DOCUMENT RESUME

ED 359 754   FL 020 819

AUTHOR   Melander, Bjorn
TITLE   Main Worlds and Supplements. An Attempt at Using a
                Qualitative Approach to Quantitative Data.
PUB DATE   Apr 92
NOTE   13p.; In: Nordic Research on Text and Discourse.
                NORDTEXT Symposium (Espoo, Finland, May 10-13, 1990);
                see FL 020 811.
PUB TYPE   Speeches/Conference Papers (150) -- Reports -
                Research/Technical (143)

EDRS PRICE   MF01/PC01 Plus Postage.
DESCRIPTORS   *Cognitive Processes; *Content Analysis; Foreign
                Countries; *Languages for Special Purposes;
                Semantics; *Text Structure; Uncommonly Taught
                Languages; Written Language
IDENTIFIERS   *Schema Theory; Text Patterns; Textual Analysis

ABSTRACT
        One of the analyses carried out within the University
of Uppsala (Sweden) study, "LSP Texts in the 20th Century,"
classified the cognitive text content into five different cognitive
worlds: the scientific, the practical, the object, the private, and
the external. This paper investigated patterns of distribution in the
texts of these cognitive worlds. A classification of the function in
the textual structure of the cognitive worlds results in four
different types: main world, secondary world, excursus, and
supplement. This model makes it possible to describe content patterns
at a macrolevel and also to correlate these patterns to various
contextual factors, thereby making it possible to relate text and
context. (Author/JP)

ERIC
Full Text Provided by ERIC

Main worlds and supplements

An attempt at using a qualitative approach to quantitative data

BJÖRN MELANDER

Abstract

One of the analyses carried out within the Uppsala study 'LSP Texts in the 20th Century' is an analysis of the cognitive text content. The basic part of this analysis is a classification of the text content into five different cognitive worlds, a notion that has similarities with the schema concept of cognitive psychology. The cognitive worlds - the scientific, the practical, the object, the private, and the external worlds - correspond to different contextual factors, mainly sectors of society. In this paper, results from an investigation of the patterns of distribution in the texts of these cognitive worlds are presented. Based on these patterns, a classification is made of the function in the textual structure of the cognitive worlds into four different types called main world, secondary world, excursus, and supplement. This model makes it possible to describe content patterns at a macrolevel and also to correlate these patterns to various contextual factors, thereby making it possible to relate text and context.

1. Introduction

A project called 'LSP Texts in the 20th Century' is currently being carried out at the Department of Scandinavian Languages at the University of Uppsala, Sweden. The aim of the project is to study both synchronic and diachronic differences between Swedish LSP texts from two different genres, science and popular science, and from three periods of time, 1895-1905, 1935-1945, and 1975-1985, within the fields of economics, medicine, and technology. The text corpus comprises 90 articles, five in each group of texts. The articles are on an average six to seven pages long and have been analysed in their entirety. (The material is presented in detail in Gunnarsson, Melander, and Näslund 1987.)

The analyses are directed towards the textual aspects that are relevant to the message structure of the texts. They focus, therefore, on deeper textual levels, namely the pragmatic and cognitive levels. One of the key parts of these analyses is a study of *the cognitive text content*.

The basic part of this analysis is a classification of the text content into one of five different types or *cognitive worlds*, a notion that has similarities with the schema

concept of cognitive psychology. The cognitive worlds used in the analysis are the *scientific*, the *object*, the *practical*, the *private*, and the *external* worlds. These different worlds are all related to various factors in the social context of the texts that one has reason to believe influence their content patterns. (See Gunnarsson 1989a and 1989b for a detailed presentation of the analytical model and its theoretical foundations.)

The central elements of *the scientific world* are theories, classifications and experiments or observations of different kinds. A typical example of a classification is:

> (1) Men inte sällan ser man pneumotorax som resultat av diagnostiska eller terapeutiska åtgärder eller som komplikation till komplicerad eller avancerad lungsjukdom.

> [But pneumothorax is quite commonly interpreted as a result of diagnostic or therapeutic measures or as a complication of compound or advanced pulmonary disease.]

The practical world is characterized by the relation to the profession that is connected to each of the three fields, that is the work of the economist, the physician, and the technician. The description might concern things like diagnostic measures or treatments of diseases, or interaction between an expert and a layman, as in the following example, taken from a text dating from 1905, which is an account of a conversation between a doctor and a patient:

> (2) En 60-årig kvinna frågade jag för en kort tid sedan hänsynsfullt följande: "gumman har väl inte badat på 10 år?"

> [A short time ago I quite politely asked a 60-year-old woman the following question: "I suppose it must be ten years since you last had a bath?"]

The parts of the text classified as falling within *the object world* deal with the very objects of the different fields, for example a disease, a technical device, or an economic phenomenon. This example describes the economic situation in Sweden during the second world war:

> (3) Under tiden fram till och med aprilkrisen i fjol skedde andra och mera omfattande omflyttningar av köpkraft än de som förmedlades över budgeten.

> [During the period up to and including the crisis of April last year, substantial transfers of purchasing power were effected over and above those caused by the budget.]

The essential component of *the private world* is the individual person and the way he or she is affected by the phenomenon that is discussed. One may for example find descriptions of the consequences of a certain illness for a patient's private life or, as in

this example, of the good results that an economical way of living is said to have upon one's inner life:

> (4) Den sparsamme är alltid förtänksam, och hans besparade medel skänka honom trygghet och ingifva honom en känsla af oberoende.

> [The thrifty person is always prudent and his savings provide him with a feeling of security and independence.]

Within *the external world*, finally, the relation is to public and social factors of different kinds. One type of relation is to various social conditions that might be of importance for the matter at hand, for example a proposed solution to an economic problem. Another type of relation is to social actions of different kinds, for example political decisions, as in this sentence, where the author recommends that a law regulating the treatment of tuberculosis should be introduced and points to Sweden's neighbouring countries as good examples:

> (5) I Norge finnes en verklig tuberkuloslag och i Danmark har regeringen nu framlagt lagförslag i liknande riktning.

> [In Norway there is a proper Act of Parliament dealing with tuberculosis, and in Denmark the government has just introduced a Bill of similar scope.]

In the cognitive analysis we have gone through the texts and classified the content of their different parts as falling within one of these different types. With the exception of certain formal headings, and supplementary material such as pictures, diagrams, and tables every part of the text has been classified as belonging to one single cognitive world. The categories used are thus totally discrete.

In order to get textual parts that can, at least to some extent, be understood without access to the context, in all the examples above I have chosen instances where a whole sentence has been classified as belonging to a certain cognitive world. The basic unit of analysis is, however, the *semantic content unit*, which is a concept that could approximately correspond to a proposition or a sentence atom (in Näslund 1989:11-12, a detailed description is given, see also his article in this volume). This makes it possible to do a very detailed analysis where even small shifts in the content are accounted for (in Gunnarsson 1989b:48-65 an example of an entire analysis of a text is shown). Perhaps it should be pointed out here that the fact that we have been working with such small units of analysis does not mean that we believe that the content units can be understood and interpreted in isolation. In the analysis we have, instead, tried to gain a grasp of the more global text content. Of course the parts of a text have a meaning only in relation to their context.

2. A qualitative approach

The close analysis of the texts has naturally produced a lot of data. In this paper I will present an attempt at using a more qualitative approach when analysing some aspects of these large sets of data. More precisely, I have tried to make a classification of the function of the content belonging to the five different cognitive worlds, based on the way it is distributed in the texts.

This classification is made on the basis of three different aspects. The first aspect concerns the tendency of the content units falling within a certain cognitive world to appear scattered in small pieces in several different parts of the texts or held together in a few large clusters.

In order to investigate this aspect, each content unit has been assigned a value indicating its relative position in the text. This value is simply the ratio of the consecutive number of the content unit in question and the total number of content units in the text. If, for example, a text consists of 1,000 content units, the final unit has been assigned the value 1,000/1,000, the middle one the value 500/1,000 and the first one the value 1/1,000. In this way, the position of all content units can be determined, and it is also possible to compare texts of different length.

Different measures of variability can be calculated on the basis of this value, for instance the standard deviation of the relative position of the content units belonging to a particular cognitive world. A high figure would then indicate a high degree of splitting-up. This method would not, however, give an accurate picture of the situation, since one would also get a high value where a certain cognitive world constitutes a large part of a text, although it may not be at all scattered, but instead held together in, for example, one single large cluster.

To create a more accurate measure, I have instead calculated the standard deviation that one would get theoretically if all content units of a certain cognitive world were completely evenly distributed in the text, in strings as long as their mean length in the total material. This theoretical value has then been compared with the standard deviation that one actually gets for the world in question. The difference between the theoretical value and the observed one has then been computed and finally the ratio between this difference and the theoretical value has been determined. This quotient could perhaps be called *the ratio of variability*.

In order to give a better idea of what these calculations represent, the distribution of the text content in one of our texts is illustrated in Figure 1. Along the abscissa are marked the content units (ranging from 1 to 851) and along the ordinate axis, the

different cognitive worlds; from top to bottom the scientific (SC), the practical (PR), the object (OB), the private (PI), and the external (EX) worlds.
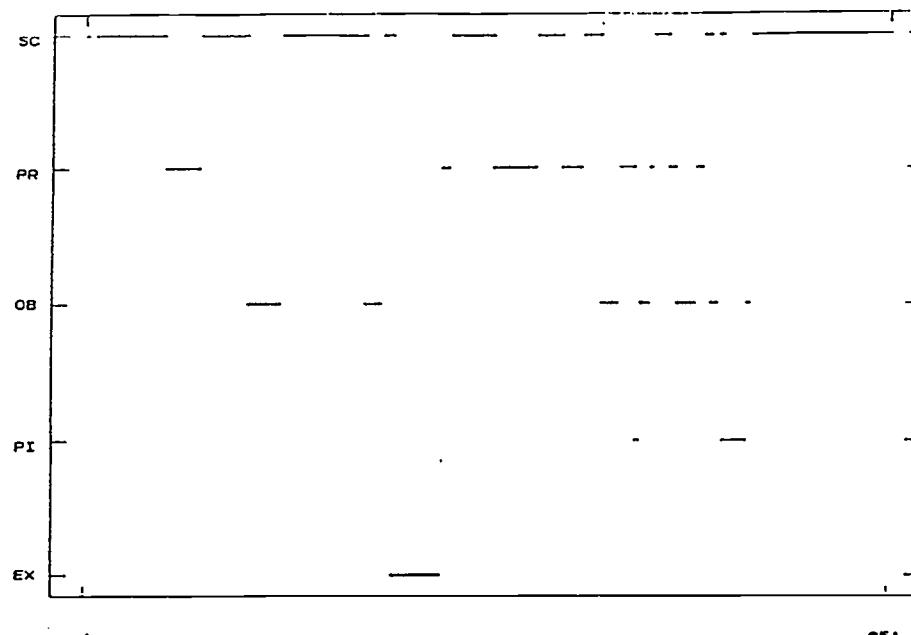

Figure 1: *Distribution of the different cognitive worlds*

In Table 1, the values of the theoretical and the observed standard deviations, the difference between them, and finally the ratio of variability (in percentages) are shown for the text in question.

Table 1. *Theoretical and observed standard deviation, difference between these values, and ratio of variability*

	THEORETICAL	OBSERVED	DIFFERENCE	RATIO
Scientific world	30.0	32.5	-2.5	-8.3
Practical world	32.6	22.2	10.4	31.9
Object world	36.2	23.2	13.0	35.9
Private world	17.3	4.6	12.7	73.4
External world	17.4	1.7	15.7	90.2

The private and the external worlds have the highest values for the ratio (73.4 and 90.2 respectively). This means that, in relation to the theoretical value, they have a small

6

112

observed variability (and consequently a big difference), and, as is shown in Figure 1, the content units belonging to both these worlds are clustered in comparison with the ones falling within the three other worlds, which are scattered over larger parts of the text.

On the basis of the ratio of variability, a dichotomization has been made of the cognitive worlds into the types *clustered* and *scattered*. The limit between the two types has been set to a ratio of variability of 50% (which means that the observed variability is only half of the theoretically expected one). It cannot be denied that this limit is somewhat arbitrary, but generally it seems to be in accordance with the intuitive impression one gets from, for example, an illustration like Figure 1.

This is the first basis of classification. The second one is related to the fact that a certain cognitive world may be regarded as *dominating* different parts of a text.
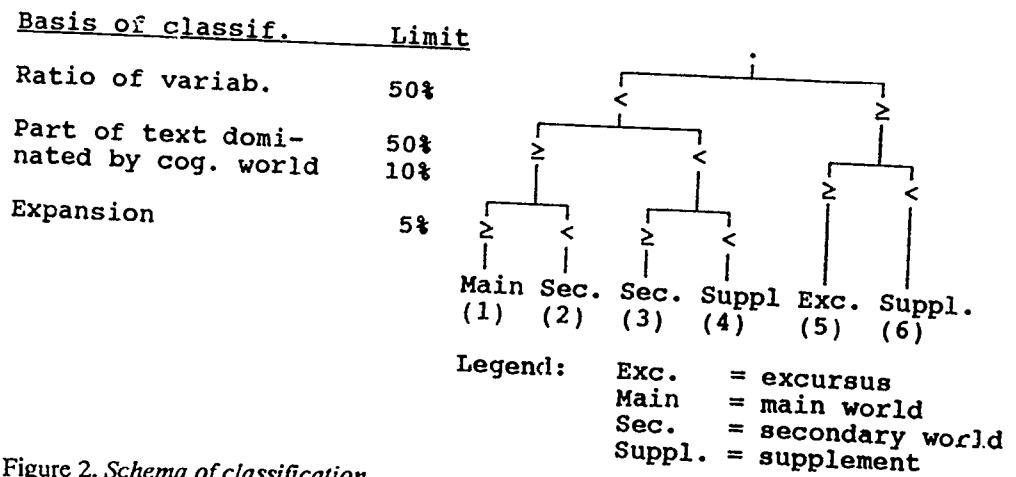
From one point of view, a text can be seen as a chain of strings of content units of different types. If one follows a text from the beginning to the end, one sees that certain parts of the text are dominated by a particular cognitive world, while the other worlds are just small insertions into the bigger structure built up by the dominating world. In other parts of the text, one finds longer insertions of other cognitive worlds, insertions which add up to a more considerable change in the text content. One could say that the first type does not constitute a real change, because the small insertion is 'bridged' by the dominating world, while the second type represents a real shift in the cognitive text content.

To determine which parts of a text can be said to be dominated by a certain cognitive world, the percentage that the world in question amounts to within this part has been calculated, following the strings of content units in the text from beginning to end. If the percentage of the world in question drops below 50%, due to a longer insertion of content units belonging to another world (or other worlds), this has been taken as indicating a major shift in the text content and the calculation has begun again from scratch at the next occurrence of this world. In this way, it is possible to calculate how much of the text is dominated by a certain world (that is where this world amounts to more than 50%).

The percentage of the text dominated by the world in question can then be compared with the percentage this world makes up of the text if one just counts the number of content units belonging to that particular cognitive world. Such a comparison shows that some worlds expand a lot compared to the original figures - which of course indicates that they bridge quite a lot of smaller insertions - while others expand very little or not at all (indicating very few bridges). This *rate of expansion* forms the third ground for classification.

In the example shown in Figure 1, the scientific world expands by 17% while all other worlds have small figures of expansion. This world also dominates a large part of the text (76% of the text is dominated by this world). The scientific world is thus scattered, dominating a large part of the text and finally expanding. It can be said to form the theme, basis or foundation of this text while the other worlds function more as supplements in the bigger structure constituted by the scientific world.

On the basis of the calculations just described, a general classification into four classes has been made of the functions of the different worlds in the texts. These four classes I have chosen to call *main world, secondary world, excursus*, and *supplement*. The following schema (Figure 2) has been used in the classification:

Figure 2. *Schema of classification*

As can be seen from the schema, the worlds in the first group, *main world*, are scattered in our sample text, dominating more than half of the text, and expanding. These worlds make up the basic structure of the text. As mentioned above, the scientific world forms the main world. In this text, one also finds examples of two other groups, namely numbers 4 and 6. The worlds in group 4 are scattered but neither dominate the text nor expand. This means that they are spread in small chains in various parts of the text. In our example, the practical and the object worlds exemplify this group. Worlds with this distribution are classified as *supplements*. The remaining worlds, the private and the external worlds, both fall into group 6. Here are found the worlds that are clustered, but only make up a very small part of the text (less than 10% of the text is dominated by the cognitive world in question). Worlds with this distribution have also been classified as supplements.

The three remaining groups are not represented in the example. Of these groups, number 2 is extremely infrequent (only two instances in the total material) and will therefore not be considered here. Group 3 is made up of the worlds that are scattered and also expand, just like the ones that have been classified as the main world, but in contrast to the main worlds these worlds do not dominate more than half of the text. Group 3 worlds dominate some parts of the text but they are also sometimes interwoven with the base, and have been classified as *secondary worlds*. The last group, number 5, resembles the secondary worlds in as much as they also dominate a part of the text, but in this case almost all instances of the world are located in a relatively limited part of the text. They therefore form a kind of digression or excursus in the text and in accordance with this, they are classified as *excursus*.

This model for classification has generally worked very well. Of course, there are some instances of borderline cases, but the absolute majority of the worlds have fallen clearly into one of the categories. In most cases the distance to the limiting values has been clear.

3. Results

3.1. *Main worlds*

In Table 2, the number of instances of a certain world having the function of main world in the different groups of material is shown. In order to get higher figures and more unequivocal tendencies the figures for the scientific and the object worlds have been merged into one group and the figures for the remaining worlds into another. The three latter worlds are the ones that are the most 'open', taking into account things from professional life, from society in general or from private life, while the two former cognitive spheres are more 'closed'. Other results, for example regarding the collocational patterns of the cognitive worlds, have also shown that there is a close connection between these two worlds (Melander 1990).

Both within science and popular science, the 'closed' group increases over time within economics (1-1-3 for popular science and 1-5-5 for science) and medicine (1-3-4 and 2-3-4). The 'open' group shows a less clear-cut diachronic development, but the overall trend is clearly towards lower figures. The results thus indicate that these fields are subject to a development towards letting a more closed type of content form the basic structure of the texts.

Table 2. *Number of instances as main world*

Genre	Field	Period	CLOSED (SC+OB)	OPEN (PR+PI+EX)
Popular science	Economics	1	1	3
		2	1	2
		3	3	0
	Medicine	1	1	1
		2	3	0
		3	4	0
	Technology	1	4	0
		2	3	1
		3	2	2
Science	Economics	1	1	3
		2	5	0
		3	5	0
	Medicine	1	2	2
		2	3	2
		3	4	0
	Technology	1	5	0
		2	2	0
		3	4	0

Abbreviations: SC = scientific world, PR = practical world, OB = object world, PI = private world, EX = external world.

Within technology, science and popular science exhibit a different trend. Popular science shows an increase (0-1-2) for the open group and a drop (4-3-2) for the closed one, while the science texts have no clear diachronic development. This difference between popular science within technology and the other groups of texts is in accordance with other results obtained within the project, especially a rise in the percentage for the external world which characterizes the texts of technology in popular science (see Melander 1989a: 33-41).

There is probably no self-evident explanation for this deviant development within technology. Tentatively, however, I would like to suggest that it might be connected to a change of attitude towards technology. The present century, especially the post-war era, has seen an earlier enthusiasm for technology and a belief in technological progress give way, at least in part, to a more negative attitude (cf Eriksson 1986: 232, 303). It might he sufficient by way of example to point to the debate about envirunmental pollution. Factors like these may explain the fact that the external world

116

more frequently functions as the main world in the popular science technology texts (cf Melander 1989b, 1989c).

The figures in Table 2 do not really add up to 90, which, as mentioned, is the total number of texts analysed. This is due to the fact that some texts, a total of 20, lack a clear main world. Of these 20, the majority - 13 - fall within popular science, a result which might indicate that the scientific texts are characterized by a stronger organization of the text content and a more marked focus on one single type of content within each text.

3.2. Supplements, excursuses, and secondary worlds

In Table 3, the figures for the supplements, excursuses, and secondary worlds are shown.

The figures are once again combined into two groups; the scientific and the object worlds on the one hand and the practical, the private, and the external worlds on the other, and also in other respects, the table is buil. up the same way as the preceding one.

First of all, one might notice that supplement is the most frequent type (a total of 239 instances) and excursus the most unusual (19 instances). The group secondary world has 63 occurrences.

It is mainly the second, the open group, of cognitive worlds that functions as the supplements in the texts. In all cases, this group has higher figures than the closed group. However, there seems to be a difference between economics and the two other fields in this respect. The difference between the figures for the two groups of cognitive worlds is smaller within economics (totally 20) than within medicine (totally 40) and technology (45). This divergence is stronger during the first two periods than during the last one. During period 1, economics has in total a difference of 4 and the value for period 2 is 5. Medicine has a total difference of 13 (period 1) and 14 (period 2) and the figures for technology are 13 and 19. The figures for period 3 are more even, 11 for economics and 13 for both medicine and technology. Maybe this indicates that economics has established a clear functional division of the different types of text content later than the two other fields, a fact that could perhaps be related to the comparatively short history of economics as an established academic subject.

Table 3. *Number of instances as supplement, excursus, and secondary world*

			SUPPLEMENT		EXCURSUS		SECONDARY WORLD	
			CLOSED (SC+OB)	OPEN (PR+PI+EX)	CLOSED (SC+OB)	OPEN (PR+PI+EX)	CLOSED (SC+OB)	OPEN (PR+PI+EX)
PS	E	1	5	7	1	3	2	0
		2	3	6	0	1	5	3
		3	5	10	1	1	1	1
	M	1	2	10	0	1	6	1
		2	4	11	0	1	2	2
		3	4	12	0	2	2	0
	T	1	3	10	1	1	2	1
		2	3	12	0	1	4	1
		3	5	11	0	1	2	0
S	E	1	4	6	1	1	3	1
		2	4	6	0	0	1	1
		3	3	9	0	0	2	0
	M	1	4	9	1	0	2	2
		2	5	12	0	0	1	0
		3	5	10	0	0	2	2
	T	1	4	10	0	0	1	0
		2	2	12	0	0	6	0
		3	2	9	1	0	3	1
TOTAL			67	172	6	13	47	16
			239		19		63	

Abbreviations: PS = popular science, S = science, E = economics, M = medicine, T = technology, SC = scientific world, PR = practical world, OB = object world, PI = private world, EX = external world.

The excursus is, as mentioned, a rather infrequent type. The closed group of cognitive worlds has very few instances (a total of 6) of this functional type. The occurrences within the open group of cognitive worlds are almost all located in popular science (of 13 in total, one finds 12 within popular science). This type of digression or excursus thus seems to be a characteristic of popular science.

The secondary worlds, finally, are mostly made up of the scientific and the object worlds (47 instances, while the three other cognitive worlds have 16 occurrences). The figures are relatively evenly distributed over the different groups of texts, and there are no clear differences between, for example, popular science and science, or between the

different periods of time. This result further underlines the fact that the scientific and object worlds form the basic structure in the texts.

To conclude, one might say that this analysis of a corpus of Swedish LSP articles shows that different types of content seem to play different roles in the macrostructural content patterns of the texts. There are, however, clear differences both from a diachronic and a synchronic point of view, differences which can, at least to some extent, be related to various factors in the social context in which the articles can be placed. An analysis based on the cognitive world concept therefore seems to make it possible to get a better understanding of the relationship between a text and its social context.

References

Eriksson, G. (1986). *Västerlandets idéhistoria 1800-1950*. Stockholm: Gidlunds.

Gunnarsson, B.-L. (1989a). LSP texts in a diachronic perspective. In *Special Language: From Humans Thinking to Thinking Machines*, C. Laurén and M. Nordman (eds.), 243-252. Clevedon and Philadelphia: Multilingual Matters Ltd.

Gunnarsson, B.-L. (1989bc). The LSP text and its social context: a model for text analysis. In *Learning, Keeping and Using Language: Selected Papers from the 8th World Congress of Applied Linguistics. Volume II*, M.A.K. Halliday, J Gibbons, and H. Nicholas (eds.). Amsterdam: John Benjamins. (In press)

Gunnarsson, B.-L. (1989:c). *Facktexter under 1900-talet 2. Metoder för textanalys på makro- och mikronivå*. FUMS Rapport nr 145. Uppsala universitet.

Gunnarsson, B.-L., Melander, B., and Näslund, H. (1987). *Facktexter under 1900-talet 1. Projektpresentation och materialbeskrivning*. FUMS Rapport nr 135. Uppsala universitet.

Melander, B. (1989a). *Facktexter under 1900-talet 3. Resultat från kognitiv textanalys*. FUMS Rapport nr 148. Uppsala universitet.

Melander, B. (1989b). Kunskapsvärldarnas variation. Resultat från en kognitiv analys av facktexter. In *Svenskans beskrivning 17*. Åbo. (In press)

Melander, B. (1989c). Diachronic Changes in the Cognitive Content of Swedish LSP Texts. In *Proceedings from the 7th European Symposium on LSP*. Budapest. (In press)

Melander, B. (1990). *Patterns of Content Structure in Swedish LSP texts*. Paper read at the ADLA-AScLA international conference in Copenhagen, August 23-24, 1990. (MS)

Näslund, H. (1989). *Facktexter under 1900-talet 4. Resultat från kognitivt inriktad referentanalys*. FUMS Rapport nr 149. Uppsala universitet.

Björn Melander is a research assistant at FUMS, the unit for advanced studies of modern Swedish, at the department of Nordic languages at the University of Uppsala. He is currently working on a Ph.D thesis dealing with patterns of content structure in Swedish LSP texts.