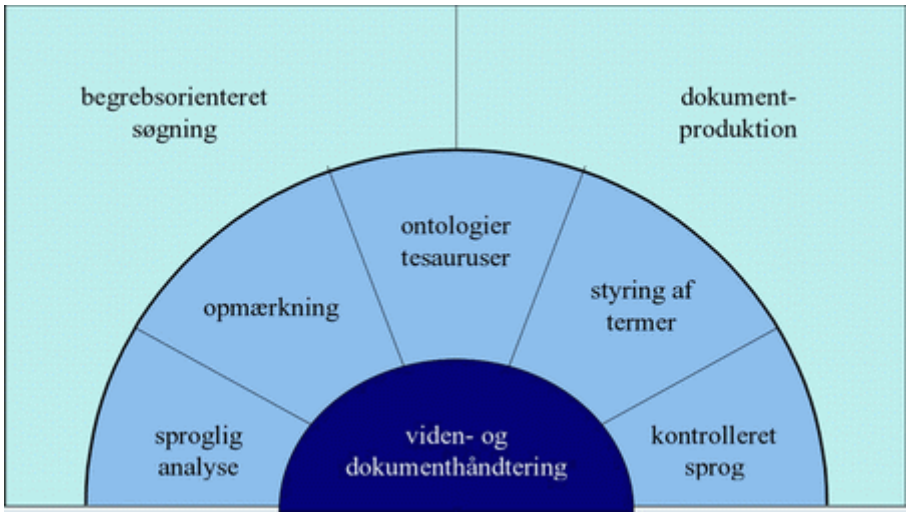

A comparison of HyperSE and Nordea's and Bang & Olufsen's requirements

Lina Henriksen, Bart Jongejan, Bente Maegaard

VID report no. 5

Center for Sprogteknologi, University of Copenhagen
December 2004

If you want to obtain the report, please contact CST, cst@cst.dk, or download from CST's web site by following this link: http://www.cst.dk/VID/.

The VID project is supported by Center for IT-forskning (now the Danish Research Agency).

VID: Knowledge and document handling with language technology

Danish companies have a marked need to supplement their existing linguistic competence and knowledge with language technology tools and methods. Such tools and methods should support the employees in their work, as well as make it possible to store knowledge about facts and processes in the company information systems, and form the basis for the evolution necessary for the survival and growth of the companies in an ever more global economy.

The VID project is an R&D project with the purpose of investigating the various possibilities offered by language technology for information retrieval and document production, and of supporting the participating companies in developing tools for better exploitation of their own knowledge, as well as in better and more efficient production of documentation, including multilingual documentation. In addition to CST, the project participants are 1) the companies Bang & Olufsen Audio Visual, Zacco A/S and Nordea AB who act as the technology users in this project, and 2) Navigo Systems A/S and Ankiro who are technology providers. The project comprises the following research tasks:

- Analysis of the text data the companies need to handle in order to create thesauri/ontologies for the relevant semantic domains, and investigation of the best suited formalism for representing these.
- Creation and further development of language technology components for automatic classification and concept based information retrieval, including tailoring of language technology basic modules for tagging of text.
- Research into ambiguities in texts, obstructing information retrieval; also the opposite problem: that the same content may be expressed in different ways and consequently be difficult to retrieve in large text corpora.
- Research into Controlled Language – also in a multilingual perspective – as a tool for document production, including an analysis of the style and Tone of Voice the companies wish to use, and a setting up of models for this sublanguage.
- Investigation of language technology methods suitable for the quality assurance of this aspect of document production, e.g. through terminology control and grammar control.

The project runs 2003-2004, and is supported by the Danish Research Agency.

Contents

1 Background

This VID report is the last report on the investigation of the document writing rules for Nordea Bank AB and Bang & Olufsen. As in previous reports, we discern between two types of document writing rules: readability rules (cf. VID report no. 1) and recommendations for obtaining the companies' Tone of Voice (cf. VID report no. 4).

During the project it became clear that many of the rules specific to these companies (hereafter also called "the company rules" for short) are similar to Controlled Language (CL) rules described elsewhere in literature and implemented in software. Therefore it seemed worthwhile to have a look at existing CL tools and see whether they have the breadth in their design to also encompass the company rules.

Many CL rule sets aim at avoiding complex language in written texts, so that any reader understands every aspect of the texts' content. The widely acknowledged set of CL-rules, AECMA[1] Simplified English[2] (AECMA SE), is a good example of such rules. AECMA SE is specifically designed for writing life-critical documents in the aerospace industry that are read by people for whom English is not their native language. There are several mature CL tools that support the AECMA rules.

Another group of CL rule sets try to restrict the use of language constructs that hamper automated translation. Such rules are not necessarily good for improving human readability. Therefore we have paid little attention to such sets of rules in the VID project and concentrated on AECMA SE instead.

Although in many ways similar to AECMA SE, the company rules brought forward in VID have a different purpose than AECMA SE. That does not mean that the resemblance of the company rules and AECMA SE is incidental. Many of the AECMA SE rules are generally acknowledged as promoting good language and therefore also embraced by the companies.

The difference between AECMA SE and the company rule set is as profound as their similarity. First and foremost AECMA SE allows, with a few exceptions, only one part of speech per word. For example in AECMA SE, *overlap* is a noun and therefore its use as a verb is not allowed. The companies do not in any way want this restriction, as the company rules mentioned above can only be implemented in a CL tool that allows for more than one reading of a word.

As mentioned above, control of a document's Tone of Voice is also important for Bang & Olufsen and Nordea. VID report no. 4, 'Tone of Voice', discusses rules that underlie

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

[2] AECMA SE was developed in the 1980s and is a list of writing rules originally used by the aerospace industry.

control of Tone of Voice and also gives some concrete recommendations for implementation of these rules in a tool?

The much greater freedom in the choice of words required by the companies also means that a CL tool must be able to handle much more extensive lists of approved and unapproved words than an AECMA SE complying tool. Also, these lists will be subject to large mutations. Therefore it is very important that a CL tool (or the tool creating the CL tool's dictionary) is able to import word lists from external sources, converting spelling, part of speech and other attributes to their internal equivalents, without the need to enter each word manually.

2 Introduction to HyperSE

In this report we will compare Nordea's and Bang & Olufsen's controlled language (CL) rules with the rules implemented in the HyperSE tool, and we also briefly discuss the additional requirements arising from Tone of Voice rules. HyperSE is an authoring tool developed by Tedopres[3] that focuses on controlled language in relation to readability, clarity and precision. HyperSE is partly developed on the basis of the AECMA Simplified English rule set.

Tedopres has provided CST with a basic version of HyperSE (version 2.0, July 2004) and of the Dictionary Maintenance Tool (DMT) (version 1.0, July 2004) for mutual inspiration: CST could get hands-on experience with a CL tool and develop ideas that Tedopres eventually might find interesting and implement in its products.

It is not the purpose of this activity to evaluate HyperSE. That would not even be possible, given the fact that HyperSE is not sold as a shrink-wrapped standardised product, but always is adapted to the needs of the customer. Rather this document describes which rules are present in the HyperSE tool, which are not and which rules might need customisation to fit Nordea's and Bang & Olufsen's needs. When we write that "rule *x* of company *Y* is not supported by HyperSE" it means that *x* is not implemented in the examined version of HyperSE, and not that *x* cannot be implemented in a version tailored to company *Y*'s needs.

HyperSE can be plugged into a number of commercial word processors. The version put at CST's disposal is a plug-in for Microsoft Word. The DMT is a stand-alone program that gives a sufficiently privileged user full control over the dictionary employed by all installations of the CL tool in a company. Text writers who use HyperSE will normally not have permission to use the DMT, but can use HyperSE to add nouns to a user dictionary. In this report the distinction between end users and 'power users' is not essential. For that reason we use 'HyperSE' as a shorthand to denote both the CL tool and the DMT.

This document discusses the company CL-rules in (more or less) the same order as described in the first VID report, "Kontrolleret sprog"[4] (KS). For ease of reference the paragraph numbers from the first report are cited in this document.

Note: 1) Below we are always referring to the HyperSE tool, but the comments would in general apply to any CL tool, with differences in the details. 2) As stated we are referring to HyperSE 2.0, July 2004, but of course new versions will be appearing and consequently some of the cases of 'Not in this version of HyperSE' will disappear.

---

[3] http://www.tedopres.com/

[4] http://cst.dk/vid/public/rapport-1-kontroll.zip

3 Lexical rules

(KS section 4.3)

3.1 English – which version?

(KS 4.3.1)

Both Bang & Olufsen and Nordea require British English. In special cases, Bang & Olufsen uses American English.

HyperSE supports both custom lists of approved and of unapproved words. For example, with the dictionary maintenance tool it is possible to disapprove the noun *color* and add the hint to use British English instead of American English.

3.2 Hyphenation and compound words

(KS 4.3.2) (KS 4.6.6)

Both companies discourage manual or automatic hyphenation of words.

Besides that, Nordea has some rules for the correct use of hyphens. There are rules for fractions (*four-fifths*), quarters of the compass (*South-East Asia*), compound adjectives used in front of a noun (*long-term growth*, but *in the long term*), nouns formed from phrasal verbs (*set-up*, but *the company set up operations …*) and other rules regulating when - and when not - to use hyphens.

The HyperSE tool does not have a general rule that recognises fractions, compass directions etc., but it is possible to disapprove explicit groups of words, such as *four fifths*, *long term growth*, etc. It is also possible to have explicit entries for the verb *set up* and for the noun *set-up*, both of which are approved and each with a hint text.

It seems feasible to automate these rules and implement them in a CL tool. First, words would be ascribed parts of speech (POS tagging) and then a shallow syntactic analysis together with hyphenation rules could decide whether hyphens would be appropriate. In this way endless look-up lists of hyphenated terms and of unapproved word combinations lacking hyphens would be avoided and it could be decided whether *set-up* is correct or *set up*.

3.3 Abbreviations and acronyms

(KS 4.3.3 and 4.3.4)

Bang & Olufsen does not use abbreviations at all. Only in Bang & Olufsen's logo the acronym B&O is used. Nordea has a number of rules:

1) If you want to use abbreviations that are not generally known, spell out the word on first use, with the abbreviation following in brackets.

Not in this version of HyperSE. Implementation of this rule in software could to some extent be automated, but in many cases the user would have to be involved to elicit a confirmation. A simple and perhaps preferable implementation would leave out the automatic scanning of the text for definitions of abbreviations and ask the user right away on the first occurrence of an unknown abbreviation. The user's answer would then have to be stored to avoid that the same question was posed on other occurrences of the same word.

2) Do not use full stops to identify abbreviations.

Known abbreviations containing one or more non-closing full stops (*e.g*) can be put in the list of unapproved words. This type of abbreviations can be handled by HyperSE (but see below).

Abbreviations with closing full stops are not covered by this version of HyperSE. The examined version of HyperSE always interprets a full stop following an abbreviation as a sentence delimiter. As a consequence, the tool cannot comment the full stop as an unwanted part of the abbreviation.

A simple rule can be implemented that probably catches many abbreviations closing with full stops. Consider the following sequence: an abbreviation closing with a full stop, a blank and a word starting with a lower case character. In this situation it is very probable that the full stop is intended as part of the abbreviation and that it not also plays the role of a sentence delimiter. So the tool can detect this.

In other situations it is harder to detect the intention with a full stop.

- The text writer may have invented an abbreviation and put a full stop after it. Dictionary look up cannot tell that there is an abbreviation and heuristic rules for detecting abbreviations may fail as well.

- Even in mid-sentence the first word following an abbreviation may start with a capital letter.

- A new sentence may erroneously start with a lower case letter.

3) Reduce the use of *eg* and of *ie*.

HyperSE can display a hint text on every occurrence of a word. In this simple way the user can be warned against overusing these abbreviations.

4) Abbreviations like *m* for *million* and *bn* for *billion* are appended after the number without space between the number and the abbreviation.

Not in this version of HyperSE. This is easy to implement in software: display a hint text when a number is followed by a space and *m* or *bn*.

5) The abbreviation *k* for *thousand* is seldom used.

Not in this version of HyperSE. The tool does not know that a token consisting of a number followed by *k* very probably is in fact two items. It must be fairly easy to implement this in software. The second item, the *k*, could then be provided with a hint text.

6) There must be a space between currency symbol and amount.

Not in this version of HyperSE. This should not be difficult to implement as currency symbols are relatively few and easy to recognise.

7) There is no period after an initial in a name.

Not in this version of HyperSE. This could perhaps be handled by putting all capitals followed by a period (and also the initials that consist of two or more characters, such as *Th.*) in the list of unapproved words, but this has probably averse side effects. For example, the token *A.* can be an initial, but it can also play the role of a label in a list, like this:

A. Curry is doing well.

B. Broccoli made my day.

This example shows that a name-recogniser would not always suffice to spot names: *Curry* and *Broccoli* are family names but also food stuffs. If the food stuff interpretation is correct, then A. and B. are surely list item labels. If the family name interpretation is correct, then A. and B. might be initials.

8) Nordea has a list of abbreviations for units of measurement.

This list can easily be incorporated in HyperSE's list of approved words.

9) Acronyms are written with capital letters.

This list can easily be incorporated in HyperSE's list of approved words.

10) Abbreviations that can be pronounced and are composed of bits of words rather than just initials should be spelled out in upper- and lower-case.

This list can easily be incorporated in HyperSE's list of approved words. HyperSE can handle words that need special capitalisation, warning the user if she uses the wrong capitalisation.

11) Acronyms do not require the definite article *the*. The definite article is required if the abbreviation cannot be pronounced as a word (initialisms).

HyperSE's 'count nouns' and 'slight mass nouns' require articles, whereas 'mass nouns' and 'strong mass nouns' do not. So the rule may be implemented by defining acronyms as mass nouns and initialisms as count nouns. This has the side effect that

initialisms can have plural *s*, which in general seems quite right (for example *SMEs* for *Small and Medium-Sized Enterprises*). The plural form can be disallowed in the entry in HyperSE's word list, if needed.

3.4 Numbers

(KS 4.3.5)

Nordea requires that

1) In English numbers, commas signify thousands and a full stop the decimal place.

This is in HyperSE.

2) Sentences do not start with figures.

Not in this version of HyperSE. This must be very easy to implement.

3) Simple integer numerals (1-10) are written in words, except when followed by the percent symbol % or by units of measurement.

HyperSE supports this rule except for the percent sign.

3.5 Pronouns

(KS 4.3.6)

Bang & Olufsen forbids possessive pronouns at the end of a sentence and discourages the use of the pronoun *one*.

Not in this version of HyperSE. Finding the pronoun *one* requires some shallow syntactic analysis of the text.

Nordea demands to write *you* instead of *the customer* and *we* instead of *the bank*.

HyperSE can only partially take care of this. We can attach a hint text to the terms *the customer* and *the bank* (and to other terms covering the same concepts). However, this has the unwanted side effect that *the customer* in *the customer relations* also is marked, even if *the customer relation* is put in the term list without a hint text.

There may be cases where *the customer* and *the bank* are acceptable terms. If the acceptability of these terms merely depends on the type of document, then different HyperSE profiles – one allowing *the customer* and *the bank* and one not allowing these terms – can handle this distinction.

3.6 Use of *please*

(KS 4.3.7)

Bang & Olufsen advises against the use of *please*.

HyperSE can take care of this. Attach a hint text to the word *please*.

3.7 Modal auxiliaries

(KS 4.3.8)

Nordea forbids the use of *shall* and favours the use of *must* and *is/are to*.

HyperSE can take care of this. Put *shall* in the list of unapproved words together with a text hinting to the alternatives.

Can, *must* and *may* must be used in precise ways.

HyperSE can cater for this by putting hint texts on each of these words describing their correct use.

3.8 Sentence connectives

(KS 4.3.9)

Bang & Olufsen advises against the use of discourse connectives to bind sentences together, for example *However, …* and *Nevertheless, ….*

HyperSE goes some way to support this by means of hint texts.

3.9 Currencies

(KS 4.3.10)

Nordea prescribes the use of ISO standards for currency.

HyperSE supports this if all known non-ISO currencies are put in the list of unapproved words. New non-ISO ways of denoting currencies will not be detected.

According to Nordea's rules, the currency must precede the amount.

Not in this version of HyperSE. This should be fairly easy to implement.

3.10 Pun, slang and humour

(KS 4.3.11) (KS 4.6.5)

Bang & Olufsen discourages the use of pun, slang and humour.

This is not supported by this version of HyperSE and seems to lie outside the reach of current technology. It is completely in the hands of the text writer.

3.11 Department names

(KS 4.3.12)

Nordea has rules for the names of divisions and departments of a company. These should be capitalised.

As mentioned before, HyperSE can handle words that need a special capitalisation, issuing a message to the user if she uses the wrong capitalisation.

3.12 Jargon

(KS 4.3.13)

Nordea has a policy to write clear, plain language.

This is, in fact, the main objective of all CL tools, and therefore also of HyperSE.

3.13 Negative lists of words

(KS 4.3.14)

Both companies have lists of words that should not be used.

This is fully supported by HyperSE.

4 Simple syntax rules

(Section 4.4 in VID report no. 1)

4.1 Contractions

(KS 4.4.1)

Bang & Olufsen forbids contractions like *isn't* and *don't* in user guides. Contractions are allowed in marketing material.

HyperSE supports document styles, with different negative lists for each.

4.2 Noun phrases

(KS 4.4.2)

Both companies restrict the number of consecutive nouns.

This rule is implemented in HyperSE.

4.3 Inflected verbs

(KS 4.4.3)

Bang & Olufsen requires that there be an inflected verb in each sentence, preferably present or past tense or imperative.

This does not seem to be part of this version of HyperSE. The grammar checker which is standard in Microsoft Word marks sentences without inflected verbs. As HyperSE works as an add-on to Word there is not really a need to implement this facility in HyperSE.

4.4 Genitive

(KS 4.4.4)

Bang & Olufsen encourages the use of *of* with the names of products instead of genitive-*s*.

This is supported by HyperSE.

4.5 Prepositional phrases

(KS 4.4.5)

Bang & Olufsen allows at most two consecutive prepositional phrases.

Not in this version of HyperSE.

An algorithm for finding prepositional phrases can be based on a shallow syntactic analysis, which need not comprise the whole sentence. We expect that even a rather simple grammar can catch most (but not all) sequences of three or more prepositional phrases.

(KS 4.4.14)

Bang & Olufsen cautions against the use of a prepositional phrase at the start of a sentence.

Not in this version of HyperSE. The rule can be implemented by looking for a preposition as the first word of a sentence.

4.6 Date and time

(KS 4.4.6)

Nordea has a number of formatting rules for date and time, some of which may be specific for Nordea or the financial sector. Examples:

Tuesday 12 January 2001
Jan, *Feb*, *Mar*, ...
31/12/04 (but only where space does not permit the date in full)
16.45
Q1 2001 (but in tables *2001Q1*)
FY02, *2001/02* (broken financial year)

Nordea cautions against using relative time expressions like *this year*.

Most of these rules must be tailor-made in any CL tool. The algorithm must find deviating formatting and issue warnings where necessary. Only a complicated algorithm, probably involving a syntactic parse of the sentence, can avoid false positives and at the same time cover all wrong formatted dates and times, but a simpler algorithm may already do a good job at the expense of some precision and recall.

4.7 Tense

(KS 4.4.7)

Bang & Olufsen cautions against the use of present and past participles.

This is supported by HyperSE.

4.8 Imperative

(KS 4.4.8)

Bang & Olufsen encourages the use of imperatives, especially in instructions.

HyperSE supports this rule.

4.9 Grammatical number of verbs

(KS 4.4.9)

Both companies have the rule that with collective nouns, sometimes plural is the most appropriate, while singular is better in other cases, depending on whether emphasis is on the individual members or not.

HyperSE supports this by allowing hint texts on entries in the dictionary.

4.10 -ing forms of verbs and nominalisation

(KS 4.4.10) (KS 4.4.12)

Both companies put restrictions on the use of the present participle.

HyperSE supports this, but may need to take an extra step away from AECMA SE and acknowledge that for example *drawing* is a present participle in some cases and a noun in other cases. Only the use of *drawing* as present participle should be warned against. Full support of this rule requires a simple syntactic analysis.[5]

The situation becomes even more complicated because both companies restrict the use of the gerund (*drawing* as an act, not as a product). The only viable way to incorporate this distinction in a CL tool is to make the user aware of the ambiguity presented to the tool when it finds words like *drawing* and *painting*. Words ending on *–ing* and not

[5] The choice between *drawing* as a noun and *drawing* as verb would not occur in a strict AECMA SE setting, because AECMA SE does not allow homographs. Slackening the AECMA SE rules has the potential to introduce considerable complexity to a checking tool.

marked as both noun and verb in the dictionary can be assumed to be verbs (gerund or present participle).

4.11 Passive

(KS 4.4.11)

Both companies discourage the use of the passive form.

This is part of HyperSE.

4.12 Superlatives

(KS 4.4.13)

Both companies caution against the use of superlatives.

HyperSE can take care of this rule by means of hint texts. The support could be extended to words that are not in the dictionary by an algorithm that recognises regularly formed superlatives.

4.13 Sentence initial infinitives

(KS 4.4.14)

Bang & Olufsen discourages the use of an infinitive at the start of a sentence.

Not in this version of HyperSE.

This should be not too difficult to implement in a CL tool. It requires the presence of a sentence initial ‘to’ and a dictionary lookup of the next word, not a syntactic parse.

5 More elaborate syntax rules

(Section 4.5 in VID report no. 1)

5.1 Word order

(KS 4.5.1)

Nordea recommends having the subject near the beginning of a sentence.

Not in this version of HyperSE.

Checking compliance with this rule requires quite some syntactic analysis. It can be done, though.

5.2 Subordinate clauses and parenthetical phrases

(KS 4.5.2) (KS 4.5.3)

Bang & Olufsen allows at most three consecutive subordinate clauses, but at most two if the subordinate clauses are parenthetical.

Not in this version of HyperSE.

This requires a complicated parsing of the whole sentence.

Often the distinction between parenthetical and non-parenthetical subordinate clauses can only be made based on world knowledge. This is outside the reach of current technology.

5.3 Parentheses

(KS 4.5.4)

Bang & Olufsen discourages putting explanations and sentences in parentheses. It is allowed to put single words in parentheses.

Not in this version of HyperSE.

It should be rather straightforward to implement these rules in software.

6 Layout and text level

(Section 4.6 in VID report no. 1)

6.1 Sentence length

(KS 4.6.1)

Both companies restrict the maximum number of words per sentence.

This is supported by HyperSE.

6.2 Paragraph length

(KS 4.6.2)

Bang & Olufsen allows max 13 sentences in a paragraph. Nordea recommends short paragraphs.

HyperSE partly fulfils these requirements in that it can put a maximum at six sentences per paragraph, which is the number of sentences recommended by AECMA SE. This number cannot be changed by the user in the examined version of HyperSE. It should be easy for Tedopres to change the number or let the user specify it.

6.3 Amount of information per sentence

(KS 4.6.3)

Both companies restrict the number of ideas per sentence. Bang & Olufsen recommends having one main point plus, at most, one subtopic.

HyperSE supports this partially. It has the option to mark sentences with more than one topic[6].

It is very difficult to define what represents one idea, let alone to implement automatic counting.

[6] 'Topic' is a term used in the AECMA SE rules in the following prescript: "Keep to one topic or one instruction per sentence". AECMA SE goes even so far as to state "Each paragraph must have only one topic." Here we assume that 'idea', 'point' and 'topic' refer to the same concept.

6.4 Headings

(KS 4.6.4)

Nordea recommends the use of short headings and to have paragraph headings at regular intervals.

Not in this version of HyperSE.

Given that headings in Microsoft Word normally are marked up with styles dedicated to headings, it should be relatively easy to measure their length and frequency.

6.5 Italics and capital letters

(KS 4.6.7)

Bang & Olufsen does not recommend the use of italics and capital letters in general.

Nordea recommends that only the first word in a heading is written with a capital. Italics must be used for references to literature and in Nordea's "rules, terms and conditions". They are also used for words in other languages not commonly used in English.

Not in this version of HyperSE.

It is feasible to automate the correct handling of capitals in headings. All but the first word must be written without capitals, unless a word requires special capitalisation according to the dictionary. It is more difficult to spot literature references, although a heuristic mechanism might be able to find patterns common to literature references, such as a high density of punctuation signs, dates and person names. If Nordea's rules, terms and conditions always have the same wording (or nearly so), then these are easy to spot. Finally non-English words can be defined as words that are not in the dictionary or that are marked as foreign words.

It may be difficult for a CL tool to spot italics in a text, depending on whether the word processor program that hosts the CL tool allows the CL tool to have a look at the mark-up codes.

6.6 Punctuation

(KS 4.6.8)

Nordea has adopted the rule that acronyms and initialisms in plural do not have an apostrophe: *SMEs*. Neither do years: *1990s*.

This rule is to some degree supported by HyperSE in that it can warn against genitive *'s*.

Nordea prescribes that one should not write a capital as the first character after a colon.

Also, Nordea recommends not to put commas before *and*. Nordea has more rules for the correct use of commas.

This version of HyperSE does not support these last two rules. It should be fairly straightforward to implement these rules, but care must be taken if the first word after a colon always starts with a capital - a name, for example.

6.7 Lists

(KS 4.6.9)

Nordea has a very precise prescription for how to make bulleted or numbered lists. In general, the introductory sentence concludes with a colon and all but the last list items end with a semicolon. The last list item ends with a full stop. If any list item consists of more than one sentence, all list items start with a capital and end with a full stop. The introductory sentence also ends with a full stop.

Not in this version of HyperSE.

It seems feasible to ascertain the correct layout of lists by automated means, but it requires that the CL tool has access to the host program's (Microsoft Word) internal text representation. A mere flat text representation would not suffice, because the positions of the text elements relative to each other is lacking in such a representation. Also, some syntactic parsing is needed in order to find the finite verbs in the list items.

7 Tone of Voice

Tone of Voice recommendations, as described in VID report no. 4 “Tone of Voice”, regulate the language from a slightly different perspective than the readability rules mentioned above.

The above CL rules promote readability and generally concern linguistic phenomena that should be avoided when possible whereas the Tone of Voice recommendations promote the company image and more often concern certain language characteristics that should be encouraged.

The Tone of Voice aspects relevant in a particular text are also more heavily dependent on the text type and the receiver type than the readability rules. The nature of readability rules implies that they are relevant more or less as a whole set in many contexts (though not all). Tone of Voice rules can however not apply as a whole set, but should be seen as a set for which the single parts may vary as the selection of one Tone of Voice recommendation in a particular context will naturally exclude another recommendation. If, for example, it is decided that user guides should have special focus on the company value *Excellence* this automatically means that the other company values should be less pronounced. Similarly, if marketing material describing the virtues of the company and its products should have many sentences containing company or product name as the grammatical subject, this also means that we should expect fewer sentences where the grammatical subject is the receiver.

Therefore it is an absolute necessity that an authoring tool supporting Tone of Voice allows the user to set up a number of parameters which together constitute different profiles (this is already a feature in many CL tools, also in HyperSE). A company might for example wish to set up profiles for annual reports, sales material for private customers, product documentation for technical staff, manuals for end users etc. The analysis of the text, or possibly text fragment, performed by the tool should then establish whether the individual parameters of the profile are adequately handled.

In the following we will take a closer look at the main Tone of Voice aspects described in VID report no. 4 and suggest some overall ideas for implementation of these. Basically, we suggest that the Tone of Voice phenomena are all anchored in the dictionary of the authoring tool, with features describing the different kinds of verbs and value-laden words combined with rules that encourage the use of these words in specific contexts.

7.1 Value-laden vocabulary

It has been established that company values are partly reflected by particular value-laden word collections. For example the Bang & Olufsen personality value *Excellence* is reflected by words as *elegant, flawless, genuine, intelligent, perfect, powerful* etc. A particular profile should therefore define the importance or relevance of for example the

Excellence value in the particular context and for example recommend that 5% of the total number of words are from the *Excellence* word collection.

As mentioned in the introduction, the Dictionary Maintenance Tool contains the lexicon and specifies, for example, approved and non-approved words, parts of speech and other significant features about words. This tool could perhaps be extended with other features specifying the type of personality value for the relevant words. A rule stipulating a particular amount of value-laden words from particular word collections should of course also be developed.

7.2 Modality

Modal auxiliaries are sometimes used to establish a relation between sender and receiver. The work with company texts in this project has suggested that only certain kinds and only a limited number of modal auxiliaries should be used in connection with company name/product name as the grammatical subject. Furthermore, it may be desirable to encourage the use of particular modal auxiliaries in sentences where the receiver is the grammatical subject.

It is probably possible to implement a rule that regulates the use of modal auxiliaries in relation to certain types of grammatical subjects. Modal auxiliaries, company names and receiver paraphrases are relatively small fixed sets of words and though product names are not a fixed set of words, changes and extensions are controllable. For optimal functionality it is of course a prerequisite that the tool can identify the grammatical subject, but otherwise all sentences containing for example a modal auxiliary as well as a known company or product name could trigger a hint text.

7.3 Mental verbs

Mental verbs expressing hopes, expectations etc. support the establishment of a mission and of the sender as a person with personal feelings. These verbs are very important to use in some text types whereas they are insignificant in other text types.

Regarding implementation, mental verbs should be marked with a suitable information type in the dictionary. They are comparable to modal auxiliaries as they constitute a small set of words. If the authoring tool can identify the grammatical subject, then the use of different types of mental verbs in relation to specific grammatical subjects can be controlled.

7.4 Verbs and processes

The remaining verbs should also be grouped in different categories in order to control sender/receiver relations or, in other words, to control ‘Who does what to whom’. Verbs play a very important role when describing what the sender does, how this is done and what the receiver is expected to do. Possible categories for the verbs could for example be *action, ambition, effort, help, offer, suggestion* etc.

As with the other word types mentioned above, it seems feasible to develop an information type describing the different verb types in the dictionary. As with modal auxiliaries and mental verbs it is also important to investigate these remaining verbs in relation to the grammatical subject. A parameter in a particular profile could for example suggest that 10% of the sentences should have a company name/product name as the grammatical subject together with verbs from the categories *ambition, help* and *action*.

7.5 Tool functionality

In order to support the author with respect to Tone of Voice, we have seen that a tool should have a certain number of functionalities.

1. It is important to be able to distinguish various types of vocabulary through the marking in the lexicon. This was already discussed above.

2. It is important to be able to set a level of desired frequency, such as at least x % of the total number of words, or at most x % of the total number of words. It should also be possible to express this for specific parts of speech, e.g. the number of passives should be at most x % of the total number of inflected verbs. The use of this feature was also discussed above.

3. A feature which was not already discussed is a suggestion feature that will support the creative writing through vocabulary suggestions. It should always be possible for the user to get a list of value-laden words belonging to a specific type of personality value, e.g. *Excellence* for inspiration. This feature has something in common with the use of synonym dictionaries in a normal word processor, but it will be organised, not through synonymy, but through the personality value and can also be structured in hierarchies. The value *Excellence* may e.g. be expressed through a number of basic excellence values and each of these may have sub-values in return. This gives the user a better overview of the proposals. It should be possible to call the feature during authoring, but it should also be part of the reporting at the document control, that suggestions are made for instances of the classes of words that should be more frequent.

8 Conclusions

We have seen that the Nordea and Bang & Olufsen writing rules can largely be implemented using a CL tool.

In our work with Tone of Voice rules we have concentrated on those that would most easily be implementable in an authoring tool, and we have seen that such rules cannot directly be implemented in a current CL tool. However, the important result is that the functionality extensions that would be necessary are within the reach of a CL tool. We think that this is a very promising result, and we hope that in the future such extensions will be made by CL providers.

9 References

AECMA Simplified English, PSC-85-16598, *A Guide for the Preparation of Aircraft Maintenance Documentation in the International Aerospace Maintenance Language*

British National Corpus, http://www.hcu.ox.ac.uk/BNC, Oxford University, 2000

Delin, Judy: *Language and Brand Position*, Enterprise IDU, 2001

Fairclough, Norman: *Language and power*, Longman, UK, 1989

Henriksen, L., B. Jongejan, B. Maegaard: *Kontrolleret sprog*, VID report no. 1, CST, 2003 (in Danish)

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