

Corporate Voice, Tone-of-Voice and Controlled Language Techniques

Lina Henriksen, Bart Jongejan, Bente Maegaard

Center for Sprogteknologi, University of Copenhagen
Njalsgade 80, DK-2300 Copenhagen S, Denmark
{lina, bart, bente}@cst.dk

Abstract

Corporate documentation and publications should signal the company's key personality values. In this paper we show that the concept of Controlled Language may be extended to also cover Tone-of-Voice, and we see similarities and differences between Readability and Tone-of-Voice.

Background and Purpose

Introduction

Many companies are realising the opportunities offered by language technology for supporting and streamlining document production, not only at the level of using correct terminology etc, but also at the level of supporting the right Tone-of-Voice of the corporate language. The present paper presents results in this area achieved in the Danish VID project (Viden og Dokumenthåndtering med sprogteknologi – Knowledge and Document Handling with Language Technology).

VID is headed by Center for Sprogteknologi (Centre for Language Technology). The project group includes three large Scandinavian companies with high demands for quality and efficiency regarding document production, as well as two Danish technology companies specialised in search and knowledge organization participating as technology providers.

The aim of the project is to examine, develop and/or refine HLT techniques for acquiring and representing relevant parts of domain knowledge and corporate language in the participating companies, specifically oriented towards information retrieval as well as efficient and high quality production of documents.

Background

Two of the Danish companies mentioned above, namely Bang & Olufsen and Nordea, are both working in the domestic as well as the export markets, and they need to express themselves in many languages, their corporate language being English.

The companies want automatic support for their writing of user documentation, corporate communication texts etc.

Both companies aim at high quality texts. Both companies are high profile, and they have developed a corporate image which should be supported by the language used in the written documentation. I.e. if the company image signals 'solidity', 'loyalty', 'quality', 'fun', 'creativity' or 'design', such characteristics should also be signalled by the language of their documentation.

Objectives

Consequently, the aim of this study is twofold: on the one hand a tool is sought which can help produce high quality text, with the amount of consistency, clarity etc. necessary, and on the other hand a tool is sought which can support the specific corporate voice of each company.

The challenge is to combine these two objectives in one tool. The type of tool which is the result of this investigation is depending on Controlled Language (CL) techniques to a very high degree. However, the language it defines, is not an ordinary CL.

Controlled Languages

A CL is a subset of a language with a restricted grammar and a domain-specific vocabulary. It is used by writers to achieve a consistent and clear style and terminology within a particular text type.

Traditionally CLs enhance simplicity with for example simplified sentence structures and sometimes a very constrained vocabulary where the semantics of each word is limited to one meaning. The primary goal of CLs is to improve readability (also for foreign readers) and sometimes also facilitate translation.

AECMA¹ SE is an example of a CL guide. AECMA was developed in the 80'ies and is a list of writing rules originally used by the aerospace industry to produce manuals and other documentation. Later the guide has been used as a basis for other CL guides and tools (there are several tools on the market today developed on the basis of AECMA).

AECMA SE is classified as a Human-Oriented Controlled Language (HOCL)² and the aim of AECMA SE is to produce clear, precise and unambiguous texts.

Tone-of-Voice

Traditional CL rules and Tone-of-Voice rules are usually not integrated in one single authoring guide or tool. One reason is that CLs are - at least traditionally - used in connection with documentation and manuals where readability is considered the prime factor. Another reason is that Tone-of-Voice phenomena are not universal by nature, but very company specific.

Whereas CL rules are usually concerned with simplicity, Tone-of-Voice rules are concerned with emotion and not necessarily simplicity. It is often the words and phrases surrounding the core message that constitute an original Tone-of-Voice and cutting away these elements may simplify the text, but may also mean a loss of emotion and nerve.

¹ AECMA is the acronym for Association Européenne des Constructeurs de Matériel Aérospatial

² A HOCL's primary goal is to improve readability whereas the primary goal of Machine-Oriented Controlled Languages (MOCLs) is to improve translatability

Tone-of-Voice in this context concerns definition of a style of communication reflecting a company's corporate image taking the receiver and type of message into consideration as well.

Text production, two case studies

The companies mentioned above, Bang & Olufsen and Nordea operate within very different fields. Nordea is a financial services group in the Nordic and Baltic Sea region and Bang & Olufsen is a manufacturer of high quality audio/video/lifestyle equipment. As mentioned both companies wish to develop a tool supporting the production of high quality texts expressing their specific corporate images.

Readability Rules

In the VID project these companies started out by defining rules for "good language". They did this on the basis of their own texts using their long experiences as text- and copywriters, but without research-based or formalized knowledge of the field. The primary goal of this task was identification of rules that improves readability.

The companies developed about 40 main rules (some with several sub-rules) which can be classified into the categories of lexical, syntactic and textual rules (as proposed by O'Brien (2003)).

The below classification into sub-categories is inspired by O'Brien (2003) and gives an overview of the rules allowing us to compare the two rule sets developed by the companies.

Sub-categories of lexical rules	Applied by company
Restrictions on use of vocabulary	Bang & Olufsen and Nordea
Restrictions on use of abbreviations and acronyms	Bang & Olufsen and Nordea
Rules regarding currencies and numbers	Nordea
Restrictions on use of superlatives	Bang & Olufsen and Nordea
Ruling out word division	Bang & Olufsen
Rules regarding use of pronouns	Bang & Olufsen and Nordea
Rules regarding date and time format	Nordea
Restrictions on use of verb forms	Bang & Olufsen and Nordea
Restrictions on use of slang, idioms and metaphors	Bang & Olufsen

Sub-categories of syntactic rules	Applied by company
Restrictions on use of verb tenses	Bang & Olufsen and Nordea
Rules regarding use of modals	Nordea
Restrictions on use of participles	Bang & Olufsen and Nordea
Rules regarding use of prepositions	Bang & Olufsen and Nordea
Rules limiting the number of subordinate clauses	Bang & Olufsen

Restrictions on the length of noun phrases	Bang & Olufsen and Nordea
Rules regarding sentence structures	Bang & Olufsen and Nordea
Restrictions on the use of passive	Bang & Olufsen and Nordea

Sub-categories of textual rules	Applied by companies
Restrictions on sentence length	Bang & Olufsen and Nordea
Rules limiting the use of parentheses	Bang & Olufsen
Restrictions on paragraph length	Bang & Olufsen and Nordea
Restrictions on information load in a sentence	Bang & Olufsen and Nordea
Restriction on length of headings	Bang & Olufsen and Nordea
Rules urging the author to be clear and precise	Bang & Olufsen and Nordea

As an example one of the rules controlling the use of pronouns can be mentioned. This rule says that the personal pronouns as e.g. *you* and *we* are preferred to *the customer*, *the company* and other third person paraphrases as these paraphrases tend to make the text more complicated. For the same reason, the pronoun *one* should not be used.

Comparison of Rule Sets

It is an interesting observation that the two companies developed nearly identical rule types and even the exact contents of each rule set are very similar. A few differences: Nordea specifies rules dealing with formats of currencies, numbers and dates in detail, and Bang & Olufsen is a bit more restrictive regarding sentence structures.

Another interesting observation is that most of the above rules are identical or very similar to the rules of the AECMA SE guide though the companies had no previous knowledge of AECMA.

Tone-of-Voice

The companies' corporate images are however not adequately expressed through readability and consistency alone. Expressing a corporate image as an inherent part of a text requires a Tone-of-Voice analysis. The factors that must be considered in this connection are:

- Sender
- Receiver
- Message

A definition of the Sender involves distillation of a company's key personality features. The personality of the Sender is static and constitutes the core of the Tone-of-Voice phenomenon. These personality features should be expressed in the texts through choice of words, textual and syntactic structures, and possibly through other means.

The Receiver and Message are dynamic elements as the tone used in a text to a colleague may be very different from the tone of a text to a shareholder. Similarly, the tone used in a reminder could be very different from the tone of a press release or an annual report. The Receiver/Message

types determine which personality values will dominate the text and to which degree they will dominate.

Identification of Sender

The companies already had well-established and very clear descriptions of their corporate images, i.e. their values, tasks, goals, missions and the promises they want to make to their customers and the society.

Nordea's key personality values are:

- Nordic
- Informal
- Committed to People
- Straightforward
- Embraces Change and Challenge

Bang & Olufsen's key personality values are:

- Excellence
- Originality
- Passion

Identification of Receiver and Message

For a given Sender, the Receiver and Message are to some degree interdependent. Some types of messages are only used in connection with some types of receivers and vice versa. This was very much the case in the VID project and we therefore identified the decisive factor for each company. Nordea's focus is on the Receiver: International Customers, Internal Staff, Investors and Public at large. Bang & Olufsen's focus is on the Message: User Guides, Training material and Marketing material.

Corpus analysis

We performed corpus analyses to support our investigation of Tone-of-Voice.

The companies selected a number of texts (Nordea: 196748 words in 92 texts, Bang & Olufsen: 87301 words in 25 texts) which were considered good texts with the right Tone-of-Voice. The texts were created for different media: for printing on glossy paper, in brochures or in booklets, or for electronic delivery via Internet or slide shows. Many documents are richly illustrated and care has been taken for a good text layout. All these aspects that certainly contribute to expressing personality values are of course lost during a corpus analysis.

The texts were organized in the above categories. Then the texts were converted to a common format that the analysis tool supported. The texts provided by Bang & Olufsen are in PDF-format, but Nordea's texts had several formats: HTML (50), PDF (39), Word (2) and PowerPoint (1). All texts were converted to RTF-format, thereby retaining as much text layout information as possible, and assembled into text corpora, one corpus per category. The layout information could in many cases help to find sentence boundaries. Although we were only interested in words, a proper segmentation into syntactically coherent text chunks is important for a good part of speech (POS) tagging.

We used Eric Brill's part of speech tagger with rule files trained on the Brown corpus. The resulting POS tagged corpora were further analyzed. Frequency lists were created arranging adjectives, nouns, pronouns and verbs in POS groups. Word counts included all kinds of text tokens, except punctuation characters and parentheses. We accepted numbers as words. The tables below illustrate how words are distributed over the four POS for each

category and company. The most distinctive figures are the low frequencies of pronouns in the categories Investors (mostly annual reviews and quarterly reports) and Public at large.

Nordea:

Category	# words	Adj %	Noun %	Pron %	Verb %
International customers	13841	9,28	29,53	4,84	14,79
Internal Staff	16414	7,61	25,76	6,14	18,52
Investors	118170	8,12	25,86	0,68	11,63
Public at large	48323	8,72	24,83	1,20	12,53

Bang & Olufsen:

Category	# words	Adj %	Noun %	Pron %	Verb %
User Guides	63170	5,91	23,98	4,14	17,21
Training material	17057	6,81	25,43	3,72	14,94
Marketing material	7074	6,98	22,83	5,15	15,32

From these frequency lists we manually extracted value-laden words expressing the companies' personality values and each word was validated by checking the use and context in the corpus. None of the words extracted belonged to the companies' terminology collections as these do not contain this kind of value-laden words.

Nordea

In this first, explorative investigation we examined two receiver types: International Customers and Internal Staff, in both cases the volume of text was modest.

Through context-checking in the corpus it was established that the personality values are very nicely represented in the customer related material. Committed to People is expressed through words as *secure, efficient, reliable, trusted, true, thorough, benefit* and Embraces Change and Challenge through e.g. *forefront, innovative, future, idea, vision*. Informality is expressed through personal pronouns as *your, you, own* and through words as *personal, private*. For the Internal texts we found close to the same vocabulary, with one interesting exception: the frequent personal pronouns include *our, we, us* and *I*.

Bang & Olufsen

Similarly, for Bang & Olufsen we examined two text types: Marketing material and User Guides. The personality values came very nicely through in the Marketing material. Excellence is expressed through words as *perfect, exact, flawless, exclusive, outstanding, spectacular*. Passion is expressed through words as *enjoy, magic, dreams* and through the use of personal pronouns *you, your, we* (as passion in this context is also about creation of a close relationship to the customer). Originality is expressed through a long list of words as *idea, new, unique, original, intuitive, progressive, stunning* and *striking*.

In the other text type, User Guides, the direct visibility of the key values was slightly less. E.g. the concept of originality is not an important value in a user guide, so the list of 'originality words' is shorter, and the same goes for Excellence although to a lesser extent.

Analysis and comparison

The Personality Value and Receiver/Message Effect on Vocabularies

For both companies it is the case that the Receiver/Message combination has an effect on the vocabulary. The key personality values are expressed directly through the vocabulary in marketing material and customer related material, whereas other means are probably used to express the values in e.g. User Guides.

In the Bang & Olufsen User Guides, the personality value Excellence is best expressed through the excellence of the language, the text structure etc, i.e. the readability of the user guide. This is one very clear example where Tone-of-Voice is actually expressed through CL type rules.

Comparison of the Company Vocabularies

As the list of personality values is not the same for the two companies, it is to be expected that the words describing these values will be different. The extraction of such words confirmed this. Nordea have Committed to People as a key personality value and Bang & Olufsen have Excellence; these do bring out different words. It can be observed that most of the personality values are expressed through adjectives, and very few through verbs.

Use of Personal Pronouns

One POS we would like to focus on, is the use of personal pronouns in the second person. For Nordea this is part of the personality value Informal, and for Bang & Olufsen it is part of the value Passion. It is an important part of the way the two companies address their readers that they both want to establish a close and personal relationship with their readers/customers, and it is part of their writing rules that second person personal pronouns should be used. This is a very clear example of Tone-of-Voice, and yet it is also part of traditional CL rules. The reason it is part of CL rules is that in theory a text is more readable if it is direct. So, again we see that one rule can serve two purposes.

A closer look at the corpus and frequency information about pronouns reveals that the percentage of pronouns is more or less the same, with two notable exceptions as mentioned above: the share of pronouns in Nordea Investor material and Public material is smaller than average. Again this is an effect of the Receiver/Message combination. The Investor material and Public material are mostly annual reports and the like, and as such not meant to be personal.

Nordea Invest.			Nordea Customer			B&O User G.		
No	Oc.	Pron	No	Oc	Pron	No	Oc	Pron
39	220	we	1	250	your	1	1371	you
56	172	our	2	225	you	2	831	your
77	126	its	26	44	we	14	231	it

In fact, not only are pronouns less frequent in these text types, but different pronouns are used. *you* and *your* are normally the most frequent pronouns in all the text types analyzed, and this is in full agreement with the writing rules for Readability. However, in the two types of Receiver/Message dealing with annual reports etc, *our* and *we* are predominant as can be seen from the above table (the rank numbering includes only the POS studied):

Here we have a case where the Tone-of-Voice in a certain way overrules the writing rules for Readability. An authoring tool should be able to support the Tone-of-Voice of an annual report, and not request a high number of occurrences of *you*.

These observations stress the fact that Readability and Tone-of-Voice should be studied together and that the authoring tool should cover both, but at the same time it is important to distinguish the two phenomena and analyze them separately in order to understand them well.

Tone-of-Voice rules are not as intuitively and easily identified as readability rules. Another difference is that Tone-of-Voice rules are firmly anchored in the vocabulary and to a lesser extent in syntactic and textual structures. Tone-of-Voice aspects do however enforce the importance and necessity of CL rules.

Conclusion and future work

In this preliminary investigation we have shown that the CL techniques that traditionally serve an AECMA type simplified CL, can be expanded to serve a high quality, company specific corporate voice text production, and that a focus on personality values and their integration in the language of company communication can strongly improve the quality. The techniques that have to be applied are mostly already known from controlled language, as integration of Tone-of-Voice is primarily an expansion of the CL vocabulary rules.

Corpus work has shown that Tone-of-Voice can be distinguished and that it can be distinguished from Readability. An inclusion of a Tone-of-Voice feature will highly improve current authoring tools.

Future work will focus on the methodology to determine the vocabulary expressing personality values, and on other ways of expressing personality values.

Acknowledgements

We thank all VID-participants, but especially Sys Bundgaard and Lone Bræmer from Nordea and Anita Vestergaard and Anette Johansen from Bang & Olufsen and Dorte Haltrup, Costanza Navarretta and Bolette Pedersen from Center for Sprogteknologi for their helpful suggestions and comments. VID is funded by the Danish Research Councils.

References

- AECMA Simplified English, PSC-85-16598 A Guide for the Preparation of Aircraft Maintenance Documentation in the International Aerospace Maintenance Language Fairclough, Norman: Language and power, Longman, UK, 1989
- Kamprath, C, Adolphson, E, Mitamura, T and Nyberg, E: Controlled Language for Multilingual Document Production: Experience with Caterpillar Technical English, in: Proceedings of CLAW, Pittsburgh, 1998
- Mitamura, Teruko: Controlled Language for Multilingual Machine Translation, in: Proceedings of Machine Translation Summit VII, Singapore, 1999
- Ng, Sik Hung and James J. Bradac: Power in Language, Sage Publications, Newbury Park, 1993
- O'Brien, Sharon: Controlling Controlled English. An Analysis of Several Controlled Language Rule Sets, in: Controlled Language Translation, Proceedings of the EAMT-CLAW 03 Conference, Ireland, 2003