AN INTELLIGENT WORD-BASED LANGUAGE LEARNING ASSISTANT

John NERBONNE* **Duco DOKTER****

Résumé - Abstract

L'analyse morphologique peut fournir d'aide intelligente aux étudiants débutants de français. Elle peut servir à les informer du sens grammatical de la morphologie, à leur donner accès à des dictionnaires en ligne, et à leur permettre la comparaison de toutes les occurrences d'un mot qui leur est inconnu. Cet article traitera d'un système dont l'implémentation s'est complètement réalisée et dont le prototype s'est déja avéré un succès dans une première étude de son utilisation.

Morphological analysis can provide intelligent assistance to beginning students of French. It can be used to provide information about the grammatical meaning of morphology, to enable access to online dictionaries, and to allow users to compare other occurrences of unfamiliar words. This paper reports on a fully implemented system whose prototype has proven successful in an initial user study.

Mots clefs - Keywords

Morphologie, Dictionnaire, Texte bilingue, Apprentissage des Langues à l'Aide des Ordinateurs

Morphology, Dictionary, Aligned Corpus, Computer-Assisted Language Learning

*Alfa-informatica, BCN, University of Groningen, P.O.Box 716, NL 9700 AS Groningen, The Netherlands. E-mail: nerbonne@let.rug.nl

**Alfa-informatica & Wisdom Research, P.O.Box 116, 9400 AC Assen, The Netherlands. E-mail: dokter@wisdom.nl

INTRODUCTION

Computer-Assisted Language Learning (CALL) is sometimes cited as an application area for Natural Language Processing (NLP), but in fact little use is currently made of NLP in this large and growing market for language software. Instead, CALL gets along with the standard techniques of Computer-Assisted Instruction (CAI), i.e. hypertext, database technology, digital audio and video, and networking. Some authors go so far as to conclude that advanced technology is simply too unreliable (Salaberry M. R. 1996). This judgement seems to be based on overly ambitious uses of artificial intelligence in CALL.

The present paper reports on the successful development of NLP-based CALL software designed to support the learning of reading, especially the bothersome problem of vocabulary acquisition. The project was based on morphological analysis software to assist Dutch learners of French, a language with elaborate morphology. The software was developed sufficiently to allow complete dictionary access to a medium-sized dictionary ($30,000$ entries), from arbitrary texts. Further, it provided useful information to learners on the grammatical meaning of morphology and also in the form of access to an extended text base containing examples of words in real contexts.

1.1. Background, Project and Goals

Linguistics scholars, but also school children, are fond of *glossed* texts, i.e. foreign language texts in which “glosses” (word-by-word, nearest native-language equivalences) together with morphology are provided. An example may stir memories:

Arma	virumque	canō,	Trojae	quī	prīmus	[...]
arm-Pl	man-Acc-&	sing-1s	Troy-Gen	RPro	first-N-Sg	

‘Arms and the man I sing who first from Troy [...]’

The second line contains the glosses, the nearest English equivalents of the Latin originals coupled with the grammatical status of the Latin inflections (and the clitic *-que*). While the largest market for such texts may well be school children cramming for exams they might better prepare for by learning Latin, the texts serve a legitimate purpose in allowing less experienced readers to approach natural, even challenging texts more quickly than they otherwise might. Glossed texts furthermore allow the deeper understanding of the grammatical patterns of a language which is valuable in adult language learning. These texts are sold to language learners across the globe, and are appreciated by them. If we can automate the provision of the information that makes them valuable, this should be useful.[1]

Glosser-RuG is one of the demonstrators developed within the GLOSSER project, a cooperation between the Universities of Groningen and Tartu, the

[1] *Pace* objections from occasional language teachers that dictionary lookup time is the motivating factor behind lexical learning.

Bulgarian Academy of Science, and research departments within Morphologic (Budapest) and Xerox (Grenoble). Glosser-R*u*G was developed at the Rijksuniversiteit Groningen. In many ways it is a modern variant of the old idea of text annotated for language learners, illustrated above for Latin.

The basic idea of the glossed text has been recast using modern means, and has been both restricted and extended. It has been recast by using automatic morphological analysis to provide the glosses—both the grammatical information carried by the morphological inflections and the dictionary equivalent. This means that essentially *any* French text is now available with Dutch glosses, for essentially the low cost of computer processing (ignoring the amortization of development). A further modernization of the idea has been to move the glosses to a hyper plane, so that readers control how many words are glossed. Practically, this just means that the glosses are supplied only on request.

The idea of the classical glossed text has been extended by providing other examples of word use, drawn automatically from corpora. In the example above, this would mean that a reader who wished to see further examples of the word *virum* could obtain these readily, perhaps examples such as these.

Paucī	virī	sapientiae	student
few-N-Pl	man-N-Pl	wisdom-Dat-Sg	eager-for-3-Pl

'Few men are eager for wisdom'

Note that the program is capable of finding alternative inflected forms of words, just as in this example, in which the string *virum* does not appear, only the (inflectional) alternative *viri*. This extension to the fundamental concept of glossing was intended to supplement dictionary explanation for advanced users.

The restriction of the software that's been realized (vis-a-vis the older glossed texts) is that the third line—the coherent translation—is not available. This is not technically feasible unless a humanly prepared translation is accessible. The latter option is explored in the corpus of examples (wherever bilingual corpora could be found).

We have perhaps overworked the metaphor of the glossed text in this introduction because it suggests why Glosser-R*u*G is successful—just as these texts have been. Simple, quick dictionary access alleviates the tedium and wasted time of dictionary lookup by hand (or by an online dictionary that isn't integrated into a reading browser).

1.1.1. 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 E. L. 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 E. & Feldweg H. 1997) had a similar focus to our own—that of providing “COMPprehsion 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*.

1.1.2. Project

Glosser-R*u*G aimed at applying state-of-the-art techniques in NLP to the paradigm of *communicative* CALL (Warschauer M. 1996) applications. The paradigm is “communicative” since the skill of reading written communication is central. Software was developed to facilitate the task of reading a foreign language by providing information on words. Techniques that were applied in this project include morphological analysis, part-of-speech (POS) disambiguation, aligning bilingual corpora, World-Wide Web technology, and indexing. The project vision foresaw two main areas where Glosser-R*u*G applications can be used. First, in language learning and second, as a tool for users that have a bit of knowledge of a foreign language, but cannot read it easily or reliably. The latter group might not be trying to learn, only to cope with a specific text. A user might, for instance, need to read a software manual that contains a number of unfamiliar words. Glosser-R*u*G provides the user (or learner) with a means of looking up information on unfamiliar words in a straightforward and user-friendly manner. Software has been developed by other partners for English/Estonian, English/Bulgarian, English/Hungarian. This paper describes only the French-Dutch work and demonstrator.[2]

1.2. This paper

The remainder of this paper reviews further the motivation for the project, its technical realization, and its functionality from the developer’s and from the users’ perspectives. Brief final sections characterize the contributions of the project and avenues for improvement.

2. MOTIVATION

If a rudimentary level of instruction in foreign-language grammar is assumed, then a great deal of the learning required in order to be able to read texts in this language is simply vocabulary learning, which is best pursued in lexical context (Mondria J.-A. 1996; Krantz G. 1991). Glosser-R*u*G aims to make this as easy and accurate as possible: for virtually all words that frequently occur in texts, Glosser-R*u*G aims to provide contextually sensitive in-

[2] The demonstrator for the other language pairs is described in (Glosser 1997).

formation in the form of dictionary entries, explication of morphology and examples of word use in especially collected (bilingual) corpora. Moreover, this information is to be accessed in a quick, straightforward manner within an integrated environment. Naturally, one needs to analyze the final result to assess the degree to which these aims have been achieved.

Glosser-R$\underline{u}$G aims to distinguish itself from many CALL programs by its emphasis on language use as opposed to drill and test, by its ability to support nearly any level of text difficulty, and by its emphasis on effectively removing the tedium of dictionary use from intermediate language learning. Other CALL software has focused primary on providing exercises, answer keys, and links to grammar explanations (Last R. 1992). Glosser-R$\underline{u}$G on the other hand, focuses on providing assistance to novice readers — whether these are actively involved in educational programs or not, and the focus is clearly on the level of *word*, including the grammatical information associated with inflectional endings. We therefore regard traditional CALL software as complementary in purpose.

Although Glosser-R$\underline{u}$G was primarily designed to serve the needs of people trying to learn language, it may also be of service to people trying to avoid learning a language (as noted above). (Zaenen A. & Nunberg G. 1995) note that, even as fully automatic machine translation has receded as a reasonable mid-term goal for natural language processing, several goals have emerged which are less ambitious, but useful and attainable. These focus less on eliminating language barriers and more on assisting people in learning and understanding the wide range of languages in current use. It is still the case that language differences form a substantial barrier to the free flow of ideas and technologies: ideas are effectively only accessible only to those in command of the language they are expressed in. But since an ever increasing number of people encounter texts electronically, automated methods of language processing may be brought to bear on this problem. Glosser-R$\underline{u}$G can help people who know a bit of French but cannot read it quickly or reliably. It allows a native Dutch person to learn more about French morphology, it removes the tedious task of thumbing through the dictionary and it gives examples from corpora. Some of these informational services are of use not only to language learners, but also to those trying to cope with foreign texts.

3. TECHNICAL REALIZATION

Glosser-R$\underline{u}$G makes it easier for Dutch speaking students to read French texts. It provides four kinds of information on words: morphological analysis, POS-disambiguation, automatic access in a bilingual dictionary, and examples of word use in especially collected corpora. The current demonstrator (implemented on the UNIX platform, in particular, HP-UX and Linux) is implemented completely in the Tcl/Tk scripting language (Ousterhout J. K. 1994), ensuring easy rewriting of parts of the program, rapid prototyping and portability of the source. Also, although the use of a scripting language obviously slows down

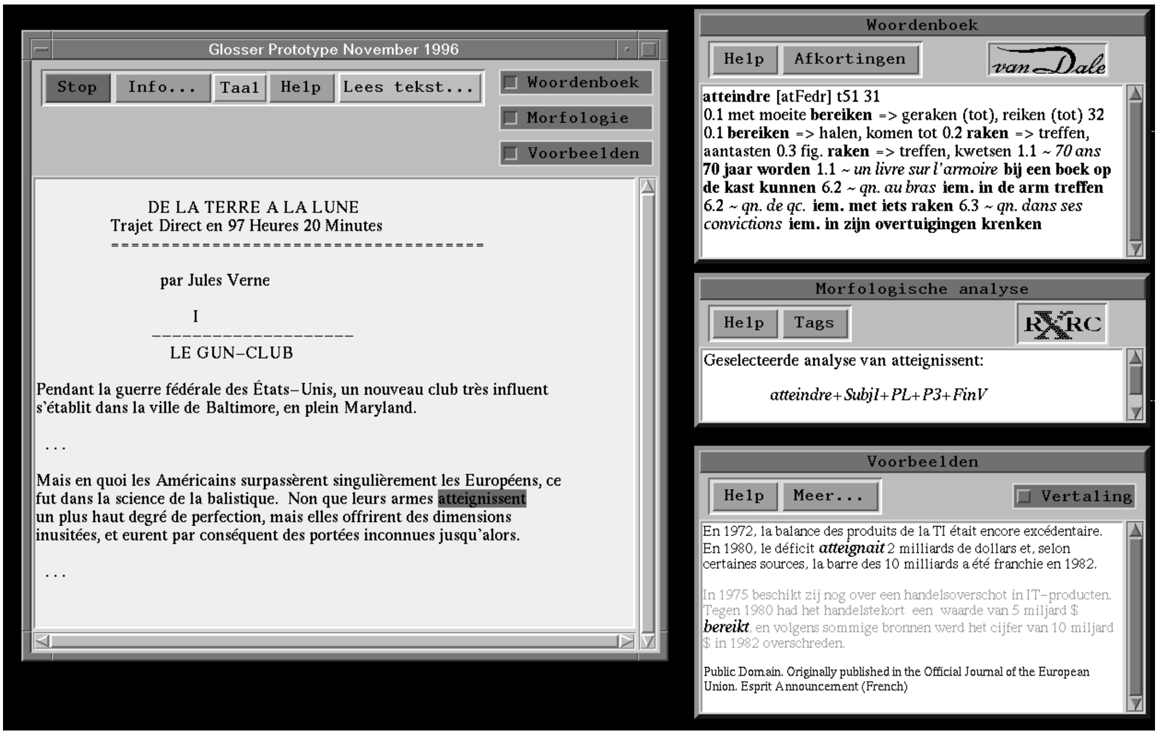

Figure 1. The Glosser-RuG front-end as it is displayed on the screen when morphology, dictionary and corpus examples are sought. The window providing help is omitted. In this snapshot, the user has sought information on the form *atteignissent* from Verne's *De la terre à la lune.* Note that the intermediate user might well fail to identify the lemma *atteindre* in this case. The correct dictionary entry from the bilingual dictionary is offered, some grammatical information on the form (e.g. that it is subjunctive) and a further example of its uses (in this case from a European Union document). Note that the example found in the corpus was an alternative inflected form of the verb—the corpus is indexed by lemmata, not strings.

processing in relation to compiled code, lookup speed is still sufficiently fast: a single lookup including all sources of information takes approximately 2 seconds (for details see (Dokter D. 1997a)). Most of this time is consumed by morphological analysis, which is an external, compiled program. Two seconds lookup time satisfied users. The same speed may be expected on other platforms.

3.1. Front-end

The front-end of Glosser-RuG, displayed in Figure 1 has four separate windows. The main window that is popped up when Glosser-RuG is started provides the general control of the application. The same main window contains a browser (read-only editor), a facility for loading files for reading, and three on/off-switches for controlling the specific sources of information that are to be used for word-lookup. The other windows are used for display of the three different sources: a dictionary, morphological analysis/ POS-disambiguation, and examples. The window providing examples actually consists of two separate windows, one for display of the examples, the other for the aligned translation (if any). Apart from these main windows, there is a separate window that provides help in a hypertext fashion especially on the interpretation of the information in different components of the application.

The sources to be accessed for a given lookup can be specified by the user at any time during a session. Another feature is a pop-up menu that shows the files in the current directory. Clicking on a file loads it into the editor to be processed.

3.2. Morphological Analysis

This module is Glosser-R$\underline{u}$G's intelligence. It parses the word selected by the user, and thus determines the stem (or lemma) and endings. Morphology is naturally called whenever a user asks for morphological information about a word. If the user has asked in particular for morphological information, then stem and endings are displayed in an accessible fashion. In this case, we assume that morphological analysis/POS-disambiguation is directly informative to the user.

But note that the morphological parse is indispensable for dictionary access. It is used to find the underlying lexemes of words as they appear in the text, since in general dictionaries do not provide entries for inflected forms. For example, if a beginning user wishes to look up the form *cru,* (s)he has probably encountered a form of *croire,* but a direct string search in a dictionary cannot find this lemma, only the homographic nouns. Similarly, the results of the morphological parse are indispensable to the *examples* modules. Without this, the search for other occurrences in the text would amount to string search. Related forms of the word would never be found. Given a morphological module, we enable users to search for occurrences of *words,* and not just *strings* (see subsection 3.4).

The *morphology* module finds all of the possible morphological parses of a word, in general several. Since language learners in general have little benefit from appreciating this degree of potential ambiguity in language, it was necessary to find some means of cutting it down. This task is performed effectively by part-of-speech disambiguation. POS disambiguation is precisely the correct level given a dictionary which provides this information (as most do). Naturally it is necessary to coordinate the categories of the POS disambiguation machinery with the dictionary for most effective deployment. The POS disambiguator assigns a part-of-speech category to each word in the text. This additional information is used to choose the right entry in the case of homographs in the dictionary and corpus of examples.

Glosser-R$\underline{u}$G was fortunate to be able to use a state-of-the-art morpho-

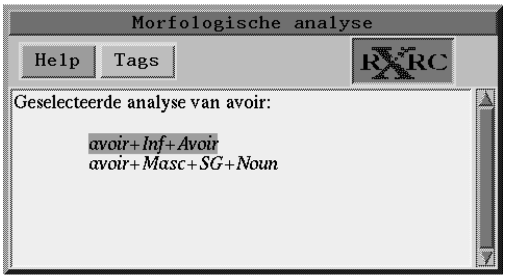

Figure 2. Morphological analysis window. Note that here, just as in general, there is more than one possible analysis for a word. In that case, statistical disambiguation machinery is brought to bear. The most likely analysis is highlighted, as it is here, but alternative analyses are available to users who wish to pursue information about these.

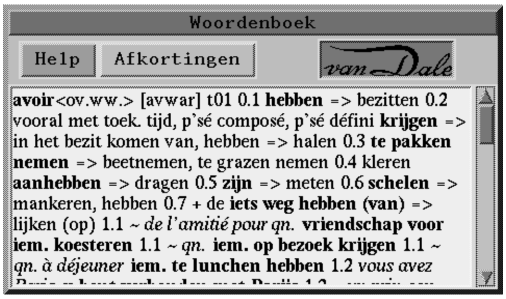

Figure 3. A dictionary entry as displayed in Glosser-R$\underline{u}$G. In order to make the program maximally self-explanatory to users, the dictionary information is presented in its customary format.

logical analysis/POS-disambiguation package from Xerox's Grenoble Research Centre: *Locolex* (Bauer D. *et al.* 1995). *Locolex* is built on finite-state machinery (Sproat R. 1992; Roche E. & Schabes Y. 1997) and is capable of recognizing all the approx. $300,000$ forms of nearly $50,000$ lemmata. An example analysis is shown in Figure 2.

Because of the large number of words which can have different grammatical functions (noted above), *Locolex* incorporates a stochastic POS tagger which it employs to disambiguate. In case *Locolex* is wrong (which is possible, but quite unlikely), the user is free to specify an alternative morphological analysis, which is then looked up in the dictionary and examples index.

3.3. Dictionary

Glosser-R$\underline{u}$G incorporates the Van Dale dictionary *Hedendaags Frans* (van Dale 1993), a bilingual French-Dutch dictionary containing $30,000$ lemmata (base forms corresponding to approx. 180.000 inflected forms). Figure 3 illustrates the front-end of the dictionary within Glosser-R$\underline{u}$G. It was essential to the success of the development effort that the dictionary was available in a well-structured format. It was not necessary to become involved in the intricacies of converting typographical source files.

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. A further refinement implemented a simplest form of word sense disambiguation: whenever the dictionary listed fixed, contiguous expressions (e.g., *idée fixe*) the lookup process was sensitized to check for these in texts.

In one demonstration version of Glosser-R$\underline{u}$G, audio files were added to a small number of dictionary entries to suggest the value of multimedia dictionaries as foreign language aids.

3.4. Examples

To provide users with a variety of examples, a large corpus was collected, consisting of a number of different kinds of texts, for instance, literature, technical, political, etc. Also, as many bilingual examples as possible had to be provided, to ensure easy understandability of how words can be used. For the corpus Glosser-R$\underline{u}$G has relied on specialized corpus projects, such as the ECI (ECI) and MULTEXT (MULTEXT) for bilingual corpora, although some work in (re)aligning the texts was needed. It was difficult to find unencumbered, aligned, bilingual texts in French and Dutch.

In order to determine the size of corpus needed, experiments were originally conducted with a frequency list of the $10,000$ most frequent word *forms* in French (information on *word* frequency not being available). A corpus of 2 MB contained 85% of these, and a corpus of 6 MB 100% (van Slooten A. 1995). A reasonable goal for a Glosser-R$\underline{u}$G-like product would be 100% coverage of the words (lemmata) found in the $30,000$-word dictionaries, and 100% coverage of the most frequent $20,000$ words. This proved overly ambitious. The final corpus size for Glosser-R$\underline{u}$G was 5 MB in monolingual, 3 MB in bilingual text (that is, the French text), accounting for $16,701$ different lexemes.[3]

As the corpus grows, the time for incremental search likewise grows linearly. When the average search time grew to several seconds (on a 70 MIPS UNIX server), it became apparent that some sort of indexing was needed. The texts were then indexed by determining the lemmata and POS of the individual words using the same morphological analysis method as described in section 3.2, and creating an index of N occurrences of each lemma and POS thus

[3]A back-of-the envelope characterization suggests that 10 times as much text would be required for full coverage. This is based on Zipf's famous observation that frequency is inversely proportional to rank frequency and the statistics to-date (including the overlap with the dictionary). This would be feasible, but was regarded as too time-consuming for the original project.

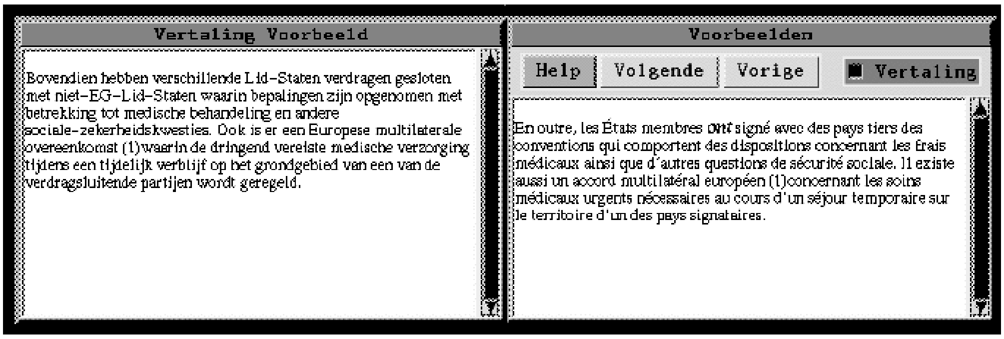

Figure 4. A bilingual example in Glosser-R<u>*u*</u>G.

found (Dokter D. 1997b). The index thus provides a link between the lemmata and the full, possibly inflected forms. Lexeme-based indexing provides not only further occurrences of the same string, but also inflectional variants of the word. If the selected word is *livre*`+Masc+SG+Noun` for instance, the search should find other tokens of this and also tokens of the plural form *livres*. It is clear that 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 4. If the example has been found in a bilingual text, the user can 'pop-up' the translation from the examples window.

3.5. Architecture

The modules described above are connected in a straightforward way. A central control module handles communication between the user and the auxiliary modules and data resources. The control module itself communicates with four further modules, namely morphology, dictionary, examples, and help, the first three of which are described above.

The help facility contains explanations and instructions of major facilities, and a glossary of the abbreviations for grammatical terms used in the *morphology* and *dictionary* components. Only a rudimentary help module is realized.

A demonstrator with reduced functionality is available on the World Wide Web at `http://www.let.rug.nl/~glosser`. In order not to expose valuable dictionary resources to theft, only a limited number of texts (and hence, words) may be looked up.

4. FUNCTIONALITY

Once the demonstrators that were developed had been found sufficiently robust to support reading of essentially all non-specialized texts, they were subjected to performance analysis and user studies.

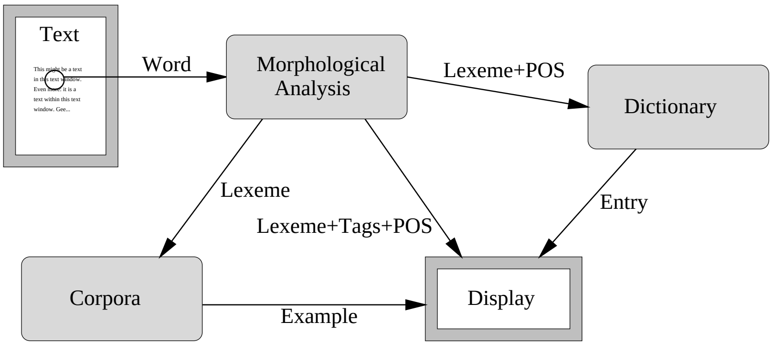

Figure 5. A sketch of Glosser-R$\underline{u}$G architecture.

4.1. Performance Analysis

The performance analysis tried to *evaluate* how well the software met its specifications, and *assess* whether shortcomings would be important in this type of application. To begin, 500 words in 100-word samples were selected at random from five different texts. These words were fully checked for accuracy in analysis.[4] The texts varied in genre: official European Commission prose, (soft) pornography, poetry, and political opinion.

The mistakes were distributed unevenly in the texts and can be grouped into four types: mistakes of input due to incorrect selection by testers or incorrect keyboarding in the text itself (misspelled words); words missing from morphological analysis or dictionary; incorrect linguistic analysis; and irrelevant corpus examples. We illustrate and discuss each of these in turn.

input errors Some of these arose because testers expected cliticized elements (e.g., the *l'* in *je l'ai lu*) to be included in the analysis, but the application's tokenization excluded them. The exclusion was motivated by algorithmic convenience, but also by our intended user, the intermediate-level language learner, who should not need assistance for these words. (We are aware that advanced language learners find subtleties in function word use challenging, but that is a separate matter. Glosser-R$\underline{u}$G automates dictionary access, but not grammar explication or even access.)

Misspelled words also fell into in this category. While it is easy to dismiss these as falling outside specifications, they might still be bothersome in an application of the Glosser-R$\underline{u}$G type. Suppose, for example, that one wished to attach a Glosser-R$\underline{u}$G -like facility to an email browser. Spelling errors would then naturally be expected. And spelling errors occur to some degree even in carefully edited text. We interpret this to mean that

[4]We gratefully acknowledge the work of 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.

a facility to correct spelling would be welcome in an application of this type.

Finally, several “errors” resulted in applying Glosser-RuG to ASCII text, because Glosser-RuG expects Latin8 encoding. Some of the errors were invisible to the eye, in which accented capitals had been encoded as unaccented (which is a common typeface), e.g. *Église*. Naturally, ASCII encoding is less than optimal for French, particularly in texts for learners, but it would be preferable for the software, not the learner, to be made sensitive to this.

missing words 12 words did not appear in the dictionary or could not be analyzed by the morphology software correctly. Seven of these missing words were brand names and the like, e.g., *Collier's, Vargas* and *Life*. Naturally, neither the morphological analyzer nor the dictionary can or should be so complete as to include *all* French words, let alone foreign ones. Moreover, we suspect that this sort of error is negligible in the language learning application, since brand names, etc., will be easily recognized by intermediate level users.

Five words were missing from the dictionary. This can in no way be regarded as a shortcoming of the dictionary, which was chosen exactly for its limited coverage. A more comprehensive dictionary would be less useful to intermediate-level students.

Finally, two missing words were *fréquemment* and *généreusement*, which in fact *are* in the dictionary, but listed under the adjectives they are derived from, *fréquent* and *généreux*, respectively. This suggests that a second level of dictionary indexing would be useful.

linguistic errors 17 “misanalyses” fell into this category, which was surprising since users hadn't seem to encounter any. There were no errors of morphological analysis: in every case the analysis selected was a linguistically possible one (which was incorrect given the context). All the errors were faulty preferences for particular morphological analyses (5) or POS categories (12). As an example of the first type, *redouble* was analyzed as an imperative when it was in fact used as a third-person indicative. As an example of the second type, *droits* was analyzed as an adjective, when it was used as a noun.

These errors resulted in a preference for a possible, but incorrect analysis. It is worth noting that few of these errors are “carried” into subsequent processing. Faulty morphological analysis naturally does not result in incorrect dictionary lookup. Even the incorrect POS assignments usually do not result in faulty dictionary look-ups, at least in the dictionary used, since this dictionary lists such homonyms in a single entry (e.g., *droit*). In fact, none of these errors placed the user in the incorrect dictionary entry. The rather higher number of errors in POS assignment here in contrast to user studies 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 (again, we ignore advanced users' problems with the subtleties of using function words). This is a point at which the intended application is forgiving vis-à-vis shortcomings in the underlying technology.

irrelevant examples We found the most errors in seeking examples from corpora. 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 more serious. This naturally suggests that Glosser-R$\underline{u}$G'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, a much larger corpus is needed (see section 3.4 above).

To sum up, four major mistake types appear rather more frequently than one would wish. None of these mistakes surface in extensive user experimentation, which provides both a reason for them not being solved during development, and a prove of practical usability even with these flaws.

5. A USER STUDY

A user study was conducted with a group of second-year students of French (advanced beginners).[5] The goal of this study was to evaluate Glosser-R$\underline{u}$G in comparison to a traditional method of instruction in reading comprehension in which hand-held dictionaries are used.

We were interested in the effect of Glosser-R$\underline{u}$G on text comprehension and the facility with which dictionary information could be accessed. In addition, we asked about overall satisfaction with the program. The group of 22 subjects was divided into a group that used Glosser-R$\underline{u}$G, and a group that used a hand-held version of the same dictionary. In addition to the factors noted above, the subjects using Glosser-R$\underline{u}$G were asked to comment on the system, to give us a clearer picture of users' demands on this sort of application and suggestions for improvement.

5.1. Results

The results of the study concern three issues: comprehension, functionality and the subjective satisfaction of Glosser-R$\underline{u}$G users. We include speed

[5]The material from this section is reported on at more length in (Dokter D. *et al.* 1998; Nerbonne J. *et al.* 1999).

in addressing functionality. We address these in turn.

First, Glosser-R$\underline{u}$G subjects understood the text better than the users of the hand-held dictionary, but the result was not statistically significant. It was interesting to note that Glosser-R$\underline{u}$G users estimated their own comprehension much higher—the program appeared to boost confidence.

Second, since virtually all subjects used all of the time allotted for reading—defying our expectations, we were not able to find contrast in reading speed (although it was apparent from log files that Glosser-R$\underline{u}$G subjects re-read the text). This is a point at which the experimental design bears revision. Still, Glosser-R$\underline{u}$G users looked up three times as many words as paper dictionary users. This ignores repeated lookups of the same word, something which also occurred (see below). (Krantz G. 1991) argues that the number of dictionary look-ups is a reasonable guide to a low level of vocabulary acquisition, so the sheer speed is bonus.

Third, all Glosser-R$\underline{u}$G subjects were keen on re-using the program. The overall judgement of the program was very positive, 4.2 on a scale of 5.

5.2. Additional Insights

In addition to issues we specifically addressed in designing the study, several further aspects of this program came to light in the course of the user study.

Log files showed that the dictionary contained the most important information for users, a fact which informal remarks corroborated (although the teacher of the course was enthusiastic about morphological analysis being available).

Users prompted several modifications of Glosser-R$\underline{u}$G through their "additional remarks". In the first prototype (the one used in the user study), 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 errors, since the procedure asked for some precision: non-words (parts of words, etc.) were often selected and not filtered adequately by "tokenizing". In view of this and user comments, newer versions of Glosser-R$\underline{u}$G automatically select *words* under the mouse (highlighting to make this transparent to the user). A single mouse-click now starts a search action. This automated selection ensures that no erroneous words can be submitted for lookup; a degree of control has been surrendered to the application.

The fact that users often looked up the *same* token of a given word several times prompted us to recall the original metaphor on which Glosser-R$\underline{u}$G is based, that of the "glossed text". A near kin to glossed text is self-annotated text in which marginalia are recorded precisely in order to obviate repetitions of dictionary look up. All beginning readers of foreign languages have well-scribbled practice texts attesting to the usefulness of this practice. This led us to explore means of allowing users to record dictionary equivalents directly in the text being processed. In newer versions the user may click on dictionary translations in order to insert them into the text (directly after the original word). A mouse-click on a translation removes it if necessary. In both cases a selection "highlight" follows the mouse in order to make the actions as transparent as possible to the user.

5.3. User Study Conclusions

Glosser-R*u*G improves the ease with which foreign language students can approach a foreign language text. The most important difference is simply the speed of lookup, consequently 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 text comprehension improvement was not significant, further experiments may demonstrate modest gains. Perhaps most significantly, users were enthusiastic about continuing with the program. A (future) longitudinal study might focus on actual vocabulary acquisition.

6. FUTURE DIRECTIONS

In addition to the obvious possibility of implementing further versions of Glosser-R*u*G or similar programs for other languages, several improvements suggest themselves:

digitalized pronunciations Even dictionaries with careful IPA renderings of pronunciations cannot compete with high-fidelity recordings. Arguably, this should be a requirement on the dictionary used.

multi-word lexemes Many dictionary senses are tied to more than one (typographical) word. *mettre qn. à la porte* (lit. 'to put someone at the door', i.e. to throw someone out) depends for its meaning on the verb *mettre* and the prepositional phrase *à la porte*. As the COMPASS project demonstrated, there is both a need for better processing here, and an opportunity to recognize a reasonable percentage of such examples efficiently (Breidt E. & Feldweg H. 1997).

word-sense disambiguation A very primitive notion of word-sense disambiguation was implemented: whenever the dictionary provided a string of several words, this would be highlighted in case the user selected the same string in text. For example, the entry *mondial* in the Van Dale dictionary provides the example *guerre mondiale*, translated as Dutch *wereldoorlog* ('world-war'). Glosser-R*u*G exploits this feature by providing the correct sub-entry if *mondiale* is selected in the context immediately preceding *guerre*. This reduces look-up time for the user.

There remain serious obstacles to providing reliable sense discrimination, in particular sense discrimination based on a given dictionary scheme.

pedagogical additions Teachers who examine Glosser-RuG would like more opportunity to prepare texts with their own annotations. Others have asked about the opportunity to derive test material from texts used for reading. Both of these are feasible, the latter within limits.

add-ons for browsers or email The demonstrator suggests that it might be useful to support reading directly by people who are not engaged in formal language instruction, or perhaps not even primarily interested in improving their foreign language ability. Given our emphasis on automatic methods applicable to arbitrary texts, a spin-off in support for translations is conceivable.

CONCLUSION

Glosser-RuG was developed with the philosophy of exploiting available NLP technology wherever possible. Glosser-RuG shows that valuable tools for communicative CALL are feasible given the current state of NLP technology.

More generally, the success of the effort shows that current, imperfect NLP techniques can be put to good use if the application is appropriate. In this case, the difficult task of *understanding* the text is left to the language learner, and NLP concentrates on the routine tasks, such as dictionary access, which most language learners find tedious and unproductive. The focus on tasks humans *wish* to avoid is beneficial.

Acknowledgements

Petra Smit worked successfully on early prototypes until illness demanded her full attention. Auke van Slooten collected French texts on the Internet and developed the first indexing programs. Edwin Kuipers donated code for the Web demonstrator, and Lily Schurcks-Grozeva collaborated in the user study. Felice Portier of the Centre Culturel Français, Groningen, and her students provided valuable criticism on earlier versions of this prototype. Lauri Karttunen, Elena Paskaleva, Gábor Prószéky, Tiit Roosma collaborated in the design—which was common to the French-Dutch assistant described here and an English assistant developed for Bulgarian, Estonian and Hungarian speakers. The work reported here was funded by Copernicus grant 94-343 to the GLOSSER project.

RÉFÉRENCES

ANTWORTH, Evan L. (1992) : “Glossing text with the PC-KIMMO morphological parser”, *Computers and the Humanities*, vol. 26, n˚ 5-6, pp. 389–98.

BAUER, Daniel , SEGOND, Frederique , ZAENEN, Annie (1995) : “LOCOLEX: Translation rolls off your tongue”, *in Proceedings of the Conference of the ACH-ALLC'95*, Santa Barbara, USA.

BREIDT, Elisabeth , FELDWEG, Helmut (1997) : “Accessing foreign languages with COMPASS”, *Machine Translation*, vol. 12, n˚ 1-2, pp. 153–174, spec. iss. on New Tools for Human Translators.

DOKTER, Duco , NERBONNE, John , SCHURCKS-GROZEVA, Lily , SMIT, Petra (1998) : “Glosser-R*u*G: A user study”, *in Language Teaching and Language Technology,* S. Jager , J. Nerbonne , A. van Essen (ed.), pp. 169–178, Lisse.

DOKTER, Duco (1997a) : *Glosser-RuG, Prototype December 1996*, Techreport, University of Groningen, Alfa-Informatica.

DOKTER, Duco (1997b) : *Indexing Corpora for GLOSSER*, Techreport, University of Groningen, Alfa-Informatica.

ECI, *European Corpus Initiative (ECI) Multilingual Corpus I*, `www.elsnet.org/resources/eciCorpus.html`.

GLOSSER (1997) : *Glosser, Final report*, Final project report, University of Groningen, Alfa-Informatica, avail. at `www.let.rug.nl/~glosser`.

KRANTZ, Gösta (1991) : Learning Vocabulary in a Foreign Language, Thèse de PhD, Göteborg.

LAST, Rex (1992) : “Computers and language learning: Past, present — and future?”, *in Computers and Written Texts,* C. Butler (eds.), Oxford, Blackwell, pp. 227–245.

MONDRIA, Jan-Arjen (1996) : Vocabulaireverwerving in het vreemdetalenonderwijs: De effecten van context en raden op de retentie, Thèse de PhD, University of Groningen, Groningen, The Netherlands.

MULTEXT, *Multilingual Text Tools and Corpora*, `www.lpl.univ-aix.fr/projects/multext/`.

NERBONNE, John , DOKTER, Duco , SMIT, Petra (1999) : “Morphological processing and computer-assisted language learning”, *Computer-Assisted Language Learning*, vol. 6.

OUSTERHOUT, J. K. (1994) : *TCL and the TK Toolkit*, Reading, Mass., Addison-Wesley.

ROCHE, Emmanuel , SCHABES, Yves (1997) : *Finite-State Language Processing*, Cambridge, Massachusetts, The MIT Press.

SALABERRY, M. Rafael (1996) : “A theoretical foundation for the development of pedagogical tasks in computer-mediated communication”, *CALICO Journal*, vol. 14, n˚ 1, pp. 5–34.

SPROAT, Richard (1992) : *Morphology and Computation*, Cambridge, MIT Press.

VAN DALE (1993) : *Handwoordenboek Frans-Nederlands + Prisma, 2e druk*, Van Dale Lexicografie b.v.

VAN SLOOTEN, Auke (1995) : *Searching and Quoting Examples of Word-Usage in French Language Corpus*, Rapport technique, Alfa-Informatica, Rijksuniversiteit Groningen.

WARSCHAUER, M. (1996) : “Computer-assisted language learning: An introduction”, *in Multimedia language teaching,* S. Fotos (eds.), Tokyo, Logos International, pp. 3–20.

ZAENEN, Annie , NUNBERG, Geoff (1995) : “Communication technology, linguistic technology and the multilingual individual”, *in CLIN V: Papers from the Fifth CLIN Meeting,* T. Andernach , M. Moll , A. Nijholt (ed.), pp. 1–12, Enschede.