READING COMPREHENSION AND PERCEIVED COMPREHENSIBILITY OF LEXICAL DENSITY AT DISCOURSE AND SENTENCE LEVEL

Jarl Backman
READING COMPREHENSION AND PERCEIVED COMPREHENSIBILITY OF LEXICAL DENSITY AT DISCOURSE AND SENTENCE LEVEL

JARL BACKMAN

Abstract

A text was varied in three levels with regard to lexical density (LD=the quota lexical words/total amounts of words in a sentence). One hundred and thirty-two subjects comprehension of text (discourse comprehension) was tested by a knowledge test and by subjective ratings. From the same text were also sampled 10 sentences of which each varied in the same density levels and were judged in the same comprehension scale as the whole text (sentence comprehension). The discourse analysis gave no differential effects of comprehension defined via knowledge tests and ratings. The single-sentence analysis gave significant contrasts between different LD-levels for 7 of 10 sentences. In a factorial analysis a significant interaction effect between LD and sentence was obtained. The discrepancy between discourse comprehension and sentence comprehension was interpreted so that the subjects in the latter case were oriented towards non-lexical factors, one of the reasons being lack of context. Implications for reading research are finally discussed.
A linguistic variable named lexical density (LD) has lately attracted some interest concerning verbal learning. Lexical density is defined as the quota lexical words/total amount of words in a sentence and this variable seems among other things to be related to retention of sentences. (Perfetti 1969).

In the group of lexical words there are mainly nouns, adjectives, verbs and adverbs (see Lyons 1966 for a more explicit definition). These classes are often looked upon as the information carrying elements in a sentence and are of importance for, among other things the readability of a text (see e.g. Coleman 1971). Furthermore high lexical density is considered to lead to a greater strain for memory processes. Sentences with high density (LD→1) should be more difficult to reproduce at recall than sentences with low density. Perfetti (1969) confirmed this inverse relation between LD and recall. Perfetti's research accounts for results being valid for separately exposed sentences. However, in later research by the same author (Perfetti & Garson 1973) short prose passages were used as stimuli where retention was measured by a recognition method. The subject task consisted of choosing between a sentence which was identical with the one in the original text, together
with another sentence changed in a certain way (semantically, lexically, grammatically or unchanged). The retention of semantic information was found to be more resistant over time compared with grammatical and lexical information. The analysis of the dependent variable is, however, still made on sentences taken separately.

The investigations on sentence level which have claimed to describe or explain discourse learning and comprehension have been criticized by many, probably with some justification (see e.g. Carroll 1971, 1972 or Rothkopf 1972). There has been some doubts about the relevance and the validity of several of the linguistic and psychological concepts and units of analysis (e.g. the concept of deep structure in the generative transformational grammar) that during a long time have been used at generalizations to discourse level. The amount of verbal input and time for exposition are the factors that make inferences from sentence level to discourse level so risky. This report adds another variable to all the dimensions deduced from sentence level, but being tested with respect to the possibilities of generalization to longer whole passages. The aim here is to study if and how a linguistic variable - lexical density - has some effect on the comprehension of on one hand a whole text and on the other separate sentences. The
following restrictions were added: the subjects should in both tests be the same and the separate sentences should come from the same source as the whole text.

METHOD

Subjects

One hundred and thirty-two students at senior level served as subjects in the study. They were not paid for their participation.

Verbal Material

The text put before the subjects consisted of approximately 1800 words. The content of the text, construction and function of the computer, was judged as unknown to the subjects. The text was taken from a book by Wettermark (1968) "ADB från början" (Automatic Data Processing from the Beginning) and has earlier been used in reading comprehension tests. (Backman 1972).

The text varied with respect to lexical density in three levels: .31-.50, .51-.70 and .71-.90. These levels were brought about through supplements or elimination of the number of non-lexical (grammatical) words in the original text. Structural changes were avoided, and the sentences on the various density levels could be seen as functionally equivalent.

The knowledge test defining comprehension was of multiple-choice type (5 alternatives 4 of which were distractors), and was tested according to the principles
given by Marks & Noll (1967). They account for a statistical model for evaluating items in knowledge tests with regard to the extension of which comprehension is measured. This technique refutes the accusation of invalidity when using knowledge questions as a measure of comprehension. Here an index for every item is calculated as a function of information exposure with respect to among other things earlier knowledge, conjecture, and error in question construction. This analysis reduced the number of items from 26 to 20.

Procedure and Design

All subjects started with the knowledge test and after that they could study the text placed before them without time limit. The subjects were not informed that they were going to be tested after having read the text. After the text and the comprehension test the subjects read a shorter text of about 300 words. This formed a comparison standard for a subjective judgment of the comprehensibility of the text read through earlier. When subjective judgments of the degree of difficulty of tests are used as a dependent variable there is reason to assume that the content of a comparison standard has an effect on the judgments. Consequently, the content of the texts could be a factor of decisive importance for the subjective judgments. Therefore the content of the standard test was varied in three respects. One of the
texts was taken from a newspaper and dealt with local news. The second text was fiction, and the third text was taken from the same source (technical text) as the test text i.e. from Wettermark (1968), but not from the passage read through. The three texts were approximately on the same lexical density level =.70. The subjects marked on a rating scale the experienced comprehensibility. The scale was a seven-degree with the following verbal categories: "very easy to comprehend (1), "easy to comprehend" (2), "rather easy to comprehend" (3), "neither easy nor difficult to comprehend" (6), "very difficult to comprehend" (7). Three various density levels and three standard texts give 9 conditions, and from every density condition the subjects were randomly distributed on the three standard texts with approximately the same number on each text.

Finally the subjects were given the task of giving a subjective judgment of single sentences regarding comprehensibility. The sentences were randomly taken from the text earlier read through by the students. Ten various sentences were each presented on the three earlier used density levels: .31-.50, .51-.70, and .71-.90, i.e. the same variation as earlier. The rating scale consisted of the same categories as above. The subjects then, gave judgments on ten various sentences, each of them on three various density levels, i.e. 30 sentences in all, presented in a randomized order.
RESULTS

Discourse Level

Differential effects of LD-level at the knowledge text could not be proved. The means for the levels .31-.50; .51-.70- and finally .71-.90 were calculated to 18.8, 18.5, and 17.8 and after sorting out items according to the Marks-Noll criteria to 15.1, 15.3, and 15.3 respectively. An ANOVA gave in both cases $F(2,129) < 1$. Consequently, lexical density is not an important variable for discourse comprehension.

Table I shows mean rating values of comprehensibility at various LD-levels and standard texts. The values do not differ much and gather round the verbal categories "rather easy" (3), and the middle category "neither easy nor difficult to comprehend" (4). A 3x3 ANOVA gave no significant interaction (LD x Standard text), $F(4,119) = 1.19$, $p > .25$. None of the main effects could be proved as significant either. For standard text was received $F(2,119) = 1.88$, $p > .10$ and for LD level $F(2,119) = 2.57$, $p > .05$. The

Table I in about here

comprehensibility ratings seem to be independent of both density as well as the standard text with which the test text is compared. To sum up, the discourse analysis makes comprehension defined as knowledge test and as rating independent of the lexical density of a text.
**Sentence Level**

Every sentence was analyzed and regarded as single-factor design with repeated measures of the same elements over LD levels. In figure 1 is given for every sentence the number of significant contrasts ($p < .05$) according to the Newman - Keul method (Winer, 1962). For three of the sentences the effect of lexical density did not appear. One sentence has one significant contrast, five sentences have two, and one sentence has three maximum significant differences in pairs. Accordingly, the result on the whole (in 7 cases of 10) supports Perfetti's (1969) outcome but does not correspond to the discourse analysis where no effects of LD could be proved.

---

Data on sentence level can, however, also be seen as emanating from a 10 (sentence) x 3(density level) factorial design. Here an ANOVA gave a significant interaction between both factors, $F(18,3913) = 3.57$, $p < .001$. The effects of density level was not significant, $F(2,3913) = 2.96$, $p > .05$, while on the other hand the effect of sentence gave significance, $F(9,3913) = 52.27$, $p < .001$. Consequently this analysis does not allow any clear-cut interpretation of the factor lexical density.
Computation of central values of the rating values over LD levels (with the assumption that the sentences can be regarded as equivalent) allows another comparison with the discourse analysis. In both cases the same 7-degree scale with identical verbal categories was used. The comprehensibility ratings on sentence level gather round the value 3, e.d. "rather easy", while the corresponding ratings for discourse comprehensibility are right above 4, meaning "neither easy nor difficult to comprehend". The same subjects, then generally judges single sentences as being easier than corresponding lexical density on discourse level. The result of both rating analyses can be seen in figure 2. Here are, also presented Lix values for the various density levels defined via the formula proposed by Björnsson (1968):

\[ \text{Lix} = \frac{w}{s} + 100 \frac{lw}{w}, \]

where \( w \) = number of words, \( s \) = number of sentences and \( lw \) = number of long words (> 6 letters). These Lix values can be translated into verbal categories according to the Lix translator proposed by Björnsson (1968, p 89).

The categories are almost identical with the rating scales in this research. Lix 20 is characterized by Björnsson as "very easy", and 50 as "difficult" etc. The
COMPREHENSION AND LEXICAL DENSITY

difference is that Björnsson's scale does not define in which respect a text is difficult or easy. A reasonable assumption, however, is that the grade of comprehension is intended. Figure 2 shows that the Lix values (divided by 10) better correspond to discourse data than to sentence data.

DISCUSSION

The linguistic variable lexical density did not operate differentially on discourse level in this experiment irrespective of comprehension being measured via knowledge test or subjective rating. The result can be interpreted so that a succession of sentences in a text are worked at selectively by omitting nonessential information. The individual integrates selected information out of every new sentence with a context and with earlier gained knowledge. Discourse comprehension is moreover characterized by the reader's inferring and organizing conceptual and relational structures. This is probably not done by complete analyses of every single sentence but by extracting 'holistic' semantic structures out of the text (see e.g. Sachs 1967, Bransford & Franks 1971). Such structures seem to be invariant at an addition or a reduction of the number of grammatical words in a given sentence. Accordingly, a semantic content remains unchanged over various LD levels, and any sentence can in a given context be related to one or several other sentences at discourse integration. The individual works 'constructively' at input level and the
result can be predicted (see Andersson & Bower 1973 for an alternative model on sentence level and output). This line of reasoning is supported by a comparison of rating data at discourse and sentence level. In the former case the whole structure of the text is judged according to comprehensibility and ought to be more difficult to understand than a single sentence. This is also what happens which could be seen in figure 2.

On sentence level the prediction of the 'sentence analyst' is confirmed and the obtained results have implications for various types of readability indices or formulas. These are usually based on two simple variables viz. word length and sentence length (see Klare 1974-1975). If sentence length is calculated for sentences containing significant and non-significant contrasts the mean values 15.2 and 10.7 are obtained. This can be regarded as a remarkable difference. The result gives reason for suppositions like sentence length being an indicator of difficulty caused by density manipulations. However, such a conclusion conceals the interactive relation and the outcome in the discourse analysis where no differential effects of density variations could be observed.

Although the semantic 'load' of a single sentence is unchanged at lexical density manipulations, the comprehensibility ratings of the subjects are changed. This
COMPREHENSION AND LEXICAL DENSITY

difference between discourse comprehension and sentence comprehension can not be explained because of differences in subject groups (the same group was used) or because of differences in contents of text and sentence (the sentences came from the same text). Accordingly, the outcome on sentence level supports Perfetti (1969) although not quite clearly. It seems as if the subjects were oriented at extra-semantic factors when rating single sentences. This can, as Fredriksen (1975) has suggested, be explained by lack of contextual information. The individual then elaborates the whole sentence where lexical but even grammatical variables are allowed to operate. Here the individual does not use the language competence which facilitates a semantic elaboration, i.e. the omission of redundant information. The result of this research indicates that cognitive processes operating on single sentence level are not quite identical with the processes on discourse level. A model for the latter case should incorporate influences of among others context and previous knowledge demands not paid much attention to at single sentence comprehension.
REFERENCES


<table>
<thead>
<tr>
<th>LD-level</th>
<th>.31-.50</th>
<th>.51-.70</th>
<th>.71-.90</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fiction</td>
<td>4.53</td>
<td>4.57</td>
<td>4.50</td>
</tr>
<tr>
<td>Newspaper</td>
<td>4.21</td>
<td>3.43</td>
<td>4.32</td>
</tr>
<tr>
<td>Technical</td>
<td>3.64</td>
<td>3.53</td>
<td>4.80</td>
</tr>
</tbody>
</table>
Figure captions

Figure 1. Comprehensibility ratings of the ten different sentences at three levels of lexical density. The figures to the right of the curves marks the number of significant contrasts for every sentence.

Figure 2. Comprehensibility ratings at discourse level (D), sentence level (S) and transformed lix-values (L) for three lexical density levels.
Figure 1

Lexical Density

.31-.50 .51-.70 .71-.90

LEXICAL DENSITY
Earlier reports in this series:

1973:
1. READABILITY OF STAVES WITH DIFFERENT LINE AND NOTE DISTANCES.
   Jarl Backman and Allan Lindahl

2. LITERACY AND SOCIETY IN A HISTORICAL PERSPECTIVE - A CONFERENCE REPORT.
   Egil Johansson (Ed.)

3. THEORETICAL PROBLEMS IN CONSTRUCTION OF CRITERION-REFERENCED TESTS.
   Ingemar Wedman

4. RELIABILITY, VALIDITY, AND DISCRIMINATION MEASURES FOR CRITERION-REFERENCED TESTS.
   Ingemar Wedman

1974
5. DROPOUT AND PASS-RATE IN HIGHER EDUCATION - ARE THEY USEFUL MEASURES OF EFFICIENCY?
   Inga Elgqvist-Saltzman

6. SOME COMMON WORD ATTRIBUTES AND THEIR RELATIONS TO OBJECTIVE FREQUENCY COUNTS.
   Jarl Backman

7. ON TEST-WISENESS AND SOME RELATED CONSTRUCTS.
   Ingvar Nilsson and Ingemar Wedman.

1975
8. THE OCCURRENCE OF TEST-WISENESS AND THE POSSIBILITY OF INDUCING IT VIA INSTRUCTION.
   Ingvar Nilsson.

1976
9. THE FUNCTION OF GROUP-SIZE AND ABILITY LEVEL ON SOLVING A MULTIDIMENSIONAL COMPLEMENTARY TASK.
   Thor Egerbladh.
UNIVERSITY OF UMEA

Mailing address:
Department of Education
University of Umeå
S 901 87 Umeå
SWEDEN