

Textual Variation in a Professional Setting

Øivin Andersen, University of Bergen

ABSTRACT: This project is a textlinguistic investigation of documents related to three central areas of professional activity on Norwegian offshore oil platforms: drilling, production and maintenance. In this paper I want to present aspects of theories from different sectors of linguistics seen in the light of a set of hypotheses which were tested in a pilot project. Documents produced by professionals in these three areas seem to display consistent textual variation. The explanation of this variation has (at least) two possible sources (both having sociocultural implications): Firstly, variation can be explained as a function of differences in the special character associated with these three knowledge domains (in a wide sense). Secondly, variation can be explained as different degrees of influence of the English language on Norwegian. At this stage of the project the former source is focussed on.

The investigated variables were textual term density (ratio of number of terms and number of word forms in texts), NP density (ratio of number of NPs and number of word forms in texts) and NP internal complexity (levels of nesting and number of words contained within NPs in texts).

The initial set of hypotheses predicted that non-linguistic factors such as social status, social network relations and degree of professional specialization would correlate with identifiable linguistic variation in texts from these three areas of activity. These hypotheses were strengthened in the pilot project.

Relevant theoretical aspects in the field of terminology (term identification), LSP (genre analysis), sociolinguistics (social network theory) and text linguistics (text typology, and speech act theory) will be presented in the context of my project. In the light of this some possible additional variables will be suggested.

Background

My project is part of a more general, interdisciplinary project called *The North Sea: A Testing Ground for Linguistic and Cultural Hypotheses*. The aim of the project is to complete a multidisciplinary study of the linguistic Norwegianization process that has taken place on the Norwegian Continental Shelf over the last fifteen years, i.e. from the first part of the 1980s. The pioneering activities on the Norwegian Continental Shelf relied to a great extent upon the use of English. Following the decision of the state-owned oil company, Statoil, to implement Norwegian as the working and administrative language on the Gullfaks oil field with effect from 1985, Norwegian has been adopted on several other oil fields and also in other petroleum companies. English is therefore no longer the sole language of the Norwegian petroleum industry.

During the initial stages of the Norwegian petroleum activities there were many people who considered the use of English as a cultural threat to the Norwegian language in particular and to Norwegian culture in general. Opinions differed greatly, but it would be fair to say today that the most pessimistic views were somewhat exaggerated.

The general object of enquiry of our interdisciplinary project is the study of the consequences of the enforcement of a planned Norwegian terminology especially designed for the purpose of LSP technical communication in the offshore industry. This is studied from a sociological, a linguistic, a text-typological and a sociolinguistic point of view. The basic questions we ask ourselves are the following:

1. Which sociological mechanisms underlie the distribution of Norwegian terminological equivalents contrasted with the English original terms? Do the companies operating on the Norwegian Continental Shelf have a deliberate language policy?
2. Which attitudes to questions of terminology and Norwegianization have been typical of the Norwegian oil workers? Have they been positive to Norwegian terminology?
3. Which properties are characteristic of the Norwegian terminological equivalents contrasted with the English original terms? Which practical working principles of terminology have had the greatest influence on the Norwegian language use offshore?
4. The central question in my own project is the following: Can variation in the structure of documents be identified and explained by referring to the socio-cultural features of the discourse communities involved?

Introduction

As the first step in the process of answering this question, I have had a closer look at indicators of status and specialization in technical texts from three areas of activity: drilling, production and maintenance. I will give a brief outline of my findings and offer a possible explanation of the data in terms of semantic and pragmatic theory in the field, and concentrate on a special problem which I intend to investigate further.

The Pilot Project

My initial hypothesis was that text typological variation in the three fields of offshore activity (drilling, production and maintenance) can be related to two sociological dimensions: Degree of social status and degree of social network density (following Milroy 1980).

As LSP research has pointed out, high frequency of coding features like term density, nominalization and the passive often indicates professionalism in texts. In a pilot project, presented at the LSP symposium in Vienna in 1995, I tested the coding features term density, NP frequency and NP complexity in texts from these three domains. All the texts will have a high score when compared with comparable text types written in a nonspecialist language context,

but it turned out that texts from the drilling area had the highest score on all three coding features, the maintenance text had the lowest score and the production text had a medium score. This is taken to indicate that high social status and high network density (both typical of drilling activities offshore) seem to correlate with high scores on coding features like term density, NP frequency and NP complexity.

Explanation of Data from the Pilot Project

Some important aspects of these findings can be explained within the framework of the well-established sociolinguistic terms *speech community* (Labov, 1966 and Hymes, 1974) and *linguistic network theory* (Milroy, 1980) on the one hand, and the genre oriented concept *discourse community* (Swales, 1990) on the other hand.

Swales' concept is the basic foundation for his definition of genre. The concept of discourse community is useful, but it is also important for me to distinguish between *genre* and *text type* (following Biber, 1988). Genre refers to the classes of texts that are determined on the basis of external criteria relating to author's or speaker's purpose/intention (i.e. descriptions, instructions, etc.). Text type, on the other hand, refers to classes of texts that are grouped according to similarities of linguistic form (i.e. nominal texts, verbal texts, etc.).

According to Swales, the primary determinants of linguistic behaviour in a speech community are social, whereas the primary determinants in a discourse community are functional. Speech communities are centripetal (they tend to absorb people into a fabric with high network density), whereas discourse communities tend to be centrifugal (they tend to separate people into occupational or speciality interest groups).

The basic manifestation of these two groupings is the medium: Speech community characteristics are primarily studied in terms of oral interaction, whereas discourse community characteristics are primarily studied in written documentation (op.cit.:24f).

As Swales points out, literacy may take away locality and parochiality, but in some cases literacy may have the opposite effect. In some specialist contexts with high social network density (i.e. where many people know many other people), written documentation may underline and perpetuate special features associated with their oral professional communication in their day-to-day job situation.

The offshore drilling activity seems to be an example of this. Drilling activities are typically located at a specific and restricted area of the platform, and the personnel work in close proximity, i.e. within earshot of one another. In terms of network density and locality, the drilling personnel may be called a sub-speech community on the platform. Because of their high degree of professional specialization, their social status is high, and their terminology is very specific and almost impenetrable to outsiders.

The maintenance workers, on the other hand, have a relatively low status. Network density is low in this group, because they carry out their work individually and isolated from one another. The terminology of maintenance is

much less specialized than drilling terminology, whereas production workers again occupy a mid position.

A central point here is that to a large extent all three groups have to document their future and past activities by writing technical manuals describing the mechanical design and use of technical installations.

Consequently, the three groups may also be called discourse communities, but only the drilling workers also have typical speech community characteristics.

Degree of professionalism in texts is a genre study, but in order to elucidate these textual features, a method of text typology is needed. Biber (1988) has studied central differences between speech and writing in a functional perspective. It is a well-known fact that texts are related along particular situational or functional parameters such as formal/informal, restricted/elaborated, literary/colloquial etc. Biber terms these parameters dimensions, because they are ordinal variables, defining continuums of variation. By relating a number of functional dimensions with linguistic codings in texts, Biber claims that it will be possible for researchers to determine how many linguistic dimensions are required to account for the variation among a set of texts. Biber's method is also a discovery procedure for co-occurrence patterns of clusters of features.

In my project, the functional dimensions *degree of social status* and *degree of social network density* are linked to the linguistic coding features term density, NP frequency and NP complexity.

It is also a well-known fact that nominalizations and passives tend to co-occur in texts, especially in scientific, technological and expository texts.

Further Linguistic Features

So far we have discussed NP-like constructions. Many NPs are derived from verbal constructions in technological texts like the ones I investigate. The study of argument structure has mainly focussed on arguments associated with verbs, cf Jackendoff (1987). It has often been suggested that an important difference between nouns and verbs is that verbs may take obligatory arguments, whereas nouns only take optional arguments.

Grimshaw (1990) claims that nominals can take obligatory arguments. She makes a distinction between complex event nominals, having an obligatory argument structure, and simple event nominals and result nominals, having no argument structure. The distinction between the different types may at times be difficult to draw, and one and the same nominal may at times be interpreted as a complex event nominal and as a simple event nominal according to context. Besides, the distinction between simple and complex events is often difficult to make.

However, the most important issue here is that the external argument *subject* is obligatorily expressed with finite verbs, or, to put it in more modern terms: the obligatory presence of a non-empty external argument (i.e. the subject) is governed by the tense/aspect character of the finite verb in a clause.

Here we have to distinguish between argument structures, which are syntactic expressions, and semantic or thematic roles which are semantic functions of the arguments. It is widely agreed that argument structures are constructed in accordance with a thematic/semantic hierarchy (cf Jackendoff 1972). Grimshaw assumes a version of the hierarchy in which the agent is always highest in the hierarchy. This means that if we have an agentive verb and if an agent is available it will typically fill the external argument position (i.e. the subject function). With some agentive verbs, like *move* and *open*, the subject function may be taken over by the goal role, as in (2):

- (1) X opens the valve manually.
- (2) The valve opens at high pressure.

In (2) the agentive verb *open* is used non-agentively, because there is no agent available for the subject position. The prototypical subject in languages is claimed to be agentive, because the prototypical verb is agentive. But in technological and expository texts agentive constructions like (1) are very rare.

It has been attested that professional texts of this type are characterized by impersonal non-agentive constructions of various types. In addition to (2) we have a high frequency of constructions like (3) and (4):

- (3) The valve is opened manually.
- (4) The opening of the valve reduces pressure.

In (3) and (4) the agentive role is available for syntactic expression, in contrast to (2), and may optionally be expressed, as in (5) and (6):

- (5) The valve is opened manually by x.
- (6) X's opening of the valve will cause a pressure reduction.

Moreover, the prototypical subject is topic, expressing given information with widest scope.

From a pragmatic point of view, it is an obvious fact that agentivity cannot be expressed in an obvious manner if it is unknown. If it is known, it is very often peripheral, or unimportant to the thematic progress of the text. Since the typical subject is both agentive and topical, it would be a natural strategy for the producer of technical texts to suppress the expression of agentivity in agentive constructions. Since agentivity tends to be expressed by an obligatory subject in active agentive clauses, agent suppression is impossible to achieve with these constructions.

The most obvious way of realizing agent suppression on the syntactic level is to use constructions which make agent suppression possible. The most productive of these construction types is nominalization and the passive.

This fact may explain the static impression of these text types. From a pragmatic point of view nominalization is not only an agent suppressing device, but also a strategy for transforming dynamic content into static content. As Lyons (1977:174ff) points out, nominals are typical referring expressions. In technical texts making extensive use of descriptions, activities are described as

processes where agentivity is backgrounded. This will explain the use of the non-agentive use of 'ergative' verbs like *open* in (2). The Norwegian equivalent *åpne* cannot be used non-agentively. Consequently, in the corresponding Norwegian texts we find corresponding non-agentive constructions like the passive (*åpnes*) or the non-agentive medial-reflexive (*åpne seg*).

When these processes are realized as nominals, they can easily be referred to by anaphorics and nouns used anaphorically. The static impression of the texts can then be explained by the fact that prototypical nominals refer to countable entities or objects, including deverbal nominals denoting activities packed down as entities.

The static and descriptive character of many professional texts may also explain the fact that most technical terms are nouns. The verb *to drill* is not a term in drilling terminology. Instead we have a multitude of terms referring to various types of drilling activities, as in (7):

- (7) deviation drilling, jet drilling, rotary drilling, turbine drilling, etc.

In order to investigate these phenomena further, I am going to have a closer look at the thematic roles of the subjects in descriptive technical texts relating to the three areas of activity (drilling, production and maintenance). A more comprehensive study of the argument structure of nominal and passive constructions will be conducted with a special eye to agent suppression. I do not expect to find passives and nominals with expressed agents as in (5) or (6), and I suspect that they will be unacceptable in context even if they are perfectly grammatical.

There is thus reason to believe that agent suppressive linguistic coding properties like nominalization, passive and the non-agentive use of ergative verbs co-occur in technical texts to form dimensions indicative of professionalism of some kind.

But description is only a subtype of genre in technical texts. Technical manuals also contain many instructions. The imperative mode is a typical and central type of construction in these texts. The possibility of agent suppression is a common property of nominalizations, the passive and the imperative. A special feature of the imperative is that the agent is suppressed because it can normally easily be inferred from the situational context in an unambiguous manner. Agentivity is carried by strong conventional implicatures. Typical examples are (8) and (9):

- (8) Avoid rough handling...of the pipe.
(9) Do not unload pipe by dropping.

Instructions like (8) and (9), referring to the handling of the drill string, will have the reader as the potential agent. There is no imperative passive in Norwegian and English, but many other agent suppressive devices can be used to instruct. The passive infinitive with a modal auxiliary is a typical device, as in (10) and (11):

- (10) Care should be taken to prevent chafing of tool joint shoulders on adjacent joints.
(11) Thread protectors must be installed on both ends of pipe.

There are a number of additional linguistic features which may enter the same cluster of dimensions in these types of texts. Impersonal constructions (like presentation) will be a case in point. Use of verbal tense is another candidate. It is fairly well known that the use of present tense in technical texts functions to focus on information being presented, and remove focus from any temporal sequencing (as opposed to the use of the past tense where temporal sequence, typical of for instance narrative texts, is focussed).

The Corpus

In order to conduct this kind of research a significant electronic corpus is required. In my project we have access to the system manuals describing Gullfaks A and B. The Gullfaks A contains 67 manuals describing various sub-systems on the platform, both in English and Norwegian. Gullfaks B contains 52 manuals in both languages. Each manual has about 150 pages on average. These manuals will have to be scanned and SGML-coded for the purpose.

Moreover, the terms of these texts have been excerpted and collected in a petroleum terminology database, containing 90 000 searchable terms representing more than 30 000 terminological concepts.

This data base is currently being updated and further refined. It is planned that the deverbal nominalizations will be furnished with indication of argument structures.

To conclude, the interplay between genres and text types reveals that various types of linguistic codings both on the lexical level and the morpho-syntactic level have the function of serving a comprehensive macrostructural intention on the part of the producer of texts. The various coding features, like nominalizations and the passive, form a part of the text producer's strategy to realize his/her intentions by an act of successful written communication.

References

- Biber, D. 1988. *Variations across Speech and Writing*. Cambridge: Cambridge University Press.
Cambridge.
- Grimshaw, J. 1990. *Argument Structure*. Cambridge: The MIT Press.
- Hymes, D. 1974. *Foundations in Sociolinguistics. An ethnographic Approach*. Philadelphia: University of Pennsylvania Press.
- Jackendoff, R. 1972. *Semantic Interpretation in Generative Grammar*. Cambridge: MIT Press.
- Labov, W. 1966. *The Social Stratification of English in New York City*. Washington DC: Center for Applied Linguistics.
- Lyons, J. 1977. *Semantics*. Cambridge: Cambridge University Press.
- Milroy, L. 1980. *Language and Social Networks. Language in Society*. 2. Oxford: Basil Blackwell.
- Swales, J. 1990. *Genre Analysis. English in Academic and Research Settings*. Cambridge: Cambridge University Press.