

Øivin Andersen, University of Bergen

Degree of specialization, network density and status - a text typological approach

1. Introduction

The text typological problems discussed in this paper form a part of a major project called "The North Sea as a linguistic and cultural laboratory". The project is sponsored by the Norwegian Research Council.

One of the principle questions in this project is: To what extent can different socio-cultural norms be said to influence the production of texts relating to the oil activity in the Norwegian part of the North Sea? More specifically: Can variation in text typology structures be explained by referring to sociocultural background features of the different professional language user groups?

In other words, we are dealing with the familiar but exceedingly complex meeting between texts on the one hand, and the surrounding socio-cultural setting providing different interpretive frames on the other hand.

Since Roman Jakobson's pioneer studies of basic functional aspects of language and the Prague School functionalism, the social and the

pragmatic dimension of language have received increased focus. Basic aspects of the concept of context have been analyzed more intensely than before. The new trends in modern pragmatics (represented by the pioneer works of Austin, Searle, Grice and Sperber/Wilson), have significantly widened the scope of modern linguistics. Consequently, the complexity of the object of study in linguistics has increased.

This complexity within the field of LSP is described in Baumann (1994) and Schröder (1993). Baumann's integrative, multidisciplinary approach to LSP texts is decomposed into eight dimensions of investigation: the intercultural, the sociolinguistic, the psycholinguistic, relating to the branches of science, the functional, the textlinguistic, the linguostylistic and the lexical semantic. In varying degree these dimensions are all relevant to our project.

But in text typology studies there is also need for a theory of genre. The relationship between text typology and genre has been defined in different ways. In my project I follow Biber (1988). Biber defines text typology as a classification of texts based on formal, linguistic criteria (e.g. complex texts, simple texts, nominal texts, verbal texts etc.). Genre theory on the other hand will give you a classification of texts based on "external" criteria (e.g. fiction, conversation, scientific texts, panel discussions etc.).

Within LSP Swales (1990) has studied the concept of genre and developed a theory of genre over several years. As a theoretical basis for his genre concept he makes a crucial distinction between the well established sociolinguistic concept of speech community and discourse community. Discourse communities in Swales' sense typically apply the written channel of communication and have a sociohistorical and functional basis. More significantly, discourse communities have a centrifugal effect (as opposed to the centripetal effect of speech communities), i.e. discourse communities separate groups of people according to special occupational and interest groups, whereas speech communities absorb people into a set of social patterns. Moreover, discourse communities have their own set of mechanisms of internal communication among their participants. Consequently, discourse communities will apply their own set of participant mechanisms primarily in order to convey information and feedback. One or several genres are applied in order to reach these communicative goals in a given LSP domain.

Text typology theory and genre theory will provide the basis of my part of the project.

But text typology studies, in order to be successful, will to a large extent be dependent on linguistic statistical analysis. In Baumann (1994) there is a survey of how statistical analysis may be applied in his eight dimensions (mentioned above). Biber (op.cit.) uses both a quantitative and qualitative approach to textual variation. His use of statistics is fairly advanced (using factor analysis), but extensive quantitative studies of texts in this manner presuppose availability of textual material in a coded form (e.g. SGML-coding). In our project, which is in its initial stage, indeed, extensive SGML-coding has not been possible so far.

2. Problem specification and general hypotheses

Three basic areas of activities of the oil industry offshore are investigated as to text typology variation: The activities associated with drilling, production and maintenance.

The initial hypothesis to be tested is that text typological variation in these three fields can be related to two sociological dimensions: Degree of social status (high vs low) and degree of social network density (high vs low) (following Milroy 1980). Finally, there is an LSP-related dimension: degree of professional specialization (measured by term density in texts).

A major problem is how to correlate sociological dimensions with linguistic features. The variable concerned with term density is the simplest one, but problems relating to term identification in texts and identification of legitimate contexts cannot be overlooked. These problems are not discussed here.

The two sociological variables are more complex. This has several reasons. Firstly, the concept of social status is complex and has to be decomposed into several subcomponents.

Several studies show that high frequency of linguistic features like nominalization and passive often indicate professionalism in texts. To some extent these features are connected with agent suppression, a pragmatic communicative strategic device which may be used to create distance between the author and his message. The effect of this is often a downtoning of the non-descriptive meaning (i.e. the emotive and the

expressive meaning) of the message. Moreover, strongly nominalized texts tend to give the readers a static impression, which is typical of "objective" descriptions and explanations of facts.

In my pilot project I have looked at term density and NP density in the three areas of activity mentioned.

3. Data from the pilot project

If texts from the three different areas are to be compared text typologically in terms of a set of linguistic features they should preferably represent the same genre. Moreover, they should be of equal length, especially if the frequency of features is investigated.

The data for my initial investigation was taken from the manuals of the Gullfaks A platform. These manuals represent the first stage of planned normative terminology work connected with the use of Norwegian in the off-shore industry. As the first test I chose the first 1000 words (in terms of lexical, terminological items, not word forms) of three manuals, each representing the three activity areas mentioned: A description of the drill string for the drilling area, a description of the glycol system for the production area, and a description of a 50 ton crane for the maintenance area. The texts chosen all represent the same genre: system descriptions.

My hypothesis concerning term density predicts that drilling will have the highest density, maintenance will have the lowest density and production will have a medium density. Drilling activities are typically located at a specific part of the platform and the personnel work in close proximity to one another. The characteristic terminology associated with drilling is extremely specialized and almost impenetrable to outsiders. Maintenance work, on the other hand, is spread throughout the platform, and workers typically carry out their work individually and isolated from one another. The terminology of maintenance is much less field specific than drilling terminology. Production work is in a mid position both as to degree of specific work location and degree of specialized terminology. It is also expected that these salient differences will correlate with dimensions of text typological variation.

As indicated in figure 1 this seems to be the case in my data:

figure 1: Term density for first 1000 words:

per 100 words	Drilling	Production	Maintenance
000-100	14	11	5
100-200	4	5	3
200-300	6	9	3
300-400	13	3	7
400-500	2	6	3
500-600	6	7	4
600-700	11	4	2
700-800	12	8	3
800-900	9	4	1
900-1000	9	1	2
sum	86	58	33

The score for drilling is 86 terms for the first 1000 words, production has a score of 58 and maintenance is down at 33. The figures seem to be rather clear, but the hypothesis will have to be strengthened through the study of further data.

If there is an identifiable connection between nominalization and high status, data should reveal that drilling will have highest score on nominalization, production will have a medium score and maintenance will have the lowest score. In order to test this I extracted all the NP sentence constituents of the three texts and made a count to show NP density, as illustrated in figure 2:

figure 2: NP density for first 1000 words:

per 100 words	Drilling	Production	Maintenance
000-100	17	14	13
100-200	11	10	11
200-300	14	13	11
300-400	15	8	16
400-500	17	12	25
500-600	14	15	18
600-700	9	8	15
700-800	13	13	12
800-900	10	14	14
900-1000	11	16	16
sum	131	123	151

As can be seen from the sum of the scores, this hypothesis is only partially strengthened: The score for drilling is higher than for production, as predicted, but the score for maintenance is not the lowest, as we would expect, but in fact the highest. How can we account for this?

If term density is a reliable variable for degree of specialisation, we would expect the terms of drilling and production to be more specific and (in technical domains) more concrete than terms dealing with the maintenance and security aspects of these activities. Moreover, we would expect that many abstract terms relating to types of malfunctions and types of security actions would be described in a rather general manner. Thus, we would expect a higher number of abstract terms in the maintenance area than in the two other areas. As figure 3 shows, this seems to be the case:

figure 3: Abstract/concrete distribution for first 1000 terms:

Drilling	Production		Main-tenance		concrete	abstract
	concrete	abstract	concrete	abstract		
000-100	12	2	7	4	1	4
100-200	3	1	4	1	0	3
200-300	3	3	8	1	0	3
300-400	9	4	2	1	4	3
400-500	2	0	6	0	3	0
500-600	5	1	6	1	4	0
600-700	8	3	1	3	2	0
700-800	12	0	7	1	2	1
800-900	5	4	2	2	1	0
900-1000	3	6	1	0	0	2
sum	62	24	44	14	17	16

As can be seen, there is a significant majority of concrete terms both in the drilling and in the production area, whereas the number of abstract terms in the maintenance area is as high as the number of concrete terms. A concrete term is defined here in a very restricted manner, referring exclusively to physical objects. All other terms are counted as abstract, including descriptions of observable phenomena (like pressure), or non-observable phenomena (like security).

The high NP score for the maintenance area may be due to the fact that many abstract and general security terms are introduced and repeatedly referred to (by typical referring expressions like NPs).

However, the question of NP density must be looked further into. Amongst other things it is likely to be connected to degree of NP complexity in texts. NPs can be analysed by applying conventional labelled bracketing. The complexity of these structures can be measured by the ratio of words per NP and by the degree of NP internal nesting. As is well known NPs both in Norwegian and in English (and perhaps in any language) the internal structure of NPs can be very complex. This complexity may be an indicator of the degree of specialization of domain specific texts.

I have done some preliminary investigation work on internal NP structure of these texts, but it is not completed yet. The results so far indicate a considerable variation in complexity, both in the sense of number of words per NP and in the sense of degree of nesting. This variation is, however, probably due to the fact that introductory parts of texts were chosen for my pilot project. Here more data is needed before more conclusive evidence can be given.

4. Further investigation and expected results

This scanty evidence relating to term and NP density is sufficient to give a certain basis for my initial hypotheses. Thus, it is likely that the term density variation to some extent correlates with degree of social network density in the three areas of investigation.

A discourse group like the drilling crew on a platform at sea may perhaps be described in terms of what Milroy calls "first order zone network" (Milroy 1980), where everybody knows everybody else and socialize with each other, not only at work, but also in leisure hours at the platform. In such a dense social network it is possible to identify the various variables of the network. One possible variable is the collegue factor, i.e. working together at the platform. Common professional background, being educated at the same institution, having the same age and sex are other examples of possible variation likely to influence text typology.

Text typologically, these social variables may be describable in terms of Chafe's distinction between *integration* (such as frequent use of nominalization) and *fragmentation* (such as high frequency of asyndetic coordinated clauses), and his distinction between *detachment* (such as high frequency of the passive) and *involvement* (such as high frequency of first person pronouns).

But the most important thing is to take very cautious and small steps in this process.

Reference list

Baumann, K-D. (1994) The significance of the statistical method for an interdisciplinary analysis of professionalism of texts, in Brekke M., Andersen Ø, Dahl T. and Myking, J. (eds.) *Application and Implication of Current LSP Research* volume 1. Fagbokforlaget. Bergen..

Biber, D. (1988) *Variation across Speech and Writing*. Cambridge University Press. Cambridge.

Milroy, L. (1980) *Language and Social Networks* Oxford University Press. Oxford.

Schröder, H. (ed.) (1993) *Fachtextpragmatik* Günther Narr Verlag Tübingen.

Swales, J. (1990) *Genre Analysis. English in academic and research settings* Cambridge University Press. Cambridge.