Glosser-R$\underline{u}$G: A User Study

Duco Dokter, John Nerbonne, Lily Schürcks-Grozeva, Petra Smit
Alfa-Informatica, University of Groningen

1 CALL and Glosser-R$\underline{u}$G

Traditionally, CALL software has mainly emphasized providing drills, answer keys and grammar explanations, thus, in a sense replacing the human teacher in traditional language education (Last 1992). The position of Glosser-R$\underline{u}$G in the CALL-paradigm (Warschauer 1996), however, is focused on the improvement of *communicative* skills. The function of the application then, is mainly *assisting* novice readers in developing their comprehension and reading skills, thereby improving their ability to communicate in French.

This paper supports Last's (1992) claim that relatively straightforward programs such as Glosser-R$\underline{u}$G can achieve a great deal of success within the general framework of CALL. One of the key aspects of any learning system is that the relationship between teacher (that is, the CALL system) and student is imperiled by anything which may provoke a lack of confidence by the latter in the former. This implies that CALL systems should be extremely stable concerning software, as well as predictable in the specific educational support they provide. This implies that complex systems are more likely to cause problems in the student-teacher (program) interaction. The more one moves from the level of individual word and phrase to the semantic level, the greater is the danger that the relationship of trust between the student and computer will be inadvertently broken, given the current state of the art in linguistic technology. The focus of Glosser-R$\underline{u}$G, therefore, is mainly on the level of words. This stance is widely supported by previous research. The reading of text significantly improves the learners' vocabulary by providing lexical context, even without the use of additional sources like dictionaries (Krantz 1990). This context not only serves to clarify, but also to create a framework in which words are more easily remembered (Mondria 1996).

On the one hand Glosser-R$\underline{u}$G is a 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 increasing their vocabulary. On the other, it is also a tool for text comprehension, in that it assists people who know a bit of French but cannot read it quickly or reliably due to the presence of a number of unknown words in the text (Nerbonne & Smit 1996). Therefore, it can be used in educational tasks, and as an on-line tool for 'Just-In-Time knowledge' (Allodi 1996), creating a multitude of potential applications in education and professional use.

Although Glosser-R$\underline{u}$G was specifically designed for French texts, the architecture of the program makes it suitable for processing any language for which the informational sources used in the current application are available. In fact, the program is mainly an integrated and unified account of several previously developed linguistic tools and methods. The major informational sources are: an on-line dictionary provided by Van Dale Lexicografie (Van Dale Lexicografie 1993), morphological analysis software provided by Rank Xerox (Bauer & Zaenen 1995), and examples of the use of words in other texts. The program relies heavily on the morphological analysis software, which provides the link between inflected word forms and the 'citation forms' found in dictionary entries (Sproat 1992), and which determines the part-of-speech function a specific word fulfills in a sentence; this latter analysis is used in dictionary lookup. Each word for which information is sought is analyzed within its linguistic context (being the sentence a word occurs in) and disambiguated with respect to its base form, lexical category and the morphological properties of the word ('tags') that determine its appearance or inflections. The analysis thus provides the link between the possibly inflected form of the word as it appears in the text and the citation forms as they are listed in a dictionary. This LEMMATIZATION, therefore, plays a dual role: it is informational in that it provides the user with information about the precise inflections of a given word and their grammatical interpretations, but it is also functional for further processing, by providing the link between surface forms as they appear in the text and the dictionary forms of words.

2 A Detailed Outline of Glosser-R*u*G

The complete Glosser-R*u*G system as used in this study[1] consists of a read-only editor, a dictionary, a morphological analysis/POS-disambiguation system and an example-lookup. The program is completely window-oriented and designed to be user-friendly; little knowledge of, or experience with computers is necessary to use Glosser-R*u*G. The complete screen the user gets when all modes of information are activated and a specific word is looked up is shown in Figure 1. (The menubar at the bottom of the screen is not a part of Glosser-R*u*G.)

The left window is the read-only editor in which the text to be read is displayed, the upper-right window is the dictionary, below that is the window displaying the morphological analysis, and in the lower-right corner is the window displaying examples of the selected word in other texts. Figure 2 shows the appearance of the editor, with the highlight on the word *victime*. This highlight simply follows the mouse-cursor. When a word is highlighted, it is selected for lookup by a single mouse-click.

In the upper-right corner of the editor three on/off-switches control the specific sources that are to be used for a single lookup. Other features (buttons) include a hypertext oriented help function that is shown in a separate window, a file selector and language selector (not yet functional in this prototype). The precise functionality of these buttons on the editor window (and of the other

[1] This version is still a prototype, but relatively stable. New features have already been added according to results from this study, see section 3.3

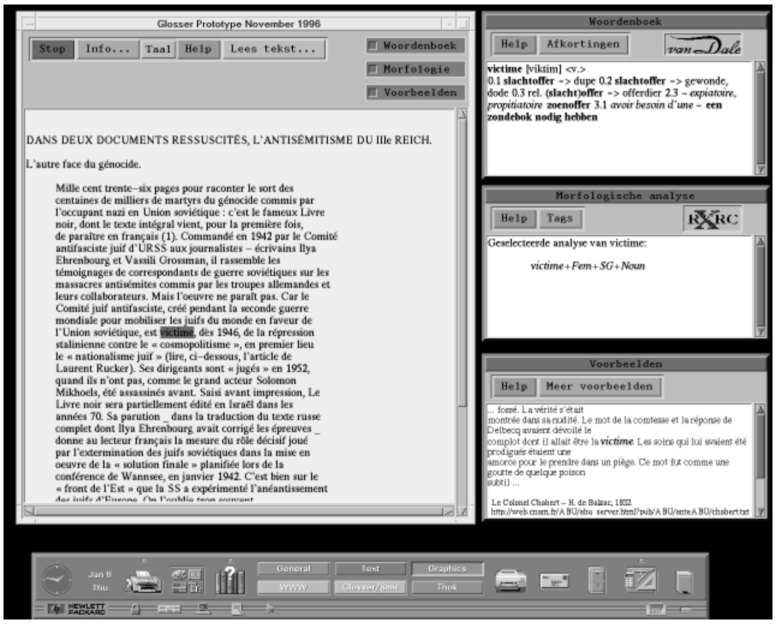

FIGURE 1 The front-end as it is displayed on the screen. The window providing help is omitted.

windows) is beyond the scope of this paper, see Dokter (1997*a*) for a complete account of this prototype.

2.1 Morphological Analysis

Morphological analysis is a necessary prerequisite for dictionary lookup, since most dictionaries do not contain inflected word forms as distinct entries. Glosser-R*u*G was fortunate to be able to incorporate *Locolex*, a state-of-the-art system provided by Rank Xerox (Bauer & Zaenen 1995). An example of morphological analysis is displayed in Figure 3. (Additional information on the tags used is available for users of the application in the hypertext help function.)

Locolex analyzes the morphological properties of a word, and disambiguates the part-of-speech (POS) which the selected word constitutes in the text, based on syntactic context. The lemma thus found is used for dictionary and example lookup. Apart from the lemma, the POS tag is also used for dictionary lookup, to disambiguate entries that occur within different syntactical categories.

2.2 Dictionary Lookup

An important source of information on words for Glosser-R*u*G (and of course, in general) is an on-line dictionary, provided by Van Dale Lexicografie (Van Dale Lexicografie 1993). This dictionary is actually a number of tagged text files, that is, no search routines are implemented. Therefore, an index has been generated,

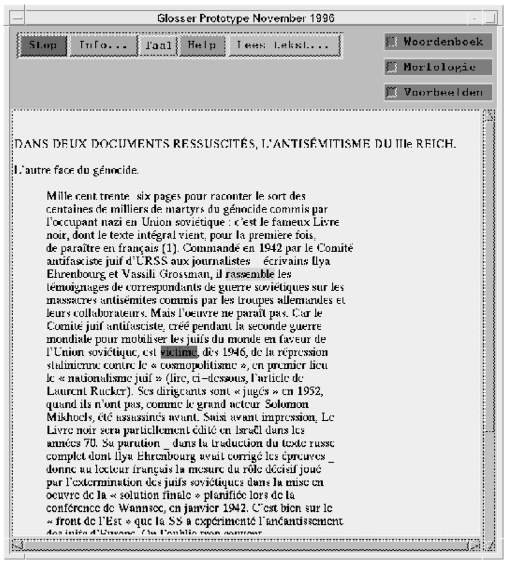

FIGURE 2 The read-only editor displaying a text

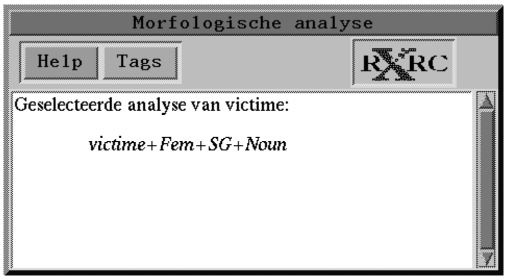

FIGURE 3 Morphological analysis

consisting of <*word, POS, file, position*> quadruples. This (sorted) index can be accessed using binary search, ensuring fast lookup of words. For dictionary lookup lemmata are used, as generated by the morphological analysis, as well as the POS of the word as determined by *Locolex*. The latter feature implies that the correct entry is found for words with multiple entries (due to different possible syntactic categories), that is, in accordance with the POS the word is tagged with. Figure 4 shows the display of the dictionary entry for *victime*.

2.3 Examples

The third informational source of Glosser-R*u*G is constituted by the availability of examples of the use of a given word in a corpus of French texts. This provides the user with several distinct lexical environments of a word, which has been

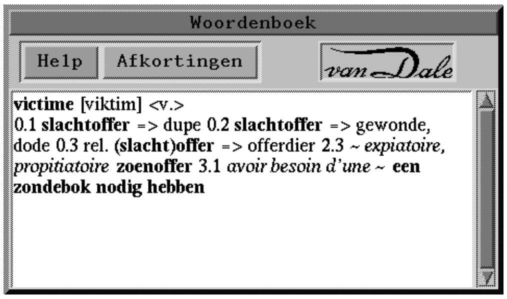

FIGURE 4 The on-line dictionary

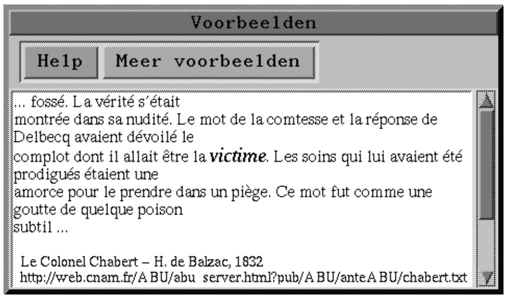

FIGURE 5 Examples in other texts

proved to be an important factor in vocabulary learning (Mondria 1996, Krantz 1990), see section 1.

Approximately 5 Mb of text is used for these examples, accounting for a coverage of 16701 different lexemes [2] of the French vocabulary. The texts are indexed by determining the lemmata of the individual words using the same morphological analysis method as described in section 2.1, and creating an index of N occurrences of each lemma thus found (Dokter 1997*b*). The index then provides a link between the lemmata and the full, possibly inflected forms, which enables the user to see examples not only of the specific word as it occurs in the text that is being read, but also of other inflections. Examples are displayed, with a reference to the source (if available), in the 'examples' window, as illustrated in Figure 5.

3 A User Study

The goal of this study was to evaluate Glosser-RuG in comparison to the traditional method of text reading and comprehension using a hand-held dictionary. The specific factors that were regarded as relevant and could be accounted for

[2]Note that these are the lexemes as found after processing the text with Locolex, that is, the number of inflected forms in the corpus is much higher.

in this study, included the overall judgement of the program, a simple measurement of the effect which the use of Glosser-RuG had on text comprehension and the functionality of the program. The group of 22 subjects was divided into a group that used Glosser-RuG, and a group that used the hand-held dictionary. Apart from the above factors, the subjects using Glosser-RuG were asked to comment on the system, so as to get a clear picture of users' demands of the application and suggestions for improvement.

3.1 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 5 minutes to get acquainted with Glosser-RuG. 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-RuG environment editor as described in section 2, the other group used a version on paper and they 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-RuG, and the hand-held dictionary for the other subjects. The second part of the questionnaire is given here. The text in italics represents the phrases that differ according to the specific group:

1. What is your level in French?
   novice– 1 2 3 4 5 –advanced
2. Do you read French texts on a regular basis?
   never– 1 2 3 4 5 –very often
3. How well do you understand the text?
   poor– 1 2 3 4 5 –excellent
4. Do you think that the information *Glosser-RuG/the dictionary* gives for a word is sufficient for good understanding?
   insufficient– 1 2 3 4 5 –complete
5. Do you like using *Glosser-RuG/ the dictionary*?
   not at all– 1 2 3 4 5 –very much
6. *Would you like to use programs like Glosser-RuG in the future?/Have you ever used a computer program for language learning?*
7. What is your overall judgement on *Glosser-RuG/the dictionary*?
   bad– 1 2 3 4 5 –very good

All questions are to be answered on a scale of 1-5 to enable statistical analysis, except question 6 which is answered by *yes* or *no*.

3.2 Results

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

1. comprehension (part one of the questionnaire, question 3 of part two)
2. functionality of Glosser-R$\underline{u}$G vs. dictionary (question 4, number of lookups)
3. subject evaluation of Glosser-R$\underline{u}$G and dictionary (question 4, 5, 6, and 7)

The other questions (1 and 2) were used for overall control. The first class can be evaluated objectively according to the questions on the text. A maximum score of 100 points was possible, for 11 questions, that is approximately 9.1 points per correctly answered question. In this study, other possible factors for measurement of comprehension are not included, like the time necessary for reading the text and testing the vocabulary after reading[3].

The average score of Glosser-R$\underline{u}$G users on the questions was 58.8, while the score of the dictionary users was 50.5. Although results show a higher average for the Glosser-R$\underline{u}$G users, the difference was not significant. It is expected that it might be significant with a larger subject group. An interesting point here is that the users of Glosser-R$\underline{u}$G scored significantly better on the question concerning self-estimation (question 3). There might be some effect of the use Glosser-R$\underline{u}$G on the comprehension users think they have, due to the relative ease with which a user can look up unfamiliar words.

The class of FUNCTIONALITY concerns the quality and usefulness of the sources used for Glosser-R$\underline{u}$G, being a dictionary, morphological analysis and examples, and of the application as a whole. This issue can be addressed in comparing the number of lookups, the number of words not found in the dictionary and by evaluating question 4. Since all subjects used the full 20 minutes to read the text, a simple ratio is sufficient to describe the difference in lookups. The average number of words looked up with Glosser-R$\underline{u}$G in proportion to those looked up with the dictionary was 45/14, which is a very significant difference ($p < 0.001$). The actual number of lookup events with Glosser-R$\underline{u}$G was even 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-R$\underline{u}$G managed to look up a much larger number of words (and read the given information) within the same amount of time. The average time needed for a single lookup for the program users is 22.6 seconds, that of dictionary users is 85.7 in this study. The content of the dictionary

[3]It was expected that times could be compared, in order to see whether the computerized lookup would speed overall reading times, but nearly all subjects used all of the time alloted—even when it meant rereading the text. The experimental design requires refinement here—some way of motivating subjects to finish quickly is needed.

given by Glosser-RuG, was preferred to the content of the hand-held dictionary, although the difference was not significant. Although both the hand-held dictionary and the on-line dictionary were identical in content, the number of translations displayed by Glosser-RuG might differ due to POS-disambiguation. Also lay-out of the displayed information was not completely identical. Another issue concerning the functionality of the program is the specific use which the subjects made of the possible informational sources:

Searches= 629		
source	**used**	**% of total**
Dictionary on	623	99.0
Morphology on	276	43.9
Examples on	261	41.5

Clearly, the dictionary is regarded the most important source for support in reading texts. An interesting point here, which does not show from these numbers, is that users often consult other sources immediately after lookup of 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. In this specific text three different words could not be found by Glosser-RuG. These included one English word and one word abbreviated with an apostrophe (*jusqu'alors*). For the dictionary users this number was four words, including an inflected form of a verb (*sait*). This offers an interesting perspective on automatic morphological analysis previous to dictionary lookup, as described in section 2.1. This issue remains to be investigated.

For the evaluation of Glosser-RuG in general, contrasted with the evaluation of the hand-held dictionary, questions 5, 6, and 7 were used. The practical use of Glosser-RuG (question 5) was judged better than the use of the dictionary, although this result is not significant. All users were keen on using future versions of Glosser-RuG (or continue using this version). The overall judgement of the program was very positive, 4.2 on a scale of 1-5.

3.3 Qualitative Data

Apart from the quantitative data obtained with statistical analysis, some useful information on the functionality of the system was provided by remarks of the users during and immediately after the sessions. This section will discuss the additional features of Glosser-RuG, that were suggested by the subjects and added after the study.

In the first prototype, a user had to select a word by dragging the cursor over a part of the text with the left button held down. This led to a large number of error-messages, since the procedure required some precision: non-words might be selected and ruled out afterwards by the program and punctuation had to be filtered out. After a selection was made, it had to be submitted for a search by explicitly clicking on a search button. This was regarded a redundant action by most users. Therefore, the new prototype automatically selects words currently under the mouse and highlights the selection. A single mouse-click now starts a search action. Automated selection ensures that no erroneous words can be submitted for lookup; control is entirely with the application.

A second important addition is the possibility of actually making notes in the original text (real 'glossing'). Data on the searches conducted by single users on a single text showed that a number of words was looked up several times, which indicates that users did not remember each word that has been previously looked up. It therefore seemed an improvement to enable the user to write translations directly in the text. Moreover, users specifically requested this feature for Glosser-RuG. Hence, a new prototype enables the user to click on translations found in the dictionary, upon which the chosen translation is inserted in the text, directly after the original word. A simple mouse-click on a translation removes it if necessary. In the first case, in order to select, the highlight follows the mouse in the dictionary window, in the same way as for selection in the text that is being read. In the latter case the cursor changes shape, so as to show the nature of the action, that is, deleting the translation-note.

During the study, it appeared that some of the elderly subjects were unable to use the mouse-driven features of the application and could therefore not comfortably interact with the program. This is considered a general contemporary problem and is not addressed further in this paper, although these results have been incorporated in the data given.

4 Conclusions

It can be concluded that Glosser-RuG 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. Although the difference in text comprehension shown by the two groups was not significant, involving larger numbers of subjects in further experiments may reliably show a (small) difference. The exact figures will be the subject of a future investigation. This work will include the topics of short/long-term retention and variable reading times in order to achieve a better insight into text comprehension. However, the overall evaluation of Glosser-RuG leaves us optimistic. All subjects judged the information Glosser-RuG gives to be sufficient, and the program in general to be user-friendly.

Acknowledgements

Glosser-RuG is part of the GLOSSER project, COPERNICUS grant 343, which aims at applying state-of-the-art natural language processing techniques to technology for *Computer Assisted Language Learning* (CALL) (Nerbonne 1994). We would like to express our gratitude to Felice Portier of the Centre Culturel Français and her students for their cooperation in this study, and to Prof. Elena

Paskaleva from the Bulgarian Academy of Sciences for supporting L. Schürcks-Grozeva, of the Bulgarian group in the GLOSSER-project, to participate in this study. Further, we would like to thank J.-A. Mondria for consultation.

References

Allodi, A. (1996), Distance learning: Information needs and technology options, forthcoming, Alfa-informatica, University of Groningen, Groningen, The Netherlands.

Bauer, F. S. D. & A. Zaenen (1995), Locolex: Translation rolls off your tongue, *in* 'Proceedings of the conference of the ACH-ALLC95', Santa Barbara, USA.

Dokter, D. A. (1997*a*), Glosser-RuG, prototype december 1996, Techreport, Alfa-informatica, University of Groningen, Groningen, The Netherlands.

Dokter, D. A. (1997*b*), Indexing corpora for glosser, Techreport, Alfa-informatica, University of Groningen, Groningen, The Netherlands.

Krantz, G. (1990), Learning Vocabulary in a Foreign Language; A Study in Reading Strategies, PhD thesis, University of Göteborg, Göteborg, Sweden.

Last, R. (1992), Computers and language learning: Past, present - and future?, *in* Chr.S.Butler, ed., 'Computers and Written Texts', Blackwell, Oxford.

Mondria, J.-A. (1996), Vocabulaireverwerving in het vreemde-talenonderwijs; De effecten van context en raden op de retentie, PhD thesis, University of Groningen, Groningen, The Netherlands.

Nerbonne, J. (1994), Glosser- technical description, Project proposal, Alfa-informatica, University of Groningen, Groningen, The Netherlands.

Nerbonne, J. & P. Smit (1996), Glosser-RuG: in Support of Reading, *in* 'COLING96', Groningen, The Netherlands.

Sproat, R. (1992), *Morphology and Computation*, MIT Press, Cambridge, MA, USA.

Van Dale Lexicografie (1993), *Handwoordenboek Frans-Nederlands, 2^e^ druk*, VanDale Lexicografie b.v., Utrecht.

Warschauer, M. (1996), Computer-assisted language learning: An introduction, *in* S.Fotos, ed., 'Multimedia language teaching', Logos International, Tokyo, pp. 3–20.