

# Grammar, Incorporated

*Lars Borin*

*Göteborg University*

*lars.borin@svenska.gu.se*

*Anju Saxena*

*Uppsala University*

*anju.saxena@lingfil.uu.se*

## **Abstract (Swedish)**

Grammatik hör till kärnämnen i lingvistik och datorlingvistik. I grammatikundervisningen borde inte vårt mål vara att få studenterna att lära sig definitioner av begrepp och konstruktioner utantill, utan de borde istället lära sig att förstå och identifiera strukturella mönster. Vi förordar användningen av autentiska texter – korpusar – som utgångspunkt för övningar i grammatikkurser. Lösningar framtagna i projektet *IT-baserat kollaborativt lärande i grammatik* används som illustration av hur korpusar (svenska i kontrast till ett "okänt" språk) kan integreras i ett IT-baserat grammatiklärostöd.

## **Abstract (English)**

Grammar is part of the core curriculum in linguistics and computational linguistics. While teaching grammar, our goal should not primarily be that students memorize definitions of concepts and grammatical constructions, but rather that they understand and learn to recognize different structural patterns. We argue in favour of the use of authentic text corpora, as the basis for exercises in grammar courses. Some of the solutions developed in the project *IT-based Collaborative Learning in Grammar* are presented to illustrate how corpora (of Swedish in contrast to an "unknown" language) can be integrated into an IT-based grammar training material.

## **1. IT-Based Collaborative Learning in Grammar**

Grammar, explicitly stated, scientific generalizations about linguistic form and its relation to linguistic content, is a necessary basic component of all linguistic studies of language and, thereby, part of the core curriculum in the subjects of linguistics and computational linguistics. In addition to this, grammar is also accorded a prominent role in many foreign language

curricula, especially if the language in question is morphologically rich or deviates strongly in some other way from the Western European "modern languages" in its structure.

It is also generally acknowledged that the goal of teaching grammar – especially at the university level – should not primarily be that students memorize definitions of concepts and grammatical constructions, but rather that they understand and learn to recognize different structural patterns<sup>1</sup>. This can hardly be achieved without giving students practical training in the skill of grammatical analysis. Research has shown that hands-on problem solving is more stimulating and thought provoking, than when the information and results are handed down to the students during lectures. Furthermore, our experience has been that students learn about grammatical constructions and phenomena more actively when these constructions are discussed by comparing the system found in Swedish (the language our students are most familiar with) with another language (Saxena 1997, 2000a, 2000b). An added factor contributing to active student participation is the choice of material for exercises and group activities. A corpus of natural language material for grammatical analysis contributes to more active student participation, as we argue below.

With these pedagogical considerations in mind, the project *IT-based Collaborative Learning in Grammar* suggests a new form for teaching courses in grammar in linguistics. We wanted to focus on grammar courses, because students often are insufficiently prepared in this subject from secondary schooling, and, as we already stated above, the ability to reason about grammar and to carry out grammatical analyses of language utterances are necessary prerequisites for all linguistic study of language. In the proposed method, interactive practical training and corpus-based exercises comprise an integral part of the students' learning process, giving

---

<sup>1</sup> The work described here forms part of the project *IT-based Collaborative Learning in Grammar*, a collaboration between the universities in Uppsala, Stockholm and Gothenburg, originally granted funding by the Swedish Agency for Distance Education (DISTUM), and after DISTUM was split up and the parts reallocated into other national agencies in 2002, continued funding by the Swedish Council for the Renewal of Higher Education, for the three years 2002 - 2004. Anju Saxena is the principal investigator for the project. See also

<http://www.lingfil.uu.se/personal/anjusaxena/distum.html>. Leif-Jöran Olsson, Department of Linguistics, Stockholm University, has carried out all software development and linguistic resources integration in the project.

them the opportunity and incentive to participate more actively in their own learning process. Using IT as a tool for collaborative work allows the students to choose the problem solving strategy which suits them best, as well as the time and place to work on the problem. Apart from the lecture sessions, the student-teacher contact, in this context, can either be in the classroom or virtual, i.e. by means of electronic communication.

The aim of this paper is to argue that the use of authentic texts in the form of corpora as the basis for exercises in grammar courses is more advantageous than using made-up or isolated single sentences or phrases. The access to corpora on the Internet allows much more flexibility, than if the same text material were available in book form. Some of the solutions developed in the project *IT-based Collaborative Learning in Grammar* are presented below to illustrate how language corpora can be integrated into an IT-based training material.

## **2. Some Advantages of Using Corpora in Teaching Grammar**

A corpus of natural language material for grammatical analysis contributes to a more active participation, as it not only presents the grammatical constructions in their context, but also gives students a greater freedom to approach the material and conduct the investigation from a perspective which suits their individual learning styles. A text corpus consists of naturally occurring language in its natural linguistic context, since it is made up of complete texts or large text fragments, as opposed to the made-up or isolated single sentences or phrases often used to illustrate grammatical points in linguistics textbooks. This accompanying linguistic context makes it possible to investigate the textual, discourse-level, and functions of the grammatical phenomena.

There has been a growing interest in using natural language corpora in teaching and in research, partly due to the growing availability of computer-readable linguistic corpora, and partly due to an increased interest among linguists in examining language in its natural context as opposed to investigating constructed language examples in isolation. Researchers, teachers and students now have access to different types of language corpora to discover facts about language, for example, which words are the more frequently used words in a language or a language-type, in which context they predominantly occur, and which grammatical

patterns are associated with a particular linguistic item (Ghadessy, Henry & Roseberry 2001).

There have been two primary approaches to the use of corpora in language teaching/learning, viz. the "*behind the scenes approach*" ("COBUILD approach"), and the *Data-Driven Learning* (DDL) approach. Until recently, the "*behind the scenes approach*" predominated. Corpora, in this approach, are used by researchers and materials producers in building dictionaries and other language learning material, and in fact most modern dictionaries of English claim to be corpus-based. The corpora involved have, traditionally, been very large. The Bank of English Corpus – the corpus resource used in compiling the Collins COBUILD dictionary (hence the alternative name "COBUILD approach") – comprises more than half a billion words, and this number is growing, as the corpus is continuously being added to. Furthermore, in this approach, the users (students, for example) do not get to use the corpus themselves in order to come up with their own analyses and learn from that. Instead, they are confined to the end products – dictionaries, textbooks, grammars, computer-assisted language learning programs, etc. – with all their pros and cons, including a sometimes fairly high cost.

In the DDL approach, students use corpora directly in their own learning. They use the corpus, for example, to discover linguistic patterns and to organize linguistic patterns which they observe, arriving at generalizations inductively and verifying deductive rules. Such exposure to corpora gives students the chance to, not only, extract relevant examples of one or the other linguistic structures, but it also provides them with material for discussion, when they find gaps to verify, extend their hypothesis, and arrive at generalizations.

One advantage of using corpora in teaching is that instead of learning about linguistic theories in a vacuum (a more passive learning method, where facts are fed to students in the form of lectures and textbooks), students have a chance to test these theories themselves against the corpora and learn about these theories or concepts for themselves (a more active learning method). When corpora are used by students as part of their learning, the distinction between teaching and research is "blurred", as students, by a kind of "discovery procedure" (thus, research), learn things for themselves (Knowles 1990). The use of corpora in teaching can, in this

way, affect both teachers' and students' roles. This approach is as relevant in a classroom setup as in self-study situations.

The corpora used for this purpose may be large or small. The purpose here is exploration and not a waterproof description of the phenomenon under consideration. It is possible that the use of small corpora may lead to different results than if large corpora had been used instead, but this difference does not take away all the advantages of using corpora in teaching.

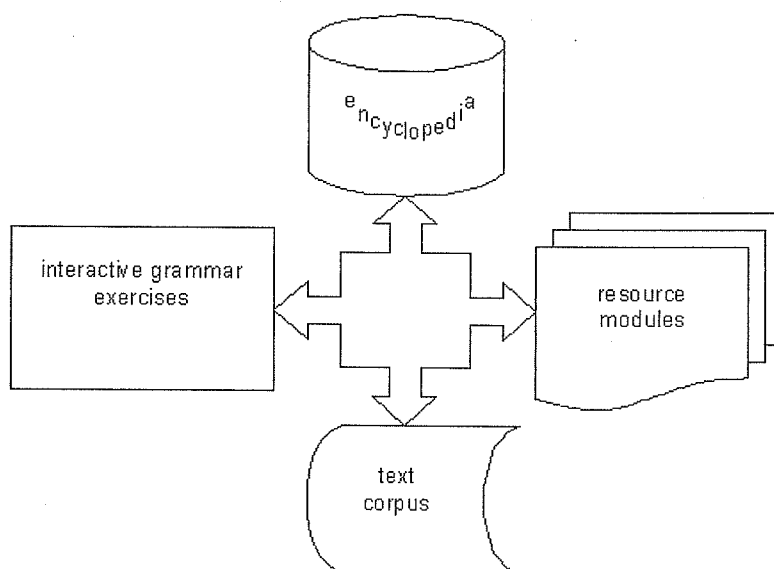
The gap between the COBUILD and DDL approaches is, however, getting smaller. More access to corpora provides better conditions for using them in producing language learning tools, as well as in using them directly in teaching/learning. Most importantly, current language technology actually allows for the design of IT-based learning tools that combine the two approaches, making possible both "pre-digested" and direct access to unadulterated corpora in the same learning environment.

### **3. An Outline of the IT-Based Collaborative Learning Model**

The project *IT-based Collaborative Learning Grammar* started January 1, 2002 with financial support from Distum/Swedish Council for the Renewal of Higher Education in Sweden with Anju Saxena, Department of Linguistics and Philology at Uppsala University, as the Principal Investigator. The collaborating institutions are Uppsala University (Department of Linguistics and Philology, Department of Scandinavian Languages, Department of Information Technology, Development and Evaluation Unit), Stockholm University (Department of Linguistics), and Göteborg University (Department of Swedish, Department of Linguistics). The main aims of the project are:

- To use web-based annotated corpora of Swedish and a foreign language as the basis for learning grammatical patterns
- To develop web-supported collaborative learning in grammar, where corpora form the basis for group activities, illustrations in lectures, and as part of the final examination
- To use the web-supported collaborative method in regular courses in grammar in collaborating university departments

A significant consideration during all the stages of this project has been that the technical aspects of using or working with the web-based system should not increase the workload for students or faculty. An outline of the training material under development in the project is presented below. It has a modular architecture, composed of four types of modules (figure 1):



**Figure 1: Organization of the proposed IT-supported grammar training application**

1. "Encyclopaedia" module (Figure 2), containing descriptions of grammatical concepts and constructions. Its content will be attuned to the contents of the course and the interactive exercises (as, in their turn, the exercises will be adapted to the "encyclopaedia" content), and at appropriate places, there will be hyperlinks to interactive exercises dealing with the current topic.

2. "Text corpus" module (Figure 3), containing at least (a) part of speech-tagged (POS-tagged, i.e. annotated with morphosyntactic categories) and syntactically annotated corpora (also known as *treebanks*) of Swedish, and (b) an annotated corpus of a foreign language. For (a), we are using the SUC (POS-tagged) and Talbanken Swedish corpora (annotated with functional syntactic labels, such as *subject* and *object*, as well as morphosyntactic categories).<sup>2</sup> For (b), we use a corpus of Kinnauri narratives,

<sup>2</sup> The Stockholm Umeå Corpus (SUC) is described by Ejerhed & Källgren (1997) and the format and annotation of the Talbanken syntactically annotated corpus of Swedish (Einarsson 1976a, 1976b) are thoroughly described by Teleman (1974).

which is hyperlinked to a morpheme dictionary.<sup>3</sup> Further, the graphical user interface provides students with a "map" of how and where one particular morpheme, word, or syntactic function occurs in the corpus, providing support in their work on the functions of grammar (Figure 3). The students will work with the same corpus as part of their group activities and as part of their examination. Furthermore, we suggest using the same corpora in grammar courses at all levels. As a result of this, when students enter C-level studies and continue working with these corpora, they can use the knowledge acquired during the B-level grammar course and, hopefully, do more in-depth analyses of complex syntactic constructions than would have been the case otherwise. At the beginners' level (the A-level), we will use only the Swedish corpora, but at higher level grammar courses in linguistics (B- and C-levels), and in the Swedish as a second language education program, we will use an annotated corpus of a foreign language along with the Swedish corpus. The corpus of Kinnauri narratives will be used during the project period.

The fact that suitable corpora are available and, to a large extent, may be used directly in the project is a major advantage. But our attempts to reuse existing language technology resources have also raised issues of compatibility and prompted a general wish for more standardization in this area (this particularly concerns the resources used in the "exercise" module, see section 5 below).

3. "Interactive exercise" module (Figure 4). Our aim here is to provide students with a set of exercises, with basic tools for computer mediated student cooperation in virtual workgroups, e.g.:

- a "spreadsheet" for problem-solving
- optional "step-by-step questions" for the grammatical topic covered
- corpus search exercises (Figure 3)
- corpus-based grammatical analysis exercises, e.g. part of speech assignment (Figure 6) and determining the syntactical function of a constituent (Figure 7)

---

<sup>3</sup> Kinnauri is a Tibeto-Burman language spoken in India. Portions of the Kinnauri narrative corpus are available on the web:  
<http://www.lingfil.uu.se/personal/anjusaxena/korpusen.html> (Figure 5).

- grammar rule writing exercises, whose basic premise is a further refinement of an idea presented by Borin & Dahllöf (1999). We propose to use grammar rules written by students (ideally using an existing grammar development tool) as search expressions in the treebank. In its simplest form, the result of the search would be expressed as precision and recall. Given an NP rule formulated by a student, we could automatically tell how many of the (maximal) treebank POS sequences matching the rule actually make up NPs, how many are not NPs, and how many NPs in the treebank are not described by the rule. There are all kinds of conceivable elaborations of this basic scheme, which could be seen as a more linguistically sophisticated parallel to the use of (un-annotated) text corpora and concordancing software in the DDL approach (Flowerdew 1996). For the computational linguistics students, there is the additional advantage of being able to work from the very beginning of their studies with the same kind of tools and resources that they will be using "for real" after graduating, in their professional life. The corpus search tool is finished (Figure 3), but the integration of a grammar writing environment still remains to be done. We have not yet been able to find a ready-made environment user friendly enough (for our linguistics students) and bug-free enough to be immediately useful for our purposes. Thus, it seems likely that we will have to put in some development effort in this area. If this turns out to be the case, the most likely kind of workbench that we will modify or build will be one within the general paradigm of unification-based feature structure grammar. There seems to be quite a few such systems available (Hammarström 2002).

The exercises will have hyperlinks to the "encyclopaedia", to the "resources" and to the annotated corpus of a foreign language (which, in turn, will be hyperlinked to the dictionary; see Saxena (2000a). As part of each theme, students will first discuss the construction during the lecture session, then again while examining the construction in the corpus and,



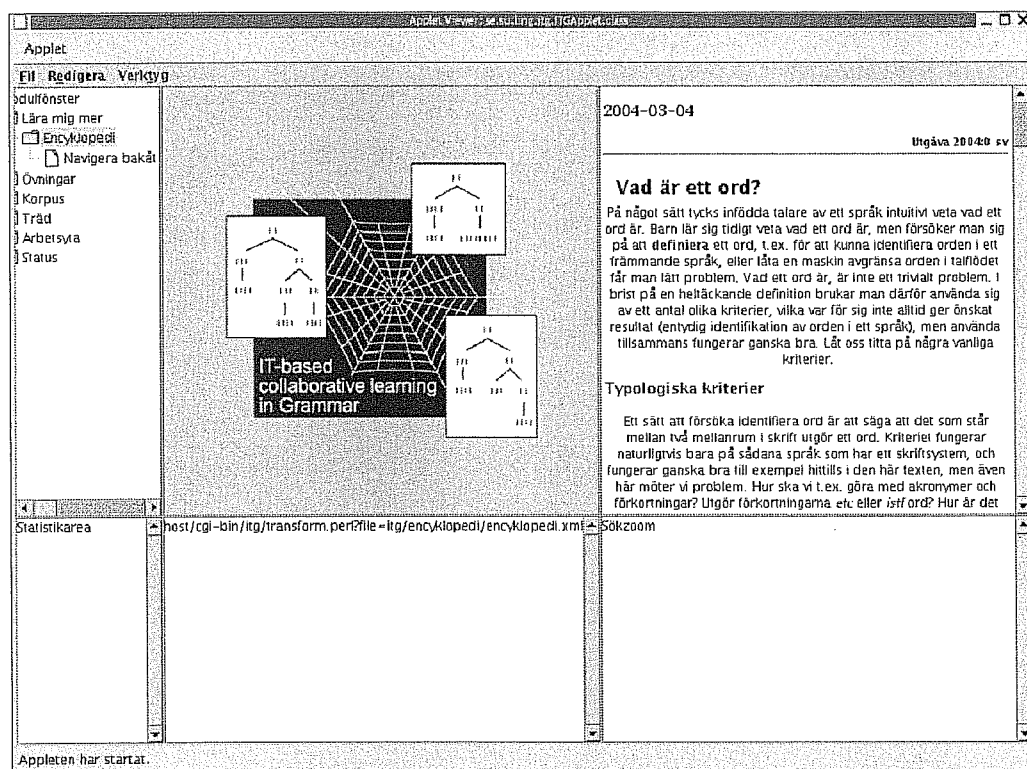


Figure 2: The encyclopaedia module

finally, also while comparing the results of the corpus-based analysis with the Swedish system and then discussing it in the group. This learning method, where the same construction is examined from a number of mutually reinforcing practical and theoretical viewpoints will, hopefully, provide the students with support and incentive in their learning process.

4. "Resource" modules will provide a pool of resources for further reading. This could be in terms of providing relevant links or in the form of "off-line references", such as references to articles in books and journals.

The architectural organization of the software described here has several advantages, the two most significant ones being *extensibility* and *conceptual decentralization*. "Extensibility" means that new functions can easily be integrated in the application. "Conceptual decentralization" is especially significant, as it allows the possibility of adjusting to individual learning styles. For example, if the student prefers to start out with the "encyclopaedia" material and go from there to the appropriate exercises, when she feels the need to do so, she has that choice. At the same time, the application allows the possibility of starting out at other entry points, e.g. "interactive exercises" with the option of calling up the relevant "encyclopaedia" material at each instant.

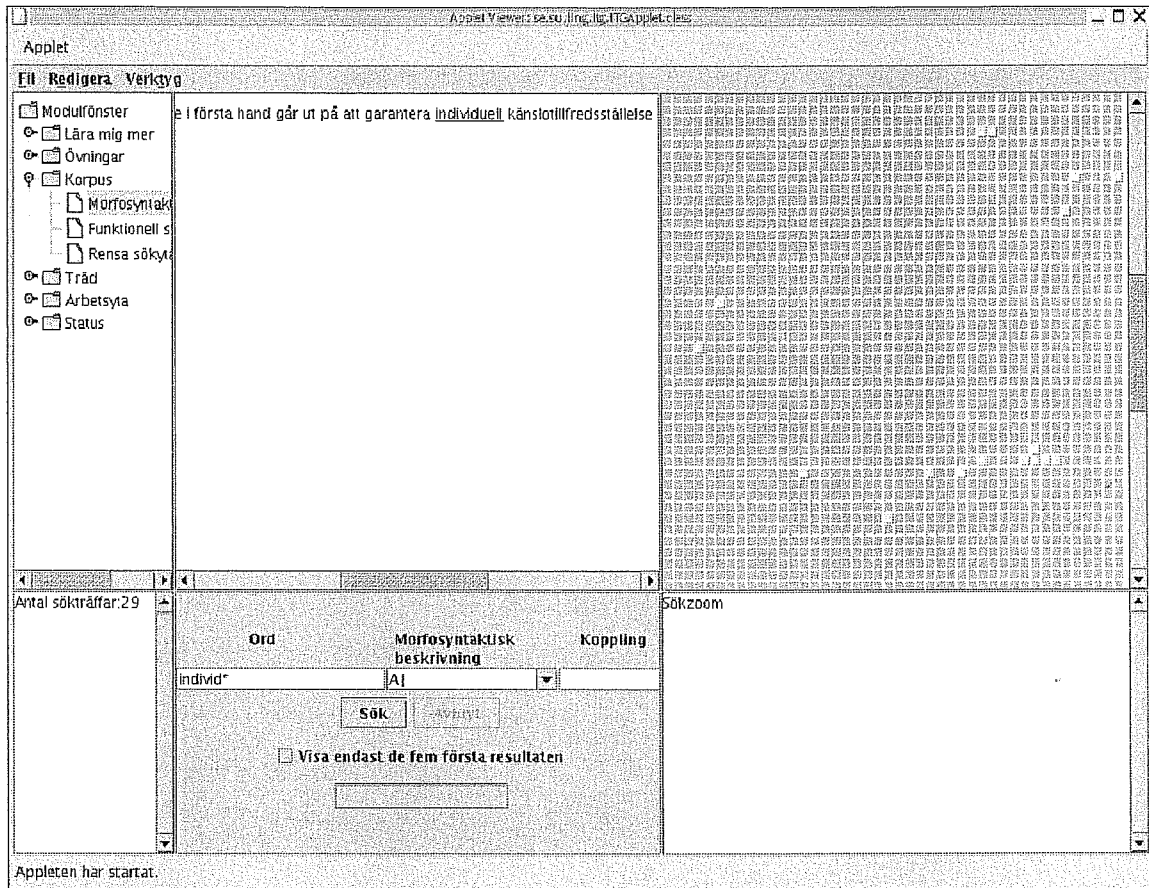
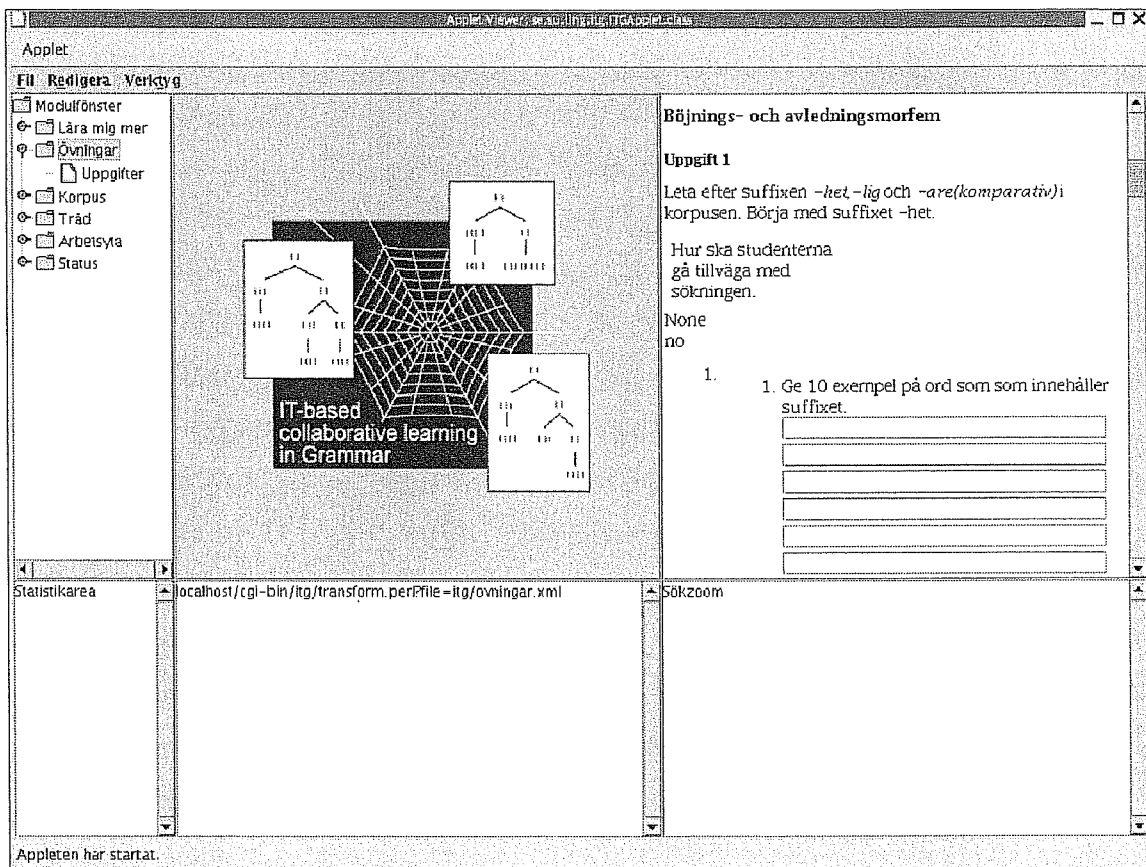
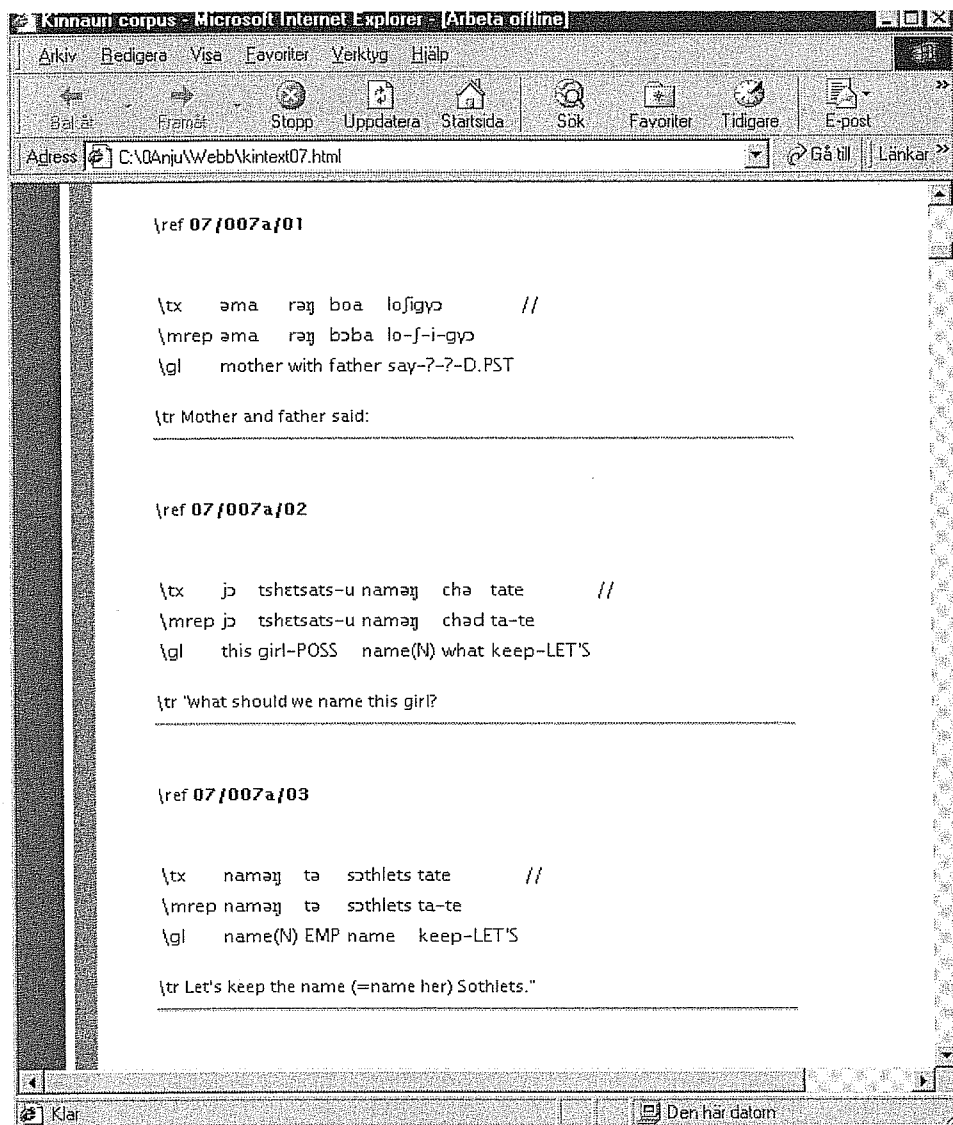
Figure 3: Text corpus module: search in *Talbanken* corpus

Figure 4: Interactive exercises



**Figure 5: Kinnauri narrative corpus**

The software is being realized using standard WWW and open source software, i.e. software which is generally free and where the source code is freely available and modifiable by the user, for implementing the modules. This design philosophy has the advantage of making the application maximally platform-independent, as well as providing a familiar interface – a standard web browser – for students and faculty. Access to Internet for both students and teachers is the only technical requirement for the collaborative learning method proposed here. Thus, most of the newly-written code is in Java, and the application is in the form of a Java applet. Corpora are converted into a uniform XML format, close to TEI,<sup>4</sup> whereas annotations are made standoff – i.e. reside in separate files – and stored in

<sup>4</sup> The *Text Encoding Initiative*, <http://www.tei-c.org> develops and maintains a kind of "de facto standard" for text corpora and similar text materials.

RDF format.<sup>5</sup> For storing and searching corpus data and annotations, we are using *eXist*, an open source XML database manager (Meier 2002, <http://exist.sourceforge.net>).

#### 4. Motivations for the Three Essential Components

We will now describe, very briefly, motivations for the three essential components, which comprise the collaborative learning method proposed here, namely the corpus-based training material, the collaborative learning, and the use of IT.

A text corpus consists of naturally occurring language in its natural linguistic context, as it is made up of complete texts or large text fragments as opposed to the made-up or isolated single sentences or phrases often used to illustrate grammatical points in linguistics textbooks. This accompanying linguistic context makes it possible to investigate the textual discourse-level functions of the grammatical phenomena. For example, the unmarked word order in Swedish is SVO, whereas in Kinnauri it is SOV. In certain contexts, however, the default word order is not maintained in Swedish and Kinnauri. Working with a corpus provides a means to examine the kinds of divergences that occur in Swedish and Kinnauri, and it also provides the basis to examine the context in which such divergences occur.

Advantages of collaborative learning have been mentioned and emphasized in a number of works. Johnson, Johnson & Smith (1991) argue that it increases the quality of education, as it can potentially bring together members of a group (students in this case), which differ in their perspectives as to how one should solve a particular problem, hence providing food for thought and discussion as to which of the possible "solutions" they should choose. Some researchers even claim that the essence of learning is in conversations (in a wide sense) among learners or groups of learners and that, consequently, knowledge, meaning, and beliefs are constructed only by collaboration (Gay & Lentini 1995).

---

<sup>5</sup> *Resource Description Framework* (RDF) is the emerging description language for realizing the vision of the "Semantic Web". RDF is being developed under the auspices of the World Wide Web Consortium; see <http://www.w3c.org/RDF/>.

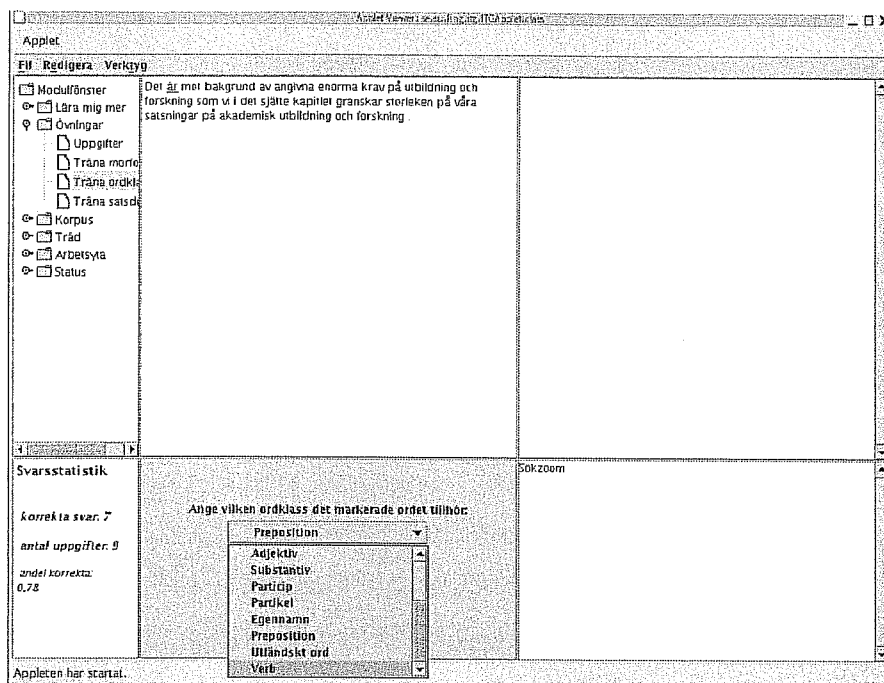


Figure 6: Part of speech exercises automatically generated from SUC

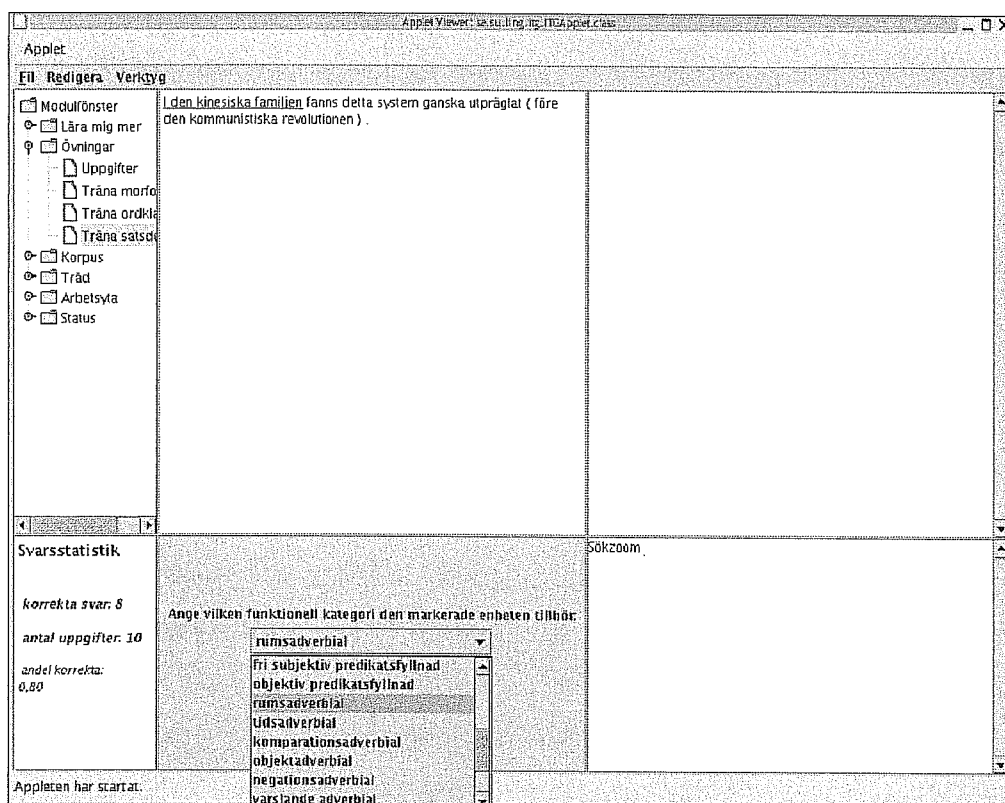


Figure 7: Functional syntax exercises automatically generated from Talbanken

"IT-based" has two distinct meanings in the context of this project. On the one hand, it refers to the computer tools for working with text corpora, which were discussed above. On the other hand, we also use the term to refer to a means of communication and collaboration. Both components are equally important to the project.

There have been a number of studies investigating and evaluating the use of *computer-mediated communication* (CMC) in university education. Many of these studies suggest that CMC shows advantage over the traditional classroom teaching. For example, Hiltz (1986) suggests that CMC in university education "facilitates discussion and debate". Shedletsky (1993 in Witmer 1998) shows that CMC provides students "opportunities for experiential learning". Even in a context where most interactions are face-to-face or by other more traditional means of communication, CMC adds something important to the learning process. McCarthy & Monk (1994) argue that additional means of communication (apart from classroom interaction) "are reassuring and psychologically important". Furthermore, there are indications that added communicative capabilities may influence how a problem is solved, even though the solution itself might not be different. As the problem-solving process itself is certainly important in an educational setting, the added communicative flexibility provided by CMC may be more than just "icing on the cake" in such a setting. (Gay & Lentini 1995)

CMC provides the means for collaboration extending beyond one physical location. This means that we will be able to test the collaborative method proposed in this project in the collaborating departments, while at the same time maintaining a flexibility to adjust to the needs of the individual departments.

CMC, furthermore, allows students and faculty to manage their time, since it comprises asynchronous as well as synchronous forms of communication. Both dialogues and group discussions can be conducted independent of place and time. (Acker 1995)

Practicing linguists have worked with text corpus material long before the computer entered the scene, but when it did, it brought about a type of slow revolution in the ways that text corpora could be compiled, stored, and processed, and in the ways that results could be presented. During the last few years, the computational tools for working with large text corpora have become very versatile and user friendly, so that now students can use the available corpora on the Internet, creating much more flexibility than if the same corpora are available in the form of books.

Thus, in the grammar training system discussed here, students will examine each grammatical theme by first discussing it during the lecture

session, next while examining the construction in Swedish, then at the higher levels while comparing the results of the corpus-based analysis of Swedish with the patterns found in the foreign language corpus and, finally, while discussing the construction in their groups. This learning method, where the same construction is examined from a number of mutually reinforcing practical and theoretical viewpoints will, hopefully, provide the students with support and incentive in their learning process.

The Internet provides a platform for an entirely new and flexible type of learning. In a traditional classroom teaching setup, the sequencing of exercises is normally linear. With the Internet, which provides the possibility of linking different sites, it is no longer mandatory that a fixed sequencing order – a "one-way street" – is maintained. The Internet provides flexibility regarding both the time and place for doing the exercises and for arriving at the problem-solving technique, which suits each of the individual users. This flexibility also allows the same grammar learning support system to be used by different departments for different (yet, related) purposes (for example, within regular grammar courses in linguistics and also as part of the curriculum for Swedish as a second language).

## 5. Taking Stock and Looking Ahead

The project has now reached a stage where much of the software infrastructure, linguistic resources, as well as the content of textbook and exercise modules are in place, and, consequently, can be tried out in real life and evaluated.

The road to the present point has not been without its twists and turns, however, and most of the complications that we have faced have been connected to *standardization* in some way. Thus, we have found, repeatedly throughout the project, that we cannot use language technology resources,<sup>6</sup> neither programs nor corpora, directly "off the shelf", as it were. Instead, we had to devote more effort to adaptation of existing resources and development of new software, than we had reckoned at the outset of the project. There has been a continual struggle with (lack of) compatibility

---

<sup>6</sup> Here, we use "language technology resources" as a cover term for both *language resources* and *processing resources* in the terminology proposed by Cunningham (2002).

and standardization of language technology resources. Some particular issues that have arisen in this connection are:

- Differences in fundamental storage and text markup formats. The three corpora that we are using in the project come in three quite different storage formats (see Saxena & Borin 2002 for details)
- Differences in POS tagging and syntactic annotations between corpora. The SUC and Talbanken Swedish corpora, although both are POS tagged, use different tag sets, with e.g. SUC having two and Talbanken four subclasses of nouns, and SUC, but not Talbanken, marking number in nouns, etc. Tag set incompatibilities, even within a language is a problem that has been noted in the literature (e.g. by Atwell et al. 2000), and there has been some work on tools for automatic tag set mapping (e.g. Teufel 1995). The problems are compounded when several languages are involved,<sup>7</sup> which would be desirable in our setting, where the linguistic subdisciplines of contrastive linguistics and language typology rely on explicit comparisons between languages at various linguistic levels. As stated above, we know from experience that students learn about grammatical constructions and phenomena more actively, when these constructions are discussed by comparing the system found in their native language with that of another language. Preferably, the other language should be one that the students do not know already, as they then will be better able to concentrate on the analysis of "pure" form. This is why we are using the Swedish and Kinnauri corpora together in our first application.
- Differences in POS categories, syntactic categories, and grammatical framework between the corpora on the one hand and the grammar writing tools and parsers on the other. Thus, the Talbanken corpus uses a fairly traditional Swedish functional grammatical framework, where, e.g. NPs are not directly recoverable, but only indirectly,

---

<sup>7</sup> The problem of cross linguistic mapping of POS tags has not been extensively discussed in the computational linguistics literature (see Borin 2000; 2002a & Borin & Prütz 2001), but in general linguistics, there is an extensive literature on the issue of cross linguistic properties of part of speech systems and the universality of proposed parts of speech, which is very relevant in this context (e.g., Anward, Moravcsik & Stassen 1996; Itkonen 2001; Pawley 1993).



through a combination of syntactic function and lexical category of the head word, while it seems that many, perhaps the majority, of the grammar writing tools freely available on the Web presuppose a phrase structure framework.

- Differences in implementation language, storage model, API, documentation, and source code availability, etc., of potentially suitable software. For an excellent overview of these issues, see Olsson (2002).

Thus, we have been forced from the outset to seriously discuss how we are to integrate existing NLP resources in our application, as well as how to make the application itself extensible, so that, e.g. new language corpora or new annotations can be added.

The state of language technology standardization is somewhat fluid at the moment. We have tried, in the project, to anticipate to the best of our ability, in which general direction the language technology community is headed, when it comes to issues of resource and annotation format standardization. The decisions that we have made in this respect, e.g. to use TEI structural markup and RDF annotations for the corpora, seem to reflect the general trend, but of course, we may turn out to have put our bets on the wrong horse in this particular case. Only time will tell. However, we remain firmly convinced that the only way, in which it will be possible to come up with combinations of language technology and computer-assisted language learning, such as that presented here, that can be viable in general and in the long run, is by actively using and promoting standards in both areas (Borin 2002b). In fact, this is a must, if we are to consider the needs of the many "low-density languages" in the world, for which such applications can be a much needed help in maintaining or even reviving the language and, certainly, will influence the status of the language in a positive direction (Saxena ms; Borin ms).

## References

- Acker, S. R. 1995. Space, collaboration, and the credible city: Academic work in the virtual university. *Journal of Computer-Mediated Communication* 1. no. 1. Online journal:  
<http://jcmc.huji.ac.il/vol1/issue1/acker/ACKTEXT.HTM>

- Anward, J. & Moravcsik E. & Stassen, L. 1996. Parts of speech: A challenge for typology. *Linguistic Typology* 1 (2). 167-183.
- Atwell, E. & Demetriou, G. & Hughes, J. & Schiffrin, A. & Souter, C. & Wilcock, S. 2000. Comparing linguistic annotation schemes for English corpora. Abeille, A. & Brants, T. & Uszkoreit, H. (eds). *Proc. of the Workshop on Linguistically Interpreted Corpora. LINC-2000*. 1-10. Held at the Centre Universitaire, Luxembourg, August 6, 2000.
- Borin, L. 2000. Enhancing tagging performance by combining knowledge sources. Byrman, G. & Lindquist, H. & Levin, M. (eds). *Korpusar i forskning och undervisning. Corpora in Research and Teaching*. ASLA, Växjö Universitet, Växjö: ASLA. 19-31.
- Borin, L. 2002a. Alignment and tagging. In *Parallel Corpora, Parallel Worlds*, Borin, L. (eds). 207-218. Amsterdam: Rodopi.
- Borin, L. 2002b. Where will the standards for intelligent computer-assisted language learning come from? *Proceedings of LREC 2002 workshop on International Standards of Terminology and Language Resource Management*. Las Palmas: ELRA. 61-68.
- Borin, L. ms. Lesser-known languages and language technology: An overview. *Globalization, technological advances, and lesser-known languages of South Asia*. Saxena A. (eds). Amsterdam: John Benjamins. Submitted manuscript.
- Borin, L. & Dahllöf, M. 1999. A corpus-based grammar tutor for education in language and speech technology. *EACL'99. Computer and Internet Supported Education in Language and Speech Technology. Proc. of a Workshop Sponsored by ELSNET and The Association for Computational Linguistics*. Bergen: University of Bergen. 36-43.
- Borin, L. & Prütz, K. 2001. Through a glass darkly: Part of speech distribution in original and translated text. *Computational Linguistics in the Netherlands 2000*. Daelemans, W. & Sima'an, K. & Veenstra, J. & Zavrel, J. (eds). 30-44. Amsterdam: Rodopi.
- Cunningham, H. 2002. GATE, a general architecture for text engineering. *Computers and the Humanities*. 36. 223-254.

- Einarsson, J. 1976a. Talbankens skriftspråkskonkordans. Corpus on CD-ROM.
- Einarsson, Jan. 1976b. Talbankens talspråkskonkordans. Corpus on CD-ROM.
- Ejerhed, E. & Källgren, G. 1997. Stockholm Umeå Corpus Version 1.0, SUC 1.0. Department of Linguistics, Umeå University.
- Flowerdew, John. 1996. Concordancing in language learning. *The Power of CALL*. Martha, C. P. (eds). 97-113. Houston, Texas: Athelstan.
- Gay, G. & Lentini, M. 1995. Use of communication resources in a networked collaborative design environment. *Journal of Computer-Mediated Communication* 1. no. 1. Online journal:  
[http://jcmc.huji.ac.il/vol1/issue1/IMG\\_JCMC/ResourceUse.html](http://jcmc.huji.ac.il/vol1/issue1/IMG_JCMC/ResourceUse.html).
- Ghadessy, M. & Henry, A. & Roseberry, R. L. 2001. Introduction. *Small corpus studies and ELT: Theory and practice*. Ghadessy, M. & Henry, A. & Roseberry, R. L. xvii – xxiii. Amsterdam: John Benjamins.
- Hammarström, Harald. 2002. Overview of IT-based tools for learning and training grammar. Project report, *IT-based Collaborative Learning in Grammar*. Department of Linguistics, Uppsala University.
- Hiltz, S. 1986. The "virtual classroom": Using computer-mediated communication for university teaching. *Journal of Communication* 36 (2). 95-104.
- Itkonen, E. 2001. Concerning the universality of the noun vs. verb distinction. *SKY Journal of Linguistics* 14. 75-86.
- Johnson, D. W. & Johnson, R. T. & Smith, K. 1991. Cooperative learning: Increasing college faculty instructional productivity. *ASYE-ERIC Higher Education Report (4)*. Clearinghouse on Higher Education. The George Washington University, Washington, D.C.
- Knowles, G. 1990. The use of spoken and written corpora in the teaching of language and linguistics. *Literary and Linguistic Computing* 5 (1). 45-48.
- McCarthy, J. C. & Monk, A. 1994. Channels, conversation, cooperation and relevance: All you wanted to know about communication but were afraid to ask. *Collaborative Computing*. 35-60.

- Meier, W. 2002. eXist: An open source native XML database. *Web, Web-Services, and Database Systems. NODe 2002 Web- and Database-Related Workshops*. Akmal, B. & Chaudri & Jeckle, M. & Rahm, E. & Unland, R. (eds). LNCS no. 2593. Berlin: Springer.
- Olsson, F. 2002. *Requirements and design considerations for an open and general architecture for information refinement*. Reports from Uppsala University, Department of Linguistics, RUUL no. 35. Uppsala University, Department of Linguistics.
- Pawley, A. 1993. A language which defies description by ordinary means. In *The Role of Theory in Language Description*. Foley, W. A. (eds). 87-129. Berlin: Mouton de Gruyter.
- Saxena, A. 1997. Nya undervisningsformer inom området grammatik. Redovisning av arbete med kvalitetsutveckling. *Paper presented at the Pedagogical workshop*. Faculty of Languages, Uppsala University.
- Saxena, A. 2000a. Corpora of lesser-known languages on the Internet: A pedagogical tool for the teaching of syntax. *Paper presented at the workshop on IT inom språkundervisningen. Uppsala University*. <http://www.lingfil.uu.se/personal/anjusaxena/symposium0303.html>.
- Saxena, A. 2000b. Developing a database on spoken language data. *Paper presented at IT and Linguistic diversity in Nordic and Baltic countries*. NorFA financed conference. Uppsala University.
- Saxena, A. (eds). ms. *Globalization, technological advances, and lesser-known languages of South Asia*. Amsterdam: John Benjamins. Under consideration for Studies in Language Companion series.
- Saxena, A. & Borin, L. 2002. Locating and reusing sundry NLP flotsam in an e-learning application. *LREC 2002. Workshop proceedings. Customizing knowledge in NLP applications – strategies, issues and evaluation*. Las Palmas. ELRA. 45-51.
- Shedletsky, L. 1993. Minding computer-mediated communication: CMC as experiential learning. *Educational Technology* 33 (12). 5-10.
- Teleman, U. 1974. *Manual för grammatisk beskrivning av talad och skriven svenska*. Lund: Liber.

Teufel, Simone. 1995. A support tool for tagset mapping. *Proceedings of SIGDAT 1995. Workshop in connection with EACL 95*. Dublin: Association for Computational Linguistics.

Witmer, D. F. 1998. Staying connected: A case study of distance learning for student interns. *Journal of Computer-Mediated Communication* 4. no. 2. Online journal:  
<http://jcmc.huji.ac.il/vol4/issue2/witmer.html>.

# **CALL for the Nordic Languages**

Tools and Methods for Computer  
Assisted Language Learning

*Edited by*  
Peter Juel Henriksen

Copenhagen Studies in Language 30

**Samfundslitteratur**

2004