

Using Language Resources in an Intelligent Tutoring System for French

Chadia Moghrabi (*)
Département d'informatique
Université de Moncton
Moncton, NB,
E1A 3E9, Canada
moghrac@umoncton.ca

(*) On sabbatical leave at GETA-CLIPS, Grenoble, France 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 bilingual 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 and sometimes cannot adequately help their children in their school work. 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.

We begin our presentation with a literature review of related work in Intelligent Tutoring Systems (ITS) particularly on Computer Assisted Language Learning (CALL and Intelligent CALL) followed by the principles that this community is now expecting from system builders. In the following sections we summarize an empirical study that helped us

define the learner model. Then, in the last section we propose the system's general architecture and an overview some of its activities; particularly those that counteract Anglicisms by double generating examples in standard French and in the local dialect using linguistic resources usually used in machine translation.

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.

1 Artificial Intelligence and Language Learning

Among the first milestones in Intelligent Tutoring Systems (ITS) was Carbonell's system (1970) that used a knowledge-base to check the student's answers and to allow him/her to interact in "natural language". BUGGY, by Brown and Burton (1978) is another system more oriented towards student error diagnostic. At around the same period researchers were starting to put also some emphasis on the teaching strategies adopted in the system such as in WEST, Burton & Brown (1976).

It's with such works and many others later, that 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 L2 learning theory (McCoy et al., 1996); Alexia (Selva et al., 1997) and FLUENT (Hamburger and Hashim, 1992) are based on constructivism, Mr. Collins (Bull et al., 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 (Menézo et al., 1996), the SAFRAN project (1997) and The Reader at Princeton University (1997) which uses WordNet, or real corpuses as in the European project Camille (Ingraham et al., 1994).

The literature review lead us to believe in the following basic principles:

- P1. Language is learned in context through communication and experience, Chanier (1994).
- P2. Language is learned in the natural order from receptive to productive.
- P3. Grammatical forms ought to be taught through language patterns.
- P4. Vocabulary learning means learning the words and their limitations, probability of occurrences, and syntactic behavior around them, Swartz & Yazdani (1992).

2 An Empirical Study for Learner Model

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 completed¹. The study consisted of one-on-one interviews where the children were presented with images having very few possible

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

interpretations. The only question that was asked was "Qu'est-ce que c'est?" (What is this?).

In the next sections, we will examine the children's answers concerning relative clauses.

2.1 Subject Relative Clauses

When the children were asked about the main subject in the picture, 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;

Some of the answers showed other elements concerning lexical use:

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

2.2 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:

6. C'est le livre *que* le garçon lit.
- *7. C'est le livre qui se fait lire par la fille.
- *8. C'est le livre à la fille.
- *9. C'est le livre qu'elle lit dedans.
- *10. C'est un livre, la fille lit le livre.

The errors seen in these examples constitute around fifty percent of the answers given by first grade children and are reduced to around thirty percent in sixth grade. Answers 7 and 10 are examples of bypassing strategies i.e.; the use of a different verb or another sentence structure as a means for avoiding relative clauses.

Answer 8 shows a common use of the preposition *à* instead of *de*. Answer 9 is also representative of the frequent use of prepositions at the end of the sentence.

2.3 Complex Relative Clauses

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

11. C'est le crayon *avec lequel* elle écrit.
- *12. C'est le crayon qui écrit.
- *13. C'est le crayon qu'il se sert pour ses devoirs.

14. C'est la branche *sur laquelle* est l'oiseau
 *15. C'est une branche que l'oiseau chante sur.
 *16. C'est une branche que l'oiseau est assis.
17. C'est le garçon *à qui* le monsieur parle.
 *18. C'est le garçon qui s'assoit sur une chaise.
 *19. C'est le garçon que le monsieur parle.
20. C'est la maison *dont* la femme rêve.
 *21. C'est la maison que la dame rêve.
 *22. C'est la maison que la madame rêve de.

2.4 Error Summary

By looking at these examples, it is evident that complex relative clauses are rather unknown to the children. They show that the easiest particles for them are *qui* and *que* even when misused as in answer 12.

It can also be concluded that they use *que* in a non standard manner every time they need to use complex relative clauses. 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 18.

3 General System Overview

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. 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 syntactic graph and the lexicon are annotated with probabilities on usually faulty expressions in order to intensify the explanation or the number of examples and exercises on those particular parts (principles P3 and P4).

We do not intend to build a fully free learning environment. The environment is partially structured. The user chooses where to start by clicking on a hot-button picture. He/she chooses the micro-domain and the wanted activities. 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.
2. Corpus or extracts from children stories are equipped with hyperlinks to word meanings or grammar usage explanations.
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;

All the above possibilities are optional. This allows the teacher to take responsibility of the degree of unstructured or of focused learning.

4 GETA's Used Resources

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).

The following modules from GETA were used in our experiment² :

- 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 multi-level analysis
 - for syntactic tree definitions and manipulations
 - for logico-semantic functions

² This work was done by Anne Sarti within her Master's degree.

The first series of experiments we realized using GETA's resources concentrate on double analysis/generation of standard French and non-standard local French. The corpus consisted of the sentences collected during the empirical study (see section 2).

Figures 1 and 2 show an example of the annotated trees created by Ariane during this

double generation of Acadian French and Standard French.

These two graphs show how straight forward was the use of language resources for highlighting similarities and/or differences in these two dialects. The same grammar can be used by incrementing its rules to include new/different sentence structures. The lexicon can be augmented similarly.

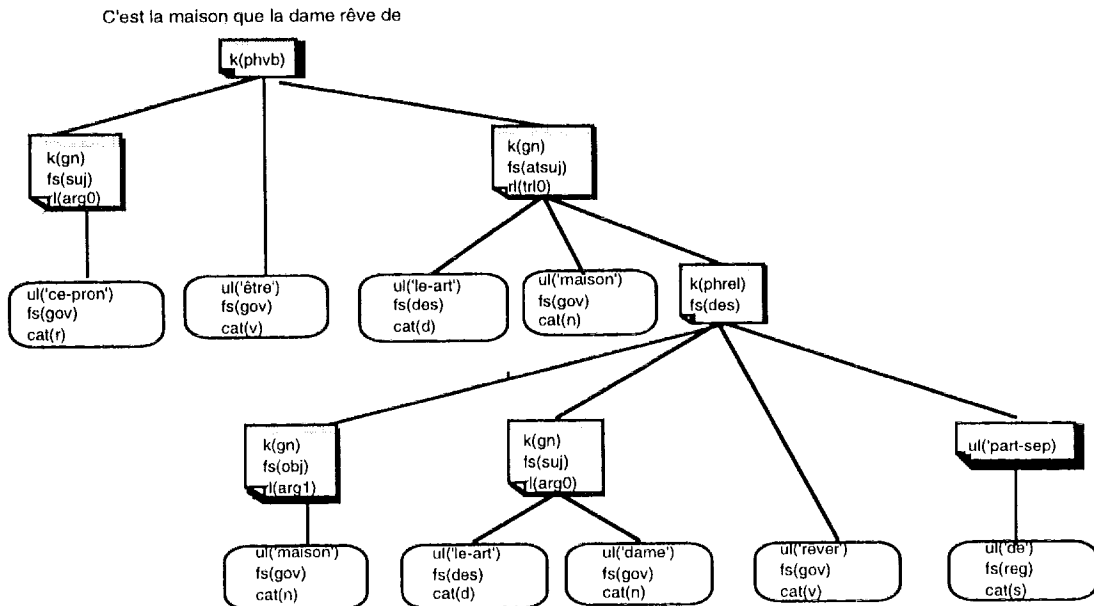


Figure 1: Annotated tree for a sentence in non-standard French.

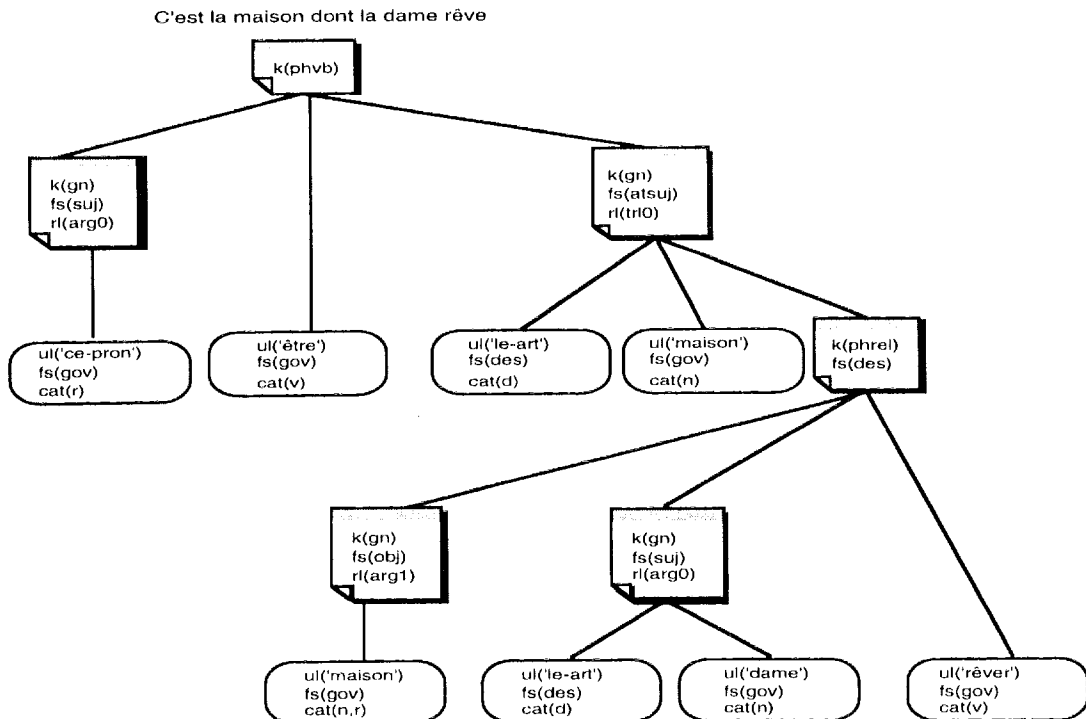


Figure 2: Annotated tree for a sentence in standard French.

Another alternative would be to consider the non-standard French as a completely new language from all points of view. In this case only the formalisms at GETA would be exploited not the existing linguistic data.

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 use 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*, Proc. 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
- Hamburger, H. & Hashim, R. (1992) *Foreign Language Tutoring and Learning Environment*, In "Intelligent Tutoring Systems for Foreign Language Learning, Swartz & Yazdani, eds., Springer-Verlag.
- Holland, V.M., Kaplan, J.D., & Sams, M.R. (eds.) (1995) *Intelligent Language Tutors, Theory Shaping Technology*, Lawrence Erlbaum Associates, Mahwah, N.J., 384 p.
- 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.
- 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.
- 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. Laval, Quebec, Canada.
- SAFRAN Project (1997) <http://admin.ccl.umist.ac.uk/staff/mariejo/safran.htm>
- 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>
- Wengers, E. (1987) *Artificial Intelligence and Tutoring Systems*. Morgan Kaufmann, Los Altos, CA.