

Using GETA'S MT/NLP Resources in an Intelligent Tutoring System for French

Chadia Moghrabi (*)

Département d'informatique
Université de Moncton
Moncton, NB, E1A 3E9,
Canada
[moghrac@umoncton.ca]

Christian Boitet

GETA, CLIPS-MAG, BP 53
Université Joseph Fourier & CNRS
38041 Grenoble Cedex 9
France
[Christian.Boitet@imag.fr]

(*) On sabbatical at GETA for 1997-1998.

Abstract

This paper presents a project that investigates to what extent computational linguistic methods and tools used at GETA for machine translation can be used to implement novel functionalities in intelligent computer assisted language learning. Our intelligent tutoring system project is still in its early phases. The learner module is based on an empirical study of French as used by Acadian elementary students living in New-Brunswick, Canada. Additionally, we are studying the state of the art of systems using Artificial Intelligence techniques as well as NLP resources and/or methodologies for teaching language, especially for bilinguals and minority groups.

Introduction

The project that we have started is intended for the minority French speaking Acadian community living in Atlantic Canada. In many families, parents used to go to English schools. Children, who now go to French schools, often switch back to English for their leisure activities because of the scarcity of options open to them. Many of these children use English syntax as well as borrowed vocabulary quite frequently. In brief, this setting of language learning is not that of a typical native speaker.

This paper presents the preliminary studies we have initiated to ensure an adequate pedagogical and technological approach to this particular language setting, and then continues to describe the use of existing linguistic resources while developing a quickly functional prototype system. Therefore, important questions have to be answered: Do computers help to enhance learning? What is the state of the art in Computer assisted language learning (CALL) and in Intelligent CALL (ICALL)? What specific language model should we be building? What system architecture should we be using? What help can we expect from NLP technology and resources?

We present our findings in the coming sections in the same order as our above interrogations. Later, we detail an empirical study that helped us define the learner model followed by the system's general architecture and an overview of some of its activities; particularly those that counteract anglicisms by double generating examples in standard French and in the local dialect.

To our knowledge, there are no systems that use machine translation tools for generating two versions of the same

language instead of multilingual generation. Another novelty is in the pedagogical approach of exposing the learner to the expert model and to the learner model in a comparative manner, thus helping to clarify the sources of error.

Enhanced Learning

The early computer assisted instruction (CAI) software packages were based on drill and practice. They had rigid environments, were not based on any learning theory, and they soon proved to be boring and to promote rote learning.

In an effort to solve these problems, and with the change in computer technology into personal computing, researchers looked for solutions in two new directions. Many of them, mostly educators and psycho-linguists, tried to implement more "pedagogical foundations" in their systems. Others, with the increasing popularity of expert systems, moved to implement more "intelligent" techniques. This latter approach was mostly adopted by AI researchers. It produced what is called Intelligent Tutoring Systems (ITS); in the specific domain of language what is known as Intelligent Computer Assisted Language Learning (ICALL).

These two approaches are not that far apart. ITS and ICALL researchers continue to solicit theoretical foundations for their systems and to submit them for evaluation by educators. Moreover, with the shift of learning paradigms into holistic, constructivist, and communicative approaches; some of the researchers are now working on a newer approach called Interactive/Adaptive Learning Environments (ILE / ALE) where the user has more freedom in the learning process.

Whatever the paradigm is, it is no more necessary to demonstrate the advantages and enhanced learning brought about by the use of computers for learning in general and for language learning in particular (Kulik & Kulik, 1991; Herman, 1994). The use of computers proves to increase motivation, to stimulate intellectual development, and to create positive attitudes towards learning.

In the case of language learning, the Cognition and Technology Group at Vanderbilt University (1996), the studies done by McKinnon, Nowlan & Sinclair (1996), by Jones (1994), by Riel (1990), and by Scardamalia, Bereiter & Lamon (1994) all show *demonstrable* language learning by computers. Students who participated in the

latter CSILE study, for example, scored systematically better on the Canadian Test of Basic Skills than the control group.

Intelligent Language Learning

Intelligent Tutoring Systems' architecture was more or less separated into four modules: an expert's model, a learner's model, a teacher's model, and an interface (Wengers, 1987). However, language learning had its own specific difficulties that were not generalized in other ITS systems. How to represent the linguistic knowledge in the expert and learner models? How to implement parsers that can process ungrammatical input? How to implement teaching strategies that are appropriate for language learning? These are some of the issues of high interest, Chanier, Renié & Fouqueré (1993).

Recent systems show how researchers are being more open to psycho-linguistic, pedagogical and applied linguistic theories. For example, the ICICLE Project is based on second language (L2) learning theory (1996); Alexia (1997) and FLUENT (1992) are based on constructivism, Mr. Collins (1995) is based on four empirical studies in an effort to "discover" student errors and their learning strategies.

Another tendency, that is very noticeably parallel to that of NLP, is the development of sophisticated language resources such as dictionaries for language (lexical) learning as exemplified by CÉLINE at Grenoble (1996), SAFRAN project (1997) and The Reader at Princeton University (1997) which uses WordNet, or real corpuses as in the European project Camille (1994).

The following section details an empirical study of local Acadian French of school age children. These results are necessary to clarify the threshold of proximity between standard French and Acadian French in creating our ITS system.

An Empirical Study

In an effort to gain some insight into the projected linguistic model, an empirical study on the population of elementary students in the City of Moncton, New Brunswick, Canada was completed. Fifteen students from every grade-level in the city schools have taken part in the study¹.

The study consisted of one-on-one interviews where the children were presented with images having very few possible interpretations :

- * A dog drinking water.
- * A dog playing with a stick.
- * A child reading a book.
- * A child writing with a pen.
- * A child kicking a ball.
- * A bird standing on a branch.
- * A man talking to a child who is sitting on a chair.
- * A woman dreaming about a house.

In the next sections, we will examine the children's answers according to three groups: les relatives en *qui* (subject), les relatives en *que* (direct object), and les relatives *complexes* (complex relative clauses).

Subject Relative Clauses

When the children were shown these pictures and were asked about the main *subject* in each of them, the answers were acceptable in standard French, showing that they had no problems in using relative clauses with *qui*. Following are some examples:

1. C'est une chienne qui boit;
2. C'est un chien qui boit du lait;
3. C'est un chien. Il boit.

It was sometimes noticed that (as in 3.) two sentences were used instead of one with *qui*, but not frequently enough to draw a conclusion about a bypassing strategy.

Some of the answers showed also other elements concerning lexical use:

5. C'est un garçon qui kick la balle. (Use of an English verb)
6. C'est une fille qui botte le ballon. (Use of an inappropriate verb)
7. C'est un papa et son garçon. (Bypassing strategy)

Direct Object Relative Clauses

In this part of the experiment, the object of the picture was the center of the questions. Following are some of the answers with the most frequent errors or bypassing strategies, they are marked with a *; the sentences with italics are the acceptable ones:

8. C'est le livre *que* le garçon lit.
- *9. C'est le livre qui se fait lire par la fille.
- *10. C'est le livre à la fille.
- *11. C'est le livre qu'elle lit dedans.
- *12. C'est un livre, la fille lit le livre.
- *13. C'est un ballon qui se fait botter par la fille.
- *14. C'est le lait qui se fait boire par le chien.
- *15. C'est le l'eau du chien.

The errors seen in these examples are more frequent in answers given by first grade children than those given by sixth graders. Answers 9, 12, 13, 14 and 15 are examples of bypassing strategies.

Answer 10 shows a very common use of the preposition *à* instead of *de*. Answer 11 is also representative of the frequent use of prepositions at the end of the sentence.

Complex Relative Clauses

The following examples give a survey of the use of indirect object relative clauses: *avec lequél / laquelle, sur lequél / laquelle, à qui*, and *dont*:

16. C'est le crayon *avec lequél* elle écrit. (It's the pen *with which* she is writing.)
- *17. C'est le crayon qui écrit.
- *18. C'est le crayon qu'il se sert pour faire ses devoirs.
- *19. C'est le crayon qu'elle écrit, la petite fille.
- *20. C'est le crayon qui aide à la fille à écrire.
21. C'est la branche *sur laquelle* est l'oiseau.
- *22. C'est une branche que l'oiseau chante sur.
- *23. C'est une branche que l'oiseau est assis.
24. C'est le garçon *à qui* le monsieur parle.
- *25. C'est le garçon qui s'assoit sur une chaise.
- *26. C'est le garçon que le monsieur parle.
27. C'est la maison *dont* la femme rêve.
- *28. C'est la maison que la dame rêve.

By looking at these examples, it is evident that complex relative clauses are rather unknown to the children.

¹ This work was done by A. S. Picolet-Crépault within her PhD thesis.

Elements of Learner Model

Generally speaking, one can conclude that students knew the individual concrete concepts, what they meant, and mostly where to use them.

It can also be concluded that they use *que* in a non standard manner every time they need to use complex relative clauses as seen in most of the above examples above. Otherwise they use a bypassing strategy by separating the sentence into two parts as in "C'est une branche et un oiseau", or by using another verb that allows *qui* as in 25.

In developing modules that answer the linguistic and/or lexical apprenticeship needs of this local population, it is important to allow the learning of proper word usage. Among the most important are words that allow the construction of more complex sentences, such as relative clauses as we discussed in this section.

General System Architecture

The system we are building has a mixed initiative, multi-agent architecture. Mixed initiative because we want the system to serve both the teacher and the student, in both teaching and in learning modes. The student is able to ask for a specific information or activity and the teacher is able to request special explanations and exercises in order to help him/her in lesson preparation. For example, the teacher could favor certain activities such as presenting examples of "non standard French sentences" and opposing them to English structures in a effort to show the children some anglicisms; or maybe choose a specific micro-world, such as Halloween or Christmas so that the exercises would be closer to children's real daily experience (principle P1).

The general approach to linguistic knowledge is that of traditional NLP techniques where each module specializes in a particular activity. Modules are coordinated by a main controller. A total of four main agents are used:

1. Morphology agent,
2. Lexical agent,
3. Syntax agent,
4. Error-specialized agent.

Each of these agents has a number of sub-agents that help in fine-tuning its work. For example, the morphology agent has a sub-agent for analysis and another for generation. The lexical agent can solicit some help from a micro-world specialized sub-agent. The syntactic graph is annotated with probabilities on usually faulty expressions in order to intensify the explanation or the number of examples and exercises on those particular parts.

We do not intend to build a fully free learning environment. The environment however is not fully structured either. The user chooses where to start by clicking on a hot button-picture on the screen and he/she chooses when to stop. He/she has the choice of the micro-domain and of the type of activities requested. However, unexpected "pop-up" activities would come up on the screen from time to time (style "Tip of the day" or "TV ad.").

As this system is being built for young children, not every single word is expected to be typed on the keyboard. Following are some examples of the look and feel of our system:

1. Children can pick activities from graphical images on the screen. Clicking on a pumpkin would favor Halloween micro-domain; or on a bunny would bring up Easter stories.
2. Corpuses or extracts from children stories are equipped with hyper links to word meanings or grammar usage explanations that the children can explore.
3. Puzzle playing where words have assigned shapes according to their functions. Fitting the puzzle means placing the words in the correct order.
4. Picking words they like and asking the system to make up a sentence; or give it a sentence and ask to "translate" it into Standard French or into local French.
5. Activities for the teacher include interactive creation of specific language structures in an interlingua formalism (such as in UNL) then automatic generation of parallel examples and exercises for his/her own use or for lectures (some teachers might have lived the same minority setting as the students).

We argue that it is better to show the learner many "good" examples of words and texts than it is to concentrate on how wrong their own sentences are.

All the above possibilities are optional. They can be deactivated, so the system would have simple menus where the main controller takes over. This allows the teacher to take responsibility of the degree of unstructured or of focused learning.

Using GETA's Resources

Building language software is not a quick process. For the last few years, the NLP community has been putting a lot of work in language resources and tools with the hope of concentrating the efforts and reducing development times.

For many years GETA has been working on MT systems from and into French. An impressive core of linguistic knowledge is available but has not yet been experimented on in building language learning software, though work is underway for integration of heterogeneous NLP components (Boitet & Seligman, 1994). Ariane for example, uses special purpose rule-writing formalisms for each of its morphological and lexical modules both for analysis and for generation, with a strict separation of algorithmic and linguistic knowledge (Hutchins & Somers, 1992). In light of the general architecture we presented above, the following GETA modules are appropriate and are being implemented for use within our system:

A. Morphological agent.

- ATEF for the morphological analysis sub-agent
- SYGMOR for the morphological generation sub-agent

B. Lexical agent.

- EXPANSF for lexical expansion
- TRANSF for translation into standard French

C. ROBRA in its multilevel analysis sub-agents

- for syntactic tree definitions and manipulations
- for logico-semantic functions

The main unavailable resource for our application is the user friendly interface for this multimedia application.

Conclusion

We have presented in this paper an ongoing software development project that is still in its early phases. In the introduction and in the first sections, we have argued for the positive effects of computers on language learning and then on some of the issues that researchers in the field are hoping to see implemented from a computational and a pedagogical point of view. We have also seen, through an empirical study, the kinds of linguistic difficulties that a minority group is encountering. In such a case one cannot help but to think about the advantages that technology can offer, especially in an era where Language resources are ready for the pick. We have opted to examine the highly formalized and parameterized resources at GETA in an effort to develop a quickly functional prototype that we can immediately submit for on-the ground testing.

Acknowledgements

Our thanks go to the Canadian Language Technology Institute CLTI, Université de Moncton and to TPS Moncton for partially financing this project.

References

- Boitet, C. & Seligman, M. (1994) The 'WhiteBoard' Architecture: a way to integrate heterogeneous components of NLP systems, *Proceedings. Coling 94*, Kyoto, 1994.
- Brown, J. S. & Burton, R.R. (1978) Diagnostic models for procedural bugs in basic mathematical skills, *Cognitive Science*, 2, pp. 155-191.
- Bull, P., Pain, H. & Brna, P. (1995) Mr. Collins: Student Modeling in Intelligent Computer Assisted Language Learning, *Instructional Science*, 23, pp.65-87.
- Burton, R. R. & Brown, J.S. (1976) A tutoring and student modeling paradigm for gaming environments . Computer Science and Education, *ACM SIGCSE Bulletin*, 8/1, pp. 236-246.
- Carbonell, J. (1970) AI in CAI: An artificial intelligence approach to computer-assisted instruction. *IEEE Transactions on Man-Machine Systems*, 11 /4, pp. 190-202.
- Chanier, T., Renié, D. & Fouqueré, C. (Eds.) (1993) Sciences Cognitives, Informatique et Apprentissage des Langues . In *Proceedings of the workshop SCIAL '93*.
- Chanier, T. (1994) *Special Issue Introduction*, JAI-ED, 5/4, pp. 417-428
- Cognition and Technology Group at Vanderbilt (1996) From Visual Word Problems to Learning Communities: Changing Conceptions of Cognitive Research. In *Classroom Lessons: Integrating Cognitive Theory and Classroom Practice*, K. McGilly ed., MIT Press, Cambridge, MA, pp. 157-200.
- Hamburger, H.& Hashim, R.(1992) Foreign Language Tutoring and Learning Environment, In *Intelligent Tutoring Systems for Foreign Language Learning*, Swarts & Yazdani, eds., Springer-Verlag.
- Herman, J. L. (1994) Evaluating the Effects of Technology in School Reform . In *Technology and Education Reform. The Reality Behind the Promise*, B. Means ed., Jossey-Bass, San Francisco, CA, pp. 133-167.
- Hutchins, W.J. & Somers, H.L. (1992) *An Introduction to Machine Translation*, Academic Press, San Diego, CA, 361 p.
- Ingraham, B., Chanier T. & Emery, C. (1994) CAMILLE: A European Project to Develop Language Training for Different Purposes, in *Various Languages on a Common Hypermedia Framework*, *Computers and Education*, 23/1&2, pp.107-115.
- Jones, I. (1994) The Effect of a Word Processor On the Written Composition of Second-Grade Pupils. *Computers in the Schools*, 11/2, pp. 43-54.
- Kulik, C. C. & Kulik, J. A. (1991) Effectiveness of Computer-based Instruction: An Updated Analysis. *Computers in Human Behavior*, 7, pp. 75-94.
- McCoy, K.F., Pennington, C.A., & Suri, L.Z. (1996) English Error Correction: A Syntactic User Model Based on Principled "mal-rule" Scoring *Proc. Fifth International Conference on User Modeling*. Kailua, Hawaii, pp. 59-66.
- McKinnon, D. H., Nolan, C. J. P. & Sinclair, K. E. (1996) The Freyberg Integrated Studies Project in New Zealand: A Longitudinal Study of Secondary Students' Attitudes Towards Computers, Their Motivation and Performance *Proc. of the 13th Int. Conf. on Technology in Education*, New Orleans, pp. 463-465.
- Menézo, J., Genthial, D. & Courtin, J. (1996) Reconnaissances pluri-lexicales dans CÉLINE, un système multi-agents de détection et correction des erreurs, *Proc. "Le traitement automatique des langues et ses applications industrielles TAL+AI'96"*, 2, Moncton, Canada.
- Moghrabi, C. & de Finney, J. (1989) PARDA: Un Programme d'Aide à la Rédaction du Discours Argumenté, *Journal Canadien des Sciences de l'Information*, 3/4, pp. 103-109.
- Picolet-Crépault, A.S. (1996) Stratégies de remplacement et de contournement chez l'enfant de 6 à 12 ans, In *"Revue de 10ièmes journées de linguistique de l'Univ. Lava"*, Québec, Canada.
- Riel, M. (1990) Computer -Mediated Communication: A Tool for Reconnecting Kids with Society. *Interactive Learning Environment*, 1/4, pp. 255-263.
- SAFRAN Project (1997) URL address: <http://admin.ccl.umist.ac.uk/staff/mariejo/safran.htm>
- Scardamalia, M., Bereiter, C. & Lamon, M. (1994) The CSILE Project: Trying to Bring the Classroom into World 3. In *"Classroom Lessons: Integrating Cognitive Theory and Classroom Practice"*, K. McGilly ed, MIT Press, Cambridge, MA, pp. 201-228.
- Selva, T., Issac, F., Chanier, T., Fouqueré, C. (1997) Lexical Comprehension and Production in the ALEXIA System, *Proc. Language Teaching and Language Technology*, Univ. of Groningen.
- Swartz, M.L. & Yazdani, M. (eds.) (1992) *Intelligent Tutoring Systems for Foreign Language Learning: The Bridge to International Communication*., NATO Series, Springer-Verlag, 1992.
- The Reader, [http:// www. cogsci. princeton.edu/~wn/current/reader.html](http://www.cogsci.princeton.edu/~wn/current/reader.html)
- Wengers, E. (1987) *Artificial Intelligence and Tutoring Systems* . Morgan Kaufmann, Los Altos, CA.