underlined portions of the following examples are themes.

- That old man bought the store yesterday evening.
Yesterday evening that old man bought the store.
The store was bought by that old man yesterday evening.
Who bought the store yesterday evening
Perhaps he bought the store
However, he bought the store

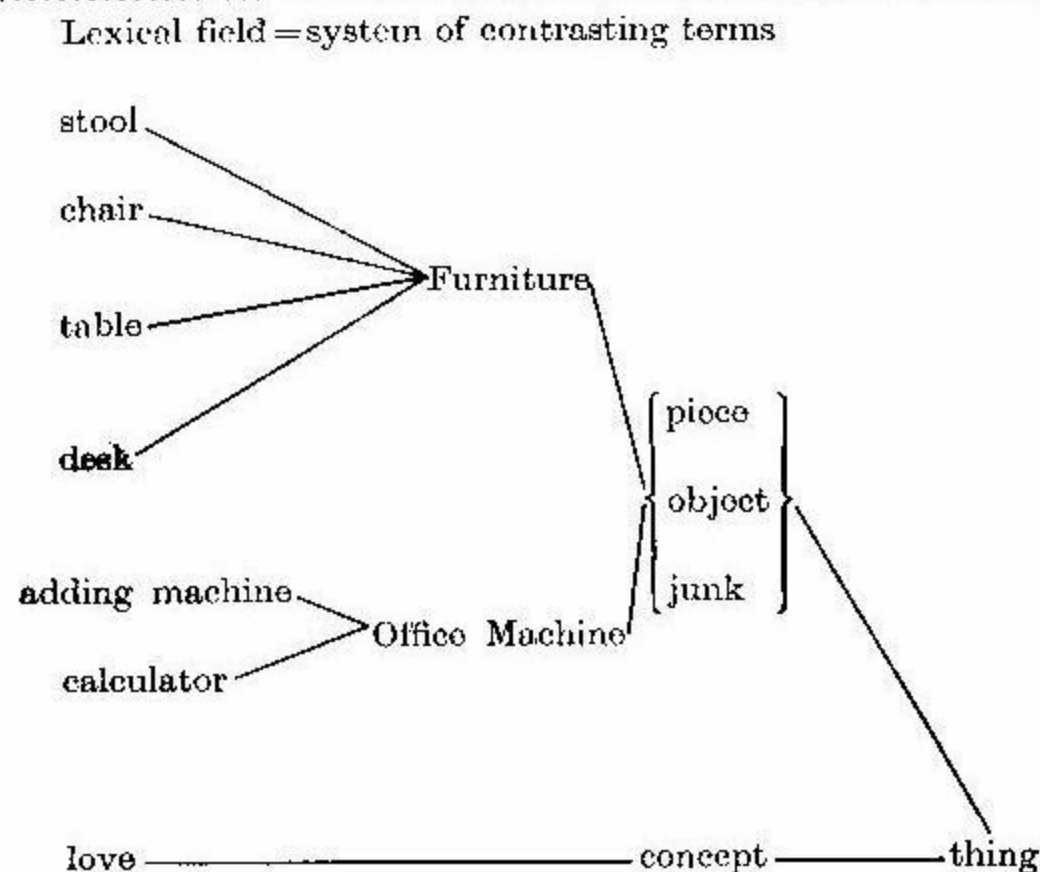
The other factor we looked at is cohesive relations. These relations are signalled by words and phrases whose very meaning relates various portions of the sentences of a text. (See Chart B).

Chart B

ILLUSTRATIONS OF SIGNALS OF COHESION

1. Reference.....Code R
 - a. Personal pronouns:
he, she, it, they
 - b. Demonstratives and definite article:
this/these, that/those, here, there, the
 - c. Comparison:
some.....the same book
such.....such books
either.....other books
2. Substitution-ellipsis
 - a. Substitution.....Code S
one/ones.....a good one
do.....Mary likes John and I do too
 - b. Ellipsis.....Code E
numerals.....*Bill bought a rose and we bought twenty*
nominal.....*Of all the dogs the best was mine.*
verbal.....*Who's coming tonight? I am.*
clausal.....*Where are you going? Out.*
3. Conjunction.....Code C
(signals of conjunction are not restricted to the grammatical class of conjunction)
 - a. Additive:
and, nor, or, furthermore, for example
 - b. Adversative:
yet, though, but, however, actually, instead
 - c. Causal:
so, therefore, because, arising out of this, as a result
 - d. Temporal:
next, then, before that, first ... then, at once

4. Lexical.....Code L



- a. Use other terms from the same field to refer to the same item:

I won't buy that chair. I've never seen such $\left\{ \begin{array}{l} \text{a thing} \\ \text{a piece} \\ \text{an object} \\ \text{junk} \end{array} \right\}$

- b. Use contrasting term to refer to a different item.

We have to get three chairs. Tables we don't need.

- c. Use subordinate term to refer to a portion of the item(s).

We don't need much furniture now. One more table would be useful, maybe.

In working with the material we took the following approach. If the theme is the point of departure of the information of the clause and if the sentences of a paragraph follow one another in a logical progression of thought, then the content of the themes of the clauses of the paragraph ought to indicate that logical progression. The signals of relations between sentences are the signals of cohesion, we therefore made the following two part hypotheses.

Hypothesis I. In a well written paragraph almost all themes will contain signals of cohesion. *Corollary:* If a theme contains no cohesive signal we ought to be able to explain why.

Hypothesis II. The cohesive signals in the theme of a given sentence ought to point to the sentences to which that sentence is related and it ought to show the type of relationship which obtains between the two sentences. *Corollary:* If the theme contains several signals of cohesion relating it to several previous sentences, this will correlate with difficulty in determining the level of relevance within the paragraph of that sentence.

In pursuing Hypothesis II we made the following five generalizations and analyzed each paragraph as in Chart C.

Generalization 1. If the theme of a clause refers to an element of a previous clause, the clause will be subordinate to that previous clause. (In terms of the coding system a code L 6, or any code beginning with R, S, or E.)

Generalization 2. If the theme of a clause contains a word or phrase which has a referent which contrasts with the referent of a word or phrase in a previous clause, then the two clauses are coordinate.

Generalization 3. If the theme of a clause is not an empty theme, and contains no signals of cohesion, the clause is subordinate to some previous clause, and is probably an example of the statement made in the previous clause.

Generalization 4. If a clause is introduced by a coordinate conjunction expressing an additive relation (*and, or, nor*), that clause is coordinate to some previous clause.

Generalization 5. If a clause is introduced by an expression indicating apposition (*for example, that is, e.g.*), the clause is subordinate to some previous clause.²

Next let us deal with each hypothesis and generalization in turn.

Hypothesis I. (a) Almost all themes will contain some signal of cohesion. Of 138 sequence sentences 120 (or 87%) contained at least one signal of cohesion. *(b)* Of the eighteen sequence clauses which contained no signals of cohesion in the theme, seven were examples of a generalization made in a previous sentence. *(c)* Two more clauses contained empty themes (themes which do not contain words which denote, e.g. *There is a book here.*) Empty themes are grammatical devices which are used to emphasize certain parts of clauses which would not normally be emphasized. Hence, these examples are evidence of the operation of a process which is outside the scope of this paper. (That is, they are explainable, but only by reference to other language processes than the ones being considered here.) *(d)* Five of the themes which contained no cohesive signal involved references to the writer and reader through the use of *we, you* and commands. The paragraphs which contained these were strongly oriented toward the reader. (It made no difference in these cases whether *we* or *you* was used. Both forms oriented the paragraph toward the reader). Even in highly structured prose, therefore, the theme is not restricted to advancing the argument. Thus, only three clauses of the 138 (or 2% of the examples) were true exceptions to the first hypothesis and its corollary.

² Chart D gives the total figures for all paragraphs. Paragraph D (given in Chart A), for example, contains 13 clauses — that means 12 sequence clauses — (all clauses other than the first clause), and of those 12 sequence clauses, 10 contain themes which have at least one signal of cohesion. Several contain more than one signal of cohesion. The generalizations made above account for the level of relevance of eleven of the twelve sequence sentences. The one case which does not work constitutes an appeal to authority.

Hypothesis II. The cohesive signals in the theme of a clause ought to point to the clause to which it is related.

Of the 8 cases in which the cohesive signals conflict three are ambiguous. (Of the ambiguous cases, three involve the word *you* which, because it refers to the reader, we expect to give problems.)

Hypothesis II. Generalization 1. If the theme of a clause refers to an element of a previous clause, the clause will be subordinate to that previous clause.

This generalization is strongly confirmed with 60 examples accurately predicting the organization, 5 real counter-examples, and 10 unclear cases. The unclear cases arise when one thing is referred to repeatedly throughout a text either through the use of a pronoun or a noun. In this case it is difficult to tell which occurrence is the relevant one for that clause. In other words, we would like to distinguish between those occurrences of a noun or pronoun which are important to the structure of a paper and those which are not. At this point, none of the hypotheses we have tried has been sufficient to do this.

Hypothesis II. Generalization 2. If the theme of a clause contains a word or phrase which has a referent which contrasts with the referent of a word or phrase in a previous clause, then the two clauses are coordinate.

This generalization is accurate in 22 instances and wrong once.

Hypothesis II. Generalization 3. If the theme of a clause is not an empty theme, and contains no signals of cohesion, the clause is subordinate to some previous clause and is probably an example of the statement made in the previous clause.

This statement is supported by 10 examples and is wrong in 4 cases. The problem in the four cases lies in the fact that these were not examples of a previous statement.

Hypothesis II. Generalization 4. If a clause is introduced by a coordinate conjunction expressing an additive relation (*and, or, nor*), that clause is coordinate with some previous clause.

One can go a bit further and say that if the *and, or* or *nor* do not begin a sentence, then the clause is coordinate to the immediately preceding clause. If, however, the *and, or* or *nor* begin a new sentence, no such statement can be made. This generalization is accurate in 13 out of 15 cases.

Hypothesis II. Generalization 5. If a clause is introduced by an expression indicating apposition (*for example, that is, e.g.*), the clause is subordinate to some previous clause and is either an example of that clause or a reprise of that clause.

This generalization is accurate in three out of three cases.

While these generalizations are only rough first attempts and make no attempt to be complete, they are surprisingly accurate. Out of a total of 148 cases they are wrong in 12 (8%) and they work poorly in 10 (6%), and 18 cases (12%) have nothing to do with the generalizations given here.

One must be very careful in drawing conclusions from these data, however. First, as we consider longer texts the percentage of cases in which identity of reference will be a poor (ambiguous) signal will increase somewhat.³

The figures given here represent an analysis of texts which have a very clear, simple structure. They were chosen by Mr Christensen on this basis. As we move to texts which have a structure which is less clear we expect to find signals which are less clear. Similarly in texts which have a more complex structure, the signals of the relations between sentences will reflect this complexity. In other words, we do not want to claim that paragraphs which can be outlined are the only good paragraphs. They are, rather, the simplest ones to deal with.

Another reservation arises out of the fact that one of the tools, the lexical signals of cohesion, was quite weak. In many cases we were forced to go beyond any coherent theory and examine our intuition to see if we felt that a given pair of words were related lexically. Similarly, in a number of cases we felt there was a definite association between two words, yet the theory forced us to consider them to be unrelated.

Finally, a problem with the analysis arises out of the fact that we ignored the information system. This system is known to be involved with the relations between the sentences of a text; the fact that we ignored it would imply that some of our exceptions, particularly exceptions to the first hypothesis, will probably be explainable by reference to this system.

In the light of all these reservations we cannot claim to have a final answer to the study of texts. We believe we can claim, however, that this approach promises to be useful in the study of the structure of texts.

³ In two longer texts we have analyzed, one about the Japanese economy and the other on projective geometry, references to Japan or economics in the one and to projective geometry in the other gave virtually no information as to the level of relevance of the clause involved. We do not believe, however, that the number of ambiguous cases will be so great as to make this first generalization of hypothesis II useless.

Chart C

ANALYSIS OF THE SEQUENCE SENTENCES OF PARAGRAPH D

Element	Code	Distance	Presupposed item	Organization
1. it	R 136	0	The other mode of thought	drop 1 from 0
2. it	R 136	1	The other mode of thought	drop 1 from 0
3. it	R 136	2	The other mode of thought	drop 1 from 0
4. it	R 136	3	The other mode of thought	drop 1 from 0
5. stars and scenery	not cohesive			drop 1 from some sentence
6. and nothing	C 111	0		same level as some sentence
And nothing	not cohesive			drop 1 from some sentence
7. As Whitehead has said, each event	not cohesive			drop 1 from some appropriate sentence
not cohesive, hence example or detail.				
8. it	R 136	7	The other mode of thought	drop 1 from 0
9. the supernatural	L 27	0	dualism	drop 1 from 8
10. body and soul	L 26	1	dualism	drop 1 from 8
11. and man	C 111			same level as 10
and man	not cohesive, hence further detail or example			drop 1 from some appropriate sentence
12. his place in the universe	R 117	0	man	drop 1 from 11
his universe	L 119	0	universal	drop 1 from 11

Chart D

TOTALS FOR ALL CHRISTENSEN PARAGRAPHS

P	# Sequence Clauses	Cohesive Themes	# Generalizations Work	# Generalizations Don't Work	Cause for not working		
					Empty Theme	Ref. to Part.	Signals Conflict Other
A	4	4	4	0			
B	7	7	7	0			
C	4	1	3	1		1	1
D	12	10	11	1			1
E	16	16	13	3		1	2
F	12	8	7	6	4		2
G	11	11	11	0			
H	12	11	11	1			1
I	8	8	7	1			1
J	10	7	9	1			1
K	4	2	3	1			1
L	4	1	0	4	2	1	1
M	6	5	5	1			1
N	10	9	9	1		1	1
O	5	5	3	2			2
P	15	15	15	0			
T	138	120	115	23	2	5	8
		87%	83%	16%			6%

REFERENCES

- Christensen, F. 1967. *Notes toward a new rhetoric*. New York: Harper and Row.
- Firbas, J. 1964. "On defining the theme in Functional Sentence Analysis", *Travaux Linguistiques de Prague 1*. 267—280.
- Halliday, M. A. K. 1967. "Notes on transitivity and theme; parts 1 and 2". *JL 3*. 37—81 and 199—244.
- Halliday, M. A. K. 1968. "Notes on transitivity and theme; part 3". *JL 4*. 167—180.
- Halliday, M. A. K. and R. Hasan. 1976. *Cohesion in English*. London: Longman.
- Hasan, R. 1968. *Grammatical cohesion in spoken and written English*. Part 1. *Programme in linguistics and English teaching*. Paper 7. London: Longmans.