

Morphological Processing and Computer-Assisted Language Learning

John Nerbonne, Duco Dokter and Petra Smit

Alfa-informatica, BCN, P.O. Box 716,
University of Groningen
NL 9700 AS Groningen
The Netherlands

Tel +31 50 363 59 74
email: nerbonne@let.rug.nl

Abstract

Contrary to most current practice and contrary to the explicit comments of some practitioners, natural language processing (NLP) can now play a valuable role in computer-assisted language learning (CALL). This paper reports on GLOSSER, and discusses the position of NLP within CALL using GLOSSER as an example. GLOSSER is an intelligent assistant for Dutch students learning to read in French. It has been fully implemented and tested, and it offers information on approximately 30,000 different words (or rather: lexemes), which may be taken from **any** text (no special preparation is required). The assistance takes the form of:

- information on the grammatical meaning of morphology;
- entries in a bilingual dictionary;
- examples of word use taken from over one million words of text (including some bilingual text).

The application has received a warm welcome in user-studies, and has been found a useful tool by students. It relies essentially on lemmatization, part-of-speech (POS) disambiguation, lexeme indexing, and bilingual text alignment—all elements of NLP technology.

Keywords: Vocabulary, Reading, On-line Dictionary Access.

The CALL Perspective

While computer-assisted language learning (CALL) is often cited as an application area for natural language processing (NLP) (Zaenen and Nunberg, 1996), Zock (1996) and others have noted the following discrepancy: most CALL programs in fact make little use of the language technology developed by NLP specialists.

This point may sound paradoxical: Aren't CALL programs by definition language technology? It is crucial to the general aims of this paper that we clarify what is understood under the term *language technology*.

Language technology is the panoply of techniques that are used to carry out tasks that are *specific* to language, in contrast to techniques that can be used for more general purposes. Examples of language technology are: speech recognition, lemmatization (finding the stem or *lemma* of an inflected word), parsing, text generation, speech synthesis (converting text to speech), or part-of-speech (POS) disambiguation (finding the syntactic category of a word)--even when this is ambiguous as in the word *left*, which can be a noun (*on the left*), verb (*She left*), adjective (*the left side*), or adverb (*Turn left!*).

Although CALL employs the computer to assist in language teaching and in language self-study, most CALL programs make little essential use of language technology, exploiting instead hypertext, digital audio and video, (simple) database technology and network communication.

There are several reasons for this. Many current CALL programs focus on drills and exercises, answer keys and grammar explanations, thus, putting self-study courses into electronic form (Last 1992). This implies that existing resources (books, exercises) be reworked for computer deployment, and the language data used for this approach can be hand-coded and “hard-wired” into an application, obviating the need for language technology with its language processing charter. Drills and exercises require that any processing of user input make allowance for learners’ errors and ideally that the processing be able to recognize and diagnose these errors, which is beyond NLP today (and it likely to remain beyond it in the foreseeable future).

The use of non-language techniques is appropriate and relatively successful, which just poses the question more insistently: shouldn't language technology be applied to CALL? This paper proposes a positive answer to this question, illustrating the advantages of NLP in CALL in a modest application, which, however, relies essentially on NLP.

What can NLP do for CALL?

A further reason for the minor impact of NLP in CALL may be that it is misunderstood by many language learning specialists. Salaberry (1996, p.12) assesses the suitability of language technology for CALL quite negatively in one of the leading journals:

“Linguistics has not been able to encode the complexity of natural language [...] That problem has been acknowledged by the most adamant proponents of Intelligent CALL [ICALL (NDS)]. Holland (1995) lists the reasons that have prevented ICALL from becoming an alternative to CALL. The most important reason for this failure is that NLP (Natural Language Processing) programs--which underlie the development of ICALL--cannot account for the full complexity of natural human languages.”¹

Although indeed linguistics has not yet been able to encode the entire complexity of natural language, this does not imply that NLP cannot be useful to CALL.

Firstly, we see Salaberry as guilty of a fallacy of division--assuming that what is true of the whole must be true of the parts. So while it is true that faithful models of human linguistic behavior are likely to remain beyond the reach of language technology for many years, perhaps decades, the same is not true of many subdisciplines. Phonological and morphological descriptions of many languages

¹ Salaberry's reference to Holland (1995) is not accompanied by bibliographic information.

are quite complete--and much more reliable than the analyses of most language teachers, so that their accuracy cannot be the stumbling block to effective CALL. Also, although CALL in the area of feedback for drills and exercises with free input needs to be fully consistent and error-proof, this is not the case for all subtasks in CALL.

Secondly, even if language technology might lack the power to deal effectively with some of traditional CALL, it seems reasonable to see where it *can* be useful. Here we agree with Salaberry and others about how the issue should be decided: it is not the technology *per se*, but the contribution it can make to teaching and learning that determines its usefulness for CALL.

It is difficult to argue from first pedagogical principles for the need for NLP in CALL. An important obstacle to such an argument is that language pedagogy experts, while agreeing on the importance of holding the attention of learners, allowing repetition, and aiming for a range of practical exercise, still differ on many further points (Larsen-Freeman 1991). Under these circumstances it is wise for CALL developers not to embrace any pedagogical theory too exclusively, but to provide tools or modules which might prove useful from various perspectives (cf. Lantolf 1996).

GLOSSER and CALL

This paper supports Last's claim that relatively straightforward programs, *inter alia* GLOSSER, can achieve a great deal of success within the general framework of CALL (Last, 1992). *Trust* is a key: students must trust their CALL systems to be right about the information they provide. That trust is threatened when systems make errors, and students lose confidence in the learning system. This means that CALL systems should be reliable linguistically, stable technically, and predictable in the specific educational support they provide. It also implies that complex systems are more likely to cause problems in the student-teacher (program) interaction. The further one moves from the level of individual words and phrases to the semantic level, the greater is the danger that students' trust will be disappointed, given the current state of the art in linguistic technology.

The focus of GLOSSER, therefore, is on words. Learners are given the task of understanding texts and may call on GLOSSER for information on the words in the text. The educational value of this focus is widely supported by research. The reading of texts significantly improves the learners' vocabulary by providing lexical context, even without the use of additional sources like dictionaries (Krantz 1990). The context provided by full text not only contributes to a better understanding of the possible uses of a specific word, but it also creates a framework in which words are more easily remembered (Mondria 1996).

GLOSSER's stance is pedagogically sound--even assessed from the wide variety of pedagogies now current in language learning. GLOSSER, the vocabulary learning assistant described in more detail below, allows students to learn language in a *communication* task (namely, that of reading). This approach enables the use of *support* tools, not merely exercises and drills, and thus shares some of the motivation for *Communicative CALL* within the CALL paradigm (Warschauer 1996). The choice of reading material is entirely up to the student and/or teacher, but it may include *authentic materials*, which Widdowson (1990) and others have argued improves the quality of learning by involving the learner more directly in the community in which the target language is spoken. Krantz (1990) emphasizes the importance of learning vocabulary *words in context*, and GLOSSER supports exactly and only that. Our more general point is a simple consequence of GLOSSER's success: NLP can improve CALL now.

The Use of GLOSSER

GLOSSER is a support tool, not a language course, and there are several ways in which support tools may be deployed. GLOSSER may be used in *individualized* instruction, following the wish for more student-centered learning, and accommodating learner differences.

So, on the one hand GLOSSER is a real CALL application, in that it facilitates language learning by providing on-line information on individual words of French texts thus helping students improve their comprehension of French texts and improve their vocabulary. On the other hand, however, it is also a *tool* for text comprehension, in that it assists people who know some French but cannot read it quickly or reliably due to the presence of a number of unknown words in the text (Nerbonne and Smit 1996). Therefore, it can be used not only in educational tasks, but also as an on-line tool for *reading assistance*, creating a multitude of potential applications in educational and professional use. Finally, its easy-to-use nature suits GLOSSER for unsupervised use (with or without accompanying instruction).

GLOSSER--Technical Realization

GLOSSER is implemented in UNIX, and facilitates the reading of French texts by Dutch students. Four sources of information are available on words: morphological analysis, POS-disambiguation, a dictionary and examples of word use in especially collected corpora. All sources rely heavily on morphological analysis and indexing techniques, which are implemented in the programming language C. Other modules, including the interface and communication are implemented in the Tcl/Tk scripting language (Ousterhout 1994), ensuring easy rewriting, rapid prototyping and portability. Script language code is slow, but overall speed is still good: a single lookup of all sources of information takes approximately 2 seconds (see the section on performance for details), mostly in morphological analysis.

Another version of GLOSSER was created for the World-Wide Web, as proof of concept and as a to demonstrate the flexibility of the support for CALL (Dokter 1997b). This version has limited functionality however, due to restrictions on its data.

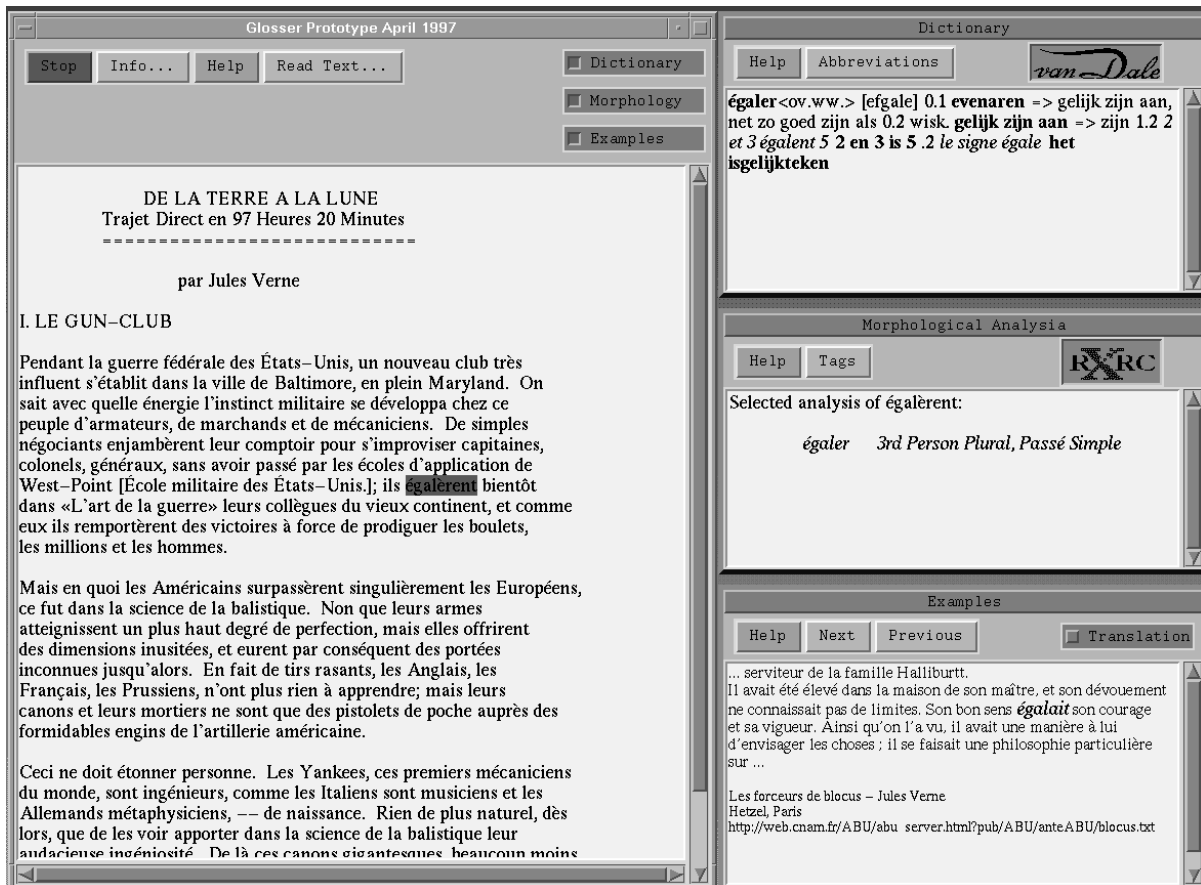


Fig. 1 Front-end and Morphological Analysis The user normally views GLOSSER in this form (we have, however, translated some labels from Dutch into English for this presentation). The large window on the left contains the text being read, in this case Jules Vernes *De la terre à la lune*. The user has clicked on the word *égalèrent*, asking for information. The smaller windows on the right show, from top to bottom, the dictionary entry for the word in a French-Dutch dictionary; the morphological analysis, including the grammatical meaning of the inflection, namely that the word is a third-person plural passé simple form of *égaler*, and finally, in the bottom window, a further example of the word as used in another text. Note that the other example is a different inflectional form.

The front-end of GLOSSER, displayed in Figure 1, consists mainly of four separate windows. The main window (left) provides the general control, a browser (read-only editor) and three on/off-switches for controlling the other three sources of information provided (these switches open the other windows). When in use, these other windows display (for any one word) a dictionary entry, morphological analysis with POS-disambiguation, and (possibly bilingual) examples. The window providing examples actually consists of two windows, one for display of the example, the other for the related translation. Finally, there is a separate help window, that provides some information on the use and interpretation of the different parts of GLOSSER.

GLOSSER's **user interface** tries to be helpful. First, words in the text that are currently under the cursor (available for look-up) are automatically highlighted. The only thing the user needs to do for a look-up, is to click the highlighted word. Second, users can add notes to the original text (insert translations), to avoid the need for more than one lookup.

Morphological analysis/POS-disambiguation is directly informative to the user but also crucial to other processes. It is used to find the underlying lexemes (dictionary forms) of words, since in general dictionaries do not provide entries for inflected forms such as *crois, croyons, crurent, cru* (all forms found under *croire*). The part-of-speech, also provided by this analysis (verb, noun, etc.), allows the program to choose the right dictionary entry in the case of syntactic ambiguity, which is very common. Finally, morphological analysis is also used in providing extra examples from other texts--this allows a lexeme-based index (instead of string-based) and increases the efficiency of the corpus. The efficiency improves because more examples of lexemes are found when all inflected forms are found (see below on examples).

GLOSSER was fortunate in having state-of-the-art software for morphological analysis and POS-disambiguation from Rank Xerox Research Centre: *Locolex* (Bauer and Zaenen 1995). A sample analysis is shown in Figure 1, middle right window. *Locolex* incorporates a stochastic POS tagger for disambiguation. In case *Locolex* disambiguates incorrectly (quite infrequently), the alternatives are listed so that the user may specify another morphological analysis, which is then used for look-up in the dictionary and examples index.

Dictionary and Examples: GLOSSER was likewise fortunate in obtaining the Van Dale dictionary *Hedendaags Frans* (Van Dale 1993). Figure 1, upper right window illustrates the front-end of the dictionary within GLOSSER. For dictionary lookup lexemes and POS are used (as generated by the morphological analysis). The availability of the POS of words greatly improves the accuracy of dictionary lookup, which otherwise may suffer from grammatical ambiguity, leading to many candidate dictionary entries, obscuring the dictionary's value.

A very rudimentary sense of *word-sense disambiguation* has been built in: if one of the examples in the dictionary matches the word context in the original text, this translation is highlighted. This is often the case in fixed expressions such as *guerre mondiale* 'world war', for the lexeme *mondial*. The user can select any translation from the dictionary, and insert it into the original text. Selection is done in the same user-friendly way as word selection in the text. To provide a rich selection of examples, a large and varied corpus was needed, including colloquial, literary, technical, political, and other prose. *Bilingual* texts were particularly attractive. GLOSSER relied partly on specialized corpus projects, such as the ECI and MULTEXT (see references for URL's) for bilingual corpora. A partner project developed a tool for aligning bilingual corpora (Paskaleva and Mihov 1998). Monolingual corpora were mainly found on the World-Wide Web, for example the Gutenberg project (see references of this paper for URL).

The current corpus size for GLOSSER is 5 MB in monolingual, 3 MB in bilingual text (that is, the size of the French text), including 16,701 different lexemes. The texts are indexed by determining the lemmata and POS of the individual words using the same morphological software described above. An index (Dokter 1997a) links lemmata to full, possibly inflected forms in the original corpora. This way, a specific instance of a lexeme can be retrieved, in its original lexical context. In the case of GLOSSER, this is in general two or three sentences, depending on the length. Lexeme-based indexing relates inflectional variants to a single lexeme (dictionary form). A search for examples of *croire* turns up the nearly 100 possible inflected variants. This improves the chance of finding examples of a given lexeme immensely.

Examples are displayed, with a reference to the source (if available), in the *Examples* window, as shown in Figure 2.

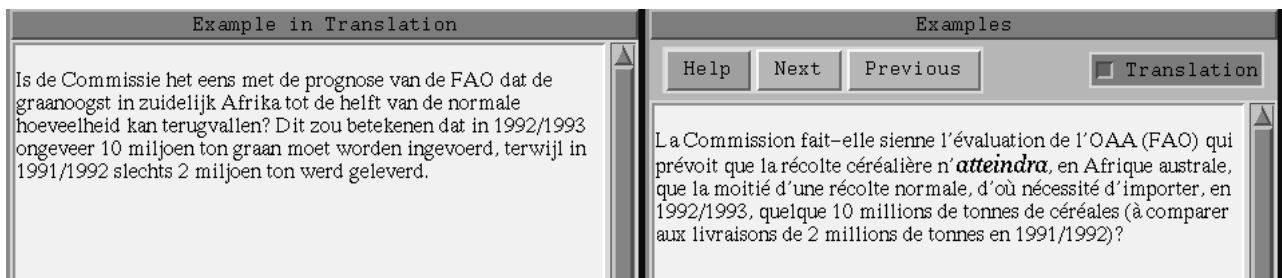


Figure 2 The window showing examples. If the example has been found in a bilingual text (shown here), the user can pop-up the translation in the parallel text directly.

Performance

The next chapter will describe the *functional* performance of GLOSSER, which allows some space for processing performance to be mentioned here. Processing times for the different modules are based on an average calculated by processing about 100 words. The words were taken from a single text, but, because the program does not cache previously looked up words, this should not introduce any bias. Times in the table are given in seconds.

	Time Used
Morphology/POS Disambiguation	1.579
Dictionary	0.134
Examples	0.139
Total	1.852

The process of morphological analysis and part of speech disambiguation clearly consumes most of the time needed to process one look-up, which is not surprising, since it is the most complex and important part of the application. The other processes, although implemented by means of scripting, take advantage of indexing techniques. A full look-up takes less 2 seconds, which users found satisfactory.

Functionality

The intended functionality of GLOSSER was to provide robust text-independent support for Dutch students of French. Once GLOSSER was sufficiently stable to support reading of essentially all non-specialized texts, the demonstrator was subjected to performance analysis and user studies.² We review this to demonstrate the maturity of available NLP technology.

² We thank Dr. Maria Stambolieva and Dr. Aneta Dineva of the Bulgarian Academy of Science, who collected data and began this analysis at the University of Groningen in April 1997.

Performance Analysis

In order to analyze performance, we selected 500 words in 100-word samples, taken at randomly chosen points in five different texts. These were checked word by word for accuracy in analysis. The texts varied in genre: official European Commission prose, (soft) pornography, poetry, and political opinion.

There were four types of mistakes, distributed unevenly in the text.

1. Mistakes in input due to incorrect selection by testers or input errors in the text itself (misspelled words and incorrect coding schemes);
2. Missing words in the morphology or the dictionary;
3. Incorrect linguistic analysis;
4. Irrelevant corpus examples.

None of these resulted in unexpected program responses, unrecoverable errors, or failures to respond. We illustrate and discuss each of the error types in turn.

(1) Input. Testers reasonably tried to view cliticized elements such as the *d'* in *d'argent* as words which might be looked up, but the program does not treat these as independent words. This decision was motivated by convenience, but also by the consideration that the intermediate level of user we aimed to help would have no need of assistance for these words.

We also include misspelled words in this category. Recalling that GLOSSER is intended for students, it might reasonably be expected that spelling had been checked, but error-correcting capabilities are still lacking in the current realization. Several errors resulted in applying the application to ASCII text, because the program expects Latin8 encoding. Some of the errors were invisible to the eye, e.g., one in which accented capitals had been encoded as unaccented (which is a common typeface), e.g. *Église*.

(2) Missing Words. Only 12 correctly analyzed words did not appear in the dictionary or morphological analysis. Seven of the missing words were brand names and the like, e.g., *Collier's*, *Vargas* and *Life*. It is difficult to know what to make of this circumstance. The morphology will never be so comprehensive that all such words are processed. The problem clearly cannot be regarded as a shortcoming of the dictionary, which was chosen for its limited coverage. A more comprehensive dictionary would be less useful to intermediate-level students. Two missing words were *fréquemment* 'frequently' and *généreusement* 'generously', which in fact *are* in the dictionary, but listed under the adjectives they are derived from, *fréquent* and *généreux*, respectively. Morphological analysis does not resolve these cases. This suggests that a second level of dictionary indexing would be useful. The remaining three missing words were not in the dictionary.

(3) Incorrect Analyses. Of 500 words, a total of 17 incorrect analyses were encountered, rather more than we had encountered in trials with users (where none were reported). There were no errors of morphological analysis in the sense that the preferred analyses were in every case morphologically possible analyses of the words. All the errors were faulty assignment of POS categories (which could result in a preference for a possible, but incorrect analysis). These result in incorrect dictionary look-ups 30% of the time (5 cases). Virtually none of these placed the user in the incorrect dictionary entry--they involved using an adjective as a noun, etc., so that the dictionary look up was the same. The rather higher number of errors in POS assignment here has to do with the fact that very frequent words tend to be ambiguous and difficult to categorize, while users are

relatively untroubled by them: they don't look up very frequent words. This is a point where the intended application very naturally tolerates shortcomings in the underlying technology.

(4) Faulty Corpus Treatment. There were 47 errors, or nearly 10%. These ranged from finding no examples (most frequent) to finding irrelevant examples, most frequently in connection with derivational morphology (which was allowed). In fact, this is a point where the performance analysis seems too forgiving. Since the random sample of words tested included substantially more frequent words than would a random sample of words users would select for look up, the problem is actually greater. This naturally suggests that the GLOSSER's corpora were too small, and indeed they were. The 4.2 MB of text contained only 16,701 different word stems. The difficulty is that, to provide coverage of, say three occurrences of the most frequent 30,000 words, we should need a much larger corpus (at least ten times as large). This went beyond the project's charter as a prototype.

To sum up, four major mistake types appear rather more frequently than one would wish. None of these mistakes surfaced in extensive user experimentation however.

A User Study

To determine the value of GLOSSER in actual education, a user study was conducted on 22 adult students. Dokter et al. (1998) provides a more complete report on this study, which we summarize here. The students were all in their second or third course (each course takes three months), and were comparable in proficiency to second- or third-year high school students. The goal of this study was to evaluate the program in comparison to the traditional method of text reading and comprehension using a hand-held dictionary. The group was divided randomly in two. The specific factors that were considered relevant and could be accounted for in this study included the overall judgment of the program, a simple measurement of the effect which the use of GLOSSER had on text comprehension and the functionality of the program. Apart from the above factors, the subjects using GLOSSER were asked to comment on the system, so as to get a clear picture of users' demands of the application and suggestions for improvement.

Setup

Each session started with an introduction that explained the major purpose of the experiment, and a short demonstration of the program. At first all the subjects were given some time to get acquainted with GLOSSER. This was done to make the subjects more comfortable with the experimental environment. Then the students were randomly assigned to two groups. Both groups were presented the same text; their task was to read this text within a limited time (20 minutes) and answer questions about this text afterwards. The text was extracted from Jules Verne's *De la terre à la lune* 1865, and contained approximately 250 words. The first group had this text displayed in the GLOSSER browser, the other group used a version on paper and were provided with the same dictionary in a hand-held format.

After the time for reading was up, the text was taken away and the subjects were given a questionnaire. This questionnaire consisted of two parts. The first, eleven questions on the text, was identical for both groups. The second part consisted of several questions concerning the evaluation of the *program* for the group using GLOSSER, and the *hand-held dictionary* for the other subjects. All questions were answered on a Likkert scale of 1-5 to facilitate statistical analysis.

Results

The results from this study can be divided into three classes, according to the specific issue addressed:

- comprehension
- functionality of GLOSSER vs. dictionary
- subject evaluation of GLOSSER vs. dictionary

Although the group was too small for very sensitive statistical analysis, we used results for further development of the prototype, and in a more intuitive way. Comprehension was evaluated by posing questions on the text. Other means of measuring comprehension examined might include the time necessary for reading the text and also vocabulary comprehension after reading. Although results showed that GLOSSER users were faster and understood better, the differences were not significant. The real time differences seemed to be obscured by the fact that users were given more time than necessary, and nearly all of them used their excess time to check and recheck their work. We expect that speed could be shown to be significantly faster with a more tightly controlled task.

GLOSSER users scored significantly higher on a question assessing confidence. They felt more certain that they'd completed the task well, perhaps because the software made the task easier.

Functionality concerned the quality and usefulness of the sources used for GLOSSER, being a dictionary, morphological analysis and examples, and of the application as a whole. Among other things, we compared the number of lookups and the number of words not found in the dictionary. The average number of lookups with GLOSSER in relation to the average number of lookups with the dictionary was 45/14, which is a significant difference ($p < 0.001$). The number of lookup events with GLOSSER was still higher, constituting a ratio of 53/14, but some words were looked up more than once, which was uncontrollable for the dictionary lookups. This figure clearly shows that users of GLOSSER managed to look up many more words (and read the given information) within the same amount of time.

	Forms Looked Up	Words Looked Up
GLOSSER	53	45
Hand-held Dictionary	14	13

Although the hand-held and on-line dictionaries were identical in content, the number of lookups displayed by GLOSSER could be influenced by POS-disambiguation. For completeness, we note that the lay-out of the displayed information was nearly, but not completely identical.

A further issue concerning the functionality of the program is the specific use which the subjects made of the informational sources:

Searches= 629		
Source	Used	% of total
Dictionary	623	99.0
Morphological analysis	276	43.9
Examples	261	41.5

Clearly, the dictionary was taken to be the most important source for support in reading texts. An interesting point here, which is not obvious in these numbers, is that users often consulted other sources immediately after looking up the word in the dictionary. This indicates that when the information a dictionary provides is regarded as insufficient for direct comprehension of a word, or a part of the text, other sources are consulted.

GLOSSER was judged superior to the hand-held dictionary in ease of use (although this result was not significant). All users were keen on using future versions of GLOSSER (or using this version further). The overall judgment of the program was very positive, 4.2 on a scale of 1-5.

Additional comments of users mainly concerned the interface. One comment which resulted in reimplementing suggested desirability of making annotations in the text. This eliminated the need for double look-ups.

Conclusions of the User Study

The user study points to one dangerous feature of any well-working tool: overuse. It seems that students tend to overuse a tool such as GLOSSER. Even though the dictionary interrupts the reading process, and even though the students did not need to look up nearly 20% of the words, they did. This suggests that students ought to be warned against this.

On the other hand, if overuse is a failing, it's the failing of an attractive system: GLOSSER improves the ease with which language students can approach a foreign language text. The most important difference is simply the number of words that can be looked up and the subsequent decrease in time needed for reading the text. Both of these may be expected to improve vocabulary acquisition (Krantz 1990). Although the difference in text comprehension shown by the two groups was not significant, we expect that a more tightly controlled task would show a modest difference. Future study would also be profitable on short- vs. long-time retention effects. The overall reception of GLOSSER was positive. All subjects judged the information GLOSSER provides to be sufficient, and the program in general to be user-friendly.

Previous Work

The idea of applying morphological analysis to aid learners or translators, although not new, has not been the subject of extensive experimentation. Antworth (1992) applied morphological analysis software to create glossed text. But the focus was on technical realization, and the application was the formatting of inter-linearly glossed texts for scholarly purposes. The example was Bloomfield's Tagalog texts.

The work of the COMPASS project (Breidt and Feldweg 1997) had a similar focus to our own--- that of providing "COMPrehension ASSistance" to less than fully competent foreign language readers. Their motivation seems to have stemmed less from the situation in which language *learning* is essential and more from situations in which one must *cope* with foreign language. In addition, they focused especially on the problems of multi-word lexemes, examples such as English *call up* which has a specific meaning 'to telephone' but whose parts need not occur adjacently in text, see *call someone or other up*.

Conclusions

Morphological processing is sufficiently mature to support nearly error-free lemmatization. This functionality can be used to automate dictionary access, to explain the grammatical meaning of morphology, and to provide further examples of the word in use (perhaps in different forms). Second language learners look up words faster and more accurately using systems built on morphological processing. It is to be expected that this will improve their acquisition of vocabulary.

This is but one instance where NLP has matured sufficiently to be of service in CALL. Nerbonne, Jager and van Essen (1998) discuss experiments with other technologies, in particular speech recognition (see Rothenberg (1998) and Witt and Young (1998)) and parsing (articles by van Heuven (1998) and Murphy, Krüger and Grieszl (1998)).

Prospects

There are many prospects for technical improvement in the GLOSSER system. Some of the ideas of the COMPASS (Breidt and Feldweg 1997) system on looking up and indexing multi-world lexemes would be a valuable addition, as would certainly be a dictionary which contained actual pronunciations. The present set of examples demonstrates the concept sufficiently, but there are too few texts and many are inappropriate. We view all of these as less pressing than finding an experimental deployment for the system in actual language teaching. This would probably call for a number of pedagogical improvements in the manner in which information is presented, in the opportunity for teachers to add to material, and to monitor use, and in the addition of facilities to provide for review. All of these would be potentially interesting technically and pedagogically.

The prospects vis-à-vis the more general point of this paper, the use of NLP in CALL will undoubtedly include disappointments. The identification of important barriers is a favorite pastime among CALL aficionados, the obstacles including exaggerated claims (and subsequent disappointments), insufficient infrastructure, competition with a staff who feels threatened by CALL, need for staff training, incompatibility of poor fit with other materials, and complex decision and purchasing structures. We refrain from developing these points beyond this list of simple reminders (but Salaberry (1996), discussed above, develops many). The obstacles are genuine, and may not be ignored, but they have received substantial comment elsewhere.

We remain confident that the long-term advantages are substantial enough for all concerned that CALL will continue growing. More extensive exploitation of language technology should make CALL more useful sooner.

Acknowledgments

The Copernicus program of the European commission supported the GLOSSER project in grant 343 (1994). The authors were the members of the project in Groningen. Lauri Karttunen, Elena Paskaleva, Gábor Prózéký and Tiit Roosmaa joined in a common design for UNIX and Windows95 versions of the program. Valuable criticism has come from Poul Andersen, Susan Armstrong and Serge Yablonsky; Edwin Kuipers joined us to make a web version of the program; and Lili Schurcks-Grozeva assisted in the user study.

References

- Antworth, E. (1992), "Glossing Text with the PC-KIMMO Morphological Parser". In: *Computers and the Humanities*, 26(5-6), pp.389-98.
- Bauer, F.S.D. and A. Zaenen (1995), "Locolex: Translation rolls off your tongue". *Proceedings of the Conference of the ACH-ALLC '95*, Santa Barbara, USA.
- Breidt, E. and H. Feldweg (1997), Accessing Foreign Languages with COMPASS. *Machine Translation*, special issue on New Tools for Human Translators, pp.153-174.
- Van Dale (1993), *Handwoordenboek Frans-Nederlands*, 2nd ed. Van Dale Lexicografie, Utrecht.
- Dokter, D.A. (1997a), "Indexing Corpora for GLOSSER". Techreport, Alfa-informatica, Groningen University, Groningen.
- Dokter, D.A. (1997b), From GLOSSER to Glosser-WeB. Techreport, Alfa-informatica, Groningen University, Groningen.
- Dokter, D.A., J.Nerbonne, L. Schurcks-Grozeva and P. Smit, (1998), GLOSSER: a User Study. In S. Jager, J. Nerbonne, and A. van Essen (eds.), pp.167-176.
- ECI, *European Corpus Initiative Multilingual Corpus I*, <http://www.elsnet.org/resources/eciCorpus.html>
- S. Jager, J. Nerbonne, and A. van Essen (eds.), (1998), *Language Teaching and Language Technology*. Swets and Zeitlinger, Lisse.
- van Heuven, V. Computer-Assisted Learning to Parse in Dutch. In S. Jager, J. Nerbonne, and A. van Essen (eds.), pp.74-81.
- Krantz, G. (1990), *Learning Vocabulary in a Foreign Language; A Study in Reading Strategies*. Ph.D.-Thesis, University of Göteborg, Göteborg, Sweden.
- Lantolf, J.P. (1996), SLA theory building: "Letting all the flowers bloom!". *Language Learning* 46(4), 713-749.
- Larsen-Freeman, D. and M.H. Long, (1991), *An Introduction to Second Language Acquisition Research*. Longman, London.
- Last, R. (1992), Computers and Language Learning: Past, Present - and Future? In C. Butler (ed.), *Computers and Written Texts*, Blackwell, Oxford. pp.227-245.

Mondria, J.-A. (1996), *Vocabulaireverwerving in het vreemde-talenonderwijs: De effecten van context en raden op de retentie*, Ph.D.-Thesis, University of Groningen, Groningen, The Netherlands.

MULTEXT, *Multilingual Text Tools and Corpora*, <http://www.lpl.univ-aix.fr/projects/mulTEXT/>

Murphy, M, A.Krüger, and A.Grieszl, RECALL--Providing an Individualized CALL Environment. In S. Jager, J. Nerbonne, and A. van Essen (eds.),. pp.62-73.

Nerbonne, J. and P. Smit, (1996), GLOSSER: in Support of Reading. *COLING '96*, Copenhagen. pp.830-35.

Nerbonne, J., S. Jager, and A. van Essen, Introduction. In S. Jager, J. Nerbonne, and A. van Essen, (eds.),. pp.1-11.

Ousterhout, J.K. (1994), *Tcl and the Tk Toolkit*, Addison-Wesley Publishers Ltd.

Paskaleva, E. and S. Mihov (1998), Second Language Acquisition from Aligned Corpora. In S. Jager, J. Nerbonne, and A. van Essen (eds.),. pp.43-52.

Project Gutenberg, <http://www.etext.org/Gutenberg/>

Rothenberg, M. (1998), The New Face of Distance Learning. In S. Jager, J. Nerbonne, and A. van Essen (eds.), pp. 146-48.

Salaberry, M.R. (1996), A theoretical foundation for the development of pedagogical tasks in computer-mediated communication, *CALICO Journal* 14(1), pp. 5-34.

Warschauer, M. (1996), Computer-Assisted Language Learning: An Introduction. In Fotos, S. (ed.), *Multimedia Language Leaching*, Logos International, Tokyo. p3-20

Widdowson, H.G. (1990), *Aspects of Language Teaching*, Oxford University Press, Oxford.

Witt, S. and S. Young, Computer-Assisted Pronunciation Teaching Based on Automatic Speech Recognition. In S. Jager, J. Nerbonne, and A. van Essen (eds.), pp.25-35.

Zaenen, A. & G.Nunberg (1996) Communication Technology, Linguistic Technology and the Multilingual Individual. In T.Andernach, M.Moll & A.Nijholt (eds.) *Proc. of Computational Linguistics in The Netherlands V*, Parlevink: Twente. pp. 1-12.

Zock, M. (1996), Computational Linguistics and its Use in the Real World: the Case of Computer-Assisted Language Learning. In *COLING 1996*, Copenhagen. pp.1,002-1,004.