Software localization into Nynorsk Norwegian

Michael Everson and Trond Trosterud
2000-06-05

Introduction

This report deals with localization of computer software into Nynorsk Norwegian, but most of what is said has relevance for other languages as well. Linguistically, speakers of (Mainland) Scandinavian understand each other, but they write four different languages (Danish, Swedish, and Bokmål and Nynorsk Norwegian). Morphologically, these four languages are equally distant from each other, but the terminological differences are smaller between Nynorsk and Bokmål than between the other two. Nynorsk is in a minority position in Norway, with approximately 12% of the users. The report was commissioned by Norsk språkråd, as part of their investigation into the feasibility of providing Nynorsk-localized versions of popular commercial software. This report begins with a simple discussion of the localization process, and proceeds to some discussion specific to the Nynorsk context.
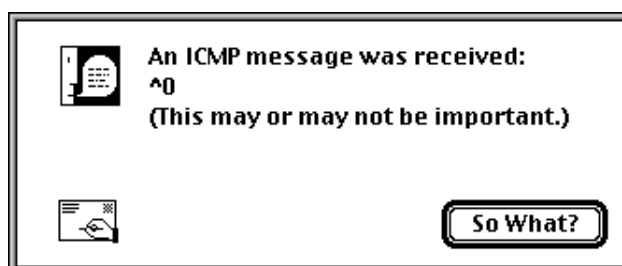
What is localization?

Software localization is the process of providing software in a form appropriate to the linguistic and cultural requirements of the user. Typically, this involves translation of the user interface, that is, of the messages a program presents to the user to enable him or her to create documents, modify them, print them, send them by e-mail, etc.

The File Menu from the e-mail program Eudora Light, in English, Irish Gaelic, and French.

The user interface also includes error messages, to which the unhappy user often may have only one response ("OK" or "Cancel"). A certain amount of cultural creativity can be found in localizations of some programs.

*An error message from the e-mail program Eudora.*

Most programs and localizations tend to more proper, less humourous messages, however.

Localization sometimes involves changing the content of pictures presented to the user. For example, the warning dialogue box often has a modified stop sign in it.
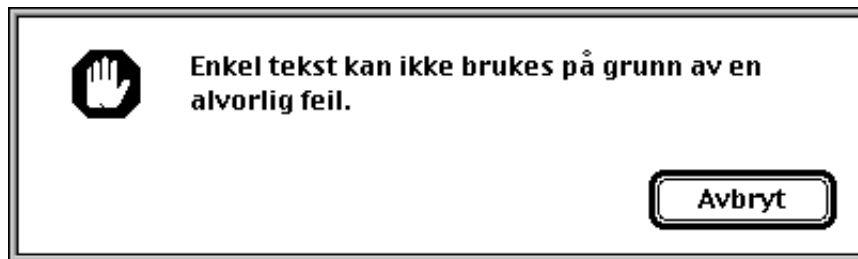
Error message from the Bokmål Norwegian localization of SimpleText.

In Greece, however, since the palm-of-hand gesture can be considered to be offensive, software localizers must replace the image with a true stop sign. Interestingly, the image used has the Latin alphabet on it even in Greece.
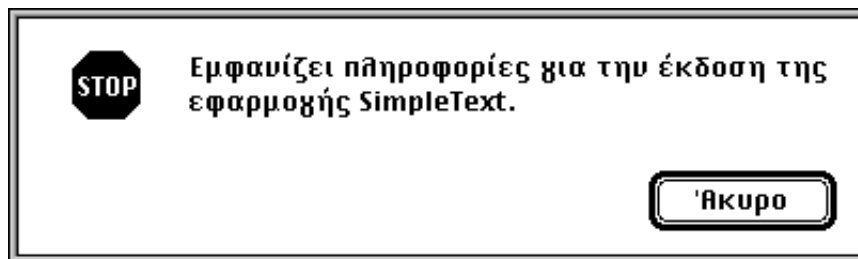
The same error message from the Greek localization of SimpleText.

The Macintosh programmers' tool *ResEdit* has a special configuration called "pig-mode" in which the program takes all available memory to run its processes. Although *ResEdit* has probably never been localized into any language, were it to be localized into Arabic or Hebrew, a different name and image would have to be chosen because pigs are not considered cute, lovable animals in Islamic and Jewish society, and the image could offend some users.

A dialogue box from ResEdit.

People working in the field of internationalization are endeavouring to educate software developers to take such cultural differences into account, but even so sometimes they may not be successful. Many e-mail programs created in the U.S. use the American mailbox with red flag , unaware that in many countries, in which a red flag is not found on letterboxes, the image is simply unfamiliar. (Especially in rural communities, the American postman will not only deliver, but will also collect mail for posting; the red flag is used when the resident wishes the postman to stop at a box to collect mail, which he might not otherwise do if there were no mail to deliver on the day to the address in question.) A good Northern European substitute might be the curled horn – though even this might not be recognized in all European countries.

How does localization work?

Not so very long ago, and still very often today, text presented to the user was written into the computer code, or "hard coded". When the program was compiled the user interface would appear. Each time a change was made, the entire program had to be recompiled, even if the change was a simple spelling correction like "wirte" to "write". (Compiling the software is the process of taking the code written by the programmer and converting it to a format which can be interpreted by the computer.) Obviously this was expensive and inconvenient for localizers, especially because many companies did not like to release the actual source code to translators for reasons of security. Apple was one of the first companies to introduce a special software resource called a "string resource", and an editor like *ResEdit* could be used to open the string for easy editing. The localizer could make changes, quit *ResEdit*, and launch the application to see the changes made. Microsoft also provides similar tools to facilitate localization.

Localization technology has become more advanced still, however. Programs exist which can extract most or all localizable text into simple plain text format or database format, where the source language (usually English) is presented on one line and a space for the localizer to enter a translation is presented on another line.

```
DITL 131 Enabled check box text 53 (6)      DITL 131 Enabled check box text 53 (6)
        <Wrap-Around Search>                         <Wrap-Around Search>
        <>                                           <Fortset frå starten>

DITL 131 Disabled static text 62 (7)        DITL 131 Disabled static text 62 (7)
        <Replace with what?>                         <Replace with what?>
        <>                                           <Erstatt med kva?>

DITL 131 Enabled button text 80 (9)         DITL 131 Enabled button text 80 (9)
        <Replace All>                                <Replace All>
        <>                                           <Erstatt alle>

DITL 601 Enabled button text 8 (1)          DITL 601 Enabled button text 8 (1)
        <Save>                                       <Save>
        <>                                           <Arkiver>

DITL 601 Enabled button text 17 (2)         DITL 601 Enabled button text 17 (2)
        <Cancel>                                     <Cancel>
        <>                                           <Avbryt>

DITL 601 Enabled button text 44 (5)         DITL 601 Enabled button text 44 (5)
        <Eject>                                      <Eject>
        <>                                           <Mat ut>
```

AppleGlot Work Glossary for SimpleText, before and after Nynorsk translation.

Tools like these are often low-cost or free, but more expensive software also exists (for the economically powerful languages at least) which can provide a suggested translation. Such "machine translation" is not always reliable; the example of the term *native speaker* (German *Muttersprachler*) translated as *eingeborener Lautsprecher* 'indigenous loudspeaker' which, while an acceptable literal translation, shows how ineffective such translation can be for localization tasks. However, many such programs include a glossary component, which can learn (or be taught) specific phrase translations, and, especially over time, this can provide a translation house with a more reliable database out of which translations can be automated.

Localization of graphics presented to the user is still problematic, because graphics must be edited individually. They are also generally quite small, and, especially if they contain text, translation can

be quite difficult. A button for sorting a database *down* (in *ascending* alphabetical order!) or *up* (in *descending* alphabetical order!) might be shown as ▢ or ▢, but if (as is the case in Irish Gaelic) the semantics of "up" and "down" do not suit the concept, the buttons might have to be localized to ▢ and ▢; in Norwegian they would have to be ▢ and ▢; in Icelandic ▢ and ▢; in Estonian ▢ and ▢; in Russian ▢ and ▢; in Inuktitut ▢ and ▢.

All such changes are time-consuming. Sometimes they are impossible. A red flag announcing on the menubar that e-mail has been received in *Eudora Light* reads "MAIL" ▢. This can be localized to "POST" ▢ in many languages, but Russian почта, Inuktitut "ᑎᑎᕋᕐᕕᒃ", and Polish "POCZTA" simply cannot be written in a 16 x 16 pixel space. Redesign of such graphics is neither simple nor quick, and hence adds to the cost of localization. Most software companies now endeavour to create language- and culture-independent graphics, though as can be seen in the sorting examples above, even the most generic of graphics may need special attention.

The size of buttons may also vary from language to language. The buttons must be (manually) resized by the localizer.

Dialogue box from the English language version of Eudora Light.
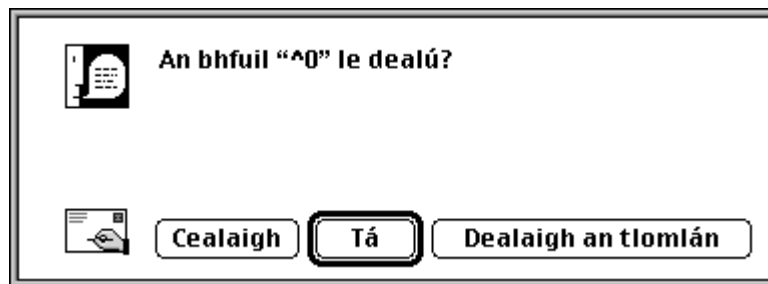
Dialogue box from the Irish Gaelic localization of Eudora Light. Note the difference in syntax between the English and Irish versions. A more literal translation showing this syntax is shown below in an artificial dialogue box made just for this report:

Sometimes the entire presentation interface must be redesigned due to linguistic requirements, if the original design is too small to permit the required localized text to fit.
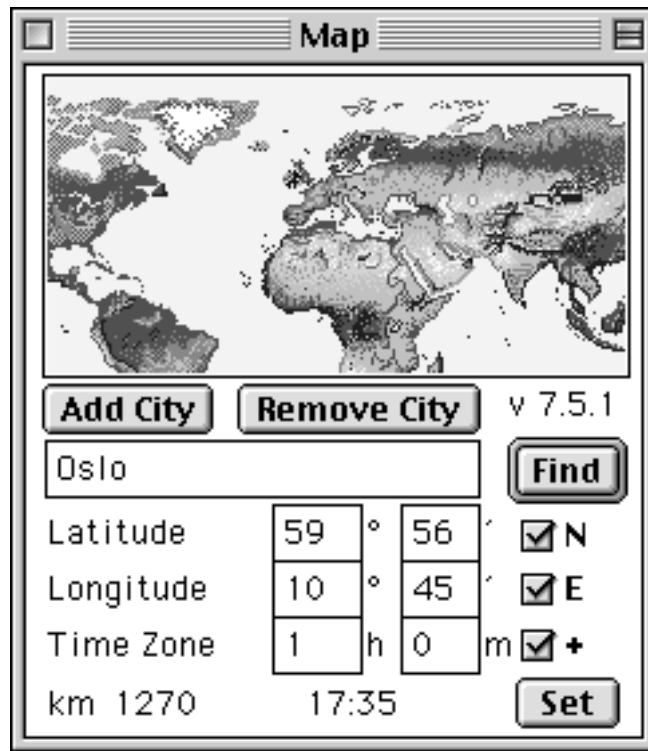
English original of the Map control panel from the Mac OS.
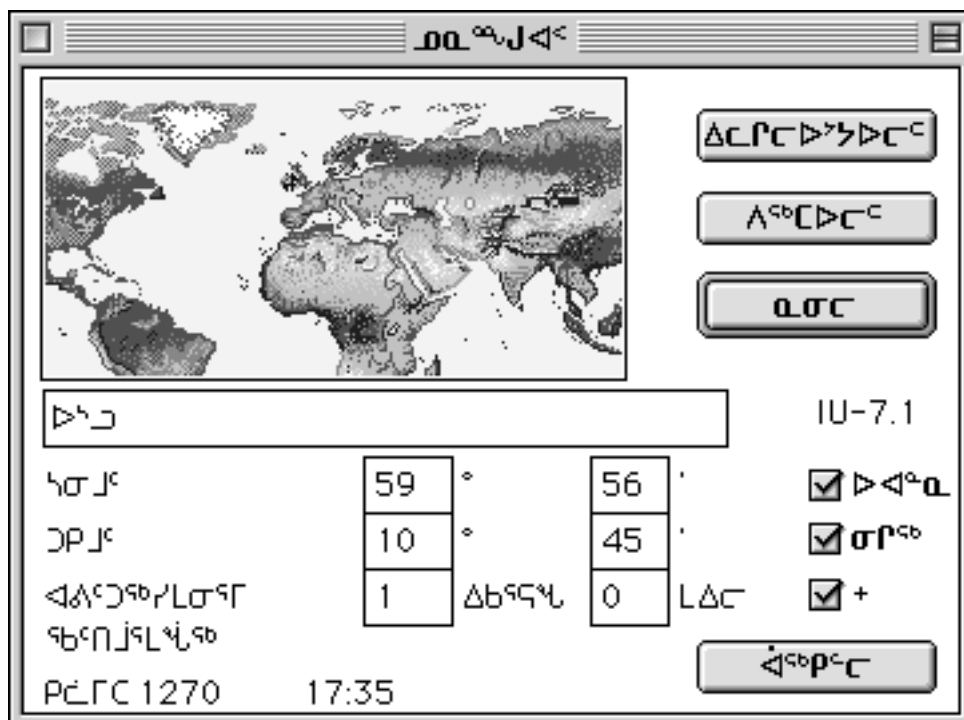
Inuktitut localization of the Map control panel from the Mac OS.

Sometimes, however, it is impossible, due to hardcoding of resources, to resize or otherwise redesign the user interface. In general such hardcoding is considered to be a programming error, and when reported by the localizer to the software developer, is corrected in the next version of the software.
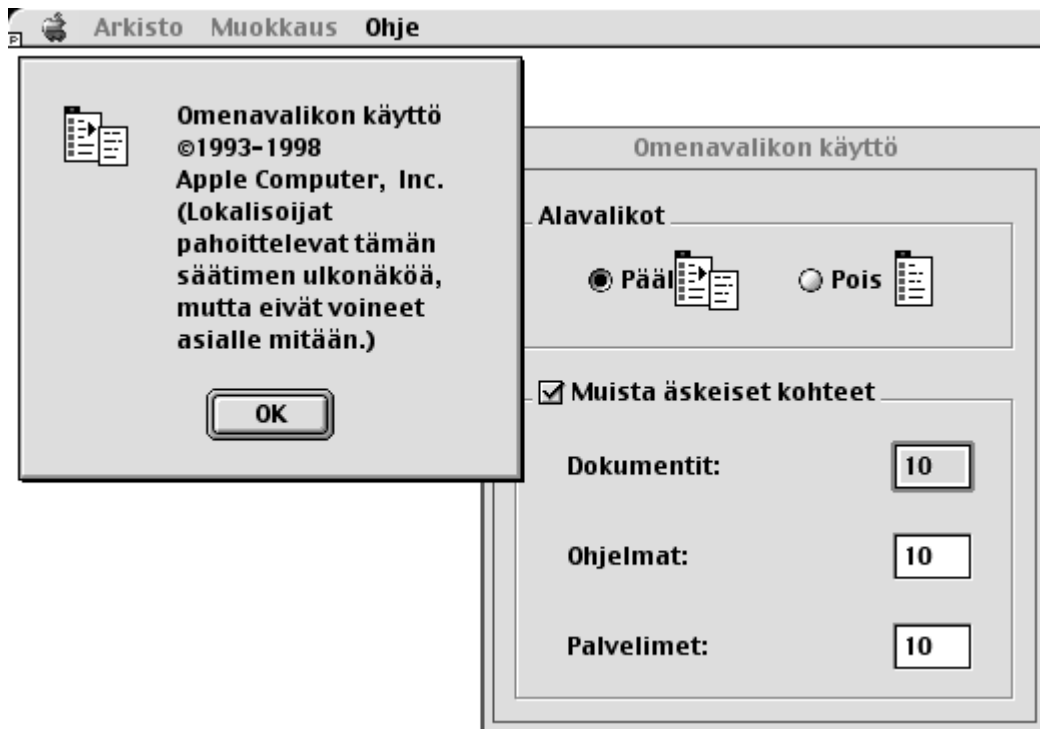
Finnish localization of Apple Menu Options control panel. Note the truncation of the "Päälle" 'On' button. The text following the copyright statement reads: "The localizers apologize for the appearance of this control panel, but there was nothing we could do about it."
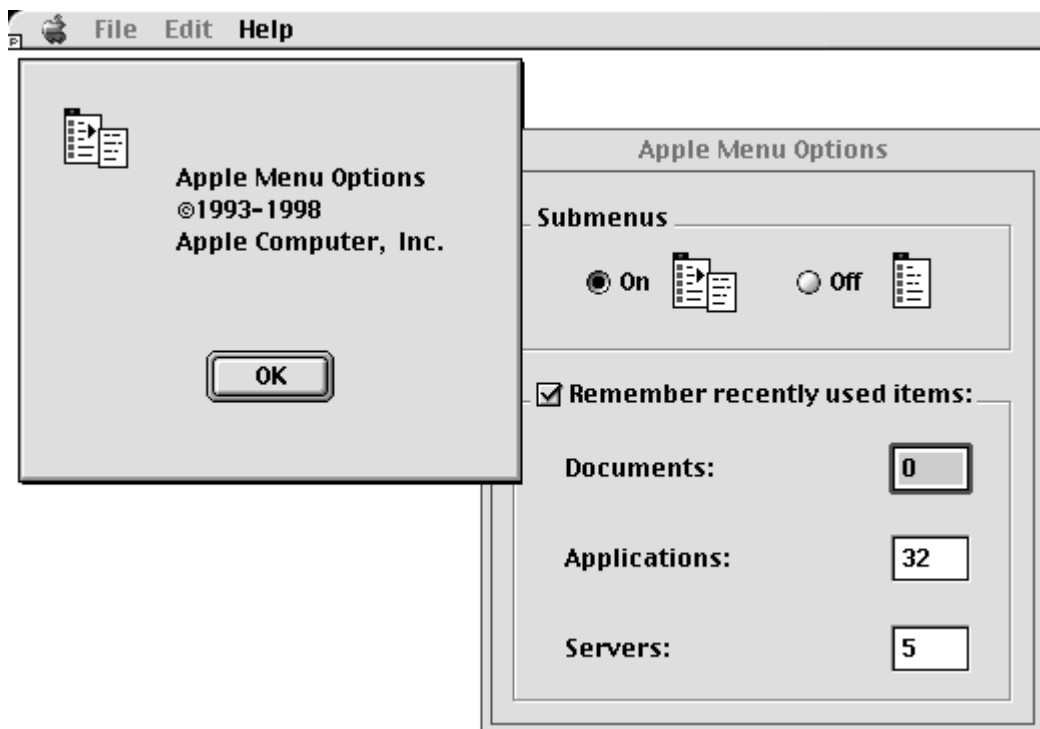
English original of Apple Menu Options control panel. Note that the localizers had to manually move the OK button down in the About Box in order to get their apology to fit.

In general, program files are shipped in one of two formats. Either the program code and graphics and text intended to be read by the user are all stored in a single language-specific compiled program, or they are stored in separate files known as resource files. If they are stored separately, the main program will point to an external file containing user strings, and the user will select which file he or she prefers via a preferences dialogue or menu.
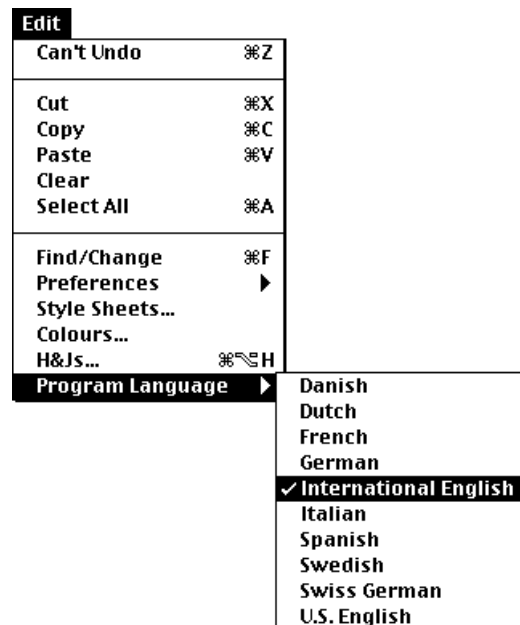
Program Language menu from Quark XPress.

Each of the localized files contains a reference code sought by the main program, and at run-time the corresponding localized string will be presented to the user. This may have implications for interface design, which may employ either dynamically-resizing buttons or buttons made generically large enough to contain strings of different length. Such programming is far more complex than that required for language-specific programs, and is *very* uncommon.

What kinds of software should be localized?

Software for office use may be divided into three main categories. Cost and complexity are the usual means of classifying the different types.

High-range software
The *crème de la crème* of software. Generally costs NOK 1540 (EUR 184, USD 200) or more. Examples: *ClarisWorks*, *Filemaker Pro*, *Microsoft Excel*, *Microsoft Word, Photoshop, Quark XPress.*

Medium-range software
Generally smaller or less complex than high-range software. Generally costs NOK 308-1540, (EUR 37-184, USD 40–200), but may be shareware or freeware. Examples: *Eudora, Graphic Converter, Microsoft Internet Explorer, Netscape Communicator, ZipIt, WinZip*, encyclopaedic software, and some of the more complex games.

Low-range software
Small, simple programs without many features. Generally costs NOK 0–308 (EUR 0–37, USD 0.00–40.00), and is often shareware or freeware. Examples: Games, plug-ins, Telnet, and other programs with very little text in them, such as the operating systems of cellular telephones, where the menus are very short.

How important is commitment to upgrading?

Software updates must also be translated. This is an ongoing cost which must be reckoned with. If Netscape 4 is localized, but Netscape 5 with far more functionality is released, users often prefer the newer improved software to the old even if the new is only available in English. This is a serious problem for *all* users of non-English localized software, whether for economically powerful languages or economically marginal ones. Software updates must also be translated. Without serious ongoing dedication to upgrade and version support, any software localization is doomed to obsolescence, as users almost invariably prefer to use the most recent release of any program.

What has to be translated?

It is hard to answer this question, because programs differ so much. But for high-range programs, the amount of text in the program itself can range from 120,000 to 1,500,000 characters, and the number of pages of text in the manuals can range from 250–850 pages. Medium-range programs may contain text in the range of 50,000 to 120,000 characters, and and between 100-250 pages of text in the manuals. Low-range programs correspondingly have less than 50,000 characters of proram text, and anywhere from 1–100 pages of text in the manual.

How is the translation to be done?

Translating technical text from one language to another is not an easy task, in the context of new software technologies. The actual technical vocabulary must be established in the target language, and new terminology must be acceptable to the language community in question. Optimally, such work is undertaken by vocabulary committees consisting of both specialists and linguists, working in close contact with the user community. A case in point is Iceland, where such committees for most major technical fields are working quite actively.

Assuming that the basic terminological work has been carried out, the cost of translating software and manuals reduces to standard translation prices. When localizers working alone have to create a greater or lesser number of technical terms on the fly, such terms may or may not be popular with users.

Machine translation

Localization is best accomplished by recourse to machine translation, where it is available. For *most* languages, it is *not* available. Where it does exist, use of machine translation can significantly reduce costs.

Software text is especially suited to machine translation, for a number of reasons. Program text differs from ordinary text in several ways. The main difference is that program text is not a single continuous text, but that each string is a text in itself, whether an announcement to the user or a query or answer to the user's actions. Since the text genre is fixed, the machine translation system may safely assume that the text is to be interpreted as *e.g.* a software command, and the number of ambiguous senses a given word may have is thus severely reduced. Because the "text" in a software program is not a single text, but rather a collection of independent texts, each text corresponding to one command, and ranging from one to several words. Ordinary texts, on the other hand, consist of sentences, but these sentences are linked to each other in non-arbitrary ways, and a translation system must be able to remember what was said in the preceding sentence(s) in order to get the references right when translating a given sentence.

In order to illustrate this point, we compared the commands of the Bokmål version of Microsoft *Word 97* with *Gaia*, a Bokmål novel by Karsten Alnæs (available in html format at http://www.hf.uio.no/tekstlab/innsyn/norsk/bm-skj.html#nta). The results are shown in the table below:

	chars.	words	sentences	words/sen.	unique sens.	unique sen./sen
Word 97	429511	56639	14570	3.9	12567	0.863
Gaia	321763	59298	5390	11.0	5342	0.994

As can be seen, the sentences found in the program are much shorter than the sentences in the novel, 3.9 words per sentence, as opposed to 11.0 for the novel. Comparing the sentences one by one reveals another important difference: Whereas 99.4% of the sentences in the novel were unique, this was the case for only 86.3% of the sentences in the program. Put differently, only 12567 of the 14570 sentences of the program were unique, 2003, or 14%, were mere repetitions of other sentences. Moreover, the sentences in the program file often differed from each other in minimal and predictable ways. Compare the following examples from Word 97:

    1 Aktiverer/deaktiverer bruk at prikket understreking i merket område
    1 Aktiverer/deaktiverer bruk av dobbel gjennomstreking i merket område
    1 Aktiverer/deaktiverer bruk av dobbel understreking i merket område
    1 Aktiverer/deaktiverer bruk av fet skrift i det merkede området
    1 Aktiverer/deaktiverer bruk av gjennomstreking i merket tekst
    1 Aktiverer/deaktiverer bruk av overskrifter i tabeller

Similar minimal variation patterns cannot be established in ordinary running text, and if they could, it would in any case be irrelevant, since running text must be translated in context and not in isolation.

Another of the main difficulties of translating ordinary text is one of cohesion: The meaning of a particular word may be determined only via the meaning of other words in the same text, some 3–4 sentences earlier in the text.

For program files, this is not the case. Rather than constituting a single continuous text, programs are made up of large collections of text, where the reference of the word is not necessarily dependent upon a neighbouring sentence. Each line is a coherent whole. Obviously this is an oversimplification, but it is very often true.

Help manuals are texts in the ordinary sense; they contain sentences which are put together to form a single coherent text. Translating manuals thus can be seen as ordinary technical translation: What is translated is texts; the vocabulary is limited and specialized, as in other kinds of technical translation.

Language-specific example: Translation into Nynorsk
Software in Nynorsk and Bokmål
A certain amount of software has been localized into Bokmål. In general this software has been high-range professional software (Microsoft, Apple, and Adobe products for instance). Medium-range software, such as freeware is in general used in English in Norway (examples: *Eudora Light* and *Netscape*). These programs are important, but because they are free, localization companies cannot localize them as their costs cannot be met by sales. Accordingly, they must be commissioned by bodies like the Norsk språkråd or the Norwegian Ministry for Education, Research and Church Affairs, or they must be localized by volunteers. (The French translation of *Eudora* is done by a dedicated volunteer; the Irish Gaelic localization was commissioned by Bord na Gaeilge and Údarás na Gaeltachta.) Low-range software (such as games) is rarely localized. Sometimes this is because the companies or individuals who make the software do not see much advantage in such localization (which generally requires some effort on the part of the software developer). In this context, Bokmål and Nynorsk both must be considered to be minority languages, despite the fact that they are natural languages used by millions of people.

We have not conducted a systematic survey, but our impression is that Nynorsk software is quite rare, and mostly confined to domestic paedagogical software, which is often released in parallel Bokmål/Nynorsk versions (cf. http://www.nls.no), and more recently to volunteer versions of shareware and other low-range software (cf. http://www.nynorsk.no). The main reason for this sad situation is the Nynorsk users' willingness to accept Bokmål software. Since they are able to understand the Bokmål software, they may have tacitly assumed that computers were not within the realm of Nynorsk. Correspondingly, software vendors have seen that Nynorsk users have accepted Bokmål products, and thus have found no reason to make Nynorsk versions. Tradition is an important factor as well. In Norway, word processors were available in Bokmål from the very beginning (the first popular PC word processor of the early 1980s, *Brum*, was made in Norway), thus subsequent word processors like *WordPerfect* were translated into Norwegian as well, whereas other program types, such as mail programs, remained in English.

From a sociolinguistic viewpoint the lack of software in Nynorsk is a case of domain loss, increasingly important as the domain grows in importance.

Non-automated translation

Translation of software to Nynorsk will in practice be either from English or from Bokmål. In the first case, manual translation is the only possible alternative at present, and the usual price for translation between Bokmål or Nynorsk and other languages is NOK 2635 (EUR 315, USD 342) per sheet (= 32000 characters) (source: Norsk faglitterær forfatter- og oversetterforening, 1 sheet = 16 A4-pages ≈ 2000 characters, whitespace not included). For translation between Bokmål and Nynorsk the price is lower – here the recommended price (there is no tariff price for translations between Nynorsk and Bokmål) is NOK 1990 (EUR 238, USD 258) per sheet. Note that these prices are standard prices; for difficult text within novel subject fields a translator may often get paid more. The table below gives estimated prices for a number of programs.

	Program Type	No. of chars.	Cost en>nn	Cost nb>nn	Manual Text Pages	Cost en>nn	Cost nb>nn
High-range programs							
MS Word 6.0	text				844	69	52
MS Word 97	text	334	28	21			
MS Word 98	text				245	20	15
Windows 98	OS	1534	126	95			
Excel	spreadsheet				248	20	15
Nisus Writer	text	120	10	7	606	50	38
Medium-range programs							
Quark XPress	DTP	57	5	4			
Powerpoint	presentation				160	13	10
Netscape	browser	117	10	7			
Low-range programs							
SimpleText	text	14	1	1			

Figures are given in thousands of Norwegian kroner. Cost is estimated from the tariff price, NOK 2635 (EUR 315, USD 342) per sheet (= 16000 characters) between different languages and Nynorsk/Bokmål, and from suggested price, NOK 1990 (EUR 238, USD 258) per sheet between Nynorsk and Bokmål. The size of the manuals is estimated at an average of 1000 characters per page. (At the time this report was compiled, NOK 100.00 = EUR 11.95 = USD 12.9875.)

It is difficult to give precise estimates of costs for localization engineering, since programs differ so much. The time-consuming part of localization engineering is resizing and tailoring of windows and

menus, and testing of the final product. The amount of work is thus dependent upon the number of clauses (e.g. pop-up windows, commands, etc.) in the program. Here are some examples:

SimpleText	304
Powerpoint 97	4078
MS Word 97	15123
Windows 98	49905

How many of these windows will have to be resized? That depends upon the number and length of the words, and the lengths of the strings. When looking at the program *SimpleText*, we found the following average length differences. The table gives number of words and number of characters in the localized text.

	words	*% of Eng*	*characters*	*% of Eng*
English	2529		13618	
Bokmål	2568	101.54	15393	113.03
Nynorsk	2566	101.46	15349	112.71
Bokmål>Nynorsk		0.8		0.32

More often than not, a translation is longer than its original. Translating from English to Bokmål increases both the total length and the number of words, and this will have consequences when adjusting the translated text to the appropriate windows. But when a program has already been localized into Bokmål, the transition to Nynorsk is far easier. The Bokmål and Nynorsk versions of *SimpleText* have practically the same number of words, and the Nynorsk text is even a bit shorter (probably due to the fact that the present tense form of many verbs is shorter in Nynorsk than in Bokmål). Thus, it is easy to predict that the effort of adjusting windows and menus from Bokmål to Nynorsk will be significantly smaller than is the case when translating between languages with different word structure. A similar comparision between English and Bokmål for *Windows 98* gives a size difference between English and Bokmål text that is approximately the same as the one shown above for *SimpleText*. Thus, one may expect that the same marginal difference between Nynorsk and Bokmål will show up for this larger corpus as well. We conclude that resizing windows when localizing from Bokmål to Nynorsk is a *much* smaller process than for the English/Bokmål pair, since Bokmål/Nynorsk string *and* word lengths are so similar, as shown in the tables above.

Engineering costs for translating closely-related laguage pairs is expected to be far smaller than compilation of translations into languages that are not that close. The only factor that will be constant is validation and testing; the same number of commands must be tested for every new localization. The difference is just that given smaller differences between the pairs, there will be fewer errors and anomalies to find.

In addition to this there is the cost of proofreading of the translated text, an average of 10% of the translation cost, as well as typesetting and printing of additional manuals. We will not estimate these costs here. At the other end of the spectrum one finds another possibility, namely that of large companies negotiating better prices for especially large contracts. We have not considered this option further.

Machine translation to Nynorsk
At present, no good machine translation systems exist from other languages to Nynorsk or Bokmål. There has been research on a system for machine translation between closely-related languages at the University of Bergen, namely, the *PONS* project, *Partiell oversettelse mellom nærbeslektede språk*. The project used Bokmål, Swedish, and English, but the basic structure was language independent, and built to be extended to other closely-related language pairs at will. The *PONS*

system works, in that it covers basic parts of the grammar of the three languages in question – but it has not been extended to a larger, working-size lexicon. Internationally, there are companies offering machine translation products between several languages, among them between English and Bokmål Norwegian (but not to or from Nynorsk; for two examples, see http://www.gy.com/home.html, and http://www.tranexp.com/). Common to these systems is that they are mainly wordform-based, with a focus on replacing words one-by-one, and with quite limited morphological and syntactic components. Still, due to the well-defined format of the language of software texts, (tailored versions of) such systems may very well be of help in translation from English into Bokmål, or indeed into any of the languages offered. We have not evaluated these systems further here.

For translators of literary texts, machine translation still has a bad reputation. Also within the field of Bokmål-Nynorsk translation, there is fear that the result will be "Bokmål with Nynorsk suffixes", instead of "good Nynorsk". In our view, several factors speak against this scepticism. As already shown, the language in software commands is fairly repetitious, not to say monotonous. The same commands are repeated over and over again, sometimes with minimal modifications and adding of conditional clauses, in order to handle slightly different situations. The syntactic patterns between Nynorsk and Bokmål are quite similar, and since many of the actual sentences are very short, much of the translation work is fairly mechanical. Rather than inviting the reader to language games, multiple ambiguity and associations (since software text is not poetry), text of the type we are dealing with here invokes one-to-one reference correspondence between the terms of the languages involved, even when there may be only a limited overlap in everyday language.

For example, although the English word "brace" has several translations in an ordinary English-Bokmål dictionary (*bånd, reim; gjord, belte; støtte, forsterkning, knekt; bøyle (tannregulering); borsveiv; parentes, klamme; bras; par (i jaktspr.); binde, gjorde, styrke, stramme, avstive, spenne; brase*), only one of these senses is relevant in the present context (*parentes, klamme*). The job of the tailored machine translation system is to choose the relevant unambiguous pairs, in this case partly from existing vendor usage, and partly from terminology lists made by Norsk språkråd. Several company-internal machine translation systems exist, one example being the one used by Scania in Sweden (cf. http://stp.ling.uu.se/~corpora/scania/ash961.html for references); in general results have been far better for systems dealing with a well-defined semantic field, than for open systems, designed for translating any text, regardless of topic.

Today there exists one program for automatic translation from Bokmål to Nynorsk, the program *Nyno*, made by the company Nynodata (http://www.nynodata.no). An independent evaluation of *Nyno*, written by Janne Bondi Johannessen, may be found here:

> http://www.hf.uio.no/tekstlab/programs/nyno.html
> http://www.hf.uio.no/tekstlab/bulletin/bull1_98/nyno.html

Since Nynodata do not have any competitors at present we give here only results from tests with their product. One should note that other companies and universities are also working on machine translation. The work done with *PONS* in Bergen on translation from Swedish to Bokmål is of course relevant, and within *PONS* there has also been some rudimentary grammar development for Nynorsk. In Helsinki, Lingsoft (http://www.lingsoft.fi) has developed morphological parsers for Nynorsk and Bokmål, and a syntactic parser for Swedish. A logical step forward from this is indexing routines for entry-based search in large databases, and if users want to search multilingual databases (such as the Norwegian governmental database http://odin.dep.no), the program has to contain routines for matching corresponding terms in different languages. But for obvious reasons we will deal with *Nyno* here.

We performed a pilot test, translating the program *SimpleText* (Bokmål: *Enkel tekst*) from Bokmål to Nynorsk. *SimpleText* is an editor that is shipped with all Macintoshes; it contained 13500 characters. We translated it in two ways, first as a raw translation by using *Nyno* as it is, and then a translation with certain parameters set in an optimal way, major lexical ambiguities resolved, etc. The work of tailoring *Nyno* took a couple of hours. Examples of this work can be seen in Appendix 3, with original text and two different translations aligned. The whole text consisted of 13500 characters, and the manual translation costs NOK 840 at a price of NOK 1990 (EUR 238, USD 258 per 32000 characters according to standard fees). In order to evaluate the machine translation we gave the two versions to a professional translator, and asked for an corrected versions of the two texts. The raw translation was corrected in 45 minutes, and the tailored one in 30 minutes. Prices for proofreading and "manuscript washing" (proofreading of bad texts, or "MT post-editing") vary from NOK 170 (EUR 20, USD 22) per hour to at least five times as much. If we assume an hourly rate of 450 (EUR 54, USD 58), proofreading of the machine translated text translated by tailored version of *Nyno* can be estimated at NOK 260 (EUR 31, USD 34), or 31% of the cost for manual translation. As a rule of thumb, publishers estimate the price for proofreading and manuscript washing at 10% and 30% of manual translation costs, and it can easily be seen that *Nyno*'s output already approaches proofreading level, especially given the real savings in that the base translation is for free. Given that more effort may be put into tailoring the program for software translation, non-automated translation costs at proofreading level or below does not seem unrealistic.

We also took a proofread version of the translated text and compiled it back into the program, the result being a working text editor in Nynorsk, which has been made available online at http://www.egt.ie/earra_bog/simpletext-nynorsk.sea.hqx.

Development of existing lexical databases and translation resources
As shown above, the machine translation system *Nyno*'s performance improved greatly even with relatively small effort. The better a system gets, the more effort it takes to improve it further. One area where the system may become close to optimal is the vocabulary. Work on this area will be twofold: existing vocabulary must be collected, approved, made into a terminology databases (sometimes called "terminology management tools") of the database and new vocabulary must be developed where there is not any, or where there are problems with existing usage. Translation systems based upon existing parallel corpora will be able to make use of existing manually translated parallel texts. As soon as larger text samples are tested, one encounters problems both with the general part of the source language vocabulary (here: Bokmål), with domain-specific vocabulary (here: software) with English loan words, with abbreviations, and with vendor-specific ideosyncrasies. In order to illustrate the problems more thoroughly, we analyzed the Bokmål vocabulary of *Windows 98*, by running it through the *Nyno* translation program, and by running it through a morphological parser provided by Lingsoft. (A morphological parser reads in text and analyzes the strings into recognized parts of speech, prefixes, suffixes, and conjugated forms.)

Lingsoft's system is based on the standard dictionary Cappelen's *Norsk ordbok*; it includes a dynamic morphological component which can recognize productive word-formation, including the compounding of roots more than three graphemes long. This system is the most comprehensive available, and was chosen here in order to report on how far automatic recognition has advanced at present. A vendor-specific system may, of course, have included several of the missing terms already; our point is just to show what must be added to the most comprehensive general-purpose parser available. Some of the words that were not accepted by Lingsoft's morphological parser are listed in Appendix 1.

A corresponding test was conducted with *Nyno*. The Norwegian words that *Nyno* did not recognize are shown in Appendix 2. *Nyno* had the same problems with ad hoc English loanwords and abbreviations as did Lingsoft's parser.

In order to look at the domain-specific vocabulary, we also compared the *Windows 98* vocabulary to Norsk språkråd's existing computer terminology database. Over the past 25 years, Norsk språkråd has done a lot of work on computer terminology, a work that has resulted in (among other things) *Norsk dataordbok*, a dictionary of 5000 entries (both in Nynorsk and in Bokmål), lists of translation equivalents between Norwegian and English, Swedish, and French, and a list of abbreviations. Unfortunately, this dictionary is more geared towards computer science than towards end-user products. The test report in Appendices 1.1. and 1.4. below will illustrate this point. We assumed that the words and abbreviations that were not accepted by Lingsoft's general parser were rejected partly because they belonged to a specialized computer vocabulary (some of the words in 1.1. were rejected for other reasons, as explained in the appendix). Of the 753 abbreviations we recorded, only 18, or 10.8%, were found in *Norsk dataordbok*, and of the 407 Bokmål words that Lingsoft's parser rejected, only 26, or 6.4%, were found in *Norsk dataordbok*. Reality is a bit brighter, due to some error sources in the material: the Norwegian word list contains some of Microsoft's spelling errors as well, and the list of abbreviations includes all wordforms written in capital letters, forming a small number of non-abbreviations. In some cases the terms covered by *Norsk dataordbok* were among the most frequent terms of the material (the appendix includes frequency information), but equally often terms frequently appearing in the text were missing from the dictionary. Our test results show that the vocabulary in *Norsk dataordbok* is concentrated on computer science terminology – but this terminology is largely irrelevant to the vocabulary required in end-user software. Clearly, work is needed in this field, both for concrete purposes, in order to develop working machine translation systems, but also in order to find good Norwegian equivalents to the haphazard direct import and ad hoc translation efforts which are common today.

Due to the way the material is structured, as a ready-made parallel corpus, the time-consuming basic work on term registration and usage context may be done automatically. Special computer terms may be abstracted automatically from the texts, in several ways:

Morphological parsers may parse Bokmål texts and extract all terms not previously found in the existing database. Given a good general parser, a large part of the extracted material will probably be domain-specific terminology. Further manual processing will then evaluate the extracted terms, and eventually provide Nynorsk cognates. In most cases the terminological problems are independent of differences between Bokmål and Nynorsk. They are typically either questions of recognizing compounds involving short stems, or English loan words and abbreviations.

Another way of extracting domain-specific vocabulary is the following. Any specialized corpus (as in the case of a software corpus) may be turned into a frequency wordform list. This frequency list may be compared, wordform by wordform, to a frequency list for ordinary prose, and each wordform's relative frequency in the two texts may be subtracted from each other. The words may then be sorted according to frequency difference, and the result is a list of words with great relative difference in frequency in ordinary and specialized texts: in other words, a list of special terms. This is work that Norsk Språkråd should give priority quite independently of the work connected with Nynorsk software, since terminology for end-user software is one of the unrepresented terminological domains in the existing dictionary. The outcome of such a work will be better lexicographical databases, both for translating between Bokmål and Nynorsk, and for proofreading and parsing both Bokmål and Nynorsk texts.

Summary
For ordinary users the program *Nyno* costs NOK 3800 + VAT (EUR 454, USD 494) for a single-user licence, and the usual discount system for multiple-user licences. How high the development costs of a tailored version of *Nyno* will be is naturally dependent upon how much one is willing to invest in further improvements with limited effects. At the end of the day this will also be decided upon

between vendor and buyer, in this case between Nynodata and different software producers. In our opinion at least parts of this work should be conducted via public funding, and thus be open to all. When it comes to terminology, it goes without saying that Nynodata, as any other actor in the market, will be interested in the termbank of Norsk språkråd, and even more so in the recommendations Norsk språkråd may have on coining new terms. Norsk språkråd, again, will be interested in enlarging its existing database of computer terminology to include also the interface between machine and end user, and not only between machine and programmer, as is mainly the case today. We recommend a form of exchange trade, where the actors in the market make their term lists available to Norsk språkråd, and thus to the public, and receive terminological aid in exchange,

Our report has shown that there is a great benefit in invoking machine translation systems between closely-related languages. As the research on machine translation between closely-related languages progresses, we may expect the threshold for developing such systems to reduce. As a result, localization for language clusters (such as the Scandinavian written languages, the Slavic, Bantu, and Turkic languages, etc.), may be handled by primary manual translation into one key language (say, Swedish, Russian, Swahili, and Turkish), and then subsequent machine translation processing will take care of the versions needed for the other languages in each language family.

Other aspects

PR consequences of broad language coverage

During 1999 several Nynorsk software products have been released, e.g. an internet browser and the user interface of a cellular phone. These products have received a good deal of attention in the media, and a quick calculation indicates that the costs to the producers in translating these products into Nynorsk have been earned back many times in terms of free advertising. (The first Nynorsk internet browser, *Opera*, was released in 1996 but did not receive much media attention.) When more and more products are released in Nynorsk, the novelty interest will of course drop, but what remains is an image of seriousness. Today a word processor vendor who offers the customer proofreading software in many languages signals seriousness and solidity. Eventually multilingual localization will be linked to quality expectations as well. The fact that governmental customers (including the important school sector) may demand fulfilment of language legislation, multilingual program version will be industry's way of signaling its ability to deliver products meeting the quality requirements of governmental customers, in this and in other respects.

Nynorsk as a language in the world

In Norway there are approximately 450,000 users of Nynorsk. This places Nynorsk around number 440 on the list of the world's 6,500 languages (http://www.sil.org/ethnologue). If we discount languages without any official status, the languages of some African countries, and some Asian minority languages, Nynorsk remains as one of the 65 most economically powerful languages of the world (40 of these being situated in Europe). Norway is a rich country with a high computer density, and it goes without saying that this is an interesting market. Even though most Nynorsk users presently use Bokmål programs, the demand for Nynorsk programs will probably grow as users realize that they can in fact be obtained.

Machine translation between other closely-related language pairs

Although this report has been on the relation between Nynorsk and Bokmål, the discussion is clearly relevant to other settings as well. Development of good machine translation systems is no cheap task, regardless of chosen methodology; neither is it impossibly expensive, however. But the fact that software texts represent a well-defined, more or less closed set of terms, occurring in standardized format, makes machine translation an option for other language pairs as well.

Conclusion

Translation of software into Nynorsk is possible, and in fact, it is a relatively straightforward task. It is far cheaper to translate a Nynorsk version of a computer program than it is to translate a school textbook. Firstly there is generally far more text in a textbook, and secondly the language of the textbook will be textually more coherent and lexically more varied; all this makes it harder to make machine translation systems for textbooks.

The total sale of products localized in Bokmål and Nynorsk will probably be larger than it will be if a vendor only is able to present Bokmål localizations. Multiple-user computers will have to have two program versions instead of one, and larger customers will buy two different multiple-user product licences instead of one.

The challenge for the near future, both for (machine) translation circles and for lexicographical circles, will be to develop the tools that are needed in order to reduce costs connected to multilingual software. Nynorsk users must understand that is indeed is possible to get software in Nynorsk, and also begin to act according to that fact. Software vendors must welcome the challenge of providing multilingual software for a multilingual world.

Appendix 1

1. Word forms in *Windows 98* which were not recognized by Lingsoft's morphological parser
The words are grouped into five different sections, according to the reason they were rejected by the parser. In each appendix the leftmost column indicates frequency. Word lists with a frequency of 1 (*i.e.*, 1.1 and 1.4) are exhaustive lists; in lists where the lower frequency numbers are missing (*i.e.*, 1.2, 1.3, and 1.5), only the most common words are included.

1.1. Norwegian words
1.1. contains an exhaustive list of the word forms that met the following criteria: they did not contain hyphens or numbers, and in the case of English loanwords they were assumed to appear in Norwegian texts as well. English words that clearly occurred in English sentences were not included. As expected from the parser design, a large number of native Norwegian words were rejected because they contained roots shorter than 4 letters (*arkformat, bakmating*). Other words are included because they formed compounds from infinitives rather than from nouns (*Skrivebeskyttede* pro *Skriverbeskyttede*). A large number of the words are neologisms, the list thus underscores the creative nature of this lexicosemantic field. Obvious misspellings were retained in the list. Wordforms in **boldface** represent wordforms which are inflectional forms of entries found in *Norsk dataordbok*. Compounds missing from *Norsk dataordbok* are indicated as missing, even when the compound parts are present ("*overflyt*" is represented in *Norsk dataordbok*, but "*overflytsfeil*" is not). Naturally, we do not want to commit ourselves to any judgements on the terms included; the list has been made to indicate the extent of overlap between terminology work done so far and the terminology found in end user products of this type.

1	adskilte	3	Defragmenterer		programmer	1	gråtone	1	Kalleren
3	**aksessere**	1	defragmenterer	3	Diskomprimering	1	gråtonebilder	**3**	**katalogtre**
2	**aliaser**	8	defragmenteres	1	diskplass	1	gråtonegrafikk	**1**	**Katalogtre**
1	Arehangeul	3	Defragmentering	1	Diskrekomprimering	1	gråtoneskala	**12**	**katalogtreet**
1	Arkformat	2	defragmentering	1	diskvertøy	1	gråtonet	**1**	**Katalogtreet**
10	**Arkmater**	3	defragmenteringen	2	DLLIkoner	1	Gullvev	1	Katalogtrenavnet
3	**arkmater**	1	Defragmenteringen	1	dobbelportkontroller	3	Hangul	1	Kilderuting
1	Arkmating	1	Defragmenteringsmetode	1	Dokumentordner	2	Heidelberger	1	Kjøpssted
3	Arkstørrelse	1	defragmenteringsoppgaver	1	Dokumentordnere	2	Heldekkende	18	Klargjør
1	Artikkelnr	2	defragmentert	1	domeneadministatorer	5	heldupleks	4	klargjør
1	Autocad	1	defragmenterte	1	Dovregubbens	2	Hoved	1	klasseprototype
1	autofjerningsfeil	4	dekompressor	4	draknapp	1	Hovedarkmater	1	Klokkeproxy
1	Autofokus	2	Dekomprimer	1	Driverinformasjonsfiler	1	Hovedinnsignal	4	kodekens
1	autokonfigureringsmodus	6	dekomprimere	1	Duplikate	1	Hovedinnsignalnivå	2	Kodekens
1	autokorrekturoppføring	4	dekomprimerer	3	Dvorak	1	hovedopplader	2	kommandolinjeparametre
1	Bahasa	4	Dekomprimerer	4	Egenskapsark	5	huben	1	konsolldelsystem
1	Bakbrems	5	dekomprimeres	1	egenskapsark	2	Hurtigbufferbom	1	kontaktark
2	Bakmating	**5**	**dekomprimering**	5	egenskapsarket	4	Hurtigformaterer	1	kontrollsum
1	Balanseror	**3**	**Dekomprimering**	1	eierskapet	2	Hurtigsøk	1	Kontrollsum
1	Balanserorstrim	1	Dekomprimeringsenheter	2	eksluder	2	Høyrejuster	2	Konvertereren
1	Bevegelsereaktivering	2	Dekomprimeringsfilter	1	Enumeratoren	2	Høyrejusterer	1	Kopikøen
4	bgntrim	2	dekomprimert	4	EPassordet	1	Høyreklikk	1	koprosessor
1	biblioteksfunksjonene	1	Dekomrimeringsprogram	2	feilsøk	**1**	**Initialiser**	1	kortjenester
1	bidiagram	**1**	**dekryptere**	2	filmal	**6**	**initialiserer**	1	Kvalitetsstyringsproxy
4	bidiagrammer	3	Deladresse	2	filsettkatalogen	**8**	**initialiseres**	1	Køfil
1	bidiagramråd	1	delbildestrøm	1	Filsettkatalogen	**7**	**initialiseringen**	1	Kølengde
1	bildemoduset	1	deldokument	1	Filtere	**1**	**Initialiseringen**	1	Køobjekt
1	Biplseksporten	1	Delnettverksmodulen	1	Filterproxy	1	initialiseringsfase	1	køprosessoren
2	Bitfrekvens	1	deloppgave	1	Flerpunktsdigitizer	4	Initialiseringsfeil	1	køstatus
1	**Bufferet**	**2**	**Delsystem**	1	Flyttallsoverflyt	1	initialiseringsfeil	**1**	**Lasteren**
1	bufferoverflyt	**1**	**delsystem**	1	Flyttallsunderflyt	1	Initialiseringsfeil	1	Leddarm
1	Bufferoverflyt	**1**	**delsystemet**	1	Forespørselskøen	4	initialiseringsfilen	1	Legitimasjonsbeskrivelse
2	Bufferoverflytsfeil	**1**	**Delsystemet**	2	foretrukkede	1	initialiseringsrutine	3	lesebeskyttet
1	bugrep	1	Delversjonsnummer	1	Forhåndsformaterer	1	initialiseringsstrengen	1	Lesebeskyttet
14	bussmastering	1	Derastrering	3	Forhåndsvis	1	Initialiseringsverktøy	1	Leseoptimalisering
1	bussmus	3	Deskpro	2	Forløpsindikator	1	initialiserte	1	Linjenr
1	Bussmus	1	diagnosere	12	formelkonvertereren	1	initiell	2	LuxSonor
1	båndstasjonformat	2	Differensiell	1	Frihodet	3	innebygget	1	Lydkodek
1	**Circumflekstegn**	1	differensiell	2	Frihåndslinje	1	Innebygget	4	lydkodeken
3	dagtid	1	digitallydinnsignal	1	Girskifter	2	Innendørsbelysning	1	lydkodeker
3	Databiter	1	digitallydutsignal	1	Gjennomstreking	2	Installasjonprogrammet	1	Lydkodeker
3	datatabellinitialisering	1	Dirkabel	1	Gradianer	1	interoppgave	1	Lydkomprimerings-
7	Defragmenter	4	diskdefragmentering	1	Gråskalert	1	johnsmith@microsoft		kodeker
4	defragmentere	1	Diskdefragmenterings-	3	Gråtone	1	kalleren	1	Lyshet

Freq	Word	Freq	Word	Freq	Word	Freq	Word	Freq	Word
2	Lysstoffrørsbelysning	1	Omadressereren	1	porttypen	1	rotstandard	1	Systeminitialiseringsfeil
1	Løpshøyde	1	Operandstakkoverflyt	2	porttyper	**1**	**ruting**	1	Telefonrør
1	Låsendringen	1	Operandstakkunderflyt	**2**	**priviligert**	**1**	**Ruting**	3	Telesyn
1	låsetillatelse$	1	oppkalleren	2	påloggingsskriptet	8	Rutingsstatus	8	Tilbakering
1	malkategorien	1	Oppkalleren	1	påloggningsskript	2	rutingstabell	1	tilbakeringingsprotokoll
1	malkategorier	**1**	**oppringt**	1	radoverskrift	1	Rutingstabell	2	tilbakerings-protokollen
2	maskinvarekodek	**1**	**Oppringt**	6	ravgul	1	Rødfiolett	1	Tilbakeringingssikkerhet
3	Maskinvaresøk	**1**	**oppringte**	**2**	**refereranse**	1	Rødtoner	1	tildeleren
4	maskinvaresøket	1	Optimaliseringsprosesen	1	Reflektiv	1	Rørtype	**1**	**Tilgansnivå**
3	Maskinvaresøket	1	Optimaliseringsveviser	**2**	**Regneark**	1	sidedør	1	Tilordnere
6	materetning	4	Overflytsfeil	1	Regnværsdag	5	Sideror	**3**	**Udefinert**
1	Meldingskorrelatoren	3	overflytsfeil	2	Regserver	4	sideror	1	uforankret
1	Meldingskøen	3	Parallellportmodus	4	rekomprimere	1	sikkerhetskopierer	**3**	**Uformatert**
1	metafiler	1	parallellportmodus	2	rekomprimeres	1	Skjerminitialiseringen	2	Ugjenopprettelig
1	metafiloperasjon	2	partisjonere	2	rekomprimering	1	Skrivebeskyttede	2	Ukomprimerte
1	metallic	1	partisjoneres	2	rekomprimeringen	2	skrivebufring	1	Uleste
1	metategnutvidelse	1	partisjoneringen	2	Rekomprimeringen	1	Skriverkø	1	Uoppført
1	minikasett	1	Passordbeskytt	1	renummereringsproblem	2	skriverkøen	1	Uorientert
1	minimumsstørelsen	2	peropsys	1	Replikering	1	skriverportmodus	1	uovervåkede
1	miniportdriver	**1**	**piltast**	1	repositoriet	5	socketen	4	uovervåket
1	minnebl	**5**	**piltastene**	9	ringforbindelsen	5	Socketen	1	Uovervåket
9	minneresidente	**1**	**Piltaster**	1	Romaskin	3	socketer	5	Uspesifisert
1	minnesresident	1	Piltastmodus	1	Romhøyttaler	1	spillportkort	1	Ustøttede
2	minnetildeler	2	Pinger	2	romklang	1	splitterenheter	1	Ustøttet
1	Minnetildeler	1	Pistasj	2	Romklang	1	stakkoverflyt	1	Utskriftsbehandlingskøen
1	mkdir	2	Portdriver	2	Romklangseffekt	1	stakkunderflyt	**5**	**utskriftskøen**
1	Modeminitialisering	2	Portforsinkelse	1	Romtemperatur	5	stasjonsdøren	**3**	**Utskriftskøen**
2	**moduset**	1	Portinformasjon	1	rotblokkpeker	1	stasjonsparameterene	1	utskriftskøfilen
1	Monohøyttalere	3	Portinnstillinger	**5**	**rotkatalog**	1	stasjonsparametre	1	uttagbar
1	**monokrome**	1	Portklasser	**1**	**Rotkatalog**	1	Statusoppsamler	1	utvalgsbuffertildeler
1	monomodi	1	portkonfigurasjonen	**4**	**Rotkatalogen**	1	Stemmeinnsignal	1	Vekslebart
1	mottakende	1	Portkonflikter	1	rotkataloppføring	1	Stereomultiplekser	**1**	**Venstrejusterer**
2	muliggjør	3	Portnavnet	1	rotklasse	1	stillisten	2	videokodeken
2	Muliggjør	1	Portnummer	4	rotmappe	**4**	**Stoppbiter**	1	videokodeker
1	nettverksenumerator	1	portnummeret	9	rotmappen	1	Sukkertøyfeens	1	vinduord
1	Nummeroverflyt	1	portreferanse	3	Rotmappen	1	Svartelistet	1	Zoominnstillinger
1	Nødstopp	1	portstatus	1	rotmapper	1	systemgjenkjennelse	1	Zoomoperasjonen
8	omadressereren	1	Porttype			1	systemgjenkjennelsen		

1.2. The combination numbers, hyphens and small letters

1.2. contains a list of orthographic words in *Windows 98* with a frequency higher than 1 which contain numbers and small letter combined with hyphen. The words are sorted according to frequency, and thereafter alphanumerically.

Freq	Word	Freq	Word	Freq	Word	Freq	Word	Freq	Word
29	32-biters	3	2-mus	2	10-lydkodek	2	AWE64-kompatibel	2	S3-skjerm
19	16-biters	3	386-utvidet	2	2-portmus	2	DC25-digitalkamera	2	Twain_32-kildebehandling
7	FAT32-stasjon	3	3D-flygende	2	3D-blomsterfigur	2	EMM386-driveren	2	Win32-påloggingsskript
6	FAT32-konvertering	3	7-biters	2	3D-flyvende	2	FAT32-kompatible	2	Win98-applet
5	8-biters	3	AWE32-kompatibel	2	3D-labyrint	2	Fast-SCSI-2-kontroller	2	Windows95-kompatibel
5	MPU-401-kompatibel	3	FAT32-stasjoner	2	3D-rør	2	MPU-401-kompatible	2	Windows95-kompatible
4	16-bit	3	NDIS3-driver	2	3D-tekst	2	MPU401-emulering	2	monokrom14-tommers
4	24-biters	3	SB16-kompatibel	2	486-baserte	2	NE2000-kompatibelt		
4	FAT32-kompatibel	3	Win32-modus	2	9-pin	2	NE2000-kompatible		
3	1WordPad-dokument	2	10-kodek	2	A20-behandler	2	S3-driver		

1.3. Words with hard hyphen

Microsoft's use of hard hyphens represents a mild break with Norwegian orthographic tradition, according to which "The start menu" should be written "*Startmenyen*", whereas Microsoft thus writes "*Start-menyen*". This use of hyphen is used in Norwegian when one of the parts is an abbreviation ("*IP-adresse*" and "*SCSI-vertskort*" are correct Norwegian), and the list herebelow may be seen as an indication of how the lexical databases must be extended in order to meet actual usage. In Norwegian, the hyphen may be used for clarity (as for schoolchildren: "*lese-bok*"), and any lexicon designed to cover actual usage must thus record the cases of hyphen usage. The use of hyphens in the corpus is extensive; only cases with a frequency higher than 3 are shown.

Freq	Word	Freq	Word	Freq	Word	Freq	Word	Freq	Word
84	MS-DOS	24	CardBus-kontroller	15	CAB-filen	12	TrueType-skrifter	11	AudioDrive-maskinvaren
82	SCSI-vertskort	22	Action-feil	14	sys-filen	12	IP-adresser	10	bps-modem
52	IP-adresse	21	CD-ROMen	14	WDM-streaming	12	DECnet-adresse	10	NetWare-nettverk
48	PCI-bro	20	Aztech-lydkortet	14	Sound-driver	11	WinPad-versjonen	10	MS-DOS-baserte
32	USB-tastatur	18	Thomas-Conrad	14	NCB-forespørselen	11	VGA-skjermkort	10	MIDI-syntese
32	ISA-bro	18	INI-filen	14	IP-adressen	11	Produkt-IDen	9	dobbeltbyte-tegn
29	MS-DOS-modus	17	USB-tastaturmus	14	INI-fil	11	PCI-kort	9	XMS-minne
27	SCSI-kontroller	17	Start-menyen	13	RMM-støtte	11	MIDI-avspilling	9	PCMCIA-kort
25	USB-enhet	17	IP-protokollen	13	DVD-spiller	11	DirectX-driveren	9	PCMCIA-harddisk
25	INF-fil	16	Card-kontroller	12	sys-fil	11	DMA-kanal	9	ODBC-driveren
24	DLL-fil	15	Token-Ring	12	VXD-driveren	11	CAB-fil	9	IRQ-tabellen

8	VxD-initialiseringsfeil	7	MIDI-instrument	5	WinFrame-serverleser	4	sikkerhets-IDen	4	IRQ-verdien
8	USB-sammensatt	7	MDFAT-oppføringen	5	WINS-server	4	nurbs-kurve	4	IRQ-tabell
8	USB-Hub	7	INF-filen	5	USB-spillkontroll	4	non-MS-DOS	4	IRQ-
8	RPC-serveren	7	Hjerter-nettverket	5	USB-mus	4	maskinvare-ECC	4	IPX-tilkobling
8	Play-enheter	6	pwl-kurve	5	System-driveren	4	docWord-dokument	4	IP-konfigurasjon
8	Play-bro	6	musikk-CDen	5	SPX-kompatibel	4	digitallyd-Fx		Kommandolinjealterna-
8	PCIC-kompatibel	6	indikator-blink	5	PCI-	4	bat-filen		tiver
8	NCB-kommandoen	6	Wave-driver	5	NetWare-servere	4	USB-skjerm	4	INF-filer
8	MS-DOS-program	6	WBEM-brukerbehandling	5	MTF-formaterte	4	USB-lyd	4	HP-skrifter
8	MIDI-driver	6	PCI-standard	5	MS-DOS-ledetekst	4	USB-hub	4	HID-lydkontroller
8	Image-dokument	6	MS-DOS-navn	5	MS-DOS-feil	4	Telnet-port	4	GPS-kort
8	FAT-mediebyten	6	MDFAT-oppføringer	5	MOF-fil	4	TR-kontroller	4	EXE-
8	DMA-verdien	6	Inn-	5	MIDI-tilordning	4	Sample-kontrollen	4	DriveSpace-stasjon
8	CD-ROM-etuie	6	IRQ-styring	5	MCA-kort	4	SPX-kompatible	4	DVD-videoavspilling
8	CAB-filer	6	IPX-pakker	5	HID-kompatibel	4	RPZA-format	4	CardBus-kort
8	AudioDrive-driver	6	DVD-plate	5	GO-serien	4	PFM-fil	4	Card-kort
7	proxy-server	6	DLL-filen	5	FAT-filsystem	4	PCMCIA-bro	4	CRC-feil
7	Windows-metafil	6	D-Link	5	Card-enheten	4	PCI-vertskort	4	Blaster-emulering
7	WinFrame-serveren	6	Audio-driveren	5	B-Wave	4	PC-Card-harddisk	4	BIOS-oppgradering
7	WinFrame-server	6	ADPCM-kodek	5	Audio-driver	4	NetWare-server	4	AVI-filen
7	Weitek-koprosessor	6	-s	5	ActiveX-objekter	4	NT-domene	4	AGP-kontroller
7	SCSI-kort	6	-instance	5	ActiveX-funksjoner	4	NCP-pakker	4	AGP-grafikkort
7	QIC-formaterte	6	-class	5	ActiveMovie-kontroll	4	Moire-mønster	4	ACB-ressurser
7	PrintServer-filer	6	-A	5	-Y	4	MS-DOS-vindu	4	A-Law
7	PnP-lydsystem	5	retail-versjon	5	-U	4	MS-DOS-programmer	4	-P
7	Mixer-driver	5	musikk-CDer	5	-N	4	MFC-program	4	'-oppføring
7	MIDI-port	5	inn-	4	trim-kurver	4	MDFAT-oppføringene		

1.4. Abbreviations

Windows 98   contains many abbreviations. If we define words containing capital but not small letters to be abbreviations, we find 756 abbreviations not accepted by the parser. This somewhat unorthodox definition gives quite a good estimate, since the source text writes e.g. "BIN" instead of "bin" or "bin." for "binary"). The list also illustrates to what degree the terminology in end-user products like *Windows 98* differ from the terminology found in programming-language environment: of the 756 abbreviations in the list below only 101, or 13%, are found in the list of abbreviations included in *Norsk dataordbok*. Abbreviations found in both sources are set in boldface in the list below. We do not evaluate whether the overlapping abbreviations actually mean the same thing; the question is rather how well a system equipped with the terminology of *Norsk dataordbok* is able to cover the letter combinations in actual use.

1	ACC	1	BACKUP	3	CICT	2	CVT	16	**DMA**
3	ACCM	9	**BAT**	8	CIMOM	4	CWBAUDIO	3	**DMS**
1	ACCPAC	1	BFT	2	CIPX	5	CWBFM	9	DNI
1	ACE	2	BIN	1	CLASSIFY	1	**DAC**	2	**DOC**
2	ACER	25	**BIOS**	3	CLP	2	DACL	1	DOMAIN
1	ACF	1	BIOSXLAT	1	**CLS**	2	DAO	1	DOSKEY
2	ACL	3	BLAH	1	CLSID	2	DART	1	DOSMGR
1	ACM	1	BLSPACE	2	CMBR	2	DATE	1	DOSNET
10	ACPI	1	BM	1	CMCDD	1	DBC	2	DPMI
8	ADI	1	**BMP**	5	CMD	2	**DBLSPACE**	12	DPT
9	ADPCM	1	BOCALAN	1	CMM	1	**DCA**	2	DRACO
1	ADSC	1	BOOTLOG	2	**CMOS**	2	**DDE**	45	DRV
1	AGC	3	BRILLIANCE	2	CMPAGENT	1	DDEOLE	2	DRVSPACE
1	AGP	2	**BST**	1	CODE	1	DDK	1	DSL
5	AOC	2	BTC	11	CODEC	3	DEBUG	21	**DSM**
1	AOCB	1	BUFSIZE	1	CODEPAGE	39	DEC	7	**DSP**
1	AOL	14	CAB	1	COLS	1	DEFRAG	1	DTA
6	**API**	1	CADD	1	COMBUFF	2	DELAY	1	DTC
1	API_CONFORMANCE	1	CALL	7	COMMAND	3	DELETE	1	**DTE**
9	APM	1	CALS	1	COMMDLG	2	DELTREE	2	**DTMF**
1	APPS	2	CARDDRV	1	CON	1	DEPCAJR	2	DTR
5	**ARP**	1	CARDS	17	CONFIG	1	DETCRASH	1	DUN~
1	ARROW	1	CAV	1	CONFIGMG	1	DEVICE	5	DVD
2	ASD	3	CCB	1	CONTAQ	2	DEVICEHIGH	1	DWC
3	ASF	4	CCP	1	COPY	1	DEVMODE	6	DWORD
2	ASPI	1	CDB	2	COPYCMD	1	DEVNAMES	1	DXB
4	ASSIGN	1	CDT	2	CPE	1	DF	2	**DXF**
6	**ATA**	3	CEO	1	CPI	1	DGDFPCI	1	DXINFO
20	ATI	1	**CGM**	1	CRT	7	DHCP	1	DXTOOL
1	ATTRIB	3	**CH**	2	CS	1	DIAG	1	DZ
1	AUTH	1	CHANGE	1	CSD	1	DIAGNOSTICS	1	EA
9	**AUTOEXEC**	2	CHAP	2	CSLIP	8	**DIB**	1	EAX
3	AUTOFIX	2	**CHCP**	5	CTRL	1	**DIF**	1	EAZ
1	AUW	3	**CHDIR**	1	**CTS**	1	**DIR**	2	EB
1	AUWD	2	CHECKONLY	4	CTX	4	DISKCOPY	1	EBIOS
8	AUX	1	CHK	1	CUPID	2	DLC	1	ECC
1	AUX_CCB	12	**CHKDSK**	1	CUR	1	DLCHLP	6	ECHO
5	**AVI**	4	CHOICE	1	CUSTOM	1	DLE	1	ECM
1	AWE	1	CHOOSUSR	3	CVF	43	DLL	1	ECP

Count	Term
1	EDIT
5	EET
1	EFLGS
2	EIP
23	EIZO
1	**EMM**
1	ENABLE
1	ENABLEHOOK
5	ENET
1	EOJ
2	EOTP
1	**EPP**
1	EPROCLIM
10	EQ
1	ERASE
4	ERRORLEVEL
31	**ESC**
3	ESCD
2	ESFM
1	ESSMPORT
1	ESTALE
1	ETSI
4	EUC
1	EUSERS
1	EXECUTIVE
2	EXIST
1	EY
1	EYEDOG
1	EZ
1	FACTONLY
1	FB
2	FC
2	**FCB**
1	FCLASS
1	**FDDI**
1	FDI
1	FDISK
1	FFFF
1	FIOLOG
2	FIX
2	FLE
3	FORMATETC
1	FOURCC
1	FPS
2	FQDN
2	FREE
15	FSA
3	FST
2	FUJITSU
1	FW
11	FX
1	GCR
1	**GDI**
1	GDIEXT
1	GEM
1	**GIF**
1	GM
3	GO
1	GOTO
1	**GP**
2	GROUP
1	GRP
3	**GS**
2	GTB
1	GW_
2	GX
1	HCI
10	HCO
1	HELP
2	HIGHSCAN
2	**HIMEM**
1	HINF
1	HISTORY
1	HITACHI
1	HKEY
6	HMA
1	HOTKEY
2	HPPCL
1	HPSCNMGR
2	**HR**
4	HRC
1	HRGN
2	HT
2	HWINFO
1	IAF
2	IBMTOK
2	**IC**
22	ICL
4	ICM
3	ICO
1	**IE**
7	**IEEE**
1	IET
2	IF
3	IFEIL
2	IFS
1	IFSHLP
1	IFSMGR
1	IGES
1	IIBMTOK
1	IIPX
3	IIT
1	IIXL
6	IMA
32	**INI**
1	INIT'
1	INSERT
10	INSTALL
2	INT
1	INTERLNK
1	INVISIBLE
1	IO
2	IOCTL_SERIAL_-XOFF_COUNTER
1	IOS
27	IP
17	IPCP
19	**IPX**
10	IPXCP
1	IRENUM
36	**IRQ**
1	ISP
2	ITAC
2	ITC
1	IWP
2	IYUV
7	JIS
1	JN
1	JOIN
1	JPG
1	JPOSTD
1	JVC
1	KCMS
1	**KEYB**
1	KEYSIZE
4	KFC
4	KIR
1	KLH
7	KODAK
1	KOPIBESKYTTET
1	LABEL
6	LANA
1	LANTIC
1	LB_SETCOUNT
5	**LC**
1	LCID
2	LCIT
2	LCOS
42	**LCP**
2	LDID
1	LEVINE
1	**LF**
1	LFN
3	LFNFOR
2	LOCK
2	LOGIN
2	LOGOFF
3	LOW
10	**LPT**
1	LSA
1	LSL
1	LST
2	LTE
2	LZH
2	MACHINES
1	MACROS
1	MAP
1	**MAPI**
2	MBF
2	MBR
1	MC*
20	**MCA**
6	**MDA**
5	MDBPB
7	MDFAT
3	MEM
1	MESSAGE
1	MEX
1	MFC
1	MFCDB
1	MFCDLL
1	MFCNET
1	MFCOLE
8	MGA
2	MIC
49	MIDI
3	MIFF
1	MJ
1	MKDIR
3	MLID
1	MMIO
1	MODULE
4	MOF
1	MOVE
1	MPC
2	MPEG
1	MPI
1	MPX
1	MRCI
3	MRP
1	MRRU
2	MRU
5	MSACM
1	MSCOMPARE
10	MSCSD
3	MSDOS
4	MSG
4	MSN
2	MSPSRV
1	MSSBLST
2	MSSNDSYS
1	MTD
2	MTF
4	MVIFM
18	MVIWAVE
7	MVMIXER
10	MVPROAUD
1	N*
1	NABTS
11	**NAK**
34	NANAO
1	NBA
4	NBFCP
1	NCAA
4	NCB
13	NCR
1	NC_PAINT
1	NDC
8	**NDIS**
2	NDISWAN
5	**NDS**
54	NEC
3	NESW
11	NET
1	NETBEUI
1	NETH
1	NETX
2	NFO
1	NHL
2	NIC
1	**NIU**
2	NLSFUNC
3	NMB
6	NME
2	NMEA
1	NMI
16	NOKIA
1	NOPROMPT
2	NOSAVE
1	NRN
2	NTI
1	NTKERN
2	NTLMDOMAIN
1	NTSC
1	NTT
1	**NUL**
4	NUM
1	NVRAM
12	NVS⁻
1	NW
1	NWC
2	NWLINK
6	NWREDIR
3	NWSE
12	N⁻SV
2	OCX
4	ODBC
4	**ODI**
14	**OEM**
4	OEMSETUP
5	OHCI
4	OHP
4	OLESTREAM
1	OLEVERB
7	ON
1	OTI
2	PAGE
1	PAGEFILE
1	PAGESWAP
5	**PAP**
3	PARITY
2	PASSWORD
5	PATH
9	PATHWORKS
1	PBR
239	**PCI**
2	PCIBIOS
1	PCIC
5	**PCL**
37	**PCMCIA**
2	PCNET
3	PCTOOLS
2	PCX
1	PDF
1	PD_RETURNDEFAULT
2	PERF
2	PFM
6	PFS
3	**PGDN**
1	**PGUP**
1	PIC
1	**PICT**
2	**PIF**
1	PIO
1	PIONEER
1	PIPC
1	PISETUP
16	PLUS
1	PLX
7	POV
1	PPDS
7	**PPP**
4	PQET
1	PREPARE
1	PRINTSCRN
1	PRN
6	PROOFSCREEN
1	PROTMAN
1	PSCRIPT
1	PTL
1	PW
1	PWS
3	**QA**
1	QASM
3	QEMM
1	**QIC**
1	QP
1	QST
2	QUIET
4	QWERTY
5	RC
1	REALMODE
2	REBOOT
2	REFRESH
6	REGEDIT
1	REMOVE
1	REMREM
1	RENAME
1	REQUIRED
1	RESET
1	RESTORE
11	RET
3	RETRY
2	**RGB**
1	RL
2	RLE
1	**RMDIR**
1	RNS
1	ROHM
9	**RPC**
1	RPL
2	RPTI
1	RSVP
1	RTE
6	**RTF**
1	**RTS**
1	RUNREG
1	RWUS
1	S*
1	SABME
8	SAMTRON
4	**SAP**
2	SBAWE
6	SCANDISK
1	SCANREG
1	SCI
3	SCM
2	SCP
11	SCSI
1	SDMS
1	SDVXD
1	SELECT
1	SET
4	SETUP
5	SETVER
3	SIDM
1	SIO
4	SIZE
1	SMALL_FRAME
2	SMARTDRV
1	**SMB**
46	SMC
1	**SME**
1	SMPTE
2	**SNMP**
4	SONY
2	SO_KEEPALIVE
5	SPECTRUM
2	SPEECH
1	SPERRET
4	SPID
1	SPI_
2	SPOOLER
2	SPX
2	SQL_CONFORMANCE
2	SQL_OSC_MINIMUM
1	SRS
1	SRVAPI
2	STACKS
3	STB
1	STOP
3	STSL
3	SUH
2	SURFACE
1	SVD
2	SVEC
5	**SVGA**
2	**SW**
1	SYS$NODE
1	SYSBCKUPRBBAD
1	SYSPUBLIC
1	T*
1	TAB
8	**TAPI**
2	TCMSETUP
83	**TCP**
1	TCPIP
1	TFPT
1	**TFTP**
1	TIF
1	TIME_FORMAT_NONE
2	TITLE
1	TLOCMGR
1	TMP
1	TODO
1	TONERSAVER
5	TP
3	TP_BNC
5	TR
1	TRACERDD

1	TRID_COM	1	UNMOUNT	1	VGATEWAY	2	VSNDSYS	1	WWII
1	TRM	6	US	1	VIONA	1	VTD	4	WYSE
2	TRN	1	USANN	2	VISCA	1	VTDAPI	**1**	**WYSIWYG**
5	TSR	51	USB	1	VKD	22	VXD	**1**	**XA**
1	TSS	2	USER	1	VL	2	VXDLDR	2	XCOPY
4	TTL	1	UTA	**1**	**VLM**	12	V⁻	1	XD
6	TTY	8	UUID	**8**	**VLSI**	6	WBEM	1	XEROX
6	TVM	2	UW	1	VMCPD	1	WBEMTEST	**2**	**XGA**
4	TWAIN	1	VBI	1	VMD	15	WD	1	XH
3	TX	2	VCACHE	1	VMM	67	WDM	1	XIF
15	TXT	1	VCD	1	VMOUSE	3	WE	1	XM
1	UART	1	VCDFSD	1	VMPOLL	3	WFW	**7**	**XMS**
12	UBACKUP	1	VCI	2	VNETBIOS	8	WIN	1	XOFF
1	UCR	1	VCOMM	8	VNETSUP	1	WINDOWSSYSTEM	1	XON
1	UDF	1	VCOND	1	VPD	1	WINHELP	1	XP
2	UDP	**1**	**VDD**	2	VPE	10	WINS	2	XY
1	UGA	1	VDEF	1	VPI	1	WKS	2	YAMAHA
1	UI	2	VDMAD	1	VPICD	1	WM_MOVE	4	YES
5	UMAX	**2**	**VER**	1	VPN	1	WOL	4	YUV
5	UMB	4	VERIFY	1	VPOWERD	1	WORDPAD	1	ZDS
7	UMC	3	VESA	**2**	**VRAM**	5	WORKGRP	1	ZERO
1	UNDELETE	4	VFAT	7	VREDIR	1	WOW	**1**	**ZIP**
1	UNDO	3	VFBACKUP	1	VSB	**10**	**WP**	1	ZVPORT
7	UNFORMAT	1	VFINTD	1	VSD	**3**	**WPS**	2	^ENTER^
1	UNIMODEM	1	VFLATD	3	VSERVER	1	WRI	1	^ESC^
2	UNINSTAL	61	VGA	1	VSHARE	2	WSS		

1.5. Combinations of letters and numbers

The combination of letters and numbers is a problem for spell-checkers. To illustrate the problem, we give the most frequent hybrid words which did not make it through the parser, with frequency given in the left column.

85	1s	14	SB16	9	filnavn2	6	D3
58	1u	14	H1	9	filnavn1	6	2lu
49	1d	14	A3	9	ES688	6	1s'
48	FAT32	13	SU0350Kan	9	B4	6	1c
45	0x0414&&pver	13	F3	9	A5	6	114u
41	EMM386	13	4s	8	R3	5	bane2
31	2s	13	2d	8	NetFlex-2		
29	32-biters	13	2-10u	8	MV80		
27	1lu	13	1-10u	8	Krnl386		
23	0x	12	stasjon2	8	ES1788		
19	16-biters	12	stasjon1	8	ES1688		
18	A4	11	S3	8	A6		
17	08x	11	ProNET-4	8	08X		
17	02d	11	NT3	7	Windows0001Word		
16	F4	11	MPU-401	7	OPL3		
16	3D	11	F5	7	FAT32-stasjon		
16	1ld	11	B5	6	PF-30		
16	10&&ar	11	03d	6	Macintosh0001Word		
15	windows98clcid	10	F1	6	M80		
15	3s	10	04X	6	FAT32-konvertering		

Appendix 2

The vocabulary of *Windows 98* and *Nyno*

The list below presents the Bokmål Norwegian words of the *Windows 98* text which were not recognized by *Nyno*.

abcdefghijklmnopqrstuvwxyz
æøå
adapterlag
afrika
akseleratorfunksjoner
alfaformat
algoritme
aliaser
Andreplass
angitte
animasjon
annotbar
antivirusprogrammer
antivirusprogramvaren
API'er
APIen
APIer
apper
ascii
asynkront
autentiseringsprotokoll
automatiseringsserveren
automodus
autoreferanserer
avbildingsfil
Avbryt
avbryte
Avbryter
avbrytes
avdatasektorene
avhengighetssetning
avi
avinstallasjon
avinstallasjonen
avinstallasjonsdisk
avinstallasjonsfilene
avinstallasjonsinformasjon
avinstallasjonsverktøy
avinstaller
avinstallere
avinstalleres
avinstalleringsdisk
avinstallerte
avkorte
avkortet
avkrysningsboks
avkuttes
avkuttet
avledede
avringforbindelsen
avspilt
avvideosekvenstil
bakover
bekreftelsesinformasjons-
klasse
bekreftelsesmeldinger
beriktigelsesfil
bes
beskyttelsesfeil
betaversjonen
bgntrim
bibliotekfil
bidiagram
bidiagrammer
bildefilnavn
bildehistogram
bindingsmodus
bindingsprosedyre
bindingsprosedyretype
bindingsreferanse
bindingsserver
bindingsservere
bindingsservernavn
bindingssikkerhet
binær

binære
binærfil
binært
BIOSen
blitsinnstilling
bmp
brukerfil
bruktmidlertidig
bugrep
bussmus
bytefor
bærbare
båndakseleratorkort
båndstasjonformat
båndstasjonskontroller
caliope
charformat
charset
configurationerror
datastrømpunkt
datastrømpunktene
datastrømpunktet
datatabellinitialisering
deaktiver
deaktivere
deaktiverer
deaktiveres
deaktivering
deaktivert
deallocate
debuggingsmodus
dedikert
dekode
dekodingsrektangel
dekodingsrektangelet
dekompressor
dekomprimere
dekomprimerer
dekomprimeres
dekomprimering
dekomprimert
desimalt
deskriptor
detektering
diagnoseenhet
diagnostisk
dictstackoverflow
dictstackunderflow
digitallydinnsignal
dir
direkteavspillingsenhet
Diskcopy
diskdefragmentering
diskettype
diskettypene
diskkonfigurasjonen
diskomadressering
diskreparasjonsprogrammet
diskvertøy
disses
DLLer
DLLIkoner
dmachannel
dmapages
Dobbeltklikk
dobbeltklikke
dobbeltsidig
dokumentmateren
dokumentsnarvei
dosrep

drivereller
drivernames
drv
dupleks
duplekser
Duplekserdekslet
duplekserenhet
duplekserenheten
duplikate
dvdplay
effekthøyttaler
egenskapsark
egliste
eilkonfigurasjon
ekisterende
ekthetsgarantien
emulatorer
emuleringsklient
enalvorlig
Encoder
endcurve
endtrim
enenkelt
enhetforespørsel
enhetsbestemt
enhetsdriver
enhetsdrivere
enhetsdriveren
enhetsdriverkomponent
enhetsdriverne
enhetsfilen
enhetsfiler
enhetsforespørsel
enhetsgrensen
enhetsinformasjon
enhetsinnstillingene
enhetsklasse
enhetsklassene
enhetsklasser
enhetsnavn
enhetsnr
enhetsnummer
enhetsomgivelsene
enhetsopptak
enhetsparametere
enhetsskrifter
enhetstype
erstattingstabellen
Ethernet
ettjenestevisningsnavn
execstackoverflow
fakser
faskø
fargebehandlingsprogram
feilrettingsprogram
feilsøkingsmeldinger
filbekreftelseslogg
filesize
fillin
filmal
filmerketabell
filnavn
filparser
filsystemdriver
filtildelingstabell
filtildelingstabellen
filtildelingstabeller
filtilgangsstreng
filtilordning
filtype
filtypeverdi
flerfunksjonskort
flerspalteformat
flyttbar
flyttbare

flyttbarhet
flyvning
fokuserbar
forankringsbro
forankringsvertsbro
forbedringsalternativ
forbindelsesbrudd
forekomstdata
forekomstdeklarering
foreldrenivå
forenelig
forespørselsbufferen
forespørselskode
forespørselsmodus
forespørselsversjon
Formaterer
formateres
formatert
formtext
fragmenteres
fragmentert
fremgangsmåte
Fremkallingsenheten
French
Ftp
fullskjermmodus
gjenopprettingsdiskett
gjenopprettingsdisketten
gjenopprettingsdiskettene
gjenopprettingsdisketter
gjenopprettingsknappen
gjenopprettingsverktøy
gjenoppretting
gjenprøving
gjentatte
gjenvunnet
globbing
gmlpassord
grafikkort
grafikkortet
gråtoneskala
harddiskpartisjonen.
heksadesimal
heldupleks
heldupleksoperasjon
hjelpeboble
hjelpeskjermbildet
hodesporingsenhet
hookswitch
hovedserveren
huber
hurtigbufrede
hurtigbufringsprogram
hurtigbufringsprogrammet
Hurtigformaterer
hurtigformateres
høyttaler
høyttalere
høyttaleroppsettet
høyttalertelefon
ikkekonfigurert
implementerer
Implementeringsgrense
implementeringsgrensen
implementert
implemetert
inaktiv
inaktivitet
inaktivt
includepicture
includetext
indikatorlys
informasjonsfil
ini
inimized

initialiserer
initialiseres
initialiserte
inkompatibel
Inkompatible
inkonsistens
inndatafil
inndatastrømpunkt
inndatastrømpunktet
inneholderangitte
innenhet
innfelt
innfil
innsignallinje
innskuffnummer
innvideosekvenstil
installasjonsfil
int
interaktiv
interaktive
interaktivt
interfacet
invalidaccess
invalidfileaccess
invalidrestore
invertere
invertert
ioerror
IRQen
IRQener
isapi
iso
jpg
kabling
kalibrerer
kalibrert
kamerainnstillingene
kapabiliteter
kastiljansk
katalogfil
Katalogsektorgruppe
Kaukasus
Kjørbar
klarertedomenet
klientvindusprosess
kodakimg
kodek
Kodekens
kodeker
Kodiak
kodingskonfigurasjon
kommandofil
kommandolinjeredigering
Kommunikasjonshøyttaler
kommunikasjonskonfigura-
sjonen
Komorene
kompatibelmed
kompatilbel
kompilator
komprimeringsstrukturen
konfigurasjon
konfigurasjonen
konfigurasjonene
Konfigurasjonfiler
konfigurasjons
konfigurasjonsalternativer
konfigurasjonsavvisning
konfigurasjonsbehandling
konfigurasjonsdata
Konfigurasjonsendringer
konfigurasjonsfeil
konfigurasjonsfilene
konfigurasjonsfiler
konfigurasjonsinfo

sikkerhetskopieringsjobb
sikkerhetskopieringsjobben
sikkerhetskopieringsjobbåpne
sikkerhetskopieringsprogram
sikkerhetskopiert
sikkerhetskopierte
sikkerhetskopifil
Sikkerhetskopifiler
sikkerhetskopijobb
sikkerhetskopijobber
sikkerhetskopimappen
sikkerhetskopiserver
sikkerhetskopiservere
sikkerhetskopisett
sikkerhetskopisettet
sikkerhetsmyndigheten
sikkerhetsnivået
sikkerhetsoperasjon
Sikkerhetsoperasjonen
Sikkerhetssertifikatets
sirkulær
sk
skalerbar
skalerbare
skalering
skaleringen
Skanningsinnstillinger
skanningslinjekontroll
skanningsområde
Skanningsrammen
skanningsside
skipif
Skjermgamma
skjermhøyttalere
skjermkonfigurasjonsverktøy
skriftadresseringsfil
skrifterecho
Skrifterstattingstabell
skrifterstattingstabellen
skriftfil
skriftinformasjonen.
skriftsnittfiler
Skriptfil
Skrivebeskyttelsesfeil
skrivebordskonfigurasjonen
skrivebufring
skriverkalibrering
sletteerstattingen
smelteenhetsområdet
sn
snarveisfil
socketen
socketer
socketkontroller

Somalia
somer
soundblaster
spline
Sporingsnr
Spørringsresultat
spørringstype
stackoverflow
stackunderflow
standardverdieneog
stasjoneneller
stasjonersom
stasjonmed
statusbar
stegvis
stempelet
Stempelfil
stereohøyttalere
stoppetøk
stoppunkt
Stoppunktsfelle
strømsparingsfunksjonen
Strømsparingsmodus
styleref
styrespakenhet
størrelsesinformasjonen
størrelsestilpasning
subclassing
subst
Superklasseinformasjon
Symbios
syntaxerror
synth
synthesizeren
synthesizerkort
synthesizerlydene
Syquest
Systemenhetsdriver
systemfilenetil
Systemgjenopprettings-
  program
systemgjenopprettings-
  prosessen
systemkonfigurasjon
systemkonfigurasjonen
sø
Søkerekkefølge
Søkerute
Targa
tastaturmus
Tatung
tc
tcmsetup
tcp

tegnbasert
Tegnhastighet
tegnlesing
tegnoversettingstabellen
Tegnsett
tegnsettet
tegntabell
Tegntabellbruk
tegntabelldriveren
tegntabellen
tegntabellnummer
Tegntabellstatus
tegnvalg
tekststrengparameter
tekstuformatert
tekstur
Teksturen
Teksturer
teksturfil
telefonenkjører
telefonienheten
telefonienheter
telefoniservereren
Telnet
telnetprogram
Terminalmodi
terminert
testfil
tftp
Thinkpad
tif
tilbakeringingsprotokoll
tilbakeringingsprotokollen
Tilbakeringingssikkerhet
tilbakerullingsbuffer
tilbakerullingsbufferen
tilbakestillingskommando
tildeleoverføringsbuffere
tildeleren
tilgjengeligog
tilgjengelisg
tilkoblingerer
Tillat
timbrag
tisifrede
Tjenesteproxy
tm
tmp
toa
toc
tooltips
topptekstskrift
tps
trackbar

trackingnumber
Transducer
transiente
transparentadapter
Tredimensjonalt
trenavn
TVTuner
TVView
Twain
txt
TXTView
tymed
typ
typecheck
Tyrkia
Tøm
Ubrukte
uformatert
uken
ukjent
uklart
ukomprimert
ukomprimerte
ulagret
Ultima
ulåst
umerkede
unassemble
undefinedfilename
undefinedresult
underholdningsplattform
Unisys
unmatchedmark
unregistred
Untouch
Uoppført
uovervåkede
uovervåket
updateonly
updown
uprivilegerte
uregistrert
URLen
uskarpt
Uspesifisert
Ustøttede
ustøttet
utdatafil
utdatastrømpunkt
utdatastrømpunktet
utgitte
uthevingsverdier
utklipp
utklippsbok

utklippsboken
utklippsbokside
utklippsboksiden
utklippsprosent
utklippstavle
utklippstavlefil
utklippstavlefilen
Utklippstavlefiler
utklippstavleformat
utklippstavleformater
utklippstavlen
utklippstavleområde
utsignalkilder
utsignallinje
utskuffnummer
utvidbare
UUEncoded
Vadem
Valider
Varslingsprogram
vekslefil
vekslingsfilfeil
Ventura
verdana
verdisettingsboks
verifiering
verifyint
Vesa
vidcap
videodatastrømmen
videodekoding
videoklipp
videokodeker
Videresend
vido
vidtspennende
Viola
virtualisere
Vista
volumfil
volumserienummer
WAVjammer
Web
Winmodem
Word
Wordstar
Workbook
World
WPMail
WPOffice
Ytelsesadvarsel
østlig

Appendix 3

Outcome of two tests with *Nyno* machine translation

This appendix shows a representative sample of the commands of the word processor *SimpleText*. The text in the leftmost column is the original Bokmål version, whereas the second and third colums show two different outcomes of machine translation with *Nyno*. The first one gives Nyno "as is", with no further settings. The "ALT" options show cases where the programs signals that several options are possible, "HOM" indicates homonymy, and "UKJENT" indicates that the relevant word is not found in the *Nyno* lexicon. "‰" indicates a paragraph break in the displayed text. The rightmost column gives the output after an hour or two of tayloring, including a learning routine for the most frequent multiple choices. Even the second translation needs proofreading (cf. discussion in the main text of this report).

Original Bokmål text	First approximate translation	Second translation, tailored setting
Arkiver	Arkiver	Arkiver
Avbryt	UKJENT:Avbryt	Avbryt
Ikke arkiver	Ikkje arkiv	Ikkje arkiv
"Arkivere endringer i dokumentet ""^0"" før du lukker?"	"Arkivere endringar i dokumentet ""^0"" før du lukkar?"	"Arkivere endringar i dokumentet ""^0"" før du lukkar?"
Finn	Finn	Finn
Finn hva?	Finn kva?	Finn kva?
Skill store/små tegn	Skil store/små ALT:tegn	Skil store/små teikn
Fortsett fra begynnelsen	Fortset frå ALT:begynnelsen	Fortset frå starten
Arkiver	Arkiver	Arkiver
OK	OK	OK
Enkel tekst ‰ ‰ Dette er et enkelt program du kan bruke for å åpne Les meg-filer, ‰ tekstfiler, filmer, 3D-filer og endel grafikkfiler og for å skrive brev og ‰ notater.	Enkel tekst ‰ ‰ Dette er eit enkelt program du kan bruke for å opne Les meg-filer, ‰ tekstfiler, filmar, 3D-filer og UKJENT:endel grafikkfiler og for å skrive brev og ‰ notater..	Enkel tekst ‰ ‰ Dette er eit enkelt program du kan bruke for å opne Les meg-filer, ‰ tekstfiler, filmar, 3D-filer og UKJENT:endel grafikkfiler og for å skrive brev og ‰ notater..
Størrelse	ALT:Størrelse	Storleik
Halvfet	Halvfet	Halvfet
Tilpass bildestørrelse til vindu	Tilpass ALT:bildestørrelse til ALT:vindu	Tilpass biletstorleik til vindu
Arkiv-menyen ‰ ‰ Bruk denne menyen til å åpne, lukke, arkivere og skrive ut dokumenter fra ‰ Enkel tekst, og for å avslutte Enkel tekst.	Arkiv-menyen ‰ ‰ Bruk denne menyen til å opne, lukke, arkivere og skrive ut dokument frå ‰ Enkel tekst, og for å avslutte Enkel tekst	Arkiv-menyen ‰ ‰ Bruk denne menyen til å opne, lukke, arkivere og skrive ut dokument frå ‰ Enkel tekst, og for å avslutte Enkel tekst
"Åpner et nytt Enkel tekst-dokument kalt ""Uten navn"". Ikke ‰ tilgjengelig fordi det er en dialogrute på skjermen."	"Opnar eit nytt Enkelt tekst-dokument kalla ""Utan namn"". Ikkje ‰ tilgjengeleg fordi det er ei dialogrute på skjermen.."	"Opnar eit nytt Enkelt tekst-dokument kalla ""Utan namn"". Ikkje ‰ tilgjengeleg fordi det er ei dialogrute på skjermen.."
Viser en dialogrute som gjør det mulig å velge et eksisterende ‰ Enkel tekst-dokument. Ikke tilgjengelig fordi det er en dialogrute på ‰ skjermen.	HOM:Viser ei dialogrute som gjer det mogeleg å velje eit ALT:eksisterende ‰ Enkelt tekst-dokument. Ikkje tilgjengeleg fordi det er ei dialogrute på ‰ skjermen..	Viser ei dialogrute som gjer det mogeleg å velje eit ALT:eksisterende ‰ Enkelt tekst-dokument. Ikkje tilgjengeleg fordi det er ei dialogrute på ‰ skjermen..
Viser den angitte siden i dokumentet.	HOM:Viser den UKJENT:angitte HOM:siden i dokumentet..	Viser den UKJENT:angitte HOM:siden i dokumentet..
Lar deg søke i det aktive vinduet etter angitt tekst.	Lèt deg søkje i ALT:det aktive vinduet etter angitt tekst	Lèt deg søkje i det aktive vinduet etter angitt tekst
Spiller av lydopptaket.	HOM:Spiller av lydopptaket..	Spelar av lydopptaket..