DESIGNING INTELLIGENT LANGUAGE TUTORING SYSTEMS FOR INTEGRATION INTO FOREIGN LANGUAGE INSTRUCTION

DISSERTATION

Presented in Partial Fulfillment of the Requirements for the Degree Doctor of Philosophy in the Graduate School of The Ohio State University

By

Luiz Alexandre Mattos do Amaral, B.A., M.A.

* * * * *

The Ohio State University

2007

Dissertation Committee:
Professor W. Detmar Meurers, Co-Adviser
Professor John Grinstead, Co-Adviser
Professor Donna Long
Professor Gláucia Silva

Approved by

Co-Adviser

Co-Adviser

Graduate Program in Spanish and Portuguese
ABSTRACT

Intelligent Computer-Assisted Language Learning (ICALL) is a multidisciplinary area of research that combines Natural Language Processing (NLP), Intelligent Tutoring System development, Second Language Acquisition (SLA) and Foreign Language Teaching and Learning. So far, most of the work done in ICALL has primarily focused on the development of NLP technology for error diagnosis, and very few systems have been fully implemented to the point where they could be used in an existing foreign language program.

The work presented here proposes to develop an ICALL system focusing primarily on the needs of foreign language students and instructors. The research project started with a survey with foreign language instructors on how ICALL could support their everyday practice. The survey was followed by an analysis of the capabilities of NLP technology, and a study on how some of the NLP tools could be used to produce a system that presented activities which could be incorporated into actual language programs. The final step was to develop a system that provides intelligent feedback following the pedagogical principles outlined.

The specific context of the study was the Portuguese Individualized Instruction Program (IIP) at the Ohio State University. The main research contribution of the project is to show one way NLP technology can be used to cope with the real needs of
language learners following precise pedagogical specifications. The concrete contribution of this research is an intelligent electronic workbook that is currently being used by IIP students. The overall contribution of the project is to take a concrete step in the direction of bridging the gap between the development of NLP technology for ICALL and the actual use of such technology in real life foreign language programs.
para Clara, Flávia e Ione
ACKNOWLEDGMENTS

I want to thank:

Detmar Meurers for everything he has done as my adviser in the last five years. Anything I can say here will not be enough to thank him for his support, guidance, and friendship during this entire process. From the moment I wrote him my first e-mail message from Brazil to the day of this defense, Detmar has always encouraged and helped me, providing all the necessary support for me to succeed. It was during his ICALL seminar several years ago that I first proposed the idea of bringing pedagogical concerns to the foreground of ICALL system development. Detmar not only supported my initial thoughts, but he also decided it was worth pursuing such research project. TAGARELA is the result of great years of collaboration and learning, and I am sure many more wonderful years are ahead of us.

John Grinstead, my co-adviser in the Department of Spanish and Portuguese, for his support and helpful feedback on language teaching methodologies and foreign language instruction in several chapters of this dissertation.

Gláucia Silva for all her contributions to this project. Gláucia has always supported my work since my arrival at OSU. She is a Co-PI in the TAGARELA project, and she provided many insightful comments about exercise development, and integration of the system into the Individualized Instruction methodology. I am grateful
for having the opportunity to work with her in the Portuguese program at OSU, and for our many research projects together. I am even more grateful for having her as my friend.

Donna Long for her thorough comments on my dissertation chapters, and on the papers I wrote for her courses. Donna played a very important role in my graduate studies. After many years as a language teacher had hidden from my eyes the beauty of the language teaching process, Donna’s inspiring classes made me reconcile myself with foreign language pedagogy.

The Office of Research and the College of Arts and Science at The Ohio State University for funding this research through an Innovation Grant.

Ricardo Hasegawa, Maria das Graças Volpe Nunes, Sandra Maria Aluísio and all other members of the Núcleo Institucional de Lingüística Computacional (NILC) for their hospitality during my visit in 2005, and for providing selected lexical entries of CURUPIRA, which constitute the base of the TAGARELA lexicon. I also want to thank Dr. Gordon P.K. Chu memorial fellowship for sponsoring my trip to visit NILC.

Stacey Baily for her constant support and feedback, and for sharing her brilliant ideas with me. And the ICALL research group at OSU for many insightful discussions.

Cory Shain, Dickie Self, Jody Croley, Chris Kovach, and Scott King for their help with the design of the web pages.

Mat Schulze, Trude Heift, Terence Odlin, Rebecca Sachs and Alison Mackey for all their comments, suggestions, and bibliographical references that were very useful for several papers and presentations made during these years.
My friends and colleagues in Spanish and Portuguese, and Linguistics for their support and numerous comments during many dry-runs of conference presentations. Especially, Amanda Boomershine, Anna Feldman, Jirka Hana, Chad Howe, Vanessa Metcalf, Adriane Boyd, Illana Bromberg, Tom Stovicek, and Gilberto Velazquez.

My professors at OSU that directly contributed to my graduate studies: Chris Brew, David Dowty, Javier Gutierrez-Rexach, Robert Levine, Terrell Morgan, Carl Pollard, Wayne Redenbarger, Craige Roberts, Scott Schwenter, and Dieter Wanner.

My friends, who are or were in Columbus-OH, for their support during all those years: Lúcia and Tim Costigan, Will Boomershine, George Bondarchuk, David Neruda, and Araceli Schwenter.

My parents José Carlos do Amaral and Ione Mattos, and my brother and sisters Fernando, Glauce e Gabriela for their constant support and unconditional love.

My family and friends who have always been on my side: Ana, Nazira, José Luiz, Maria Aparecida, Marcelo, Mimi, Talita, Tatiana, Luiz Antônio, Yarinha, Angela, Cláudio, Ilana, Márcia, Dodora, Claudette, Serge, Annie, Patrick, Nicolas, Robert e Jean.

Clara for reminding me of the important things in life.

Flávia for always being by my side in different parts of the world for more than 15 years, for giving me all the love and support I needed to accomplish anything, and for making all of this make sense.
VITA

February 17, 1973  Born – Rio de Janeiro, Brazil  

1997  BA, Letters – Portuguese - French,  
        State University of Rio de Janeiro (UERJ)  

2001  MA, Linguistics, Pontifical Catholic University of Rio de Janeiro (PUC-Rio)  

2002-2007  Graduate Research and Teaching Associate, The Ohio State University  

FIELDS OF STUDY  

Major Field: Spanish and Portuguese
TABLE OF CONTENTS

Abstract .................................................................................. ii
Dedication ................................................................................. iv
Acknowledgments ....................................................................... v
Vita .......................................................................................... viii
List of Figures ............................................................................ xiii

Chapters:

1. Introduction and Motivation ................................................. 1
   1.1 CALL as a Discipline .................................................. 1
   1.2 Research in CALL ....................................................... 7
   1.3 CALL and ICALL ......................................................... 11
   1.4 Intelligent Language Tutoring Systems ......................... 15
       1.4.1 Domain Knowledge ............................................. 17
       1.4.2 Student Model .................................................... 19
       1.4.3 Instruction Model ............................................... 20
   1.5 ICALL Research and ICALL Integration ...................... 20
   1.6 Proposed Approach .................................................... 25

2. Foreign Language Instruction and SLA Research for ILTS Development . 28
   2.1 Task-Based Instruction ................................................. 31
   2.2 Focus on Form .......................................................... 36
   2.3 Awareness and Explicit Learning .................................. 38
   2.4 Error Correction ....................................................... 42
   2.5 ILTS Integration and Development .............................. 46
3. Pedagogical Background of TAGARELA

3.1 The System
3.2 Portuguese Individualized Instruction
3.3 FLTL Integration
  3.3.1 Why ask teachers?
  3.3.2 Instructional setting
  3.3.3 The Interview
  3.3.4 Results
3.4 Types of Exercises
  3.4.1 Processing Considerations
  3.4.2 Dealing with Activity Specification Issues
  3.4.3 Design Choices for TAGARELA
3.5 Conclusion

4. The TAGARELA System

4.1 NLP in ICALL – Related Approaches
  4.1.1 Intelligent Tutor of Academic English
  4.1.2 ALICE-Chan
  4.1.3 E-Tutor
4.2 Nature of Processing
4.3 Overview of the Architecture
4.4 Instruction Model
  4.4.1 Activity Model
  4.4.2 Error Model
4.5 Interface
4.6 Analysis Manager
4.7 NLP Modules
  4.7.1 Tokenizer
  4.7.2 Spell-Checker
  4.7.3 Lexicon
  4.7.4 Syntactic Processing
  4.7.5 Shallow Semantic Processing
4.8 Feedback Manager
4.9 Feedback Strategies

5. Student Model

5.1 Current Student Models in ICALL
  5.1.1 E-Tutor
# LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>CALL and related disciplines (Levy, 1997, p. 72)</td>
<td>9</td>
</tr>
<tr>
<td>3.1</td>
<td>Listening Comprehension Exercise</td>
<td>74</td>
</tr>
<tr>
<td>3.2</td>
<td>Reading Comprehension Exercise</td>
<td>76</td>
</tr>
<tr>
<td>3.3</td>
<td>Description Exercise</td>
<td>77</td>
</tr>
<tr>
<td>3.4</td>
<td>Vocabulary Exercise</td>
<td>78</td>
</tr>
<tr>
<td>3.5</td>
<td>Rephrasing Exercise</td>
<td>80</td>
</tr>
<tr>
<td>3.6</td>
<td>Fill in the Blanks Exercise</td>
<td>81</td>
</tr>
<tr>
<td>4.1</td>
<td>E-Tutor Error Detection (Heift, 2001, p. 4)</td>
<td>89</td>
</tr>
<tr>
<td>4.2</td>
<td>TAGARELA's Architecture</td>
<td>95</td>
</tr>
<tr>
<td>4.3</td>
<td>The Language Exercise Spectrum</td>
<td>118</td>
</tr>
<tr>
<td>4.4</td>
<td>Feedback Message – Word Choice</td>
<td>122</td>
</tr>
<tr>
<td>4.5</td>
<td>Feedback Message – Agreement</td>
<td>123</td>
</tr>
<tr>
<td>4.6</td>
<td>Feedback Message – Wrong Word</td>
<td>125</td>
</tr>
<tr>
<td>4.7</td>
<td>Feedback Message – Missing Verb</td>
<td>126</td>
</tr>
<tr>
<td>5.1</td>
<td>TAGARELA’s Student Model</td>
<td>139</td>
</tr>
</tbody>
</table>
5.2 Architecture with Student Model ........................................ 142
6.1 Students’ Opinions ...................................................... 148
6.2 Accentuation Error – Verbs ............................................. 151
6.3 Accentuation Error – Nouns ............................................. 152
CHAPTER 1

INTRODUCTION AND MOTIVATION

1.1 CALL as a Discipline

Computer-Assisted Language Learning (CALL) is a multidisciplinary area of research that encompasses the study of computer applications in language teaching and learning. The discipline is usually viewed as a subfield of Computer-Assisted Instruction (CAI). While CAI is a broad term used to define teaching and learning through computer interaction for all disciplines, CALL focuses specifically on language instruction. For Levy and Hubbard (2005), CALL encompasses activities such as technology-enhanced language learning, network-based language learning, Web-enhanced language learning, and information and communication technology for language learning.

In the last two decades, the number of CALL practitioners and researchers around the globe has increased drastically, following the advances and affordability of technology. Since the end of the 1990’s the increasing popularity of the World Wide Web has brought a new vitality to CALL development, and has revealed new possibilities for the use of computers in language instruction. Nowadays, computers are widely used in foreign language teaching and learning (FLTL) to help learners experience the target languages and cultures. Currently available technology allows for
multimedia presentations, web-based TV/radio/news, emailing and chatting with native speakers, among many other uses. Computer-Mediated Communication (CMC) is a field in itself which aims at studying human interaction with computer, and several studies have discussed the use of CMC in FLTL (cf., e.g., Warschauer, 1997; Paramskas, 1999; Smith, 2005). More recently, an increasing number of institutions have been offering language courses in virtual language classrooms, where students and instructors interact in a virtual environment, instead of a regular classroom (cf., e.g., Humpel, 2003; Felix, 2002). In summary, if we consider CALL in its broader sense, we could say that any time people use a computer to learn, practice, or interact in a foreign language, they are doing CALL.

In their recent book, Levy and Stockwell (2006) differentiate between two approaches to CALL research, development and integration. They call the first one “established CALL,” and the second one “emergent CALL.” According to them, the basic difference between established versus emergent CALL is in the way CALL practitioners and researchers interact with technology:

“Established CALL involves technologies that are well established and accepted. The label is used to indicate mainstream activity in contrast to more specialized activity involving new and emerging technologies. Practitioners focus on using and evaluating CMC modes for language learning and, when CALL materials are developed, well-known authoring tools such as Hot Potatoes and BlackBoard are used in a straightforward way (i.e., without advanced adaptations).” (Levy and Stockwell, 2006, p. 246)

Established CALL practitioners and researchers tend to focus their efforts on evaluating and incorporating existing technology into teaching and learning practice. As Levy and Stockwell (2006) point out, many of these practitioners believe that focusing directly on technology development is a distraction from the main activity
of language teaching. This characteristic generally distinguishes CALL professionals from language teaching professionals in Hubbard’s terms (Hubbard, 2004, p. 3), who defines the latter as being concerned primarily with the integration of well-established technology into current methodologies in FLTL. The use of technology in their everyday practice is their main goal, and their contribution to the field is based on innovative pedagogical practice through the design of new language-learning tasks, and the integration of CALL and non-CALL approaches in FLTL curricula. They also play a critical role in the evaluation of existing technology in contextualized settings with real language learners.

The great danger of focusing exclusively on the “established” approach to CALL is to put the discipline in a straitjacket. CALL in its origins was very much connected to system design and implementation. When PLATO (Hart, 1995) and TICCIT (M. David Merrill, 1980) were first conceived in the 1960’s and 1970’s, there were no text editors, no Internet, and no personal computers. The only way the field could advance was through the development of large-scale systems. Nowadays, different types of commercial software can be incorporated into CALL practice. This fact brings a great advantage to CALL practitioners, however there is no reason to believe that developing technology to cope with the specific needs of language learners and instructors is a profitless activity.

Computers can be viewed as tools to perform certain tasks. The way they perform those tasks is defined by the needs of their users. In no other discipline commercial off-the-shelf software is considered to be the state-of-the-art material. In fact, most of the time, this type of software has no use at all. When researchers of computational biology, chemistry or physics need a computer to perform a task they first specify
it, then they develop the program to handle it. It is unclear why computational language learning should be any different. CALL users have specific needs that are different from the needs of native speakers, and FLTL instructors use methodologies that reflect their beliefs on how foreign languages are acquired. If a commercial software happens to fulfill certain needs of learners, and, at the same time, fits into a given methodology, it can be incorporated into current FLTL practice. However, we do not see any arguments for limiting CALL research and development to these convenient coincidences.

The second group of CALL practitioners and researchers believe that to advance CALL as a discipline their research should focus on the development of the technology used. Levy and Stockwell (2006) call this approach “emergent” CALL:

“In emergent CALL, we are very interested in looking at the technology to see what it can do and what it cannot. When there are shortcomings, these may be addressed. The approach may involve revisiting well-established and accepted technologies and seeing if their features can be improved or redefined for language-learning purposes. . . . In emergent CALL, researchers directly engage with the technology itself. . . . As far as emergent CALL is concerned, teacher-designers, developers, and researchers are currently looking closely at language learning programs involving:

- Speech-recognition applications.
- Broadband audiovisual technologies.
- Online teaching systems (with human tutors).
- Intelligent tutors (ICALL – with computer tutors).
- Mobile technologies.
- Fine-grained design decisions (the optimal annotation).
- Hybrid solutions.
- New authoring tools and techniques.
- Compatibility of technologies (e.g., knowledge pooling, reusability issues).”

(Levy and Stockwell, 2006, p. 242)
The term “emergent” can be misleading, giving the impression that research in this area is not consolidated, or that it is a new approach to CALL. As noted before, CALL’s origin is associated directly with the development of technology. Those terms (established and emergent) are better understood if associated directly with the type of technology used, rather than with CALL itself.

Notice that researcher in emergent CALL do not form a homogeneous group themselves. At the two ends of the spectrum we would find CALL experts that have some interest in technology on one side, and computer scientists that want to develop CALL tools but who know very little about FLTL on the other end. Researchers of what could be considered emergent CALL do not even attend the same professional conferences. While members of the Artificial Intelligence community, who work with intelligent language tutoring systems, typically go to conferences such as AIED, ITS, ICALT, ICCE, and UM, members of the ICALL community, who develop ICALL systems to be used in FLTL, generally present their work at CALICO, EUROCALL, and WORLDCALL. Of course there are researchers that go to both. These different orientations inside the group of emergent CALL could be productive if ideas were shared, and one could realize the importance of the contribution of the other. However, the divisions more often seem to be obstacles to dialogue and contribution.

Another, typically self-imposed, limitation of emergent CALL researchers is the fact that they rarely acknowledge the contribution of established CALL, i.e., people that are more directly involved with FLTL practice. Correspondingly, as section 1.5 will show, research projects whose aims were to develop new technologies for CALL
have been severely criticized for their lack of integration with existing methodologies and for the absence of reliable evaluation of their products in contextualized FLTL settings.

This dissertation can be seen as an initiative for bridging the gap between these two communities; those who develop CALL tools and those who use CALL in their everyday practice. Levy and Stockwell (2006) end their considerations about emerging and established dimensions in CALL by stating that “together (they) give a much more complete picture of current work in the field” (Levy and Stockwell, 2006, p. 251). The work presented here goes beyond this observation, since it is built on the strong conviction that CALL cannot establish itself as an independent discipline without merging these two dimensions, and that CALL research, development and integration into FLTL lies with approaches that are able to integrate theory, implementation and practice in a multidisciplinary way. It is no longer enough to develop software that can be used in CALL, or to evaluate and integrate computer tools as if they were made to be used in CALL. We believe that CALL tools should be designed and evaluated as CALL tools. The needs of learners, the pedagogical choices of instructors, and the theoretical perspectives of researchers have to be the point of departure for CALL development.

The remainder of this chapter will clarify certain issues in CALL research and development that are necessary for understanding the approach proposed in this dissertation. Section 1.2 discusses how the difficulty in establishing a clear research paradigm in CALL is affecting the development of the discipline, and why it is necessary to respect the multidisciplinary nature of CALL if we want to advance the field.
Section 1.3 introduces what conventionally has been called Intelligent Computer-Assisted Language Learning (ICALL), and discusses some inconsistencies on what is generally understood as ICALL. Section 1.4 presents the basic properties of Intelligent Language Tutoring Systems (ILTS), and the possible contributions of the research in artificial intelligence for ICALL development. Section 1.5 introduces some reasons for the lack of integration of ICALL research into current FLTL, shows the limitations of systems that are being used in FLTL practice, and points to future directions for improving current work. Section 1.6 describes the approach proposed by this study to improve ICALL development, and presents the structure of the dissertation with the content of later chapters.

1.2 Research in CALL

So far we have been using the terms CALL practitioner and researcher loosely. To prevent misunderstandings about the nature of the comments made in this section, before we proceed it is important to define how the terms practitioner and researcher are used here. In general terms, there are two activities that we consider here as research in CALL: development and testing. For the sake of definition, we consider the deployment of a certain technology as practice.

We could use a metaphor from medical research to clarify the issue. In the medical domain, research is roughly done in two phases. First new drugs have to be developed, and then they have to be tested with humans to observe their therapeutical value. Doctors that use previously developed medicine with different patients in different circumstances are not necessarily doing research. Although their everyday practice
may shed some light about the real contributions of a drug, or its harmful side effects, it is necessary to design experiments that follow well specified procedures to have scientific evidence of the usefulness of a given medicine.

In an attempt to establish a common background for CALL research, several arguments for theory based research have been presented, especially by those who advocate in favor of CALL research based on second language acquisition (SLA) theories (see papers in Egbert and Petrie, 2005). However, as Levy (1997) reminds us, theories of acquisition have only been one point of departure for CALL research. Several other CALL projects have taken into primary consideration the actual FLTL practice and methodological choices, or technological development and integration.

CALL researchers look into other disciplines for theoretical foundations for their work, and it has been argued that CALL does not have a sufficient body of work to support independent research (Egbert, 2005). This may be because CALL is a new research area, or even because CALL is a subfield of other disciplines, such as applied linguistics, and will never be fully independent (see Leech and Candlin, 1986). However, if we look closely at the evolution of CALL in the last 30 years, we will see that CALL is the product of an era where multidisciplinary approaches to common problems are the rule rather than the exception. New disciplines developed during this time tend to be multidisciplinary in nature.

Levy (1997) lists some of the main disciplines that have contributed to the evolution of CALL (Figure 1.1). Among them there are: psychology, instructional technology, artificial intelligence, human-computer interaction, computational linguistics, and applied linguistics. It is important to emphasize that the contribution of each of these disciplines was not restricted to their specific body of knowledge. They also
brought with them their methodological paradigms to undertake scientific investigation. Conceptual and methodological differences dealing with the design of experiments and the formulation of scientific questions have been major roadblocks for full collaboration between CALL practitioners and researchers from different backgrounds (Levy, 1997).

Figure 1.1: CALL and related disciplines (Levy, 1997, p. 72)

More than ten years ago, Garrett (1995) already identified one of the main challenges for CALL research and development: the integration of language teacher’s
expertise in the development of CALL systems. She also criticized the lack of support for teacher involvement with technology in post-secondary education. Today, the latter seems to be no longer true, judging by the increase in participation of CALL experts (researchers and practitioners) in conferences such as CALICO and EUROCALL. Nevertheless, the development of new technology does not seem to be part of the agenda of many CALL experts, and interdisciplinary research projects with computer scientists, linguists, and foreign language teachers are rare. For example, when it comes to the use of NLP in CALL there seems to be very few research groups that are currently active. In the last version of EUROCALL (September 2006), out of more than a hundred and seventy papers presented during the three days of the conference only three were NLP related.

One consequence of this lack of interdisciplinary work is that people who develop new technologies do so without thinking about language instruction and acquisition, and people who evaluate the roles of existing technology in language acquisition know very little about its development. If a research initiatives in CALL is limited to what technology can already do, all variables in an SLA experiment have to come from the human side of the interaction, or by using two different existing tools.

Huh and Hu (2005) criticize what they call “technocentrism in CALL research”: “If researchers emphasize the technology only and conclude that CALL programs are effective for language learning, it is problematic. The question is what really matters for language instruction, is it the computer, the teacher, the learning environment, the students themselves, or some combination? . . . By appropriately analyzing participants’ interaction and behavior, for example, we can address questions of why a system is not working or how to make it work better.” (Huh and Hu, 2005, p. 17)

What Huh and Hu (2005) do not mention is that there is a further possibility for those studies. Instead of comparing systems, or educational settings with and
without systems, researchers could change the variables within the CALL system used in a given experiment to test different working hypotheses. The reason this third possibility is usually not explored is, again, the lack of interaction between those that develop the technology and those that evaluate its use.

Research projects that incorporate development, research, and testing in real life teaching environments can contribute to promote interaction among researchers from different backgrounds, and reduce the gap between people who develop new technologies and those who use them. Some researchers have perceived this need, and have developed and/or expanded their investigation programs to incorporate elements of different disciplines (cf., e.g., Dodigovic, 2005; Heift, 2005). Section 1.6 discusses how the work presented in this dissertation wants to contribute to the development of CALL research by respecting and supporting the multidisciplinary nature of the field.

1.3 CALL and ICALL

One of the areas of CALL that has received some attention in the last two decades is the development of CALL tutors (Heift and Shulze, 2003). In general terms, a CALL tutor is a computer program that evaluates a learner’s response and provides some sort of feedback. In its simplest versions, a CALL tutor deals with simple “right” or “wrong” responses to student input. The most common types of exercises used by these tutors are fill-in-the-blanks and multiple choice.¹ These simple CALL tutors work with pre-stored answers, and use pattern matching to decide if the input is right

¹The fact that some electronic tutors only deal with restricted types of exercises does not necessarily mean that they do not have AI components. As we will see in chapter 5, there are tutors that only present multiple choice questions, yet have very sophisticated student models.
or wrong. In other words, the mechanism used to check answers is comparing them letter by letter with the target answer. Simple pattern matching as a mechanism to detect errors can work well if correct answers are predictable and listable, or if there is no expected grammatical variation in student’s response, and if envisaged errors correspond directly to intended feedback. Any activity whose possible answers cannot be conveniently listed are problematic to this type of setting. Moreover, if linguistic errors can occur throughout a recursively built structure, such as embedded non-defining relative clauses, pattern matching will not function for diagnosis. Thus, the processing limitations of pattern matching input processing can create serious challenges to computer implementations of some common activity types used in FLTL (Bailey and Meurers, in prep).

Another motivation for the development of tutors that can do more than pattern matching is the need to provide personalized feedback to individual learners on language forms and rules. More sophisticated CALL tutors need to be able to perform error diagnosis, and possibly also error correction and to generate individualized learner feedback. Error diagnosis differs from error detection because in the latter the system only needs to identify if there is an error in the sentence, while in the first the system has to provide an analysis of the nature of the error. Error correction can be considered as a further step, when the system not only diagnoses what type of error occurred but also provides suggestions for correction. Systems that are designed to do error diagnosis and/or correction are usually called Intelligent CALL (ICALL) systems, or Intelligent Language Tutoring Systems (ILTS). Some authors differentiate ICALL systems from ITLSs, claiming that the latter usually reflects the

\(^2\)Chapter 2 will discuss the current SLA research in favor of viewing language awareness as an important component in the acquisition process of an adult learner.
work on intelligent tutors with all its properties (see section 1.4 for basic properties of ITSs), while the former proposes some sort of computer-assisted instruction with error analysis (see Levy and Stockwell (2006) for details). Following the tradition of most ICALL researchers (see Heift (1998)), in this dissertation both terms are used interchangeably to refer to computer-based tutors that make use of artificial intelligence techniques.3

Traditionally, the first step to create ICALL systems has been to incorporate into them some knowledge about its target domain, i.e., knowledge about language forms and rules. Since the 1980’s, several projects aimed at integrating natural language processing (NLP) techniques into ICALL development, in an attempt to make systems that are linguistically aware (cf., e.g., Weinberg et al., 1995; Rypa and Feuerman, 1995; Heift, 1998; Nagata, 2002). In most projects the emphasis was on modifying or developing parsers, i.e., algorithms licensing syntactic structures, to be used as error diagnosis tools. The idea was to take advantage of the advances in syntactic analysis in NLP to create tools that could complement grammar-focused instruction.

In the search for the best syntactic formalism to develop those tools, several proposals were made. Weinberg et al. (1995) argue that a Government and Binding (Chomsky, 1981) framework can address the needs of ICALL tutors through implementations that can be compact and easily portable to other languages. Matthews (1993) argues that GB is also a good choice since it can be coherently linked to theories of second language acquisition that are Universal Grammar-based.

Rypa and Feuerman (1995) defend the adoption of Lexical Functional Grammar (Kaplan and Bresnan, 1982) “characterization of linguistic representations in LFG

3Notice the distinction between the terms ILTS, i.e., Intelligent Language Tutoring Systems, and ITS, i.e., Intelligent Tutoring Systems for all disciplines, including language.
map into instructions as representation structures that are accessible and potentially useful to a language student” (Rypa and Feuerman, 1995, p. 61). Reuer (2003) seems to agree with them, justifying the use of LFG in his tutor for German by stating that “the concepts and structures used in LFG closely resemble the descriptive knowledge of language learners about language, and, therefore, the representation of an automatic analysis can easily be translated from a computationally tractable form to language easily understood by the learner” (Reuer, 2003, p. 497). Heift (1998), on the other hand, opts for another constraint-based formalism, called Head-driven Phrase Structure Grammar (Pollard and Sag, 1994), for the development of the German Tutor, and its successor the E-Tutor.4

Although most of the work on developing or adapting NLP technology for ICALL tutors has been done on syntax, some projects focused on other types of linguistic knowledge. Dorr et al. (1995), for example, present the Military Language Tutor (MILT) that incorporates semantic processing. MILT has a specific module to perform semantic analysis that uses lexical semantic information. This information is encoded in the form of Lexical Conceptual Structures as formulated by Dorr (1993) based on Jackendoff (1990). Semantic processing based on lexical semantic information is also used by DeSmedt (1995) in a system called Herr Kommissar. The system simulates a role-play in a detective game, and provides feedback on syntactic and semantic structures. More recently, Bailey and Meurers (2006) have explored the development of shallow semantic processing for automatic English reading exercises that make use of Wh-questions.

4Further discussion of Heift’s work can be found in section 1.5 and in chapter 5.
Incorporating linguistic knowledge into ICALL systems is one important step to develop computer-based tutors that can interact with students in a meaningful way. The next section will discuss other elements that are necessary to the implementation of systems that are able to evaluate students’ responses and provide personalized feedback on language forms and functions.

1.4 Intelligent Language Tutoring Systems

While in ICALL most of the work has been done on the integration of linguistic knowledge into the tutor, the work in artificial intelligence (AI) in education has evolved into developing intelligent tutoring systems (ITSs) that reproduce the behavior of human tutors. Burns and Capps (1988) define the evolution from computer-assisted instruction into ITSs:

“Computer-assisted instruction evolves toward intelligent tutoring systems by passing three tests of intelligence. First, the subject matter, or domain, must be ‘known’ to the computer system well enough for this embedded expert to draw inferences or solve problems in the domain. Second, the system must be able to deduce a learner’s approximation of that knowledge. Third, the tutorial strategy or pedagogy must be intelligent in that the ‘instructor in the box’ can implement strategies to reduce the difference between expert and student performance.” (Burns and Capps, 1988, p. 1)

Hartley and Sleeman (1973) were the first to establish that an ITS should have at least three basic types of knowledge: (i) knowledge about the domain, (ii) knowledge about the learner, and (iii) knowledge about how to conduct the instruction. Despite the fact that standards for the development of ITSs have been around for more than three decades, the actual implementation and research has varied considerably between different domains.
Most of the work on the development of ITSs has focused primarily on sub-domains of science that are sometimes referred to as *well-defined* domains, such as geometry, artificial languages, and Newtonian mechanics. Some of the most advanced techniques in AI for ITSs were tested in systems designed to teach fractions (Dugdale, 1993) or algebra (Croteau et al., 2004), for example.

Well-defined domains are usually based on a fully-characterized formal theory or clear-cut domain model. Such domains are typically taught through problems whose answers can be classified unambiguously as correct or incorrect. Lynch et al. (2006) explain why well-defined domains are more popular in ITS development:

> “Well-defined domains are particularly amenable to model-tracing tutoring systems. Operationalizing the domain theory makes it possible to identify study problems, provide a clear problem solving strategy, and assess results definitively based on the existence of unambiguous answers. Help can be readily provided by comparing the students’ problem-solving steps to the existing domain models.” (Lynch et al., 2006)

Domains that do not have a fully-characterized theory, or that pose challenges to the development of a clear-cut domain model are sometimes referred to as *ill-defined* domains (Lynch et al., 2006), such as ethics, music composition, and law. Usually, in these domains, problems presented to students have multiple answers, or sometimes no correct answer at all, just possible or impossible answers.\(^5\)

Although linguistics has evolved considerably in the last century, there is no theory that can deal with all aspects of human language (or the ones most relevant to language learning). However, the great challenge in developing an ITS for language learning is not necessarily related to the fact that language is so-called ill-defined domain. The biggest problem for developing ITSs for language lies in the fact that the

---

\(^5\) Recently the AI community has shown an increasing interest in exploring the development of ITSs for ill-defined domains, and workshops such as Lynch et al. (2006) are becoming more common.
metalanguage and knowledge representation mechanisms used to describe language maybe completely foreign to the learner, which may create a gap between the way a system diagnoses and reports errors and the way the learner conceptualizes them.\textsuperscript{6}

For example, Ogan et al. (2006) show the difficulties in modeling the behavior of students while learning the aspectual difference between the \textit{preterite} and the \textit{imperfect} in French. They present some problems in the use of decision trees to model the learning process of French aspect. It is important to notice that at no time they discuss models of acquisition of French aspect. They approach the problem of learning linguistic properties the same way they would approach the problem of learning properties of other domains (well-defined or not). In chapter 5 we will see why such strategy is problematic if we consider specific characteristics of second language acquisition, and discuss in detail why language cannot be treated as another ill-formed domain when developing a model of learner knowledge.

The next three sections introduce concepts related to the role of domain knowledge, student model, and instruction model in ITS development. These concepts are revisited and expanded in chapters 4 and 5.

\subsection*{1.4.1 Domain Knowledge}

The domain knowledge, as the name says, is the knowledge about the subject matter area. In the case of ICALL systems, it is knowledge about the use and nature of language. From a software engineering perspective, the knowledge about the domain is usually located in the so-called expert module(s). Anderson (1988) defines the

\footnote{In chapter 2 we discuss in more detail some theories of SLA and their applicability in ICALL.}
expert module as the “backbone of any ITS”. Knowledge about the domain being taught is the point of departure for any instructional system. Without it, the system is like an instructor who does not know the subject he is teaching.

One key issue in the development of expert modules is how to represent the knowledge stored in there. Traditionally, knowledge representation has been codified using three distinct approaches (Anderson, 1988). The first one is to design a process that produces the same input/output behavior for a certain domain without actually codifying human knowledge used to produce that behavior. This approach produces what is called a *black box model*. The second one is to extract human knowledge from an expert, and find ways to codify and apply it that do not necessarily match the way humans do. The third one is to simulate a human being using the desired knowledge. This last approach is used to generate and test scientific models of cognition.

Domain knowledge is often classified into two types (Anderson, 1988): *declarative*, such as facts, figures, and objects; and *procedural*, which deals with how declarative knowledge is used to solve a problem. The distinction between declarative and procedural knowledge can plays a role in identifying learning strategies and procedures, as well as defining teaching mechanisms to cope with learner problems.

ICALL systems (or ILTSs) have to use descriptive mechanisms that may reflect the knowledge of linguists about the nature of language, but that do not necessarily reflect the knowledge of learners. Therefore, all work in AI to develop mechanisms to describe other domains, such as mathematics, computer science, or electrical engineering cannot be easily transposed to ITS development for language learning.
1.4.2 Student Model

The second component in ITSs necessary to produce intelligent behavior is the student model. The student model is a data structure that characterizes the student’s current understanding of the subject matter. Some authors use the term learner model instead of student model. In this dissertation those terms are used interchangeably.

Together with the student model, the system needs a mechanism to manipulate, i.e., create and update, its data structure. This is usually done by a diagnosis module. The design problem of creating a diagnosis module and a student model in sync with each other is called the student modeling problem (VanLehen, 1988, p. 55).

Traditionally, techniques to model student knowledge use mechanisms to compare the expert knowledge, i.e., the knowledge needed by an expert to complete the task, with the student knowledge. Some student models may use bug libraries (i.e., a collection of misconceptions in a student’s knowledge), others may use overlay models (i.e., a description of missing conceptions in a student’s knowledge). Similarly to techniques used to develop expert models, most of the work on student models has focused on domains where we can clearly list the necessary abilities to solve the problems presented.

In recent attempts to develop user models for ICALL systems, several authors have conceptualized the acquisition of language as the learning of linguistics forms and rules (cf., e.g., McCoy et al., 1996; Heift, 2004). Chapter 5 presents an analysis of such models, discusses the ways where such perspectives falls short of modeling the learner performing a language task, and presents an alternative to such models.
1.4.3 Instruction Model

“People learn many things without benefit of instruction, but we are distinguished as a species by our ability to pass knowledge from the competent to the less competent. To endow machines with this same instructional ability is, to a large extent, to cast the principles of instruction in precise information processing terms.” (Half, 1988, p. 79)

The last recognized component of an ITS is the instruction model. To design instruction models it is necessary to understand the nature of teaching in relation to the nature of the subject matter. Some ITSs may be designed to teach factual knowledge and inferential skills, others may target the learning of skills and procedures. Whatever goal a tutor might have, its behaviors have to be guided by learning theories for its specific domain.

Instruction models constitute one of the least researched areas in ICALL. In part because its properties have been taken for granted, in part because the cognitive approaches in SLA have not been fully explored to support ICALL design choices. Chapter 2 discusses some of the theories that can be used to guide ICALL research, and chapter 3 will show how these theories are used in this work.

1.5 ICALL Research and ICALL Integration

It has been more than forty years since research on computer-assisted language instruction and learning started (Levy, 1997). Computers are widely used for CALL today, and the number of CALL practitioners grows everyday. Since the 1990’s, FLTL methodologies have shifted to classroom procedures that do include references to linguistic forms and patterns.\(^7\) We may see these changes in FLTL methodologies as an opportunity to the development and integration of ICALL systems, and yet, there are

\(^7\)See chapter 2 for details.
very few systems that use NLP that are currently being used by language learners. Ten years ago, Levy (1997) had already noticed the unfulfilled potential of ICALL. Yet, today his words are still true:

“the influence of this area (ICALL) on CALL has so far been limited, it has the potential to alter significantly the nature of CALL.” (Levy, 1997, p. 72)

Despite the fact that NLP technology has its limitations, there are other reasons that have played a major role in the lack of integration of ICALL systems into teaching and learning practice. Most research in ICALL in the last 20 years has not focused on developing ICALL tools that cope with basic foreign language teaching requirements, nor has it seriously taken into consideration basic parameters of second language acquisition theories. Correspondingly, ICALL research has advanced very little in terms of developing a real intelligent tutor with comprehensive student and instructor models.

One major problem in ICALL research is the fact that it has often dissociated the development of the expert model from the development of complete tutors that fulfill specific pedagogical purposes (cf., e.g., Reuer, 2003; Delmonte, 2003). Of course, the expert model is where computational linguists can contribute the most. However, by leaving the student and the instructor models out of sight, researchers may render their final products uninteresting to the rest of the CALL community, and, most importantly, they may develop systems that are not suitable to help real learners in their acquisition process. Projects in this area tend to focus on developing parsers to identify grammatical errors, and exercises are design to fit the area of NLP that is being tested.
Another problem common to the integration of AI technology in real life ICALL is research paradigms that aim at developing very sophisticated ideas in a small scale for a very narrow use. In ICALL it is not uncommon to see systems that have very sophisticated student models or language processing mechanisms but deal with one specific aspect of language, such as clitic placement (Bull et al., 1995) or passive voice formation (Virvou et al., 2000; Virvou and Tsiriga, 2001). The results are usually very interesting and insightful research projects, but the final product is not meant to be incorporated into a content-based language program.

A third characteristic of some ICALL projects that has hindered the development of ICALL systems for real-life language learning is the temptation to deal with all aspects of human language in an unrestricted way. Usually these projects aim at developing NLP technology to deal with totally unconstrained input. The FreeText project (L’haire and Faltin, 2003), for example, aimed at developing a ICALL software to deal with unrestricted input by intermediate and advanced students of French as a foreign language. One of its original goals was to allow teachers to introduce new exercises and documents to suit the particular needs of their students. The project was pedagogically sound and its results seemed promising; unfortunately it did not achieve all its goals. L’haire (2004) describes the final result as such:

“Par rapport à nos ambitions de départ, FreeText a vu ses ambitions réduites. La technique de la comparaison de phrases a dû être reportée après la fin du projet et les performances du système de diagnostic peuvent sembler relativement faibles par rapport aux espoirs soulevés. Il est aussi un peu frustrant de ne pas disposer de résultats sur l’accueil réel de nos outils d’aide chez les apprenants. Nous ne disposons que d’une évaluation de spécialistes et des une réactions suscitées par les diverses présentations du logiciel que nous avons faites. Une analyse plus fine de nos résultats et des imperfections de nos
techniques devrait améliorer ces résultats. L’exploitation intelligente de notre sortie pourrait aussi être d’un grand apport pédagogique.” (L’haire, 2004, p. 10)

Fortunately, there have also been some successful cases, where NLP technology was used to generate systems robust enough to be employed as a component of a language program. Robo-Sensei (Nagata, 2002), and E-Tutor (Heift, 1998, 2003) are two successful examples. There is also a third successful example: Spanish for Business Professionals (Hagen, 1999), although in this case the system was not designed to be integrated into a language program, but rather to be used as a stand alone product.

Nagata’s system for Japanese presents a series of exercises for each of its 24 lessons. The activities are contextualized, and there are always visual aids with pictures of Japan or Japanese drawings. It also presents good cross-references with current textbooks, so that instructors can choose activities according to their current classroom material. Nagata has done extensive research in the use of such CALL systems (Nagata, 1993, 1996, 1997), and it is reflected in the general sequence of activities.

Another successful case of integration of NLP technology into a system is the Spanish for Business Professionals (SBP) (Hagen, 1999). SBP is a program to teach business Spanish in 12 units. Its units are well contextualized with an excellent selection of audio material. There is an interesting progression of exercises and good visual aids. The program also presents several help tools, including links to grammar explanations and texts with hyperlinks to an electronic bilingual dictionary.

The third example is the E-tutor, a system developed by Trude Heift (Heift, 2003) that is fully incorporated into the German classes offered at Simon Fraser University. German students enrolled in regular classes have E-tutor exercises as a requirement to the completion of their courses. Among the three systems presented here, the
E-tutor is the one that has evolved the most in recent years. It has incorporated an explicit learner model that allows the system to generate reports about students’ performance, which is an excellent tool for instructors and students. It has also gained a new layout and incorporated listening passages used in German classes.

The three systems described above have one thing in common: their development was guided by the need to create a product that could actually be used by language learners. This means that in their process of implementation it was necessary to evaluate the capabilities of the technology available in relation with the pedagogical goals to be achieved.

In chapter 3, we will see that one of the major challenges for ICALL systems is to find pedagogically sound ways to elicit students’ input that are in sync with the processing capabilities of the system. If the system fails to do so, it runs the risk of receiving some input that can jeopardize its ability to provide appropriate feedback.

One of the problems with current ICALL activity types is their extensive use of translation to elicit students’ answers. In Robo Sensei, for example, despite their clear contextualization, several of the activities proposed could be viewed as simple translation exercises. Widespread English cues are given to elicit students’ input in Japanese. In SBP, the vocabulary exercises are based on simple word translations, which makes them look like old fashioned drill activities whose triggers are lexical items in L1. Another common way to elicit students’ input in existing ICALL systems is the use of decontextualized dictation, despite the fact that dictation is not currently used in communicative-based methodologies. The E-Tutor, for example, has limited its eliciting strategies to translation, dictation, or presentation of the words to be used.
Another challenging issue that ICALL systems have to deal with is deciding the desirable amount of L1 in its activities and feedback messages. For systems that deal with languages that use non-Roman alphabets, deciding when to use L1 is especially complicated. In Robo-Sensei, the choice was to convey all information related to task descriptions and exercise instructions in English. The feedback messages are also in English, and some of the answer triggers and examples use language comparisons between English and Japanese, which makes the methodology used by the system look like it is based on the outdated grammar-translation approach (Richards and Rogers, 2001).

Finally, most ICALL systems used in real life do not have explicit student or instruction models to guide their feedback strategy or their instructional techniques. Two exceptions to this rule are the E-tutor and ICICLE (Michaud and McCoy, 2004).\(^8\)

In order to deal with the current issues in ICALL development it is necessary to explore ways to bridge the gap between the research on new NLP and AI technologies and the pedagogical needs of foreign language students and instructors. One of the main goals of a research project that aims at developing ICALL systems to be used in real life has to be to reduce these pedagogical concessions in activity design by improving the technology used. See chapter 3 for a discussion about reducing pedagogical concessions in ICALL development.

### 1.6 Proposed Approach

In view of the current needs of ICALL development and research, the point of departure for the work proposed in this dissertation is the exploration of current pedagogical

---

\(^8\)See chapter 5 for a characterization of the systems and their student models.
practice in FLTL and the possibilities for integration of ICALL. The overarching goal of this work is to advance research in ICALL development so that more tools can be created to be used in real-life language learning situations.

The work begins with a review of current SLA and FLTL theories that have influenced, or that have the potential to influence, ICALL development. Chapter 2 presents some basic principles of task-based instruction (TBI) (cf., e.g., Ellis, 2003), content-based instruction (CBT) (cf., e.g., Brinton et al., 1989), and focus-on-form (cf., e.g., Long, 1991) that are sometimes used in established CALL, but usually neglected in ICALL work.

Chapter 3 introduces the pedagogical context of the TAGARELA project, both for classroom-based Portuguese instruction and for the Portuguese Individualized Instruction Program. It also discusses the opportunities for ICALL integration in view of the needs of students and instructors in these two settings. There is a presentation of the activities incorporated into the system and of the design choices adopted for their development.

Chapter 4 presents the components of the system responsible for its intelligent behavior. Chapter 4 begins with related NLP approaches used in other ICALL, and continues with a description of the approach used in TAGARELA. It presents a description of the activity model used by the system followed by the general architecture of the system and the roles of the interface, the Analysis Manager, and the Feedback Manager. A detailed description of the development of each of the NLP sub-modules, and their applications is also presented. The chapter ends with the feedback strategies used by the system to generate messages to the student.
Chapter 5 presents a conceptualization for a student model for TAGARELA. The chapter discusses some current approaches in ICALL followed by some issues on the nature of the knowledge being modelled. Then, the chapter introduces a proposed model and the changes in the system architecture that have to be made to accommodate it. It ends with some examples of how this model can help improve feedback.

Finally, chapter 6 discusses the integration of the final project into the Individualized Instruction setting, and the reactions of students to the system. It also presents future directions for the project together with a summary of what has been achieved so far.
CHAPTER 2

FOREIGN LANGUAGE INSTRUCTION AND SLA RESEARCH FOR ILTS DEVELOPMENT

As all human activities, foreign language teaching and learning (FLTL) has always been directly influenced by social and cultural patterns. Looking back at the history of foreign language instruction, one may see that approaches to FLTL were a product of their time, based on the role that language learning played in each distinct society. The grammar-translation method, for example, played a crucial role in European and Americans societies from the seventeenth century to the first part of the twentieth century (see Kelly, 1969). At the time it was very important and prestigious for an educated man to learn classical Latin. Grammar and translation were the two key elements in Latin classes, therefore they became the relevant way of learning a foreign language.

In the nineteenth century, German, English and French schools were accused of not fulfilling their roles of helping students develop an adequate level of oral proficiency to cope with the increasing needs of communication among Europeans. The focus in FLTL began to shift from written language to oral communication. In the first part of the twentieth century, British applied linguists proposed a method based on a systematically selection and gradation (i.e., classification according to linguistic
complexity) of lexical and grammatical items to be taught using specifically designed techniques for presentation and practice. This approach was originally called *Oral Approach*, and was later more commonly referred to as *Situational Language Teaching* (SLT). Richards and Rogers (2001) list the main characteristics of such approach:

1. Language teaching begins with the spoken language. Material is taught orally before it is presented in written form.
2. The target language is the language of the classroom.
3. New language points are introduced and practiced situationally.
4. Vocabulary selection procedures are followed to ensure that an essential general service vocabulary is covered.
5. Items of grammar are graded following the principle that simple forms should be taught before complex ones.
6. Reading and writing are introduced once a sufficient lexical and grammatical basis is established.

(Richards and Rogers, 2001, p. 39)

The principles of SLT have been so influential in FLTL that even today it is easy to recognize them in most textbooks. It is also important to notice that SLT already advocated the practice of language items in meaningful situation-based activities, something that is still considered important by many language teachers.

In the 1960’s SLT started to be criticized because of its structured way of proposing language instruction. Linguists were exploring functional theories of language, and Chomsky (1957) had strongly criticized structural theories of language that did not focus on the creativity and uniqueness of possible individual sentences. British applied linguists (c.f., e.g., Christopher Candlin and Henry Widdowson) started to emphasize the need to develop learners’ communicative proficiency, rather then the mastery of structures.

In the 1970’s the *Communicative Approach* became one of the most influential approaches in the history of FLTL. The main goal of the communicative approach is
to forge the learner’s communicative competence by developing procedures to teach the four primary language skills based on the interdependence of language and communication. Language teachers adopted different definitions and practices for the communicative approach for language instruction. These practices varied from incorporating communicative activities into already existing structural syllabi, to redesigning instructional settings in a way that references to language structures were abolished and examples of authentic language are used based on communicative functions. Howatt (1984) defines these two extremes as being the “weak” and the “strong” versions of communicative language teaching.

The Communicative Approach has influenced several other approaches for FLTL developed since the 1980’s. Although the stronger versions of the Communicative Approach have not played a significant role in ICALL development, some of its late developments, such as task-based instruction and content-based instruction, can contribute to the implementation of ICALL materials to be used by language learners.

The role of this chapter is neither to compare detailed characteristics of approaches for language teaching, nor to evaluate the efficiency of different approaches to argue in favor of one of them. The first goal of this chapter is to describe some basic principles of well-accepted methods in FLTL that can support the integration of ICALL systems into language instruction. As proposed in chapter 1, it is important to identify the areas where ICALL systems can contribute to language instruction in order to design a system to be used in real life.

---

9The strong version of the Communicative Approach influenced the use of more advanced technology to create games and artificial environments where learners can practice the target language. However, by nature, ICALL systems that propose feedback on language forms and rules contradict basic principles of purely communicative methods.
As we will see in chapter 3, ICALL activities are usually designed based on the type of input they are supposed to elicit. While in traditional CALL activity design is a well-studied area with strong connections with FLTL and SLA, in ICALL, activities have been subjected to the limitations of the technology, and not much has been said about their pedagogical functions and incorporation. Sections 2.1, 2.2 and 2.3 present some research findings on task-based instruction, focus-on-form, awareness and motivation in SLA that can have a positive impact on ICALL activity design. The work described in those sections will be revisited in chapter 3, when the activities of the system proposed in this dissertation will be presented.

Besides activity design, research in FLTL and SLA can also contribute to the development of ICALL systems that use NLP technology to analyze students’ input. For those systems to be useful, the results of their analyses must play some role in the methodologies used in FLTL, usually in the form of feedback to the learner. Section 2.4 discusses how some recent research findings on the role of feedback on L2 writing can help develop sound feedback strategies in ICALL.

### 2.1 Task-Based Instruction

Task-Based Instruction (TBI) is an approach to FLTL that uses tasks as a core component of language instruction. To understand how TBI works it is necessary to know what a task is. Ellis (2003) presents a list of some common definitions of task in SLA (Ellis, 2003, pp. 4-5). Richards et al. (1985), for example, describe a task as “an activity or action which is carried out as the result of processing or understanding language, i.e., as a response. For example, drawing a map while listening to a tape, and listening to an instruction and performing a command may be referred to as...
tasks. Tasks may or may not involve the production of language.” Skehan (1996) defines a task as being “an activity in which: meaning is primary; there is some sort of relationship to the real world; task completion has some priority; and assessment of task performance is in terms of task outcome.” For Nunan (1989) a task is “a piece of classroom work which involves learners in comprehending, manipulating, producing, or interacting in the target language while their attention is principally focused on meaning rather than on form. The task should also have a sense of completeness, being able to stand alone as a communicative act in its own right.” Bygate et al. (2001) argue that “a task is an activity which requires learners to use language, with emphasis on meaning, to attain a objective.”

Ellis (2003) has his own criteria to define a task in FLTL, and to differentiate it from what he calls an exercise.

1. A task is a workplan.
2. A task involves a primary focus on meaning.
3. A task involves real-world process of language.
4. A task can involve any of the four language skills.
5. A task engages cognitive process.
6. A task has a clearly defined communicative outcome.

One aspect of all definitions of tasks seems to be central: the fact that a task is an activity that forces the student to deal primarily with meaning. Willis (1996) considered TBI to be a natural evolution of communicative methods exactly because tasks promote the use of language in meaningful contexts for communication purposes.

Another important factor is that a task requires the learner to employ cognitive processes to evaluate information in order to complete it. Notice that this criterion
in itself is too vague, since it does not restrict the nature of information to be dealt with. It has to be taken in combination with the ‘focus on meaning’ requirement to be fully understood.

Authenticity is another very common criterion to define tasks. “Authenticity concerns whether a task needs to correspond to some real-world activity, i.e., achieve situational authenticity\(^\text{10}\).” (Ellis, 2003, p. 6). However, (situational) authenticity is an problematic criterion to define a task, despite the fact that many applied linguists and language teachers seem to show no restrictions to it.

The main problem with this criterion is that the notion of authenticity is based on an artificial and biased definition of what applied linguists call ‘real world’. The question is: whose real world?

Any task can be an authentic task if it is well contextualized and meaningful to the people who are performing it. At the same time, most of the tasks that are considered to be authentic can portray situations that are absolutely strange to a certain person. There are no tasks that are ‘artificial’ or ‘inauthentic’ in nature. Consider some common examples: For Long (1985) ‘painting a fence’, ‘dressing a child’, ‘borrowing a library book’ are authentic tasks because they can occur in day-to-day living. Ellis (2003) adds to this notion by stating that:

“The ‘survival tasks’, for example, filling in several kinds of official forms, which are common in ‘second’ (as opposed to ‘foreign’) language classes, are further examples of real-world tasks. However, there are many tasks that have been used by both researchers and teachers which are patently not real-world. For example, telling a story based on a series of pictures, describing a picture

\(^{10}\)This type of authenticity is called ‘situational authenticity’ to contrast with ‘interactional authenticity’, where the type of interaction between speakers is authentic (because there is negotiation of meaning), but the situation where it occurs is not.
so someone else can draw it, identifying the differences in two pictures, deciding where to locate buildings on a map are activities that language learners are unlikely to ever carry out in their lives.” (Ellis, 2003, p. 6)

Day care teachers tell children stories based on series of pictures every day, several times a day. Sketch artists spend their days listening to descriptions of people and drawing pictures based on them. Museum curators evaluate the authenticity of paintings by comparing them to photos of the original ones. Architects that work for local governments have to make decisions about where to locate new developments and build new constructions on a daily basis. Unless no one with these professions is ever going to learn a foreign language, chances are that these ‘patently not real-world tasks’ are going to be part of someone’s very real world. On the other hand, for millions of people that do not have children, ‘dressing a child’ is a very unrealistic task, and for teenagers that never go to libraries, ‘borrowing a book’ is even less ‘real-world’ task than planning a trip to the moon.

Situational authenticity is not only a difficult criterion to be rigourously defined, but it is also not necessary for meaningful interaction to take place inside or outside of the classroom. Tasks need to be meaningful, but not necessarily authentic to achieve their goals in the acquisition process. Even what may seem to be a very form-based task can be useful, if it is meaningful to the learners and if it requires them to use the target language to complete it. However, it is important to notice that situational authenticity, especially in for situations that resemble general daily activities in the target culture, may be an interesting motivating factor to engage students in exploring a foreign culture.
The working definition for *task* in the remaining of this dissertation will be of an activity (i) that can involve any of the four language skills, (ii) whose primary focus is on meaning, and (iii) that requires that the learners employ cognitive processes to search, classify, order, reason, select or evaluate information.

It is important to notice that even whole approaches based on common definitions of task may be realized very differently depending on the instructor and the instructional setting.

“It (TBI) does not constitute one single methodology. It is a multifaceted approach, which can be creatively applied with different syllabus types and for different purposes. Within TBI, programs very tremendously worldwide, as do practitioners’ assumptions about the nature of language learning.”
(Leaver and Willis, 2004, p. 3)

A detailed discussion of how tasks can be incorporated into different types of methods can be found in Leaver and Willis (2004) and Ellis (2003). For ICALL development it is important to notice that TBI is a very popular approach in the CALL community. Levy (1997) showed that almost 60% of CALL practitioners use (some sort of) TBI principles. Not only is TBI very popular with CALL practitioners, but, as Levy and Stockwell (2006) point out, tasks are at the essence of a lot of the work that has been done in CALL.

“In established CALL, language-learning task design is very much at the heart of the matter. Task design is a feature in many books and journal articles in CALL (e.g., Chapelle, 2001, 2003). Design and evaluation frameworks have been built around the task, much research has been undertaken with the task as a the focal point, and task design and structure have been written about extensively in the literature.” (Levy and Stockwell, 2006, p. 248)

Although there has been a lot of research on how to incorporate tasks into established CALL, very little has been said about the nature of tasks and task design in
ICALL. ICALL task design poses new problems for CALL practitioners because it is not restricted to the nature of the technology/media in relation to the pedagogical objectives of the proposed task. For example, how to use online chatting in communicative contexts. ICALL task design deals directly with the processing capabilities of the systems. Pedagogical considerations about the nature of the task have to be made taking into consideration the processability of the expected input, the capabilities of the NLP tools, and the ability of the system to generate certain types of feedback messages.

As discussed in chapter 1, ICALL task design has been guided primarily by the processing capabilities of systems. Little attention has been paid to the nature of the tasks proposed. In chapter 3 we will see the consequences of current choices in ICALL task design, and what can be done to put pedagogical concerns in the foreground.

2.2 Focus on Form

Focus on Form is more of a theory about the acquisition of linguistic knowledge than a method about how to teach language\(^{11}\). After the popularization of a strong version of the Communicative Approach in the late 1970’s and early 1980’s, several researchers and teachers started to observe that in instructional contexts that were purely meaning-focused several linguistic properties of the target language were never acquired by learners (c.f., e.g., Harley, 1992; Swain, 1984; Ellis, 1994). In the

\(^{11}\)The creator of Focus on Form seems to be very much averse to the idea of creating methods. “Language teacher education programs persist in presenting classrooms option to trainees in terms of methods. While many have stopped pretending that any one method is a panacea or at least that they know which one is, most nevertheless continue to use method as a unit of analysis in their professionally oriented courses, and some even give college credit for training in particular methods taught by their developers or licensed acolytes . . . Yet it is no exaggeration to say that language teaching methods do not exist – at least, not where they would matter, if they did, in the classroom.” (Long, 1991, p. 39)
late 1980’s and 1990’s, researchers started to investigate the role of small pedagogical interventions on language forms and rules during communicative activities in the classroom (c.f., e.g., Doughty, 1991; Pica, 1994). Those researchers found that focus on form “may be necessary to push learners beyond communicatively effective language toward targetlike second language ability.” (Doughty and Williams, 1998, p. 2)

The idea of Focus on Form was first introduced by Long (1991). Originally, Focus on Form was described as an “incidental attempt to draw learners’ attention to any linguistic element in context, while maintaining a primary focus on meaning.” (Mackey et al., 2004, p. 302). Spada (1997) highlighted the fact that, differently from Long’s definition of Focus on Form, form-focused instruction also included planned pedagogical interventions on form during meaning-based activities. Mackey et al. (2004) explain that, nowadays, after Doughty and Williams (1998) and Ellis (2001) have reviewed Long’s original definition of Focus on Form, the concept includes both incidental and pre-planned interventions in meaning-based activities.

For ICALL development it is important to notice Long’s original distinction between Focus on Form and Focus on Forms. According to Long (1991, 2000), while in Focus on Form activities students have to primarily focus their attention on meaning, and language forms are only highlighted if they interfere with conveying the appropriate meaning, in Focus on Forms activities learners have to focus primarily on language forms and rules to complete the task proposed. Focus on Forms activities would be similar to activities used in traditional grammar classes, while Focus on Form activities are closer to the ones used in different communicative-based approaches. This distinction is important because it shows that Focus on Form does not challenge the
importance of contextualization and meaning in the acquisition process. Focus on Form is a way to combine the advantages of meaning-based approaches with the necessary awareness of linguistic forms to help the learner achieve a desirable level of performance in the target language.

ICALL systems that provide feedback on linguistic forms have dealt primarily with Focus on Forms activities, and have not taken the challenge of incorporating meaning-focused tasks. One of the consequences is the growing skepticism by applied linguists and CALL practitioners about the possible roles of ICALL systems in FLTL. For some it may even seem impossible to incorporate ICALL activities in contextualized methodologies, or even to have systems proposing tasks (in a TBI sense):

“I will not emphasize here the role of the computer as orchestrator (or magister). This implies a view of language learning which is antithetical to the use of tasks, and is little more than an implementation of Wilkins’ synthetic syllabus: itemised and largely decontextualized presentation of materials leaving the learner the forbidding task of synthesising and applying what has been ‘learned’. This is a view of language learning which is deservedly out of favour (Long and Crookes, 1991). Worse, a computer-based implementation of this approach would be particularly inept, since the computer would lack the intelligence of the classroom teacher to make adaptations and appropriate pedagogical decisions.” (Skehan, 2003, p. 402)

In chapter 3 we will see how ICALL activities can be designed to take into consideration meaning-based approaches, and at the same time respect the processing capabilities of the system.

2.3 Awareness and Explicit Learning

One of the most debatable and controversial issues in SLA is the role of implicit and explicit learning in the acquisition process. Several books and papers have been written about it (see Ellis, 1994, for general discussion). The goal of this section
is not to extensively review the debate about this issue. It is rather to present some fundamental arguments from research that supports the claim that awareness of linguistic forms and rules is necessary for an adult learner to acquire some linguistic properties of the target language. This body of research can provide theoretical plausibility for different types of ICALL systems that use NLP technology; from systems that provide feedback on structural and semantic errors, to systems that propose contextualized activities to raise the learner’s awareness of linguistic forms.

Schmidt (1995) attempts to answer four major questions while investigating the role of attention and awareness in SLA, and examining the theoretical foundations of current pedagogical views in FLTL.

Can there be learning without intention? The answer to this question seems to be a (perhaps) unanimous yes. Anderson (1985) and Eysenck (1982) have shown that for experiment tasks, subjects learn the necessary knowledge to master the task, whether they intend to do it or not. The way the task forces the learner to process the material to be learned is what really matters. In SLA research, intentionality is not consider to be a key issue in language learning, and motivation by itself does not guarantee a successful acquisition process (Gardner, 1985).

Can there be learning without attention? Schmidt (1995) points to the orthodox position in psychology and cognitive science according to which learning is impossible without attention\textsuperscript{12}. He presents research results that show that “unattended stimuli persist in immediate short-term memory for only a few seconds at best, and attention is the necessary and sufficient condition for long-term memory

\textsuperscript{12}For a complete bibliographical references of those studies see (Schmidt, 1995, p. 9).
storage to occur.” (Schmidt, 1995, p. 9). Notice that Schmidt’s position about the role of attention is accepted by theorists that may disagree with him in terms of the role of awareness in language learning. Krashen’s Input Hypothesis claims that “we acquire language by understanding messages, that ‘comprehensible input’ is the essential environmental ingredient in language acquisition.” (Krashen, 1994, p. 46).

**Can there be learning without noticing?** First, it is important to understand that, for Schmidt, noticing means “a conscious registration of the occurrence of some event” (Schmidt, 1995, p. 29), and it is the first level of awareness. He also considers noticing and attending as isomorphic notions, which makes the answer to this question similar to the answer to the previous one. However, the topic is not uncontroversial, and there are researchers who would not agree with Schmidt’s premise. A full discussion about the notions of noticing and attending is outside the scope of this work. The reader can find more on the topic, including several bibliographical references, in (Schmidt, 1995, pp. 18–28).

**Can there be learning without understanding?** Understanding, in Schmidt’s terms, means the recognition of general rules, principles and patterns, which implies the ability to make generalizations about the structure of the target language. Awareness at the level of understanding “refers to deeper level of abstraction related to (semantic, syntactic, or communicative) meaning” (Schmidt, 1995, p. 29). There is no consensus in the SLA community about the answer to this question. The ones who argue against the role of understanding in learning (cf., e.g., Krashen, 1994; Paradis, n.d.; VanPatten, 1994) usually use arguments comparing the acquisition of L1 with that of L2, such as learnability issues related to grammar complexity or the structure
of lexical items to be acquired. The general claim to support implicit learning is based on the idea that foreign language learners can formulate the necessary abstractions to use complex linguistic forms the same way native speakers do. For example, if you ask most native speakers of Spanish to describe the use of the subjunctive, they will not be able to do so, despite the fact that they use the subjunctive.

One of Schmidt’s arguments against these claims is based on the fact that the examples used by the defenders of implicit learning do not prove what they are supposed to prove. For example, he argues that the Spanish subjunctive, English articles, and gender marking in French are notorious problems in foreign language learning for both naturalistic and classroom-based structuralist learners, and the failure of both systems counts even more heavily against arguments for the success of implicit learning. He acknowledges the existence of both unconscious processes and human abstraction, but rejects the possibility of unconscious abstractions. For Schmidt, awareness at the level of noticing is a key ingredient in foreign language learning, and awareness at the level of understanding of a subject matter can be relevant in most cases.

Another defense of the role of awareness in second language learning is presented by DeKeyser (2003):

“In spite of a large body of sophisticated research, cognitive psychologists have not been able to provide convincing evidence that people can learn abstract patterns without being aware of them. (...) SLA researchers have similarly failed to show any significant learning of abstract patterns without awareness. (...) Furthermore, L2 studies that have dealt with broader variables such as focus on form have provided evidence for the advantage of such focus compared to mere exposure or focus on meaning; the most likely interpretation of such research is that focus on form is necessary to make learners consciously notice the abstract patterns that are not easily learned implicitly.” (DeKeyser, 2003, pp. 335–336)
One of the consequences of these research findings for ICALL development is that ICALL systems can be useful if they help learners attend to certain properties of language and make generalizations of linguistic patterns. These systems can be really useful if they can perform those tasks and, at the same time, provide opportunities for learners to practice language patterns in contextualized meaning-focused activities.

2.4 Error Correction

The role of feedback in second language instruction is another area of disagreement between those that defend explicit methods of instruction and those that believe in pure implicit learning as the key to language acquisition.

Researchers who investigate the role of Universal Grammar in the acquisition process usually defend the position that negative (especially metalinguistic) feedback plays no role in second language acquisition whatsoever (see White, 2003, for a discussion of UG in SLA). However, as discussed in sections 2.2 and 2.3, there is a considerable body of research that points to the usefulness of calling the learner’s attention to linguistic forms and rules.

Different researchers have investigated the availability of negative feedback for second language learners in communicative situations. In a study with children who were non-native speakers of English interacting with children who were native speakers, Oliver (1995) found out that non-native children received some type of feedback for around 60% of the errors in their utterances. Mackey et al. (2003) showed that for adult learners negative feedback also occurs regularly in most types of second language interaction.
One of the most common types of feedback received both inside and outside of the classroom are recasts. A recast is a “response to non-target non-native speaker utterances that provide a target-like way of expressing the original meaning.” (Mackey et al., 2003, p. 36). Lyster (1998) showed that recasts were the preferred method of feedback for grammatical and phonological errors when they occur in the classroom. However, recasts were less effective for the repair of grammar errors than for the repair of phonological ones. Panova and Lyster (2002) suggest that students do not necessarily attend to recasts in communicative settings, and that more interactive types of feedback, such as scaffolding, may be more effective in making classroom learners reevaluate the grammatical structure of their utterances. For example, providing concise error messages about specific errors in student’s production, may be more effective for the learner’s evaluation of errors, than repeating the exact same sentence with the correct grammar form.

Since the majority of ICALL systems that provide feedback on student errors deal with written input, it is important to observe research results that focused specifically on written language. Ferris (2002) presents a list of key questions to evaluate the contribution of error correction for student writing. Here are some of them:

1. **What are the effects of teacher error correction on student writing?**
   - Do writing teachers give accurate and complete feedback on students’ errors?
   - Do students who attend to teacher feedback on their errors make accurate changes in their revisions?
   - Do students who receive error feedback improve in written accuracy over time?
   - Does it matter what type of feedback students receive?
   - Are certain types of errors more ‘treatable’ by means of error feedback than others?
2. **What are the effects of other types of classroom intervention on the accuracy of student writing?**

3. **What are students' views and perceptions about error treatment in their writing?**
   - Do L2 student writers value error feedback, or do they resent it and find it discouraging and demotivating?
   - Do students value feedback on errors as much as feedback on other aspects of writing (e.g., ideas, organization)?
   - What specific feedback style or mechanisms do students prefer (e.g., selective or comprehensive, direct or indirect)?

(Ferris, 2002, p. 11)

Studies on error correction in L2 writing have shown that error correction can help students improve the grammatical accuracy of their sentences in short periods of time, e.g., a semester (cf., e.g., Polio et al., 1998; Fathman and Whalley, 1990; Ferris, 1995; Lalande, 1982). Besides, other studies have shown that students like to receive grammatical feedback on their written production, and believe that it is an important tool to improve their writing abilities (cf., e.g., Cohen, 1987; Ferris, 1995; Ferris et al., 2000). Ferris (2002) reminds us that studies that criticize the usefulness of error correction in L2 writing have often shown results that were based on feedback that was incomplete, idiosyncratic, erratic, and inaccurate.

The results of this line of research indicate that feedback messages from ICALL systems that propose written exercises can be helpful for language learners. The same body of research provides some guidelines in terms of the types of feedback to be implemented in ICALL. For example, studies have demonstrated that *indirect feedback* can help students develop editing skills over time more than direct feedback (cf., e.g., Ferris and Hedgcock, 1998; James, 1998). Indirect feedback makes students self-edit their written production, which forces them to think about their errors. This
form of feedback is also called remediation, and it is considered one of the most efficient types of feedback (Reid, 1998). Direct feedback, i.e. when the teacher provides the correction to the error, is only considered advantageous when the correction process requires a specific type of knowledge that the student does not have (Brown, 1994).

There are different ways a teacher (or a system in the case of ICALL) can provide indirect feedback. The most simple one is by locating the error without providing any information about its nature. This technique gives students more responsibility in terms of error diagnosing, and provides more opportunities for reflection. This approach is claimed to facilitate long term acquisition (Lalande, 1982). Another common strategy to give indirect feedback is to provide a code that corresponds to the nature of the error, and to let the student find the error in the target sentence. This approach helps students access metalinguistic information and facilitates the development of language awareness.

Another important contribution from L2 writing research is the claim that not all error types should be treated the equally. For example, students benefit more from direct corrections for errors that are idiosyncratic and not rule governed, e.g., certain preposition usages or collocations. Indirect feedback is more relevant for errors associated with language patterns (Ferris, 1999). Ferris (2003) also suggests that more frequent errors should be treated first, since the awareness of common errors help students develop better strategies for self-monitoring.

In terms of the classification of errors to provide feedback, Ferris and Roberts (2002) argue that larger error categories that encompass most of the information about the nature of different errors are usually more useful than dividing errors into
many different discrete error types. Explicit grammatical explanations can be useful if they are short, narrowly focused, and provided in a way that will help learners identify the type of error described.

2.5 ILTS Integration and Development

The research described in this chapter provides valuable information about current practice in FLTL and relevant notions in SLA. This body of knowledge is used as a point of departure for the ICALL project described in the next chapters. It is especially important to consider such research to establish the role the system proposed is going to play in real-life settings.

Schmidt (1995) describes the three major points of view represented in FLTL today:

- The first is the most traditional, and stresses the importance of conscious understanding and study for success in learning foreign languages. In this view, mistakes in a foreign language are the result of either not knowing the rules, forgetting them, or not paying attention.
- The second major point of view is that language learning (or ‘acquisition’) is unconscious or subconscious (no one seems to make a distinction between the two terms). Language learning in natural settings (both L1 and L2) takes place through interaction and the process of input. (...) Students can achieve a high level of proficiency, including a high level of grammatical accuracy, without any explicit focus on the language itself.
- A third, intermediate view is clearly emerging in the foreign language profession. In this view, communicative, meaning-focused instruction is essential, but not all language features can be acquired when learners’ attention is focused exclusively on meaning. A focus on form appears to be necessary and desirable, especially if provided within a communicative context.

(Schmidt, 1995, pp. 2-3)

Nowadays, many believe that FLTL has evolved to a post-methods era, where no single method or approach can present a definitive solution to the foreign language
instruction problem (see Richards and Rogers, 2001, chapter 19 for a complete discussion). The third point of view presented by Schmidt is being adapted by many language teachers because it does not present a radical solution, and allows for the inclusion of different methodological mechanisms to deal with learners’ needs.

The growth of this third choice in FLTL may be seeing as evidence that ICALL systems that propose exercises for raising awareness of language forms and rules can be incorporated into common teaching practice. Nevertheless, ICALL systems that follow purely form-focused or content-focused approaches run the risk of being rejected by instructors and learners who want to find middle ground, multifaceted approaches that incorporate form and meaning. This conclusion is corroborated by Levy (1997), who found that many teachers who are CALL practitioners “consider themselves to be eclectic in the sense that they do not follow a distinct language teaching approach, philosophy, or linguistic theory” (Levy, 1997, p. 154). Section 3.4 presents the types of exercises implemented for the ICALL system developed in this work, and discusses how those exercises are in sync with the general trends in FLTL identified in this chapter.

Another key aspect in the integration of ICALL systems into current FLTL is to decide how ICALL systems can contribute most to FLTL in the current state of affairs. To do so, it is necessary not only to assess the beliefs of instructors, but also to observe their practice and the classroom settings where they work. ICALL systems have the potential to provide greater flexibility to students in terms of time and place of study. They may also take some of the burden of providing grammatical feedback from instructors. However, Weizenbaum (1984) reminds us that it is not because a computer can perform a task that it should do it. Instructors and students have to be
convinced that it is acceptable for a computer to take over certain activities, and that it is actually beneficial to the learning process. Section 3.3 will discuss the results of a series of interviews with Spanish and Portuguese instructors at Ohio State University that provide some useful insights for the development of the system proposed by this work.
CHAPTER 3

PEDAGOGICAL BACKGROUND OF TAGARELA

This chapter presents the TAGARELA system in terms of its pedagogical components. It begins with a description of the system in terms of its general characteristics and goals. TAGARELA is an ICALL system integrated into the Portuguese Individualized Instruction Program (IIP). Section 3.2 describes the characteristics of the Portuguese IIP, and discusses how an ICALL system, such as TAGARELA, can contribute to improve language instruction in such programs.

For an ICALL system to be successfully integrated into a language program, following well established pedagogical principles is an important step. However, this integration can be facilitated if teachers are actively involved in the process. Section 3.3 shows the results of interviews with Spanish and Portuguese instructors at The Ohio State University, during which the interviewees were encouraged to think about how to improve their teaching, and how ICALL systems could support the task.

Finally, in section 3.4, the exercises implemented in TAGARELA are presented. The section introduces some relevant issues in ICALL activity design, and describes the decision making process for the creation of TAGARELA activities.
3.1 The System

TAGARELA is an acronym for *Teaching Aid for Grammatical Awareness, Recognition and Enhancement of Linguistic Abilities*.\(^{13}\) As the name suggests, TAGARELA is an ICALL system whose primary goal is to help learners of Portuguese develop their linguistic abilities through language awareness and use. The system provides opportunities for students to practice their reading, listening and writing skills, and to receive personalized feedback about their performance.

TAGARELA’s activities follow the language contents presented in the Portuguese courses, and the system menu reflects the general structure of the IIP, as discussed in section 3.2. The system presents six types of activities: *Listening*, *Reading*, *Description*, *Vocabulary*, *Rephrasing*, and *Fill-in-the-Blanks*. Section 3.4 will present these activity types in detail.

TAGARELA was designed to help fill one common gap of language instruction at the university level in the US: the lack of personalized feedback students receive on their language production because of the small amounts of time instructors can spend with each individual student. Section 3.3 discusses some consequences of language instruction in academic settings in the US, where group sizes usually surpass twenty students, and describes how systems like TAGARELA can contribute to attenuate the problem. Interestingly, the same problem occurs in the IIP program, although for different reasons, as shown in section 3.2. Therefore, ICALL systems that can help maximize the amount of linguistic feedback a student receives can be of great help for both instructional settings.

\(^{13}\)The name of the system is also a Portuguese word that translates into ‘chatterbox’. Some believe that this name was chosen because it reflects a characteristic shared by the two main developers of the system, but nobody has provided scientific proof of that.
It has been suggested that TAGARELA should also be used in regular classroom settings. Although this proposal is beyond the scope of this work, it seems a natural way to proceed the development of the system, as discussed in chapter 6.

3.2 Portuguese Individualized Instruction

To understand how TAGARELA can be used by IIP students, it is important to understand the structure of the program, its advantages, and limitations.

The Ohio State University has the oldest Individualized Language Learning Center in the US (Harlow, 1987). Foreign languages have been taught at the center since 1976. Its facility is equipped with different types of hardware and educational software.

IIP is a self-paced, mastery-based program that mirrors classroom courses in its structure. Students can take a flexible number of credits (minimum of 2 per quarter), which allows them to start and complete a course (5 credits) in different quarters. Students can also take more than one course per quarter, i.e. enroll in more than 5 credits.

In IIP, students set their own pace of learning. The Portuguese program offers beginning (101.51 and 102.51) and intermediate (103.51 and 104.51) level courses. Each course is divided into five modules. Each module corresponds to one credit hour, and it usually reflects the content of one unit of the coursebook used. To complete a module, students have to (i) study and do the exercises in the workbook and in the supplementary course packet, (ii) write a composition, (iii) take an oral and a written exam, and (iv) attend one conversation session. The lowest passing grade is 80% (letter grade “B”), and it is calculated as the average of the composition, the

---

14The activity menu of TAGARELA is divided by modules to follow the IIP structure. See section 3.4.3 for a detailed explanation.
written and the oral exams. All work in IIP is done by the student at home and on appointments with the instructor. Students need to set deadlines to finish their work. The number of times students see an instructor depends on how quickly they work and how much additional help they need. Normally, meetings with instructors are used to answer specific questions about the material, grade compositions and exams, take oral exams, and discuss general progression.

In this type of program, there is little time for the instructor to provide feedback on exercises. In fact, feedback is provided by students’ request. There is also little time for instructors to observe students’ performance in activities like listening and reading comprehension. As Silva and Amaral (2005) have shown, one of the challenges when establishing IIPs is to create materials that allow students to practice the target language and receive feedback on their production. The Portuguese IIP at OSU tries to compensate by making mandatory students’ participation in regular conversation sessions. However, feedback on their written production is still limited to either assessment tools (exam and composition) or to questions students ask about their take home exercises.

As suggested by Papen (1980) and Shedinvy (2004), the success of a student in an IIP lies on self-motivation and self-discipline. However, students that have developed specific strategies to spot problematic linguistic areas perform much better in this type program, because they are aware of the problems in their production and can point them out to the instructor. Students that do not have this ability usually struggle much more, since there are fewer opportunities for the instructor to observe
difficulties and monitor their progress. TAGARELA was designed to provide more opportunities for learners to receive individualized feedback and to reflect about their language production.

3.3 FLTL Integration

As chapter 1 emphasized, to design ICALL systems to be integrated into FLTL, it is necessary to observe the beliefs and practices of professionals in the field. Chapter 2 presented relevant concepts in SLA introduced by researchers in the area. This section describes the result of a survey with instructors conducted to better understand the current practice in language classrooms, and to identify where instructors would be willing to incorporate ICALL systems in their everyday life. Even if TAGARELA is currently only being used in an IIP, it is still important to listen to instructors to learn more about possible design choices for the system. If the final product is not appealing to them, the system has fewer chances of being fully integrated into any language instruction setting.

3.3.1 Why ask teachers?

Masuhara (1998) reminds us that since the 1980’s the learner has been the focus of most SLA research, both theoretically driven and applied studies to develop new pedagogical approaches. The role of teachers has received little attention, and the few studies that exist tend to emphasize the teacher behavior as another component of the method used.

“An investigation of the psychology of teachers in relation to security and involvement could reveal a lot about teaching which enhances language learning. An investigation into teaching styles and preferences could reveal how teachers react to and implement teaching materials. However, teachers seem to be
treated in both language learning and teaching studies as passive beings who are expected to adapt flexibly to the roles determined by the objectives of the method and by the learning theory on which the method is based.” (Masuhara, 1998, p. 239)

When designing materials to be integrated into language curricula, there are at least two reasons to investigate teachers’ opinions and behavior. The first one is based on the power teachers have to influence the choice of materials. Every year FLTL materials, such as books, CDs, films, and software, are no longer used by learners not because they became outdate or did not follow the latest trends in teaching approaches, but simply because teachers decide not to adopt them. Their decision may be based on several different factors, including the role of the material in the classroom dynamics (e.g., the rejection of bilingual dictionaries for advanced students), teachers’ choices of approaches (e.g., the use of books with drills for communicative classes), the nature of the material (e.g., the replacement of tapes by cds), or even political reasons (e.g., one’s goal of pleasing the author of a coursebook). Teachers’ opinions about a given material can facilitate or prevent integration and acceptance of the material into FLTL.

The second reason to listen to teachers’ opinion is because they can inform the ICALL designer and researcher about some details of everyday instruction that SLA researchers are not aware of. Their experience with language students allows them to observe the type of activities students like and dislike. They can also observe the types of material that help students perform certain tasks, and the details about the design of that material that are more useful to students.

ICALL has been one of the areas where integration with FLTL practice has been seen with skepticism by language teachers (Levy, 1997). One of the main reasons has
been the gap that exists between the expectations of teachers for ICALL, and the actual realization of systems. As Garrett (1995) shows, it is important to take into consideration teachers’ opinions when making design choices for ICALL systems, and deciding on the role these systems will play.

### 3.3.2 Instructional setting

Nowadays, it is not difficult to enter two foreign language classrooms and find out that the only thing in common is that a foreign language is being taught. The number of methodologies available allows the instructor to choose among several different types of approaches, which is reflected in numerous different types of practice. Thus, the task of describing a typical classroom procedure becomes almost impossible. However, to identify possible roles of ICALL systems in FLTL, it is necessary to describe general patterns in language instruction that are (more or less) accepted by most teachers.

In this study, three general characteristics of contemporary language instruction were taken as common to most popular approaches and methods in FLTL. These three characteristics were used as guidelines during the interviews with instructors to elicit their opinions about possible roles of ICALL.

#### Development of Specific Skills

When choosing activities for their lessons, a main concern of instructors is to know how a specific activity can help students become better users of the target language. In doing so, instructors search to rely on some type of model of abilities speakers of

---

15 Many of the misconceptions about what intelligent tutors can in fact do is the result of years of exaggerated promises made by the A.I. community about the development of computers that can act like humans. It has not been different for ILTSS development. There have been several claims about computers that “understand” human language, and little realization of what non-experts believe “understanding human language” means.
a given language have to develop in order to be considered proficient. There are four skills that are considered the basic ones to be developed by foreign language learners: speaking, writing, listening, and reading. The first two are usually referred to as productive or active skills, and the latter are normally called receptive or passive skills. Besides these four basic skills, depending on the methodological approach adopted some other skills are considered to be important, such as: communicative skill, cultural awareness, etc.

The choice of a given activity is then influenced by its possible contribution to the development of one or more of the basic skills. In terms of activity design, this aspect should be taken into consideration by ICALL designers, since it will influence the suitability of a given tool for integrating classroom material.

**Goal-oriented Lesson Plans**

When thinking about their lessons, instructors usually decide on the type and sequence of activities based on the aims of their lesson. In a lesson plan there are usually two types of goals: the general lesson goal, and the specific subgoals to be achieved in each part of the lesson. The description of goals is strongly influenced by the methodology chosen by the instructor. It is impossible to give an example without being biased, since the goal of the lesson is what guides the whole choice of activities and students’ expected production.

Although the general goal is a key element to a successful lesson, the subgoals of a lesson are more important to the aim of this research. The subgoals are the ones that can show to an ICALL designer what types of activities can be incorporated into the lesson. By checking the subgoals of a given lesson, one can decide where in that lesson a computer tool can be better used. Moreover, by understanding how
the subgoals influence the final result of a lesson, it is also possible to better justify the reasons to use ICALL tools in a foreign language class, and propose activities for ICALL systems.

**Stages of a Lesson**

In order to achieve the general and specific goals for a given lesson, instructors tend to divide the work in different stages, so that they can have more control of language input and students output.

There is some controversy about how these stages should be established, the need to separate some of the stages, and where the focused practice should be. Despite these divergences, most instructors agree that students should not be requested to produce a given construction if they have never been exposed to any type of input that contained it (in more communicative settings), or have never received any kind of grammatical explanation about its usage (in more structuralist settings). Therefore, the idea that students’ production should follow some previous preparation is well accepted. The general disagreement about lesson stages is about what should precede the production stage.

Because the aim of this inquiry is to find out opportunities to incorporate ICALL systems, the PPP (*Presentation, Practice, Production*) approach was chosen as a schematic procedure to a hypothetical description of a general class. This choice does not imply that the ICALL system design presented below follows PPP principles. It simply means that PPP provides a structured description of classroom procedures that helps identify situations where ICALL can be beneficial.
The general idea behind the PPP is that a successful lesson, i.e. a lesson where the general goal is achieved, should follow three steps: presentation, practice, and production. Each of these steps can be divided into sub-steps (e.g., controlled practice, less controlled practice, freer practice), and can be done in different ways (e.g., productive presentation - with elicitation - or receptive presentation - teacher’s explanation). Maurer (2004) explains the three steps in the following way:

- **Presentation** means that, before we expect students to use language or structure, we present it to them. This is true in an ESL class in an English-speaking society, where students are surrounded by the language. It is all the more true in an EFL class where English language input for students is probably much less. Thus we need to provide students with a variety of models and contexts that will give them the input they need to be able to use the language productively.

- **Practice** means that, after we present the material that we want our students to learn but before we expect them to use it productively, we give them ample opportunities to practice it. That is, practice comes before the testing situation.

- **Production** means that, having presented the language that we want our students to learn and given them opportunities to practice it, we may now justifiably expect them to produce it – to use it more or less freely in real, largely uncontrolled situations.

### 3.3.3 The Interview

In 2004, a series of interviews were conducted with instructors in the Department of Spanish and Portuguese at OSU to collect information about their classroom practice. The aim was to find out where, when, and how ICALL could be used in accordance with the teaching practice employed.

In order to avoid the problem of “inflated expectations” described by Atwell (1999)\(^{16}\), there were no specific questions about how instructors wanted computers to

\(^{16}\)Atwell describes how language teachers may have unfeasible expectations towards CALL tools, and how it makes it difficult to collect ideas about developing these tools with teachers.
help them. The aim of the interviews was to make instructors talk about their classes. They were asked to analyze the stages of their classes and point out specific steps that were problematic in terms of pace, goal achievement, participation, student centeredness, elicitation, classroom management, language practice, students’ production, and communication. Then, they were asked to make a correlation between those problematic steps and the practice of specific topics. They were encouraged to analyze if these issues were related to individual differences, and if some sort of previous and/or remedial work done individually with students could help diminish the problems. During the last part of the interview, based on what instructors said, they were given possible descriptions of some computer-based tools, and were asked how they felt about the idea of asking students to do those activities in the computer, and what stages of their lessons they would not like to have a computer tool interfering with.

The results showed the stages of the lesson where CALL exercises would be most successfully integrated, and when teachers do not believe it is appropriate to have computer tools interfere with their lesson plans.

Looking at the three stages of lesson (as described in section 3.3.2), and based on teachers answers, it is easy to point to the practice stage as showing the best possibilities of interaction between the classroom activities and ICALL exercises. Teachers said that in their classes they do not like to spend much time in activities that are meant to reinforce grammar patterns. They say this type of activity can be problematic because they reduce the pace of the lesson, since individual differences make it impossible to have all students do the same tasks in exactly the same time. They also said that for some students a basic productive presentation, with a lot of elicitation, is enough to make them engage in meaningful practice, while in other cases some
students need more time and more input before they are able to produce appropriate (and accurate) sentences. Besides, it is difficult to make every single grammar practice meaningful and suitable to communicative purposes, especially because communicative activities tend to have a certain degree of complexity, and some instructors would like to have their students (or at least some of their students) focus on form and usage for a while before integrating that language item into a more complex context where students have to focus on negotiation of meaning.

Ultimately, the problem is that instructors believe that their lessons should be communicatively appropriate, but they usually have problems when grammatical accuracy gets in the way of meaningful communication. When students start producing sentences that make no sense in the target language due to structural errors related to topics that have been seen before (e.g., wrong word order, inaccurate morphology, inappropriateness of functional words, misuse of tenses), instructors face a difficult dilemma: to go on and keep doing communicative activities, or to stop and address the errors in students’ production. If they decide to go on, they might have to face a situation where students have more doubts about how to carry on the task, do not understand the purpose of the activity, and stop participating. On the other hand, if they decide to stop and address the errors, they fear that the communicative intent of the activity could be lost. As Mackey et al. (2004) have shown, it takes significant experience to properly use incidental focus-on-form in language teaching instruction.

Teachers usually find it particularly difficult to deal with errors with structures that have been presented previously, and are not the focus of the current lesson. Although instructors have to deal with this type of situation every day, remedial work is not a very popular topic in teacher training programs, and, as discussed in
chapter 2, some SLA researchers do not believe that negative feedback can play a role in the acquisition process. Even in situations where the teacher has full control and addresses those issues in an ideal way, the problem of reinforcement will always exist, since the needs of each individual student are unique, and some learners need more time, more input, and more chances to review and practice the target language than others.

Another area that instructors pointed out as problematic to classroom dynamics is integrating the practice of receptive skills into lesson plans. They said listening and reading activities sometimes create problems in a lesson. Reading activities are problematic because students read at different paces, and if they ask students to read at home, and plan activities based on that, they risk having a lesson plan that does not work, either because students are reading with no help and may have problems understanding the text, or simply because they do not read the text. Going over the text in class is not really a good idea, since it can be time consuming and students may benefit very little from that. They all agree that it would be good to assist students individually while doing a reading activity, but that is not feasible in most classes with more than 20 students. Listening activities are not very different, and they can be even worse in big groups, where some students may understand the text the first time they hear it, while others may need to listen to it several times. Again, coping with individual differences is not an easy task.

Although exercises dealing with receptive skills present some problems to classroom dynamics, they are very important to the overall process of learning a foreign language. Without the appropriate input, students may never be able to build up the cognitive apparatus necessary to be proficient users of the target language.
A third area that instructors pointed out as problematic is the design (and the amount) of activities that helped students incorporate new vocabulary into their practice. Books in the U.S. usually bring lists of words (sometimes) with their translation, some activities are proposed to use those words, but students never have enough time to practice, and the activities are sometimes not very interesting. Besides, instructors complain that in some units in textbooks there is no clear (or easy) connection between the vocabulary and the grammar topic. Students try to compensate for that deficiency with flash cards, and mechanical memorization without any context. Instructors say that activities that present vocabulary in a creative and contextualized way and that integrate vocabulary with structure would be welcome.

By the end of the interview instructors received different descriptions of ICALL systems that provided feedback on language forms. They were then asked about situations where those systems would not be useful and/or helpful. Most instructors pointed to the production stage where activities such as discussions and role plays occur. A great number of instructors believe that activities at this stage are better done with the instructor and other students, since the presence of other human beings is a key element to make them fulfill their goal. Learning to communicate in a foreign language implies learning to negotiate meaning, understanding social behavior, observing different body language strategies, in summary, it means dealing with other people; instructors were sceptic about the potential of ICALL systems in replacing humans in this respect.
3.3.4 Results

Based on the results from the interviews, it is possible to make a list of areas where ICALL activities are more useful and can have a greater impact on the learning process.

First, ICALL exercises can be used for reinforcement. Usually, reinforcement is done through workbook activities, or some extra material. The advantages of doing it using an ICALL exercise is that the student can have access to immediate and individual feedback, which will help prevent repeated errors throughout the exercise, and help diminish the time spent in class with correction.

Second, ICALL activities can be used for remedial work. This is an area where ICALL can be extremely helpful to language learners. The need for remedial work appears when students start making repeated errors with constructions that should not present problems at that specific level. At this point students need to be exposed to accurate forms of input, and to practice the problematic structures, review errors, compare answers, and edit their own production. In some cases, they may need different types of activities, or maybe need to redo the whole process from presentation stage. Some types of errors associated with negative transfer could be the target of some activities that point out differences between L1 and L2. Remedial work has to be done as early as possible, it has to be compatible with the level of the student, it has to integrate appropriate feedback, it has to be done on an individual basis, it has to reach for the biggest number of skills as possible, it has to give the chance for the student to think about the problem, and it has to give an extensive amount of practice on the targeted structure. ICALL design can be explored in many ways to
create activities to deal with remedial work. Depending on the source of the problem, different techniques have to be used, which will have a great influence on technical aspects of what modules are necessary for the system to process students’ input.

Third, ICALL systems can be used to facilitate the enhancement of linguistic knowledge and the awareness of language structures. Classroom settings sometimes do not provide opportunities to discuss some linguistic issues that help students increase their awareness of linguistic structures and develop their linguistic competence. With appropriate feedback and interaction between what is seen in class with what is done at the computer, students can have the chance to review language concepts and develop a better understanding of the linguistic phenomena involved.

Finally, ICALL systems can be used to help instructors plan a more efficient lesson, taking some of the burden of certain activities from the lesson plan into the individualized practice. This would be specifically useful in the case of instructors that would like to make students have further practice with activities that develop their receptive skills, at the same time provide further practice of the targeted language topic. Pre-listening and pre-reading activities, for example, are two cases where ICALL systems could help learners prepare for further work.

### 3.4 Types of Exercises

The SLA research described in chapter 2 and the survey with teachers from the previous session provide important information to guide the development of ICALL activities. However, there are other factors that have to be taken into consideration in ICALL activity design.
So far we have presented important pedagogical issues that can influence ICALL system development, but nothing has been said about the technological aspects of this enterprise. As discussed in section 1.5, ICALL system designers have to find a balance between pedagogical specifications and processing limitations. In practice, ICALL developers tend to either focus on the development of the technology and not take pedagogical issues seriously into consideration, or make many pedagogical concessions to develop systems to work with a given technology.

This section presents some of the issues that guided the design and implementation of the activities in TAGARELA. Some of these issues are central to the approach advocated by this work.

First, TAGARELA is not designed to deal with 100% unconstrained free input. It is neither relevant nor necessary for ICALL integration that systems can deal with unrestricted input for all possible contexts. Not even human beings deal with this type of input. In human interaction, interlocutors have expectations about their partners' utterances, and their processing mechanisms rely on these expectations. One way linguists describe such expectations is by using Gricean Maxims (Grice, 1989). Violations of some of these maxims can generate incomprehensible input for human beings. TAGARELA’s expectations about the input come from its activity model, which not only provides information about the nature of the expected input, but also advises the system on how to process that input. Chapter 4 presents a full description of TAGARELA’s instruction model, which contains its activity model, and section 3.4.1 discusses why it is important to constrain student’s input from a processing perspective.
Second, activity design takes seriously some pedagogical guidelines based on the SLA research presented in chapter 2 and the teachers’ survey discussed in section 3.3. Section 3.4.2 presents these guidelines in detail.

Third, although informed by the availability of NLP technology, the development of TAGARELA’s processing components is not technologically driven. Several ICALL systems were created to use a previously implemented parser, or to justify a given choice for syntactic processing of ill-formed input, such as constraint relaxation or the use of mal-rules (see section 1.5 for references). As chapter 4 shows, TAGARELA’s processing tools were designed to suit the processing needs established by the types of activities implemented and by the feedback strategies employed. In fact, at this stage, TAGARELA does not have a full-blown parser. The system does syntactic disambiguation and morpho-syntactic error diagnosis using a combination of shallow processing and feature comparisons based on detailed lexicon entries (see chapter 4 for details).

Finally, feedback strategies used by TAGARELA do not prioritize morpho-syntactic analysis. ICALL systems that propose form-based activities limit their feedback messages to morpho-syntactic properties of the learner’s input. TAGARELA has NLP modules to check the semantic appropriateness of the input which allows the system to prioritize meaning-based over form-based feedback. It does not mean that TAGARELA does not deal with morpho-syntactic errors. Much to the contrary, form-based errors are diagnosed and reported, and especial emphasis is given to situations where linguistic forms can interfere with the meaning of the utterance. The feedback strategies used by TAGARELA are discussed in section 4.9.
3.4.1 Processing Considerations

As noted in section 1.5, ICALL systems that have been successfully integrated into FLTL have mechanisms to constrain learner input so that expected students’ entries are processable by their NLP components. Before further considerations on possible consequences of these restrictions on ICALL activity design, it is important to understand why these restrictions are necessary to develop reliable systems.

A key aspect in the development of efficient ICALL systems is the reliability of their feedback messages. Research has shown (Tschichold, 2003, 1999) that, differently from writers-aids tools for native speakers, ICALL systems have to be precise and reliable when providing feedback. Most grammar checkers for native speakers rely on the user’s evaluation of their analyses to confirm if the modifications suggested fit the text in terms of style and grammatical accuracy. ICALL systems, on the other hand, are expected to behave as experts that provide unequivocal analysis, and precise feedback to the learner. It is important to notice that from the learner’s perspective it is better for the system to provide no analysis for a given input, than to provide a wrong analysis that is taken by the learner as a correct one. Overflagging errors is a problem system developers have to deal with, since it should be avoided at all costs.

Although NLP technology has advanced a lot in the last two decades, it still has its limitations, especially when dealing with ill-formed input, such as in the case of ICALL.

The first set of limitations of NLP technology that affects ICALL is related to morpho-syntactic processing. In order to deal with recursive syntactic structures, a parser has to handle structural ambiguities that are common in natural languages,
such as PP attachment. When processing ill-formed input, the parser has to deal with an even greater number of possible ambiguities, since it has to generate all possible parses using either a greater number of rules (e.g., augmenting the grammar with mal-rules), or limiting the scope of restrictions imposed on its original set of rules (e.g., relaxing grammatical constraints). A great number of rules can expand the search space to point that it prevents the parse from diagnosing the errors in an ill-formed sentence. Besides syntactic ambiguity, NLP technology for ICALL has to propose new solutions to deal with lexical ambiguity. Word choice by foreign language learners can pose intractable problems to error detection, since it can further complicate syntactic and semantic processing.

The second set of limitations comes from semantic processing. For an ICALL system, it can be more difficult to determine if an answer is correct (in terms of meaning) than to determine if it is wrong (in terms of form). Most NLP technology used in ICALL deals with morpho-syntactic processing. Current systems check for spelling, agreement, subcategorization, and other types of errors that a spell-checker and a parser can identify. Meaning-based processing is usually restricted to string or token matching. This limited (or non-existent) semantic processing may cause problems to evaluate if an answer is appropriate to a given question. Problems not related to lexical, morphological, or syntactic properties are rarely identified and treated. On the other hand, developing world knowledge to make a system fully aware of semantic and pragmatic features is still not feasible.

This situation shows that ICALL systems need to incorporate some sort of mechanism to process the student’s input in terms of meaning. Because trying to develop a system that can deal with unrestricted knowledge about the world would be as
difficult (or as impossible) as developing a system that can deal with all types of ill-formed syntactic constructions, ICALL designers have to find ways to restrict the search space of semantic possibilities to ensure the processability of the input in terms of semantic appropriateness. Usually ICALL developers impose these restrictions by designing activities that are purely form-based and providing explicit information about how linguistic forms should be manipulated, such as: ‘conjugate the verb in present simple’ or ‘ask the question by translating the following sentence’.

Section 3.4.3 presents some ideas about how to avoid this type of explicit form-based restrictions in activity instructions. However, most importantly, these restrictions can be attenuated by providing more information about the activity to the system, and by incorporating NLP tools to deal with meaning. Chapter 4 shows how the work developed by Bailey and Meurers (2006) with shallow semantic processing for ICALL is incorporated into TAGARELA to allow the system to draw inferences about the appropriateness of students’ input.

3.4.2 Dealing with Activity Specification Issues

As noted above, the content of ICALL activities is heavily influenced by the processing capabilities of the system. It is necessary for the ICALL designer to specify the limits for this influence. It means that, it is necessary to establish the basic characteristics of ICALL activities that will not be changed to cope with processing limitations. These characteristics are usually motivated by pedagogical principles that can guarantee a sound integration of the system into FLTL methodologies. The first step to create guidelines for ICALL activity specifications is to identify the properties to be avoided for any given activity.
Chapter 2 has presented the debate in SLA about the role of meaningful interaction in the acquisition process (cf., e.g. Krashen, 1987, 1988b; DeKeyser, 2003; Long, 2000; Ellis, 2005). One of the key ideas emphasized in the chapter is that feedback on form is more efficient if the learner is engaged in tasks that demand the manipulation of meaning. Based on this principle, activity types that deal exclusively with translation and dictation have received severe criticism by language teachers that adopt more content-based methodologies. Similarly, activities with repetition or substitution drills, and decontextualized fill in the blanks have been left out of teaching materials for their lack of communicativeness. It is important to acknowledge the fact that many language teachers and policy makers have strong restrictions against these type of activities, and that in order to increase the acceptability of ICALL systems it is necessary to avoid mechanisms to constrain learner input that make the activities proposed look meaningless.

As discussed in section 2.1, it is important to notice that making activities meaningful, does not always mean complying with authenticity criteria. Authentic activities require students to use world knowledge to produce authentic solutions for tasks that mirror the priorities and challenges of real life situations (see Wiggins, 1998). ICALL activities have to be valid, i.e. they have to allow us to “infer real performance results for specific standards” (Wiggins, 1998, p. 141), but they do not have to be authentic.

The second step in the specification of guidelines for activities is to decide which activity types will help achieve the pedagogical goals of the system. In the case of TAGARELA, there is a clear need for activities that expose students to original listening passages. There is also a need, identified by instructors, to make students
practice reading comprehension skills. Moreover, there is a need for on the spot feedback while students practice certain written skills, such as describing a picture, and rephrasing sentences. Some properties of the target language should also be taken into consideration. For example, Portuguese is a language that presents rich verb morphology, nominal and adjectival inflection, and several contraction patterns of prepositions with nouns, adverbs, and pronouns. Some activity types can be proposed to deal specifically with these problematic linguistic properties for language learners.

The last step is to decide how to design the activities so that they become technically feasible. In other words, we have to make sure that the expected input is processable. This is not a trivial step, because decisions made here have to take into consideration the processing capabilities of the system. It means that design choices for activities and for the system architecture have to walk hand in hand.

In sum, the final decision about activity design has to take into consideration pedagogical and technical restrictions. In TAGARELA, we avoid translation, dictation, or any other method to restrict students’ input that could present problems to the methodology used. We also try to minimize the use of L1, whenever possible. At the same time, we constrain tasks so that the expected input is in sync with the system capabilities. To do so, we make extensive use of page layout, cognates, and consistency in task specifications, as we will see in the next section.

3.4.3 Design Choices for TAGARELA

TAGARELA activity design follows the guidelines described above. Its activities are divided into six different types: reading, listening, description, rephrasing, fill in the blanks, and vocabulary. These types of activities are commonly found in paper-based
workbooks and can be easily designed to content-based methodologies. TAGARELA makes use of page layout to avoid some of the problems described above, such as excessive use of L1 and long, tedious instructions. Each activity type has one specific page layout that matches the specifications of the tasks proposed. Different parts of the page have different functions that help students understand what is expected from them. Moreover, different page colors and icons are used for each activity type, so that students have visual cues about the nature of the task. Notice that the layout of the reading comprehension activity in figure (3.2), for example, share some similarities and differences with the listening comprehension in figure (3.1) and description in figure (3.3). They all present the same sequence of information: instructions; body of the activity; question items. However, their layout emphasizes the different tasks.

English (L1) instructions are not avoided, but students always see instructions in Portuguese (L2) first, and if they want to read them in English they have to place their mouse point over the American flag. As soon as they move their mouse elsewhere again the Portuguese instructions return to the screen. All buttons are in Portuguese and page layout is used to help learners understand the function associated with each button. As section 4.9 discusses, feedback messages at this point are provided in English, but when more advanced modules are incorporated into the system, they will gradually shift into Portuguese as well.

An activity menu is presented to students after they login. The menu is structured using a hierarchical structure with course level, module (modules 1 to 5), and type of activity (the six types listed above). Students have to choose their level and module, and then they can decide what type of activity they are going to do. They are encouraged to do all activities in a module, but they can decide the order in which
they are going to do them. The reason to allow students to choose the order of activities is based on the IIP type of interaction. Students in this program receive the material and the tasks, and they can decide how they are going to cover them. To use TAGARELA in a classroom setting, instructors might have to determine the order students should do the activities, so that the homework progression follows the contents covered in class.

Once students have chosen the type of activity they will do, the system displays the first activity listed corresponding to the level and module of the student. On the top-right corner of the page, a menu allows students to choose a different activity or a different module. If they want to choose a different activity type, they can use the menu on the top of the page. In the same menu, they also find links to the home page, and to logout. Students can also skip individual questions for a given activity using the menu for questions provided on the top-right corner of the question field. Similarly to other ICALL systems (cf. e.g., E-Tutor, Heift, 2003), TAGARELA has a sequence of buttons with diacritic symbols for capital and low letters. Clicking on one of those buttons inserts the accented character at whatever place the cursor is positioned in the input field (as opposed to only at the end of the input).

Listening comprehension activities (Figure 3.1) have a yellow border, and show an icon that looks like a human ear. They display a space where instructions appear followed by a visual element, which is usually a picture or a video. Below the visual element, the page is divided into two. On the right side there is a space for questions and on the left there is an area where feedback messages and help buttons appear. Listening activities usually present 'wh-' questions where students have to identify specific pieces of information from the listening passage to provide an appropriate
Figure 3.1: Listening Comprehension Exercise
answer. Expected answers are always in the form of full sentences, and sentence fragments are usually not accepted. Listening passages are usually of three types: dialogues, monologues, or descriptions. They use vocabulary presented in the lesson, as well as new content words related to the topic. The goal for this type of activity is to provide further opportunities for learners to improve their ability to listen for specific information.

Reading activities (Figure 3.2) have a similar layout to listening activities, i.e. instructions are followed by a text, and at the bottom students find the questions to the right and the feedback box to the left. Reading pages have a purple border, and the action buttons are exactly at the same place as in listening activities. Reading activities also target the learner’s ability to find specific information to answer (most of the time) ‘wh-’ questions. Texts often introduce new vocabulary items that are semantically related to the topics explored in the coursebook. Texts appear in different styles, such narration, description, and argumentation.

Description pages (Figure 3.3) have a red border and similar layout. Below the instructions there is an area that presents both the picture to be described (together with vocabulary items to be used, if necessary) and the question filed. Description activities usually present situations, people, or scenes students have to describe. The input expected is usually a full sentence, and students are discouraged to use fragments.

Different from description activities, vocabulary activities (Figure 3.4) can elicit input messages that range from a single word to a complete sentence. Instead of having to give a description of what they are seeing, students are encouraged to provide the lexical items necessary to complete the task. For most exercises, students
Figure 3.2: Reading Comprehension Exercise
Figure 3.3: Description Exercise
Figure 3.4: Vocabulary Exercise
have to write a nominal phrase that includes the name of the item in the picture within an appropriate construction. Sometimes, students are asked to write full sentences that reflect situations studied in the coursebook. Vocabulary activities have an orange border, and their layout is similar to the one used in description activities.

The last two types of activities have a more direct focus on language forms than on content. The goal of rephrasing activities (Figure 3.5) is to make students use different syntactic constructions with no or very little change in meaning. Students should be able to correlate two linguistic forms, or to use the linguistic forms introduced to produce appropriate sentences that share some semantic properties. Another type of exercise used in rephrasing is typical unscramble of words to make sentences that correlate with the topic. Rephrasing pages have a gray border, and their layout is made of instructions at the top and two parts at the bottom, one with the question item and the other for the feedback message.

Fill in the blanks (FIB) (Figure 3.6) are used in TAGARELA to cope with the need for activities that target specific properties of Romance languages, such as rich inflectional morphology. FIB are used to help students increase their awareness of some verb forms, nominal and adjectival agreement, and contractions of prepositions and other parts of speech (articles, pronouns, and adverbs). Whenever possible, items to be used are given as part of a list and out of order, so that students have also to pay attention to the meaning of the utterance while completing the sentences. The input expected is always a word, never phrases or sentences. FIB pages have a green border, and their layout follows the same specifications for other activities.

As noted before, elicitation of input for all activities is never done through translation or dictation. TAGARELA uses other techniques to constrain students’ input,
Figure 3.5: Rephrasing Exercise
Figure 3.6: Fill in the Blanks Exercise
without jeopardizing the necessary content manipulation proposed by the tasks. The system uses pictures, lists of words, contextualized listening passages, gap-filling, written cues in L2, or a combination of two or more of these techniques to limit possible plausible answers per question item. For example, in the description activity in figure 3.3 the student has to describe the hotel room she sees in the picture using one of the expressions of place in the list provided with the instructions, and the words “vaso” (vase) and “mesa” (table) provided with the picture.

3.5 Conclusion

This chapter presented the pedagogical setting in which TAGARELA is integrated, and the pedagogical considerations behind the types of exercises used by the system. It is important to notice the emphasis that on giving activity specification in terms of having pedagogical goals in sync with processing capabilities. The discussion presented in the next chapter brings a description of how activities and NLP capabilities are closely connected. Also in the next chapter, we find a description of the implementation of the system. The emphasis changes from the necessary characteristics for an ICALL system to be acceptable by foreign language practitioners to the necessary components to realize such a system and make it fully functional.
CHAPTER 4

THE TAGARELA SYSTEM

In the previous chapter we saw the description of the exercise types in TAGARELA, together with the pedagogical motivation for the use of such exercises. In section 3.4.1, we also saw that activity design for ICALL has to take into consideration the capabilities of the system’s processing modules, i.e., system and activity design have to walk hand in hand.

This chapter presents the components of the TAGARELA system that were designed to fulfill the needs of the activities previously described. We begin the chapter by providing a brief overview of related approaches in ICALL in section 4.1. Section 4.2 presents the rationale behind the way TAGARELA processes the student input, and how the general mechanism it uses to diagnose errors works. Section 4.3 presents an overview of the system architecture, while sections 4.4, 4.5, 4.6 and 4.7 present the different components of the system. Finally, section 4.9 presents some examples of feedback messages given by the system.

All components presented in this chapter are fully implemented as part of version 2.0 of TAGARELA. This version is currently being used by students in the Portuguese individualized instruction program at The Ohio State University.
4.1 NLP in ICALL – Related Approaches

One of the main difficulties in situating a work in ICALL is due to the very different goals of existing or previous research projects. Section 1.5 introduced some of these differences, and described some of the approaches used in ICALL research. Not all ICALL projects propose to develop tools to be integrated into language programs. As seeing in section 1.5, some projects aim at developing new NLP techniques or exploring how current techniques can be extended to deal with ill-formed input.

In this section we situate the nature of the NLP processing of TAGARELA in relation to other systems by briefly describing how some systems deal with student input. The selection of those systems was primarily based on three criteria that could indicate that the overarching goals of the projects bared some similarities with the goals of the TAGARELA project.

First, we took into consideration the ability of the system to deal with a variety of linguistic properties that are commonly seen in student input. This criterion excludes systems that were designed to deal specifically with one property of language, such as the Passive Voice Tutor (Virvou et al., 2000), and the Portuguese Clitic Pronoun Placement tutor that accompanies Mr. Collins (Bull et al., 1995).

Second, we selected systems proposed similar types of exercises used in TAGARELA. The idea is to exclude systems that function as writer-aids, such as ICICLE (Michaud et al., 2001), since those systems play a different pedagogical role that is not necessarily comparable to the one of TAGARELA, and systems that focus on role playing, such as Herr Kommissar (DeSmedt, 1995), because those systems usually do not incorporate NLP components for morpho-syntactic analysis, such as in TAGARELA.
The last criterion is the availability of published materials about the NLP components of the system that allows for a general comparison with the NLP components of TAGARELA. This criterion excludes systems like Robo-Sensei (Nagata, 2002), for which little is available about its NLP capabilities.

### 4.1.1 Intelligent Tutor of Academic English

Dodigovic (2005) presents the Intelligent Tutor Of Academic English, whose goal is to help students develop their writing skills for writing texts in academic English. The system presents one type of exercise that resembles a rephrasing activity. It shows the learner an excerpt from a text usually made of one or two sentences, and then asks the student to answer a question using the information from the excerpt.

The parser uses a top-down, left-to-right algorithm (Matthews, 1998) to parse the student input. The parser uses a phrase structure grammar augmented with feature specifications. The grammar has a set of “correct” grammar rules, i.e., grammar rules that combined can generate accurate sentences in the language, and a set of mal-rules, i.e., rules that allow for variations in the grammatical pattern of the language. The version of the system described in Dodigovic (2005) has a lexicon with words related to Malaria (the tropical disease). If the parser does not find a word in the input in the lexicon, the analysis cannot be completed.

The feedback varies if the answer is correct or not. If the answer is correct, the system shows a representation of the parse tree found by the parser and tells the student that the answer is correct. If the answer is wrong, the system verifies if it complies with a pre-specified error rule. If it does, the system let’s the student try again or ‘get a hint’, which provides a pre-specified feedback message. If no error
rules are found, the system provides the correct answer. Specific feedback is reported for errors that are considered “important” according to their frequency in the data previously analyzed and used to create the system.

The architecture of the system is based on the parser and the interface. There are no specific modules to handle spelling errors, for example. The student is referred to an on-line dictionary to solve a spelling error problem.

Although the Intelligent Tutor of Academic English has only one type of exercise, this system presents an example of how NLP can be used in systems to propose feedback on written input. Dodigovic (2005) presents several arguments for the benefits of the type of interaction presented by the system. However, further studies would be necessary to test how the feedback messages improve students performance, and what other feedback strategies could be used to help learners develop their editing skills.

4.1.2 ALICE-Chan

Although the goal of the ALICE-Chan project (Levin and Evans, 1995) was to create a multimedia environment that could support the development of many different types of language learning activities, the ALICE-Chan system provides exercises for students to practice certain complex structures of Japanese. One advantage of the system is the fact that the designers do not expect neither the student nor the instructor to have deep knowledge of NLP or linguistic theory. Feedback messages are always in plain English and do not require the learner to interpret syntactic trees, or more complex representations such as Lexical Functional Grammar structures.
Because ALICE-Chan works with Japanese, a language where spaces are not used to identify word boundaries, the system performs segmentation together with morphological analysis. The system uses an algorithm that performs a left-to-write detection of words and a write-to-left filtering of generated words to perform segmentation. Once the morphemes are identified they are associated with a feature structure described in the lexicon. Syntactic processing uses a Left-Right parser with pseudo-unification (Tomita, 1984), and their approach to syntax is based on the theory of Lexical Functional Grammar Kaplan and Bresnan (1982). The grammar has about a hundred rules, and in addition to the grammar the system has ‘mapper’ that assigns semantic functions to for each phrase that the parser identifies, filling the predicate’s subcategorization frames. Ambiguities found during processing are handled by the disambiguator that asks questions to the user to clarify issues in semantic role assignment, syntactic constructions, segmentation, and homographs.

To diagnose errors in a student input, the system uses an architecture with a pipeline model, i.e., the input is passed through a pre-specified sequence of modules that analyze different linguistic properties. In ALICE-Chan the input first undergoes morphological analysis, then it is parsed, and then the mapper assigns grammatical functions (cf., e.g., subject, object) and semantic roles (cf., e.g., actor, thing-acted-on). If during this sequence processing fails the system reports to the student the message: “Your answer could not be understood”. If there are ambiguities in the end of the analysis the disambiguator starts a dialogue with the student. When ambiguities are resolved the ‘Matcher’ tries to match the answer with an authored template, and if it fails a feedback message is generated based on the parser, mapper and matcher outputs.
4.1.3 E-Tutor

As presented in section 1.5, the E-Tutor (Heift, 1998, 2001, 2003, 2004) is an electronic work book for German developed by Trude Heift and used as part of the German language program at Simon Fraser University. The system architecture has four modules responsible for processing the student input (Heift, 2003, p. 536): the Domain Knowledge, which encodes the knowledge about the language, the Analysis Module, which produces instructional feedback based on the analysis of the input performed by the Domain Knowledge, the Student Model, which stores information about the student and decides on the student level, and the Filtering Module, which decides the order in which feedback has to be presented to the student when multiple errors are found. For the remainder of this section, let’s focus on the type of processing performed by the Domain Knowledge module, in chapter 5 we will discuss the system’s Student Model, and its applications.

The E-Tutor processes the student input through a pre-specified sequence of sub-modules shown in figure 4.1. First, the system performs a string matching with all pre-stored answers, if one match is found the student receives a positive feedback message. Then the system checks for punctuation errors and spelling errors. If any errors are found they are reported to the student to fix them before processing continues. The spell checker returns a list of bare forms for the words in the input together with the spelling analysis. The example check compares the bare forms of the input with those required by the exercise, and it returns an error if one of the bare forms are missing. The Missing Word and Extra Word Checks make sure that all and only the words required by the exercise are used in the answer. The Grammar Check is a parser running on a Prolog server that uses a Head-driven Phrase Structure
Grammar to diagnose syntactic errors in the student input (see Heift, 2003, pp. 536–540 for details). Finally, the Catch-All module is called to provide a generic message to the student about the sentence if none of the other error detection modules is able to diagnose a problem in the student input.

As section 4.7 will show, some of the NLP modules used by Heift bare some similarities with the NLP modules of TAGARELA. For example, her Missing and Extra Word Checks perform a similar function to the one performed by TAGARELA’s Token Matcher. Her Example Check also uses bare forms of words to identify missing
concepts, similarly to TAGARELA’s Canonic Matcher. However there are some differences in the approaches as far as checking for the appropriateness of the input sentence.

Besides the fact that TAGARELA has some more modules to perform content checking, the main difference between both approaches is the fact that the E-Tutor reports content errors as they are found by the modules, while TAGARELA (for exercises that require necessary and major content manipulation, as described in 4.4.1) collects all the information from the content checking modules before it decides which error to report. The difference between both approaches can be justified by the variation on the type of activities presented by each system. Most E-Tutor activities require that the student use the words presented in the question to formulate an answer, while in TAGARELA, some activities such reading and listening comprehension may allow for a greater variation in the expected answer, thus making the process of checking for content appropriateness more complex. A more detailed explanation of the mechanisms used by TAGARELA are presented in section 4.7.

4.2 Nature of Processing

Before we review the structure of the architecture of TAGARELA, it is necessary to understand the role of language processing in the system in relation to the type of linguistic information that is being searched. As argued in previous chapters, ICALL activities do not have to be restricted to form-based exercises that solely require the manipulation of structural components. This means that ICALL systems have to be able to identify linguistic properties of the student’s input that are not purely grammatical or form-based. At the same time NLP technology that deals
with unrestricted world knowledge is not available. Dorr et al. (1995) identified that problem common to most ICALL systems, and argued that some level of semantic representation is necessary to evaluate the appropriateness of students’ utterances.

The solution used for TAGARELA resembles a common approach used by Intelligent Tutoring Systems (ITSs) to evaluate the knowledge structures of students. ITSs for other domains commonly diagnose problems in answers by mapping the procedures presumably used by a student to solve a given problem with either the procedures used by a machine to solve the same problem, or with a general descriptions of the necessary steps involved in solving the proposed problem.

Because ICALL systems have to deal with natural language, it is not trivial to access and represent the necessary knowledge used by a student to produce an appropriate utterance. Usually, ICALL systems evaluate the form of students’ sentences, and use its analysis as a representation of the misconceptions and missing conceptions in a student’s knowledge base. In other words, systems diagnose the morpho-syntactic errors in students’ sentences and provide feedback based on the errors found. Typically students’ utterances are not evaluated for appropriateness, i.e., the semantic content of the never checked. As a result, ICALL systems have a very difficult time presenting activities that go beyond form manipulation, since they lack the capability of evaluating if the answer answers the question being asked.

In TAGARELA the domain knowledge is not limited to the knowledge used to perform a morpho-syntactic analysis of the student’s input. The representation of the domain knowledge necessary to complete a task comes from a sequence of linguistic

---

17I am using the notion of domain knowledge as described in Fink (1990). According to her, the domain knowledge refers to “the subject matter material that is to be imparted to the student as it relates to the task to be trained. Thus, domain knowledge can span the range of formal and informal fields of knowledge and skill that might need to be assimilated by a student.” (Fink, 1990, p.199)
analyses performed on all pre-stored answers to the exercise proposed. The main idea behind this approach is that it is possible to access and represent the domain knowledge used to perform a task by extracting linguistic information from target answers. In other words, through the observation of lexical, morphological, syntactic, semantic, and pragmatic characteristics of an answer proposed by an expert (in this case the instructor) it is possible to describe (at least in part) the necessary knowledge to complete the task.

TAGARELA evaluates the student’s answer by performing a sequence of comparisons among its various linguistic properties with the linguistic properties obtained from the analyses of the target answers. This technique has the advantage of reducing the complexity of the linguistic processing to detect certain types of errors or deviations, especially the ones that are related to pragmatic and/or semantic information. It is also useful to detect cases where there are certain types of linguistic variations in the student’s input that are perfectly acceptable. Look at the example below.

(4.1) **Question:** Onde mora Maria?
Where lives Maria?
‘Where does Maria live?’

(4.2) **Target:** Ela mora no Brasil.
She lives in the Brazil.
‘She lives in Brazil.’

(4.3) **Student:** A Maria mora no Brasil.
The Maria lives in Brazil.
‘Maria lives in Brazil.’

The target answer in (4.2) to the question in (4.1) uses a nominative pronoun as subject, instead of the proper noun used in the question. The student’s answer in
(4.2) instead adds a definite article in front of the proper noun, which is perfectly correct in Portuguese. Through the comparison of the syntactic trees of (4.2) and (4.3), TAGARELA can isolate the subject NP of both sentences, and analyze their structures. In a situation like this, where one NP is made of a pronoun and the other of a full NP, the system tries to match their person, number, and gender features, essentially performing a simple case of anaphora resolution. If matching succeeds, the system diagnoses (4.3) as a candidate for a possible answer. This analysis is passed on to the module that generates feedback messages that can decide how to provide feedback for such variations.

Another key element in the evaluation of students’ responses is the type of task the student is performing. Depending on the task, certain linguistic properties are more important than others, and some analyses could be irrelevant. The most obvious example is the unnecessary syntactic analysis of the student input in a fill-in-the-blanks activity where the target answers are always one word long. Information about activities is not only important in analyzing the input, but it is also essential to evaluate students’ strategies and knowledge structures.18

TAGARELA uses information about the activity to determine how to proceed with input processing, to decide on which feedback message is most appropriate, and to evaluate a student’s current state of knowledge.

Throughout the remainder of the chapter, it is important to keep in mind that TAGARELA evaluates students’ input by comparing its linguistic structures with

---

18Chapter 5 discusses how task description can be used in student modelling.
the ones found in pre-stored target answers. We now proceed to an overview of the system’s architecture followed by a more detailed description of its modules and sub-modules.

4.3 Overview of the Architecture

TAGARELA’s architecture is made of four major modules; the interface, the analysis manager, the language processing module, and the feedback manager. There are also two models that provide the necessary information about activities and learners; the learner and the instruction models.

The analysis of the learners’ input starts when the Analysis Manager receives it from the Web Interface. The Analysis Manager is responsible for deciding how to process the input and the target answers, for calling the necessary submodules of the Expert Module to provide linguistic information about the sentence, and for annotating the input with the output of those language processing sub-modules. The Expert Module is a collection of submodules that can be called to provide specific types of information about the linguistic properties of the input and of the target answers. Section 4.7 describes the submodules that are currently implemented.

After the input and the target answer are annotated with different types of linguistic information, they are sent to the Feedback Manager. The Feedback Manager decides on the best feedback strategy, generates the feedback message and passes it back to the Web Interface which displays the message to the student. The next sections describe in more detail how the different components of the system work.
4.4 Instruction Model

To understand how the processing modules work in TAGARELA, it is necessary to know what type of information is available to them, and how they use it. In TAGARELA, the Instruction Model is the repository of all information needed by the Interface to dynamically generate the exercise web pages, the Analysis Manager to decide on the way it should process the student’s input and annotate it with the output of the NLP modules, and by different submodules of the Feedback Generator.
to decide on the best feedback strategy and to generate the feedback message. The Instruction Model is the repository for two types of information: information about activities, and information about error taxonomy used by the system. The next two sections describe what is in each one of these repositories, and why this information is there.

### 4.4.1 Activity Model

In section 3.4 we saw that TAGARELA has six activity types: listening, reading, description, fill-in-the-blanks, vocabulary, and rephrasing. These activity types represent the division of tasks presented to the student when he accesses the system. However, the Activity Model uses further criteria to classify the activities in the system.

First, all activities are classified in terms of level: The level of an activity is based on its course number (101, 102, 103, 104), and the module number (1, 2, 3, 4, 5).19 At this point, we take the progression of the course material to be an indicator of the complexity of linguistic knowledge necessary to perform specific tasks.

Second, activities are classified based on the three types of input accepted by the system: (i) word, (ii) phrase, (iii) sentence. This classification is not only important to determine how the input is to be processed, but it must also be taken into account when inferring the level of proficiency of the learner in a given grammatical structure. As we will see in chapter 5, the system can use the type-of-input classification to determine how well a student knows a given grammatical structure. For example, a

---

19See chapter 3.2 for the Individualized Instruction Program description.
student that does not make agreement errors in questions where only an NP is required as the target answer has not shown that she is able to write full sentences without agreement errors.

Third, activities are classified based on the amount of content manipulation necessary to fulfill the task. There are four categories that encompass the six activity types: (i) little content manipulation (Fill-in-the-blanks), (ii) some content manipulation (Rephrasing), (iii) necessary content manipulation (Description and Vocabulary), major content manipulation (Reading and Listening). Content manipulation is an important measure because it is a main distracter to specific focus on grammatical accuracy. This classification implies that the greater the amount of expected content manipulation by the student in a task, the more likely she will be to make grammatical errors.

Besides the classification mechanism described above, activities in TAGARELA may also bring some extra information about the nature of the task performed by the learner. This extra information can be used by the system to make certain inferences about the type of non-linguistic skills necessary to complete the task. For example, while a rephrasing question may provide enough information for the system to observe if the student can deduce the necessary syntactic rules to re-write the target sentence (student’s deduction skills), a WH-question in a reading comprehension may allow the system to observe the inferences a student needs to make to correctly answer the question (student’s inferencing skills). We specifically provide the system with information about how to observe student’s production based on the type of task the student is performing. This information cannot be deduced from the classification above, because different questions for the same activity type may or may not allow
the system to draw inferences about a given knowledge unit. Information about the activity that indicates how the system can infer the acquisition of non-linguistic skills is essential for the implementation of the user model proposed in chapter 5.

4.4.2 Error Model

Section 2.4 argued for the importance of feedback on errors in L2 writing, but it did not say anything about the nature of the errors to be treated.

To develop ICALL systems that provide feedback in language to learners it is necessary to specify the types of variation on learner language (i.e., errors) for which the system is going to provide feedback. The ability to provide appropriate feedback on certain language features is necessarily interconnected with the NLP mechanisms available to the system to detect and diagnose these features.

As discussed in section 1.5, research on NLP components for ICALL has focused primarily on the development of parsing techniques for ill-formed input. The consequence of this orientation is that most ICALL systems are designed to primarily or exclusively identify morpho-syntactic errors. Section 4.1 has presented some examples of systems whose feedback messages focus primarily on morpho-syntactic analysis. Another example is the BRIDGE system as described in (Weinberg et al., 1995, pp. 39–40). The types of errors the system can diagnose are:

- subject-verb agreement (person, number, gender)
- case of sentence subject
- article-noun agreement
- preposition-noun agreement
- verb-preposition agreement
- word order errors (position of finite verb in main and subordinate clauses)
- auxiliary selection in compound tenses
Prioritizing syntactic processing in ICALL development has a direct consequence on the type of systems that can be developed. In chapter 2, we saw that even the contemporary theories of SLA that defend the use of language forms and rules to facilitate the learning process agree that manipulation of meaning has to be present for feedback on form to be effective. If we seriously consider the arguments presented by the SLA theories described in chapter 2, we can ask ourselves if ICALL systems that can only provide feedback on form, or that prioritize feedback on form run the risk of not positively influencing the development of students’ interlanguage. It does not mean that the diagnosis of form-based errors is irrelevant. It only means that we may want to reconsider the role of error diagnosis to find the right balance between form and meaning.

For example, the same type of form-based error may have different implications in meaning-based activities. One of the most targeted types of errors in ICALL is agreement. However, not all agreement errors have the same impact in the learners ability to communicate in L2. In fact, some may even give positive indicators that the learner has been acquiring what native speakers are actually saying. Look at the examples below.

(4.4) *Tu* come *maçã* de *manhã*?
You(2ps) eat(3ps) apple in the morning?
‘Do you eat apples in the morning?’

(4.5) *Onde* está *meu* caneta?
Where is my(masc) pen(fem)?
‘Where is my pen?’

(4.6) *A* menino *bonita* deixou *este* pacote.
The(fem) boy(masc) beautiful(fem) left this package.
‘The beautiful boy left this package.’
In all three examples there is an agreement error, but only in (4.6) the meaning of the utterance cannot be fully recovered by native speakers. The agreement error between a pronominal subject in the second person singular and the main verb in the third person singular in (4.4) is very common among native speakers of Brazilian Portuguese, especially young people and people from lower social classes, and does not interfere with communication. In (4.5), there is an agreement error with the grammatical gender of the word ‘pen’. Although native speakers do not usually make errors with grammatical gender, it is still possible to recover the meaning of the sentence, since there is no variation in grammatical genders of nouns. In (4.6), on the other hand, it is impossible to know if the person who left the package was male or female because there is an agreement error that involves natural gender, and morphologically both forms of the noun are possible (‘menino’ - boy; ‘menina’ - girl).

It is important to notice that all three examples represent deviations from prescriptive grammatical rules, and the learner should be advised about them. However, they clearly represent three distinct phenomena with different practical consequences to the L2 user.

More importantly than reinterpreting form-based errors using meaning-based criteria is the true incorporation of the diagnosis of meaning-based errors in ICALL. For systems to present activities where meaning plays a more central role, it is not enough to change the design of the exercises. If questions allow for more open-ended types of answers, where there are different ways to convey the information that is being expected, the system should be able to handle these possible variations in the student input. Section 4.7.5 describes in more detail the different types of activity.

---

20 A similar phenomenon occurs in some dialects of English with certain expressions, e.g., ‘there’s several cars parked over there.’
requirements in terms of semantic processing by presenting the Language Exercise Spectrum by Bailey and Meurers (in prep). One consequence of incorporating these activities is that not only the NLP modules should handle meaning processing, but also that the error taxonomy should reflect this new approach. Thus, the list of errors that a system can identify should include meaning-based errors that are in sync with the capabilities of NLP modules, the types of exercises, and the feedback strategies.

TAGARELA’s error taxonomy is divided into two types of errors: form-based and meaning-based. It reflects two groups of analysis submodules that are included in the system.\(^\text{21}\) Among the errors TAGARELA diagnoses there are:

- meaning-based errors:
  - missing content words
  - extra content words
  - wrong content selection (selectional restrictions)
  - wrong collocations
  - wrong word choice (e.g., _ser_ vs _estar_)

- form-based errors:
  - agreement (subject–verb, subject–predicative, and within an NP)
  - wrong verb form
  - missing functional words
  - extra functional words
  - wrong contractions
  - word order
  - wrong subcategorization
  - other types of lexical dependencies

Listing the errors to be diagnosed by the system is the first step to create a feedback mechanism. The error taxonomy provided above is used by the system when deciding the type of feedback to give. Together with the taxonomy there is

\(^{21}\)See section 4.7 for more details about the NLP analysis submodules of TAGARELA.
another component that plays a major role in feedback: the activity. In section 4.9 we will see the strategies used to generate the feedback messages that are provided by TAGARELA, and how an error taxonomy that explicitly incorporates meaning-based errors facilitates that strategy.

4.5 Interface

As noted in chapter 3, TAGARELA uses a web-based interface. There are several advantages of choosing this type of interface for an ICALL system (Felix, 2002, 2001; Hughes et al., 2004). The first one is the fact that ICALL developers can avoid dealing with issues related to specific operating systems. It does not matter if the student uses Windows, MacOS, Linux, or Unix, since there are free, (more or less) standard compliant web-browsers for all major operating systems.

Second, web-based systems provide flexibility for users to access the system. Because the ICALL system does not need to be installed on the user’s machine, since all NLP processing can be done on the server, students can access the system from home, computer labs on campus, or even internet cafes.

Third, web-based technology allows for good multimedia integration. The user can listen to sound tracks, see slide shows, and watch videos on-line, and it readily supports hyperlinks to relevant information internal or external to the site.

Fourth, web-based interfaces facilitate navigation and interaction with the system, since most computer users are familiar with resources of web interfaces, such as buttons, clickable texts, forms and menus.
Finally, one of the great advantages of choosing a web-based interface is the fact that the client-server configuration allows for greater storage and processing capabilities. Because processing can all be done on the server side, without relying on the hardware used by the student, it allows for the use of different types of NLP technology, and the storage of large language resources and the learner models.

While web-based interfaces have clear advantages, there are some problematic aspects that require some precautions when designing the system. The most important one is to make sure that the web-browser used by the learner offers full support for the web standards used by the system. In the case of TAGARELA, we are currently limiting access to browsers that provide reasonable support for javascript and cascading style sheets (CSS), such as Firefox 2 and Internet Explorer 7.

TAGARELA uses a standard Apache 2 (version 2.2.0) web-server. For dynamically generating web pages and interfacing with the NLP modules, we are using Mod-Python (version 3.3.1) to integrate the Python programming language (version 2.4.2) into the Apache web-server. In TAGARELA, all web-pages that display exercises are dynamically generated by combining HTML templates with information from the Activity Model. Section 4.4.1 describes in detail the information in the activity models.

Another issue that system designers have to take into consideration is the shortcomings of the stateless HTTP protocol. For the system to interact intelligently and adaptively with the learner, it is necessary that it keep track of the learner’s production and store information about the diagnoses and feedback messages provided. Following standard practice, TAGARELA uses cookie-based sessions to overcome this problem as it stores the relevant information in two different databases; one that
stores the user’s id and password for authorizing him/her to access the system, and the second, a database for each individual user for storing the student model that includes the input for all the activities the learner has done, the analysis of the student’s input for those activities, and various other relevant information used by the system to diagnose the learner state of knowledge.

TAGARELA uses asynchronous partial page updates (AJAX) in cases where a reloading the whole page would disrupt the student interaction with the system. For example, the system displays feedback messages without having to reload the web page every time a student submits an answer. The use of this technology helps implement the scaffolding technique used in the feedback messages since nothing in the page the student is interacting with changes after he clicks on the submit button, except for the feedback message displayed. This allows the student to rephrase his answer as many times as necessary following the different feedback messages received, and it is particularly important in listening comprehension exercises, where the student can answer questions and receive feedback without having to interrupt the listening passage or the video included on the page.

4.6 Analysis Manager

The Analysis Manager is responsible for the processing of the input and of the target answers. There are several factors that the module can take into consideration during this process. Currently, its decisions are based on the type of activity, type of input, and the output of certain NLP modules.

As discussed in chapter 1, research in ICALL has often emphasized the development of NLP tools, paying little or no attention to the design of the exercises
presented to the student. This approach has provided interesting results in terms of the development of NLP components, but it has not forced ICALL researchers to develop system architectures that are flexible enough to deal with different types of input in different contexts. Section 4.1 showed how systems like the Intelligent Tutor Of Academic English (Dodigovic, 2005), ALICE-CHAN (Levin and Evans, 1995), and E-Tutor (Heift, 1998, 2003) have architectures that use a pipeline model that allows it to call the NLP modules in a pre-defined order, transforming one data structure into another and terminating when specific conditions are met (e.g., when the learner response matches a pre-stored target response, or when spell checking fails).

Such a pipeline architecture works well as long as the system deals with learner input from activity types that are uniform with respect to the required NLP processing. Problems with this type of approach can appear if the analysis of the learner input has to vary based on the nature of the activity and the type of input (cf., e.g., words, phrases, or sentences). In some cases not all NLP modules are necessary to provide the necessary analysis, for example no parsing is necessary when the input is a single word.

As presented in section 4.4.1, TAGARELA six activity types require different forms of input, from words to phrases. More importantly, the activity types differ with respect to the information needed to evaluate whether the form and meaning of the student input is correct and, when this is not the case, the information needed for error diagnosis and feedback generation. For example, the reading and listening activities require morpho-syntactic and semantic analysis, rephrasing requires only syntactic analysis, and FIB activities only morphological analysis. Interestingly, even within the same type of activity, the properties that need to be identified by the
NLP modules can differ - some reading exercises target forms explicitly given in the text, others require more semantic analysis or inferences. It thus seems to be well-motivated to have a demand-driven, annotation-based architecture that allows the activity models to specify which information needs to be provided by the NLP processing.

Considering the wider NLP context, the perspective we are arguing for an ICALL system can be viewed as an instance of an application required to deal with heterogeneous input and different information needs based on that input. Our approach thus is reminiscent of current approaches to information extraction, where, e.g., IBM’s OmniFind makes use of an Unstructured Information Management Architecture (UIMA – Ferrucci and Lally, 2004) to obtain a range of annotations depending on the specific information needs.

4.7 NLP Modules

The NLP modules can be schematically divided into three groups: *pre-processing*, *form-based processing*, and *meaning-based processing*. The next sections present the modules currently implemented in TAGARELA.

4.7.1 Tokenizer

Tokenization is the process of dividing the continuous string of characters dividing the text into tokens, i.e., loosely speaking, the words. It is one of the first steps in the pre-processing of the student’s input in ICALL systems.

TAGARELA’s tokenizer receives as input a sentence or a phrase, and produces as output a tuple with two lists in it. The first list is just a sequence of the tokenized words, and the second one is a list of lexical entries with or without information about
the part-of-speech of the lexical item. To understand the function of the second list, and why the tokenizer can assign part-of-speech in certain cases, we have to review three characteristics of the Portuguese language that can play a role in the tokenization process.

The first and most simple one is the fact that abbreviations are usually marked with a dot (‘.’) after their last letter. For example, ‘abrev.’ means abreviação (abbreviation), and ‘Dr.’ means doutor (doctor). TAGARELA has a list of approximately 1300 abbreviations used in Portuguese\(^{22}\), and the tokenizer relies on this list to identify if a dot following a word indicates the end of an abbreviation or the end of a sentence.

The second characteristic of the Portuguese language that plays a role in the tokenization process is the use of contractions between prepositions and other parts of speech, such as articles, pronouns and adverbs.

\begin{align}
(4.7) \quad a + o &= ao \\
&= \text{to the} \\
&= \text{to the} \\
(4.8) \quad de + ela &= dela \\
&= \text{of her} \\
&= \text{of her} \\
(4.9) \quad de + aqui &= daqui \\
&= \text{from here} \\
&= \text{from here}
\end{align}

TAGARELA uses a list of these contractions to separate the preposition from the second element. As part of this process, it is also possible to assign a part-of-speech

\(^{22}\)The list of abbreviations as well as the lexicon used in TAGARELA was derived from the lexicon of CURUPIRA, a project developed by NILC in São Carlos Brasil. See section 4.7.3 for details about the lexicon.
of the second element unambiguously. The Portuguese word ‘a’, in example (4.7) has five different lexical entries in the lexicon: determiner, pronoun, preposition, noun, acronym. However if ‘a’ is contracted with a preposition it can only be a determiner. So, the tokenizer already assigns the part-of-speech of the word ‘a’ during the tokenization process, which will facilitate the disambiguation of lexical entries explained in section 4.7.4.

The third characteristic of Portuguese taken into consideration during the tokenization process is enclitization.\textsuperscript{23} This phenomenon occurs when clitic pronouns come after the verb attached to it with a hyphen. Most clitic pronouns keep their form when in an enclisis position. However third person accusative clitics may suffer some morphological changes. The following changes occur: (1) when the verb form ends in ‘-r’, the ‘l’ is added to the clitic pronoun and the ‘r’ is dropped in the verb form (an accent is added to show that the tonic syllable does not change), see example (4.10); and (2) when the verb form ends with a nasal sound, an ‘n’ is placed before the clitic, see example (4.11). In all other cases the clitic stays the same, see example (4.12). Disambiguation of the part of speech of the clitic involved in these cases is also straightforward, since only pronouns appear in this position.

\begin{center}
\begin{tabular}{llll}
(4.10) & comprar & + & a = comprá-la \\
         & buy & + & it = buy it
\end{tabular}
\end{center}

\begin{center}
\begin{tabular}{llll}
(4.11) & compram & + & a = compram-na \\
         & buy & + & it = buy it
\end{tabular}
\end{center}

\textsuperscript{23}Because TAGARELA is to be used by students of Portuguese in the 21st century, the tokenization process does not deal with mesoclisis phenomena, which is considered to be archaic and only used in literary texts.
There is only one case of encliticization where lexical ambiguity cannot be resolved. It occurs between the first person plural accusative and dative clitic ‘nos’, and the masculine third person plural accusative ‘os’ when preceded by a verb form ending with a nasal sound. As examples (4.13) and (4.14) show, it is impossible to reconstitute the original form of the pronoun only by its enclitic form. In cases like this, the ambiguity can only be resolved by anaphora resolution.

(4.13) compram + os = compram-nos
buy + them = buy them

(4.14) compram + nos = compram-nos
buy + us = buy us

If one of the three phenomena described above (abbreviation, preposition contraction and encliticization) occurs in the sentence, the tokenizer reports the part-of-speech of the tokens involved. If not, the tokenizer simply returns the lexical item without any specification for its part-of-speech.

4.7.2 Spell-Checker

The main function of spell-checking in ICALL is to report to the learner the words that have been misspelled in a sentence. Spelling errors are commonly classified into two categories: non-words, also known as misspelling (Rimrott and Heift, 2006), i.e., a combination of letters that creates a word that is not part of the target language (e.g., He saw the swaaw), and existing words spelling errors (e.g., I bought
**there** car.). Most spell-checkers only deal with the first type of error. In line with this, TAGARELA only considers non-word spelling errors as spelling errors. existing word spelling errors are either treated as subcategorization errors or as wrong word/expression errors\(^{24}\).

The system uses a combination of two resources to check for spelling errors. First it uses Ispell for Brazilian Portuguese (Karpischek, 2003). Ispell is a spell-checker for Unix that suggests corrections for misspelled words based on a Damerau-Levenshtein distance\(^{25}\) of 1 (Kuenning, web). If a word is identified as misspelled by Ispell, TAGARELA checks its own lexicon to see if it can find the word there. Even if br-Ispell has a larger database, there have been cases where the word is only present in TAGARELA’s database.

If the word is flagged by Ispell and not found in TAGARELA’s lexicon, the system reports the spelling error to the student, and uses the suggestions made by Ispell to indicate possible replacements. We are currently developing a filter for spelling suggestions that uses the words from the activity model as possible targets to reduce the list of words presented to the student.

### 4.7.3 Lexicon

In TAGARELA, the lexicon has a very important role, since the information stored in it is used by different modules for both syntactic and semantic processing. The lexicon is stored in a standard database, and it is keyed by the full form of the word. The lexicon is accessed by a module called Lexical Look-up, which receives as input a list

\(^{24}\)See section 4.4.2 for an explanation of the error types.

\(^{25}\)Damerau-Levenshtein distance is the minimal number of insertions, deletions, substitutions and transpositions needed to transform one string into another. See Damerau (1964) for details.
of tokens and produces as output a list of lexical entries. We generated TAGARELA’s lexicon from part of CURUPIRA’s lexicon, which was kindly made available to us by the Núcleo Interinstitucional de Lingüística Computacional (NILC) research group in Brazil. CURUPIRA (Martins et al., 2006) is a general-purpose robust parser for Portuguese with a lexicon that includes 1.5 million inflected and derived forms. Each lexical entry is annotated with its relative frequency in the Lácio-Web Corpus (Aluízio et al., 2003).

TAGARELA’s lexicon has approximately 43000 entries. The set of entries corresponds to: 30% of nouns which are the most frequent in the Lácio-Web Corpus, 25% of the most frequent verbs plus all verb forms for the verbs that were listed in the course book used in the Portuguese Program, 40% of the most frequent adjectives and adverbs, and 100% of all closed-class words.

The lexicon is fully disambiguated, i.e., there is no underspecified information. All lexical entries include the canonic form of the word and the relative frequency of the word in Lácio-Web Corpus. All other features are specific to each part-of-speech. For example, in the lexical entry of the noun ‘menina’ (girl) we would find the following information:

- PoS: noun
- gender: feminine
- number: singular
- degree: null
- regency: null
- canonic: menino
- frequency: 30
- source: lexicon
The degree is only specified when the noun is in the augmentative or diminutive forms. The same occurs with regency, which is only specified as non-null if it is a deverbial noun. The canonic form for nouns as well as adjectives is its masculine singular form, except for nouns that have feminine grammatical gender. The source of the lexical entry is indicated because, as discussed in section 4.7.1, the original part-of-speech can be assigned by the tokenizer, which generally eliminates lexical ambiguity.

A complete description of lexical entries for all parts of speech with explanations of how they were derived from NILC’s lexicon can be found in appendix A.

4.7.4 Syntactic Processing

Syntactic variation both in erroneous and in well-formed student production can present several challenges to any NLP technique that aims at producing a structural analysis of student input (Levin and Evans, 1995, p. 92). In order to deal with recursive syntactic structures, a parser has to handle structural ambiguities that are common in natural languages. When processing ill-formed input the parser has to deal with an even greater number of possible ambiguities, since it has to generate all possible parses using either a greater number of rules, e.g., augmenting the grammar with mal-rules (cf., e.g., Heift, 2003, pp. 537–540), or limiting the scope of restrictions imposed on its original set of rules e.g., relaxing grammatical constraints (cf., e.g., Vandeventer Faltin, 2003, pp. 66–78). In either case the search space increases, and it becomes harder to come-up with one unambiguous analysis. Besides syntactic ambiguity, NLP technology for ICALL has to propose new solutions to deal with
lexical ambiguity. Unusual or erroneous word choice by foreign language learners can pose untractable problems to error detection, since it can further complicate syntactic and semantic processing.

The approach used to perform form-based diagnosis in TAGARELA uses of two distinct techniques to obtain different types of linguistic information from the learner input. First, we use an approach based on Constraint Grammar (Kaklsson, 1990) to disambiguate lexical entries with local rules, the resulting representation facilitates the shallow semantic processing described in 4.7.5. Second, we use a chart-parser to identify trees and subtrees that can be used to compare the learner input with possible answers.

It is important to keep in mind that form-based errors are neither the main focus of the linguistic analysis nor are they the primary target of feedback messages. The diagnosis of form-based errors is only one element among others taken into consideration by the system to generate feedback messages. As we will see in section 4.9, feedback on form is only provided if there are no major problems with the content of the input.

Following this principle of deemphasizing feedback on form, TAGARELA’s morpho-syntactic analysis ends up being used more often to provide further information for content analysis than to diagnose form-based errors that will be reported to the learner as such. We now turn to the two techniques that TAGARELA uses to collect information about morpho-syntactic properties of the student input and of target answers.
Disambiguator

The identification of morpho-syntactic properties of the learner input begins with a module that uses some of the principles of Constraint Grammars (Kaklsson, 1990) to resolve lexical ambiguity in input sentences.

The disambiguator receives as input a list of tokens, each of which with a list of lexical entries. For each token that is licensed by more than one lexical entry the disambiguator uses local disambiguation rules to decide which lexical entry is more appropriate for that token in a given context. Disambiguation rules are based on parts of speech, but they may also use morphological information in certain cases. For example, the word ‘o’ can be a either a determiner, a pronoun or a noun. In (4.15) its local context (between a pronoun and a verb) is enough to determine that it is a clitic pronoun. However, the word ‘a’, which can be a determiner, a preposition, a pronoun, or a noun, appears preceding a noun, which is not enough to determine its part-of-speech even if we consider that it is following a verb. In order to do so, it is necessary to take into account the gender feature of the noun ‘João’. Because ‘a’ is preceding a masculine noun, it is possible to unambiguously determine that it is a preposition in this sentence.

(4.15) Eu o entreguei a João.
I it delivered to John.
I delivered it to John

The advantage of this approach over parsers that work with phrase structure rules is that by using disambiguation rules locally, even when student sentences are completely ungrammatical and no parse trees can be found, it is still possible to recover some of the lexical meaning of the items of the sentence after getting rid of
(or at least diminishing the degree of) lexical ambiguity. As we will see in section 4.7.5, the output of the disambiguator directly increases the quality of some content matching modules.

Parser

As mentioned before, TAGARELA uses a chart parser to identify syntactic trees and subtrees that could represent the constructions being analyzed. The only information about the input used by the parser is the part-of-speech; there are only lexical features that are only checked once the structural possibilities have been clarified. The parser uses a variation of the Cocke-Younger-Kasami algorithm (Kasami, 1965). All well-formed substrings found by the parser are stored in a chart. The grammar used is a context-free grammar in Chomsky Normal Form (CNF), i.e., all non-lexical rules are binary branching. The grammar does not include any mal-rules, i.e., rules licensing ill-formed constructions. Every subtree found by the parser, thus, is the representation of a grammatical sequence of words in the language.

The advantage of this type of approach is that, different from the mal-rules and the constraint-relaxation approaches, we do not have to predict the syntactic variations in students’ sentences to identify problems. The disadvantage is that we do not actually perform error detection or diagnosis with the parser itself, we only collect information that will allow us to observe where the structure has matched syntactic patterns of the language, and where it does not. We then need other modules to analyze those structures to identify any problems. We have not yet tested this approach against approaches where the parser incorporates feature unification to see which one could provide the best results in ICALL contexts.
The Agreement Checker is one of the modules necessary to perform diagnosis based on the parse tree (or subtrees) found by the parser. It receives as input a list of possible syntactic trees (or subtrees if no parse that covers all elements of the sentence are found), and the list of lexical entries. Its output is a data structure keyed by the types of construction where agreement features were checked. The values for each key is the resulting analysis of agreement checking.

There are several types of agreement phenomena in Portuguese that are handled by the Agreement Checker.

(4.16)  
\[ \text{As menino levada comeu.} \]
\[ \text{The(pl-fem) boy (sg-masc) naughty (sg-fem) ate.} \]

In Portuguese, as in other Romance languages, determiners and adjectives have to agree in number and gender with the noun. Example (4.16) illustrates some agreement errors within an NP. The values plural and masculine for the number and gender features of the determiner ‘a’ do not match values singular and masculine of the noun ‘menino’. A similar problem occurs with the features of the adjective ‘levada’. TAGARELA takes the noun to be the head of the construction, so it tries to match the features of other the other tokens against those of the head-noun. In this case it diagnoses three agreement errors, since the number feature of the adjective does match that of the noun.

The second type of agreement error that TAGARELA diagnoses is subject-verb agreement. As in many Western languages, subject and verb have to agree in person and number. In example (4.17), the subject being first person plural does not agree with the verb that is third person singular.
(4.17) Nós comeu.
    We(1pp) ate (3ps) ate.
    The boy ate.

In Portuguese, subject and a predicative headed by an adjective. To check for this
type of agreement, TAGARELA takes the values of the features of the head-noun as
the agreement feature values of the whole NP and tries to match them with the ones
of the head-adjective. In example (4.18) the adjective ‘bonito’ is masculine singular
while the NP is feminine singular. In this case TAGARELA reports one agreement
error between the subject and the predicative.

(4.18) As menina é bonitos.
    The(pl-fem) girl (sg-fem) is beautiful (pl-masc).
    The boy ate.

The last type of agreement error caught by TAGARELA is the agreement be-
tween the subject of a passive voice and the verb in the past participle. As example
(4.19) shows, the past participle in Portuguese varies in number and gender, and its
agreement features have to match those of the subject. TAGARELA uses a similar
mechanism to subject-predicative agreement check, since both constructions require
the use of a linking verb (e.g., ser, estar).

(4.19) O livro foi compradas.
    The(sg-masc) book (sg-masc) was bought (pl-fem).
    The boy ate.

4.7.5 Shallow Semantic Processing

Chapters 2 and 3 presented the debate about current approaches in FLTL and the
need to incorporate the manipulation of meaningful structures into ICALL activities.
The problem of incorporating such changes is that in order to process student responses to activities that require content manipulation the system needs to be able to evaluate the appropriateness of students’ answers.

As Bailey and Meurers (in prep) reminds us, ICALL developers have to be aware of the relationship between exercise properties, possible response variation, and processing needs for content assessment before one can incorporate activities where there could be a great variation of learners’ responses. Figure 4.3 shows the language exercise spectrum in Bailey and Meurers (in prep) with respect to how restricted the responses to exercises can be. On the left end of the spectrum there are decontextualized activities where there cannot be any variation in terms of meaning, while on the right end of the spectrum there are open ended activities where it is impossible to predict the content of the outcome. Bailey and Meurers (in prep) argue that there is place in the middle ground that constitutes a pedagogically interesting and computationally realistic target for ICALL activities. Section 3.4.1 has already presented some of the issues in processing capabilities ICALL developers have to take into consideration when designing exercises.

Figure 4.3: The Language Exercise Spectrum (Bailey and Meurers, in prep)
TAGARELA follows the basic principle used in the Content Assessment Module (CAM) system of Bailey and Meurers to evaluate the appropriateness of students’ answers. The idea is to match “increasingly complex linguistic representations of the concepts and relations expressed in the learner response to those expressed in the target response.” (Bailey and Meurers, in prep, p.13).

TAGARELA has six modules that analyze the student’s answer in comparison with the a group of target answers.

First, TAGARELA calls difflib python module (http://pydoc.org/2.1/difflib.html) that has several function for comparing strings. TAGARELA uses some of these functions to decide which of the possible target answers is the one that is closest to the student input.

After selecting one target answer, TAGARELA calls a module that performs a “correct answer check”. This module submits the student input to a sequence of tests comparing the student input with the selected target structure to decide if it can be the correct answer with or without certain types of variation, such as problems with capitalization and arbitrary numbers of spaces between tokens.

If the correct answer check fails, TAGARELA performs a sequence of comparisons between the tokens of the target answer and the tokens of the student’s answer. For these comparisons TAGARELA uses three modules: Token Matcher, where the full form of the tokens are matched, Canonic Matcher, where the canonic forms of the unmatched full form the words are matched, and Part-of-Speech Matcher, where the part-of-speech of the unmatched canonic form tokens are matched. After the three modules are called, a fourth module is used to analyze the results, and report which tokens were matched at which level, and which tokens were not matched at all.
4.8 Feedback Manager

The Feedback Manager is the module responsible for deciding on the best feedback strategy, and composing the feedback message to be reported to the student. It performs those two actions based on its analysis of the information provided by the various NLP modules called by the Analysis Manager.

There are two submodules that perform the basic operations: the Feedback Generator and the Message Generator. The Feedback Generator receives the annotated input and decides on the best feedback strategy. For this process it takes into account not only the linguistic analyses provided but also the type of activity and the type of identified errors. Section 4.9 discusses the strategies used by the Feedback Generator to make its decisions about the most appropriate feedback.

The Message Generator receives the feedback strategy together with other necessary information and generates feedback messages in HTML that are returned to the interface to be shown to the student. The module uses a list of pre-stored feedback message templates, which are instantiated with parameters specific to the instance under consideration. Section 4.9 shows some examples of the feedback messages that are generated by this module.

4.9 Feedback Strategies

Having received the linguistic analysis of the student input, which identifies possible errors discussed in section 4.4.2, the Feedback Manager could simply report back to the learner all linguistic deviations that were found during the input analysis.
However, as Heift (2003) has shown, reporting multiple errors at a time can be a very inefficient strategy in ICALL, since the learner may not attend to all the information presented.

We already mentioned the fact that TAGARELA prioritizes feedback on meaning over feedback on form, based on the pedagogical principles laid out in chapter 2. Another general strategy TAGARELA uses to provide feedback is scaffolding, because of the reasons presented in section 2.4. In general terms, TAGARELA’s feedback messages are more general the further away the student’s answer is from the target answer, and more specific the closer the student is to writing an appropriate answer to the question.

For example, in figure 4.4 the example shows that when asked about the age of the girl from the text the student made two errors. First, there is an agreement error between the subject (third person singular) and the verb (second person singular) in the sentence. Second, the student used the verb ‘ser’ (to be) instead of ‘ter’ (to have) to say how old she is, despite the fact that, differently from English, in Portuguese the verb ‘to have’ should be used when talking about someone’s age.

TAGARELA diagnosed both problems, but because word choice is a meaning based error, and agreement is a form-based error, the system decided to report the problem with word choice first. In the example in figure 4.5, the student fixed the choice of word but kept the same agreement error. Since there were no more meaning based errors, the system reported the mismatch of agreement features between the subject and the verb.
Figure 4.4: Feedback Message – Word Choice
Figure 4.5: Feedback Message – Agreement
Whenever the student has mistakenly used one word instead of the another, TAGARELA checks if they have the same part-of-speech, and suggest that the student change them, as illustrated in the example in figure 4.6.

When there are words missing in the student input, if the word is a noun or a verb, TAGARELA reports that there is a word missing with that part-of-speech, such as in figure 4.7. For other parts of speech, if the number of words missing is inferior to two, the system lists them, if there are three or more words missing the system reports the number of words missing, and says that it is difficult to evaluate the student’s answer. When there are extra words in the student answer the system reports them to the student and advises the student to exclude them from the answer.

In general, the feedback strategy used by TAGARELA was designed to make the student think about different characteristics of the input sentence, and improve their editing skills by making progressive changes to their input sentences to answer the question appropriately and accurately. The general procedure is to report errors that are content-driven first, and when the sentence is semantically comparable to the target answer morpho-syntactic errors are reported if found.
Figure 4.6: Feedback Message – Wrong Word
Figure 4.7: Feedback Message – Missing Verb
CHAPTER 5

STUDENT MODEL

This chapter introduces the current development of TAGARELA’s student model. After reading the description of the system in the previous chapter, the reader may wonder if the work presented in this chapter is really necessary, and what are the advantages of having a student model incorporated into TAGARELA. This question is not new, and has been asked several times in the ITS community. Self (1990), in a seminal paper, presents the disagreement in the ITS community about the usefulness of student models, and argues for the importance of such models. Building on Self’s arguments, we believe that the discussion presented below provides evidence that our current efforts in designing a student model for TAGARELA can in fact contribute to making ICALL systems better perform their role in foreign language instruction, which is the main goal of this dissertation.

5.1 Current Student Models in ICALL

As first introduced in chapter 1, the student model is a data structure that characterizes the student’s current understanding of the subject matter. For an intelligent tutor to interact with a student in an personalized way, it needs information about the individual using the system. Student modeling is a sub-domain of an area of AI
called user modeling. While user modeling is a general term that includes models of any user interacting with any type of application (cf., e.g., web pages, flight booking systems, search engines), student modeling is more specific to tutoring systems that are involved with the instruction of a subject.

In foreign language instruction it is not difficult to see when information about students is useful. Human tutors use it all the time when preparing lesson plans, providing feedback, deciding on assessment mechanisms; in sum, when performing most daily activities. What is not always obvious is the type of information that is being used to perform those tasks, and how that information was obtained. For example, what are the personal characteristics of a given student that are important when deciding how to provide feedback about an error? How do instructors know that a student has these characteristics? What mechanisms has the instructor used to observe those characteristics?

The classical approach to student models is based on ITSs for different subdomains of science (VanLehen, 1988). In most cases the domain knowledge is a construct of human intellect (e.g., math, programming languages, electronics, etc) and the learning process can be represented by the appropriate employment of a finite set of knowledge units. Corbett et al. (1997) list the classical components of a student model as being “an overlay of the domain expert knowledge and a bug catalog”. The first is a collection of knowledge units tagged with information about how well a student has learned it. The second is a catalogue of misconceptions or incorrect rules (also known as mal-rules) with indicators to show if the student has been using them to solve problems.
More recently, statistical methods have been introduced to estimate the state of students’ knowledge, and different types of representational networks have been used to link atomic cognitive chunks, i.e., to draw inferences about students’ knowledge based on behavioral atoms (cf., e.g., Conati et al., 2002; Gertner et al., 1998). Despite the statistical turn in the methodologies to represent the acquisition process, the way knowledge units are represented is still primarily influenced by the classical approach presented above.

In ICALL following the traditional ITS approach for student models will generally presuppose a model of the acquisition of the grammar rules described in the domain knowledge. To illustrate this point, let’s now turn to some examples of user models that have been developed in ICALL research.

### 5.1.1 E-Tutor

The student model of the E-Tutor26 (Heift, 2003, 2004, 2005) keeps track of individual “grammar skills” a student is acquiring (E-Tutor’s knowledge units). It collects positive and negative evidence from the student’s production, and uses it to keep track of the student’s knowledge of each individual grammar skill. There is no overall classification of a student’s grammatical competence. Instead, it keeps track of the student’s level of performance per grammar skill. There are three levels to classify a student’s knowledge in each grammar skill: beginner, intermediate, and advanced. When the system identifies one type of grammatical error in the student’s input, it

---

26The general description of the E-tutor was included in section 1.5.
checks the level of proficiency of that student in the student model for the particular grammar skill related to the grammatical error. The system, then decides which feedback message to use.

E-Tutor’s learner model is not based on a specific theory of SLA. It handles the acquisition of each of the “grammar skills” as an independent process with no interconnection among themselves. In practice it seems to be a model guided by pedagogical concerns about the learner’s linguistic competence for each independent grammar skill, since it is used in a structured pedagogical context where students are only supposed to do the activities that are directly related to their level of proficiency following the course material.

5.1.2 ICICLE

ICICLE (Interactive Computer Identification and Correction of Language Errors; Michaud et al., 2000) is a system developed to tutor deaf students on their written English. It receives English compositions by students, and provides tutorial feedback to allow students to perform corrections. The feedback consists of information about students' grammatical errors in them.

The student model used by ICICLE is called SLALOM – *Steps of Language Acquisition in a Layered Organization Model* (Michaud et al., 2001). The goal of the student model is to capture the status of the grammatical structures of English in terms of “acquired”, “being-acquired”, and “unacquired”. Knowledge Units (KUs) of the SLALOM architecture are abstract grammatical concepts composed of two sets of grammar rules: “correct” English rules and mal-rules. These KUs are grouped and hierarchically classified. The hierarchy proposed is based on studies in second
language acquisition that showed stereotypical sequences in the acquisition of grammatical structures by learners of the same L1 community (cf., e.g., Krashen, 1988a; Gass, 1979; Schwartz and Sprouse, 1996). This hierarchy is used to identify the current state of knowledge of a learner, and to predict the next grammatical structures to be acquired. The system compares stereotypical sequences of English acquisition with the student’s current production to create an analysis of what structures are being acquired. Different from the E-Tutor, SLALOM incorporates specific claims about how grammatical structures are acquired by the learner. Nevertheless, the SLA process being modeled is still restricted to the acquisition of linguistic properties. It is important to notice that in view of the system’s needs, this model may be sufficient, since ICICLE’s goal is to serve as a writer’s aid. However, more complex interactions or activity settings would require the system to have additional information about the learner in order to react appropriately.

5.1.3 CASTLE

CASTLE (Murphy and McTear, 1997) provides communicative role-play scenarios where students answer questions posed by the system. When the system diagnoses certain linguistic problems with a student’s production (i.e., when a student makes three errors of the same type), it proposes a set of remedial exercises.

CASTLE’s learner model stores information in three distinct groups. The first one is the student’s personal information, which includes, for example, native language, motivation, and background language proficiency. The second is the “student model”, which keeps track of the student’s performance by “domain topics”, his/her proneness to commit certain errors, and the likely causes of errors. The third is the
“cognitive model”, which stores information about the student’s preferred feedback media and exercise types, his/her interest in grammar, and the use of polite forms. The information encoded by the first group is entered by the student, in the second group it is inferred by the system, and in the third group it is acquired by the student’s input, but modified by the system. Note that the “student model” of CASTLE is the part of its learner model that keeps track of the student’s proficiency level in terms of his/her linguistic performance.

CASTLE also uses stereotypical sequences of language acquisition to update its student model. It uses the notion of grammatically-driven partitions (cf., e.g., present simple, past simple, present perfect) that divide the foreign language curriculum upon which CASTLE is based. Different from SLALOM’s uses of stereotypical sequences in SLA, CASTLE orders the acquisition of the linguistic content in its grammatical partitions based on common practice in foreign language teaching.

Conceptually speaking, CASTLE’s learner model differs from the two models presented above because it acknowledges the needs of ICALL systems to have information about a student that is not restricted to his/her grammatical performance. By keeping track of a student’s preferences and background, the system is able to take these characteristics of the learner into account when providing feedback. However, the additional information modeled is completely dissociated from the student’s domain knowledge so that performance errors cannot be linked to them.
5.1.4 Passive Voice Tutor

The Passive Voice Tutor (PVT) (Virvou et al., 2000) is an ITS to teach the formation of the English passive voice for native speakers of Greek. The system uses a multimedia interface to present questions in which the student has to transform passive sentences into active sentences and vice versa. The system can dynamically create new exercises of the same type.

PVT is a typical example of the application of sophisticated mechanisms for modeling learner knowledge for a very restricted domain. The domain representation of PVT is based on knowledge about how to convert sentences from one voice to another. Since the student model cannot represent more than what the system can diagnose, PVT student model keeps track of how students use certain rules to change sentences. The development of PVT's student model does not seem to use any specific theory of SLA as a general framework.

5.1.5 Mr. Collins

Mr Collins (Bull et al., 1995) is a student model that is not specifically designed for ILTSs, and it can be used by different types of intelligent tutors for different domains. Bull et al. (1995) argue for extending the scope of student models to incorporate aspects outside the boundary of the domain knowledge, which they say “constitute important factors in learning” (Bull et al., 1995, p. 46). Mr. Collins has two modules that deal with aspects of the learning process that are not directly related to the domain knowledge. The analogy module keeps track of different types of analogy that can be used by the learner in his/her production. Specifically in language acquisition, this module can keep track of transfer issues from one language
to another. The learning strategies module can keep track of different strategies developed by the learner in their production. The example in Bull et al. (1995) shows a list of strategies described by O’Malley and Chamot (1990).

Unfortunately, Mr. Collins was only used in an ICALL system with a very restricted domain: Portuguese clitic placement. This makes it difficult to evaluate how the model proposed can be adapted to systems that have to deal with different linguistic forms and more sophisticated exercise types. Mr. Collins’ architecture posits three separate modules and does not address the question of linking specific aspects of the domain knowledge (or the language tasks in which they are needed) to specific analogies or learning strategies. However, the idea that student models have to be extended with information that does not necessarily belong to the domain knowledge seems to be in sync with general principles of SLA research that point to the fact that learning language forms is not enough for a student to acquire a second language (cf., e.g., Bachman, 1990; O’Malley and Chamot, 1990).

5.2 Modeling Student Knowledge

Student models commonly describe the current state of knowledge of a student. As we saw above, a typical way of doing it is to specify the knowledge units to be acquired, and to annotate them based on the student’s performance. We also saw that in ICALL, knowledge units are usually depicted as grammar rules.

Acquiring a second language is different from acquiring the ability to perform mathematical operations. We cannot reduce the acquisition of a different language to knowledge about its grammatical units. A more comprehensive student model for ICALL should take into consideration other aspects of the acquisition process.
The question is: what are these aspects? According to Ellis (2003) “the general goal of language learning is the fluent, accurate, and pragmatically effective use of the target language”. For learners to achieve this goal, they have to acquire more than grammatical competence. For Canale and Swain (1980) this goal is reached when a student acquires “communicative competence”. Communicative competence is achieved when the learner has acquired adequately four major types of knowledge. The first one is grammatical competence, which is knowledge of the rules of grammar. The second is sociolinguistic competence, which is knowledge of the rules of language use. While grammatical competence deals with issues of grammaticality of sentences, communicative competence deals with the appropriateness of utterances. The third type is discourse competence, which is what controls main discourse properties, such as cohesion, coherence, and reference. Although Canale and Swain (1980) present all types of knowledge as equally important for achieving communicative competence, the fourth type, strategic competence, is the one that plays a key role in our approach to student models.

Strategic competence is described by Canale and Swain (1980) as a series of compensatory communication strategies. The term compensatory is used because originally the strategies studied were the ones employed by speakers to convey a given message in situations where there is a breakdown in grammatical and/or sociolinguistic competencies. Ellis (2003) argues that overemphasizing the compensatory factor of strategic competence results in an extremely narrow view of the concept. We will adopt the definition of strategic competence presented by Bachman and Palmer (1996), following Bachman (1990). They “conceive of strategic competence as a set of metacognitive components, or strategies, which can be thought of as higher order
executive processes that provide a cognitive management function in language use, as well as, other cognitive activities” (Bachman and Palmer, 1996, p.70). This conceptualization of strategic competence will allow us to develop a learner model that circumscribes not only the student’s grammatical knowledge but also his/her ability to use this knowledge to complete different tasks in different environments.

5.3 Assessing Learner Knowledge to Build a Student Model

Student models are built and modified based on observations of learner performance (or using information explicitly provided by the learner). The student model is not intended to store properties of the output as such, but rather inferred information about the knowledge the learner used to construct these sentences.

Research in ICALL has paid little attention to the validity of the inferences about a student’s current state of knowledge made by a system. Developers usually take for granted that linguistic errors are caused solely by a lack of linguistic knowledge and do not acknowledge the fact that the task being performed can play a significant role in determining the students’ production.

To build a model that takes into account the linguistic and the strategic competence of a student, it is necessary to provide mechanisms ensuring that the system’s inferences about a student’s state of knowledge are valid.

Describing the concept of validity for language tests, Bachman and Palmer (1996) state that “construct validity pertains to the meaningfulness and appropriateness of the interpretations that we make on the bases of test scores” and that “in order to justify a particular score interpretation, we need to provide evidence that the test score reflects the area(s) of language ability we want to measure” (Bachman and Palmer,
Assigning and interpreting test scores is a similar process to describing a student’s current state of knowledge. In fact, scoring is a mechanism commonly used in student models to identify levels of proficiency for specific knowledge units. The issue of the validity of test scores thus applies directly to the validity of the information in a student model.

In the case of ICALL systems that present specific exercises, there are two issues related to the validity of system inferences that we need to pay particular attention to. The first one is known as content validity, which McNamara (2000, p. 50) characterizes as the concept that explains the “extent to which the test content forms a satisfactory basis for the inferences to be made from test performance.” For ICALL system design, this means that it is important to ensure that the exercise types and contents offered by the system are sufficient to make the necessary inferences about students’ state of knowledge.

The second issue on validity of inferences we want to highlight relates to the methods used to obtain information about students’ state of knowledge. There are two ways in which properties of exercises affect the result of the system’s observations, which we can characterize using notions from assessment theory (cf., e.g., McNamara, 2000). Construct irrelevant variance occurs when a given exercise introduces factors that are not relevant to measure the ability we want to observe. Construct under-representation occurs when the exercise is too easy for the student, jeopardizing the observation of a given ability. Particular care needs to be taken when the knowledge or skill observed is embedded in contexts that are unfamiliar to the student’s experience or irrelevant to what is being assessed. Bachman and Palmer emphasize that the analysis of a student’s performance has to be interpreted with respect to a “specific
domain of generalization”. Thus, when we consider the validity of an interpretation, “we need to consider both the construct definition, and the characteristics of the test tasks” (Bachman and Palmer, 1996, p. 21).

In sum, in order to guarantee valid interpretations of student performance it is not enough to keep track of students’ production; it is vital to have information about the task environment where it occurs. Without a clear description of the exercise items that triggered the student’s input, our interpretations about levels of proficiency may not be accurate.

5.4 TAGARELA’s Student Model

In section 4.4.1 we saw how activities are classified in TAGARELA, and the type of information each activity includes. The classification of activities proposed allows us to keep track of where specific errors occur. It means that we can store the student’s performance for each specific knowledge unit, in different environments. The consequence is that we can in principle infer more accurately if a student has mastered a given aspect of the language, and is able to use it to perform different tasks. This way, our student model can not only describe the acquisition of grammatical form, but it can also model the acquisition of specific strategies to employ the targeted grammatical forms. Ultimately, the model we propose describes the student’s abilities to perform with language. For ICALL systems, a student model that has a description of strategic competence acknowledges the fact that the nature of the task plays a significant role in student’s performance.

TAGARELA’s Student Model in figure 5.1 has three distinct repositories. The first one is the student’s personal information. It comprises general information about
the student, both static (e.g., name, age, and first language), and dynamic (e.g., level, other languages spoken). This information is obtained through a questionnaire that the learner has to fill out before using the system.

The second repository is where information about the preferences of the student when interacting with the system is stored. For example, it keeps track of how many times a student has seen the answer for a question before actually writing the correct response to it, or which activities are the student’s favorite ones based on the number of exercises done in a free choice setup such as TAGARELA.

Figure 5.1: TAGARELA’s Student Model
The third repository is where information about student’s performance and the system’s inferences about the student current state of knowledge is stored. As described in section 5.2, we call the student’s general competence to use the target language *Language Competence*. Language Competence is divided into two subtypes of competence: *Linguistic Competence* and *Strategic Competence*.

Linguistic competence is divided into *form-driven* and *content-driven*, reflecting the two types of linguistic analysis performed by the expert module. The leaves of the linguistic competence subtrees are the properties that can be observed through the analysis performed by the Analysis Manager. These properties are of two types, form features (FF, e.g., spelling, determiner–noun agreement, word order) and content features (CF, e.g., use of synonyms, extra/missing content words, required concept matching).

Strategic competence is divided into three groups. *Task appropriateness* keeps track of the performance of the student relative to the task classification in terms of the level, nature of input, and content manipulation as introduced in section 4.4.1. *Task Strategies* keeps track of which language strategies were used (or were not used) while completing the task. Remember that each activity is specified for the type of strategies necessary to properly answer the question. Finally, *Transfer* stores information about indicators of structural and lexical transfer from the native language of the student into the second language production. The leaves of the strategic competence subtree are the observable phenomena identified by the specific modules incorporated into the Feedback Manager based on the information specified for each activity as discussed in section 4.4.1.
5.5 Required Processing Modules

To cope with the incorporation of the Student Model into the system, four submodules have to be added to the system as presented by the architecture in figure 5.2. The first one is a submodule that evaluates the task strategies being used. The role of this module is to verify if the student has successfully accomplished the task and to create inferences about the strategies used by the student based on the activity model. Depending on how well the task was completed, the system can infer the type of strategies used by the learner. In principle, if a question is correctly answered with no variations\textsuperscript{27} from the pre-specified target answer, the system assumes that the strategies listed under that activity were successfully used.

The second submodule to be added is responsible for evaluating task appropriateness, i.e., for evaluating how appropriately a student has answered the questions based on the analysis of the student input and the description of the activity provided by the activity model. The submodule spells out an analysis of the input in relation to the activity. Its output is used by the Student Model Updater when updating the student model.

The third submodule that needs to be added deals specifically with errors triggered by negative lexical transfer.\textsuperscript{28} Basically, the module fetches the sentence elements that have been matched for part-of-speech in the shallow content analysis described in section 4.7.5, and checks them against a database of false cognates between English and Portuguese. If the canonic form of both elements are listed in the database as false friends, the system assumes it was a negative lexical transfer error.

\textsuperscript{27}To check the types of variations that the system currently accepts, see section 4.7.

\textsuperscript{28}In this first stage we are only dealing with lexical transfer, leaving a structural transfer to further studies.
The last necessary submodule is part of the Feedback Manager. It is a Student Model Updater whose roles are to fetch necessary information for a given question from the Student Model, to provide this information to the Feedback Generator, to receive the results of the final analysis of the student input, and to update the Student Model by adding these results to it.

Besides the incorporation of these four submodules, the algorithm used by the Feedback Generator, described in section 4.7.5, has to be changed to use the information originated in the Student Model in its decision process. Once information
about the learner is incorporated into the system’s feedback mechanism, feedback messages can be improved in numerous ways. Knowing about a student’s production patterns of persistent errors and specific difficulties can alter the nature of the feedback message considerably.

5.6 Individualized Feedback

Feedback messages can be improved in numerous ways, once information about the learner is incorporated into the system’s feedback mechanism. Knowing about a student’s patterns of persistent errors and specific difficulties can alter the nature of the feedback message considerably.

The model proposed for TAGARELA above incorporated three repositories of information to depict different properties of what is referred to as strategic competence. Let’s look at three examples to illustrate the impact of this type of information on feedback and inferences about the learner knowledge.

Information about task strategies help identify specific problems with certain strategies used by the student while performing the tasks, for example, in a case where the student has difficulties in scanning a text for specific information. Without the model, every time a content word in a reading comprehension question is missing the system reports the problem for the student using messages such as: “There is an important noun missing”. With a model, the system can provide feedback to target the student’s problem in scanning instead of treating the problem as an isolated phenomena of a missing content word. The feedback messages could include specific hints such as: “before you answer the question, find the necessary words in the text”, or “where in the text can you find the word to answer this question?”
The other repository is Task Appropriateness. The student’s performance by activity type can prevent the system from drawing wrongful inferences about the student’s ability to handle certain linguistic forms. Suppose a student only answers fill-in-the-blank questions and does not make subject–verb agreement errors. Without keeping track of where the correct form was used the system could infer that the student can handle such forms in any context. If the student while answering for the first time a Wh-reading comprehension question made an error in subject–verb agreement the system would provide feedback as if the student mastered that form.

The last type of information about negative transfer phenomena can be extremely useful in cases where students have persistent transferring problems. If the system can identify and keep track of these problems, such as misusing specific lexical items due to negative transfer, it can provide more specific feedback to help the learner understand the differences in lexical and/or syntactic properties between language items in L1 and L2.
CHAPTER 6

CURRENT INTEGRATION AND FUTURE WORK

The TAGARELA project started in Spring 2005, and the first version of the system started to be used by Portuguese Individualized Instruction Program students in Spring 2007. During the last two years, we developed all necessary components to create a functional ICALL system that can be used by language learners. The system is the result of a research program that focused on advancing ICALL for integration into real life foreign language teaching and learning.

When researchers decide their research agendas, they have to make choices not only about their areas of contribution, but also the issues that are not central to their approaches. As noted before, the focus of most ICALL research has been on advancing the technology. We recognize the importance of such contributions, without which it would have been impossible to develop our current work. However, we also believe that it is time for ICALL research to pay more attention to the gap between the NLP technology that has been developed in the last three decades and the use of this technology in real life language teaching.

In a way, the work presented here was inspired by previous approaches whose aims were to develop ICALL systems for existing language programs, such as Trude Heift’s
E-Tutor and Noriko Nagata’s Robo-Sensei\textsuperscript{29}. We hope that more ICALL systems for language programs will be developed in the next years, and that NLP technology for ICALL can take a more central role in the CALL community.

6.1 Current and Future Uses for TAGARELA

Students of the Portuguese Individualized Instruction Program (IIP) are currently using TAGARELA as extra practice material. The system has not replaced the paper based workbook used in the program. The Portuguese IIP program is undergoing some changes, and new pedagogical material will replace some of the current components of the course. One of the goals of this changes is to approximate the content of the IIP with the content of regular Portuguese classes, for which the new material is being phased in.

TAGARELA will also be updated to reflect the changes in the Portuguese program. However, we plan for the proposed changes to go beyond the structure of the Portuguese IIP, allowing the system to be used in regular classes as well. When the TAGARELA project was first presented and funded by the Office of Research at The Ohio State University, it was clear that the IIP program would benefit from such system, and TAGARELA was developed to cope with specific gaps in this type of program. After two years of research and the implementation of TAGARELA, there seems to be no reason to restrict the advantages of having an intelligent electronic workbook to individualized instruction.

For its next versions, TAGARELA will have a new menu with activities classified in a way that allows the system to be used by Portuguese students in individualized

\textsuperscript{29}See section 1.5 for a description of both systems.
and regular classes. The system will no longer present activities per module, which is
the typical IIP structure. They will be classified by units, and language topics with
cross references to the pedagogical material being used.

In the academic year of 2007/2008, two other institutions will join the TAGARELA
project, the University of Victoria in British Columbia – Canada, and the University
of Massachusetts Dartmouth. Students from both Portuguese programs will have
access to the system that will be phased in as the new electronic workbook. With
more students using the system, we are looking forward to more opportunities for a
wider range of research projects both on fostering the ICALL research as such and
on the SLA research questions on which the approach is based.

6.1.1 Students’ Opinions

In Spring 2007 TAGARELA was first used by students in the Portuguese program at
The Ohio State University. We asked a group of eleven students in the program to use
the system to complement their regular activities, and, after a month of interaction
with the system, we asked them to complete a questionnaire about their experience.
The goal of the questionnaire was to collect students’ opinions about different aspects
of the system, and to use the information, together with the log of their interaction,
as a point of departure for future evaluation and improvements.

The questionnaire presented nine propositions and the students had to evaluate
them using the following scale: *strongly agree, agree, neutral, disagree, and strongly
disagree*. For each proposition the students could also write a comment expressing
their opinions in more detail. The tenth item was a question about the students’
favorite exercise type.
<table>
<thead>
<tr>
<th>PROPOSITION</th>
<th>Strongly agree</th>
<th>Agree</th>
<th>Neutral</th>
<th>Disagree</th>
<th>Strongly disagree</th>
</tr>
</thead>
<tbody>
<tr>
<td>1) It was easy to navigate within the web interface and to find the exercises I wanted.</td>
<td>10</td>
<td>1</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>2) Instructions were clear and I knew what to do for each exercise.</td>
<td>9</td>
<td>1</td>
<td>1</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>3) I understood the functions of the buttons.</td>
<td>9</td>
<td>1</td>
<td>-</td>
<td>1</td>
<td>-</td>
</tr>
<tr>
<td>4) Exercises were useful and related to the topic of the unit.</td>
<td>9</td>
<td>2</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>5) Exercises helped me review and practice the topics I needed.</td>
<td>8</td>
<td>3</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>6) The feedback provided was useful</td>
<td>5</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>-</td>
</tr>
<tr>
<td>7) The feedback provided helped me think about my errors.</td>
<td>6</td>
<td>5</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>8) I feel more prepared to answer questions in my exam after I used the system.</td>
<td>8</td>
<td>3</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>9) I prefer using an electronic workbook like TAGARELA to a paper-based one.</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>2</td>
<td>-</td>
</tr>
</tbody>
</table>

Figure 6.1: Students' Opinions

The results for questions one, two and three show us that, in general, this group of students did not have problems with navigation and instructions. Although most of the interface was in the target language, they were able to find the exercises they wanted, understand instructions, and use the buttons to navigate. Some comments by students were:

- “Everything is placed very well and easy to find.”
- “Everything was available in a logical manner.”
- “At first I was confused of how to get to the next question but then I quickly figured it out.”
As we can see from the last comment, some students tried out some buttons at first to understand their functions, and because their places and functions were consistent throughout the different exercise pages, once students realized what the buttons were for, they could use them without any problems.

The results for questions four and five show that question types were appropriate to the level of students. In general students liked the design of exercises and agreed that the level of the language used fits their expectations and needs. One student complained about the listening passage, saying that it was a little fast, and several students said that description and fill in the blanks were well developed.

Questions six and seven show one area where the system can be improved. Although all students agree that the feedback messages made them think about their errors, three students were not satisfied with the messages provided. From their comments and the log that recorded their interaction with the system we could see that there were mainly two issues that bothered them: punctuation and accents. Students complained that the system cared too much about punctuation. The algorithm used by the feedback manager chooses punctuation errors to be reported if the rest of the input is considered to be correct. Students complained about this criterion.

The problem with accented characters is more complicated and led to some wrong interpretations of feedback messages. Errors with accentuation in Portuguese can lead to two problems with the word in the input. First, the misplacement or omission of an accent can generate a non-word, i.e., a spelling error. In this case, students did not have any problems. The system pointed out that the word was misspelled, and students had the chance to fix it. The second problem created more serious misunderstandings. Accentuation errors in Portuguese can often change the part of...
speech of word. In those cases there was a mismatch between the system’s analysis and the interpretation of the error by the student. What students considered to be an accentuation error, the system classified it as a wrong word (or part of speech) error.

Figures 6.2 and 6.3 show two examples that occurred with students while interacting with the system. In figure 6.2, the student wrote the conjunction “e” (and) instead of the verb “é” (present tense third person singular of the verb to be). The student complained about the feedback sending the following message: “I forget to put an accent over ‘e’ and it said there is an important verb missing.” In figure 6.3 a similar example occurred with another student who wrote the verb “medico” (present tense first person singular of the verb to medicate) instead of the noun “médico” (doctor). The student also sent a message pointing out what he considered to be a problem with the feedback message: “I typed in “ele é medico” and forgot to put an accent in “médico”. It then said: “There is an important noun missing in your sentence.” The accent mark was missing, not the noun.”

Some other examples show how accents in Portuguese can create specific difficulties to English speaking students. In view of this fact, it seems that the system would benefit from the implementation of a diagnosis module to deal specifically with accentuation errors. Based on students interactions with the system, we are currently analyzing the cases where such errors occur to propose an approach to the development of the module.
Figure 6.2: Accentuation Error – Verbs
Figure 6.3: Accentuation Error – Nouns
Question eight shows that most students though that the interaction with the system was useful to their learning process. However, question nine shows that not all of them are convinced that an electronic workbook can fully replace a paper-based one. The three quotes bellow nicely illustrate three reactions to the proposition in nine.

- “The good news is, one is able to get more work done in less time; however, I cannot retain the information learned as well as I can on paper.”
- “I think I would like doing more electronic work but still some paper-based ones.”
- “I really enjoyed using it and found it a lot more engaging than the books. This could be a great forum for or supplement to the courses.”

When asked what their favorite type of exercise is, students were divided, and five of them said that listening activities were the best ones. Students usually enjoyed the listening passages, and agreed that this type of exercise helped them develop their language skills.

- “I think listening is the most important because Portuguese being spoken is a lot different than being read. Getting used to the pronunciation and people speaking at a faster rate is really important.”
- “The listening exercises were great!”
- “I really like the listening exercises. They provided me with an opportunity to practice the Portuguese that I know so that I can get a better feeling of how the language is spoken.”

In general, students’ reactions towards the system were very positive. Further studies are necessary to determine to which degree an ICALL system such as TAGARELA can contribute to the language acquisition process, however, the small sample presented here shows that such a system has good potential to at least motivate students.

- “I think that this is a great program, and it definitely would be helping in individualized instruction.”

153
• “Everything was very useful. I wish they would have had something like this when I first started learning Spanish.”
• “I wish I could use Tagarela for all of my Portuguese classes that I am taking at Ohio State.”

6.2 Current and Future Research

Because of its multidisciplinary nature, the implementation of an ICALL system such as TAGARELA allows for a range of research projects in different areas. Here are some areas where we envisage research can evolve in the future.

6.2.1 Improving NLP components

We envisage that future NLP research will take two distinct directions. Firstly, the evaluation and improvement of current NLP modules. For syntax processing, for example, we want to independently test the two currently used approaches. We plan to expand the disambiguator module to perform the full syntactic analysis based on a Constraint Grammar approach instead of the current hybrid approach making use of a bottom-up parser. After implementing those changes, we can explore which of the two techniques can produce better results for error diagnosis in ICALL, or if a hybrid model such as the one used here is really the best choice.

Secondly, we plan to develop new modules to explore some of the deficiencies in current processing. For example, our shallow semantic analysis produces linguistic information about the nature of the tokens between the student’s answer and the target answer. However, nothing is said about the semantic relations of the tokens within the same sentence. We are currently exploring the effects of this deficiency for certain feedback strategies, and we believe that the system will be able to provide better feedback messages if it has information about some basic semantic roles and
relations in sentences. The first new module to be implemented uses the matching techniques developed by Bailey (2007) to match relations of elements in a single utterance.

6.2.2 Future SLA Research in ICALL

The feedback messages the system generates is the final result of all the processing and pedagogical choices the ICALL system designer has to make. The nature of these messages is influenced by exercise types, processing capabilities, and feedback strategies. Currently there are some studies in ICALL about the advantages of intelligent feedback (cf., e.g., Nagata, 1993), about how to deal with multiple errors (cf., e.g., Heift, 2003), and about the role of student models in ICALL feedback (cf., e.g., Heift, 2005). However, there are no studies about the nature of feedback messages and their role in the acquisition process.

For the development of TAGARELA, we took for granted the results of SLA research that show that feedback on form is more efficient if provided in meaningful contexts while the learner is performing a task.³⁰ To which extent these results can be transposed to ICALL is open to debate. There are no studies that actually show that ICALL activities and feedback are more effective following the same standards as oral classroom activities.

Because TAGARELA provides feedback on meaning and form, and because it offers a variety of activities that go from more form based to more content based, it is possible to evaluate the effectiveness of different types of feedback for different types of activities in an ICALL environment. For example, the system can be programmed

³⁰We are using the definition of ‘task’ as discussed in section 2.1.
to: (i) only provide feedback on form, (ii) only provide feedback on meaning, (iii) provide feedback on form first or on meaning first, or (iv) combine different feedback strategies for different activities types to evaluate the results. There are several experiments that have to be done using the variables at hand to better understand the effect of ICALL exercises and feedback on learners, and TAGARELA provides a sound environment for such research.

6.3 Conclusion

This dissertation presented the development of an ICALL system taking as a point of departure the needs and practices of foreign language students, instructors and researchers. As argued in the introduction, the approach presented here aims at reducing the gap that exists between the NLP technology that has been developed for ICALL, and the use of this technology in real life foreign language contexts. The TAGARELA project contributes to reduce this gap by providing a fully implemented system that uses current NLP techniques to process students’ input and provide feedback. There is still a lot of work ahead of us to popularize ICALL systems and make them more useful in CALL settings. The TAGARELA project is another important step in that direction.


Bailey, Stacey and Detmar Meurers (in prep), Exercise-driven selection of content matching methodologies.


Chapelle, Carol A. (2003), *English Language Learning and Technology: Lectures on teaching and research in the age of information and communication*, John Benjamins, Amsterdam.


Egbert, Joy L. (2005), Conducting research on call, in CALL Research Perspectives

Egbert, Joy L. and Gina Mikel Petrie (2005), CALL Research Perspectives, Lawrence Erlbaum Associates, Mahwah, NJ.


Ellis, Rod (2003), Task-based Language Learning and Teaching, Oxford University Press, Oxford, UK.


Ferris, Dana (1995), ‘can advanced ESL students be taught to correct their most serious and frequent errors?’, *ATESOL* **8**(1), 41–62.


Ferris, Danna, S. Chaney, K. Komura, B. Roberts and S. McKey (2000), Perspectives, problems, and practices in treating written error, in ‘Workshop Presented at the International TESOL Convention’, Vancouver, BC.


URL: citeseer.ist.psu.edu/gertner98procedural.html


Hughes, Jane, Claire McAvinia and Terry King (2004), ‘What really makes students like a web site? what are the implications for designing web-based language learning sites?’, ReCALL 16(1), 85–102.


Murphy, Linda (2005), ‘Attending to form and meaning: the experience of adult distance learners of French, German and Spanish’, Language Teaching Research 9(3), 295–317.


Polio, Charlene, Catherine Fleck and Nevin Lender (1998), ‘If only I had more time: ESL learners’ change in linguistic accuracy on essay revision’, *Journal of Second Language Acquisition* 7, 43–68.


169


Shedivy, Sandra (2004), ‘Factors that lead some students to continue the study of foreign language past the usual 2 years in high school’, System 32(1), 103–119.

Silva, Glaucia and Luiz Amaral (2005), Challenges for establishing a Portuguese individu¬

Skehan, Peter (1996), Second language acquisition research and foreign language in-
struction, in J.Willis and D.Willis, eds, ‘Challenge and Change in Language Teaching’, Heinmann.

Skehan, Peter (2003), ‘Focus on form, tasks, and technology’, Computer Assisted

Smith, Bryan (2005), ‘The relationship between negotiated interaction, learner up-
take, and lexical acquisition in task-based computer-mediated communication’, TESOL Quarterly 39(1), 33–58.

Spada, Nina (1997), ‘Form-focused instruction and second language acquisition: a
review of laboratory research’, Language Teaching 30, 73–87.

Suri, Linda (1991), Language transfer: A foundation for correcting the written english
of asl signers, Technical Report 91–19, Department of Computer and Information
Sciences, University of Delaware, Newark, DE.

Suri, Linda and Kathleen McCoy (1991), Language transfer in deaf writing: A correc-
tion methodology for an instructional system, Technical Report 91–20, Department
of Computer and Information Sciences, University of Delaware, Newark, DE.

Swain, Merrill (1984), A review of immersion education in canada: Research and
evaluation studies, in ‘A Collection of US Educators’, California State Department
of Education.

Tomita, Masaru (1984), Efficient Parsing for Natural Language, Kluer, Boston.


Tschichold, Cornelia (2003), ‘Lexically driven error detection and correction’, CAL-

Tsiriga, Victoria and Maria Virvou (2003), Evaluation of an intelligent web-based
language tutor, in ‘Knowledge-Based Intelligent Information and Engineering Sys-

Vandeventer Faltin, Anne (2003), Syntactic error diagnosis in the context of computer assisted language learning, Thèse de doctorat, Université de Genève, Genève.


APPENDIX A

Glossary of Acronyms

- AI – Artificial Intelligence
- CALL – Computer-Assisted Language Learning
- CBT – Content-Based Instruction
- CF – Content Feature
- CMC – Computer Mediated Communication
- FF – Form Feature
- FIB – Fill-in-the-Blanks
- FLTL – Foreign Language Teaching and Learning
- GB – Government and Binding
- HPSG – Head-driven Phrase Structure Grammar
- ICALL – Intelligent Computer-Assisted Language Learning
- IIP – Individualized Instruction Program
- ILTS – Intelligent Language Tutoring Systems
- ITS – Intelligent Tutoring System
- L1 – First or Native Language
- L2 – Second Language
- LFG – Lexical Functional Grammar
- NILC – Núcleo Interinstitucional de Lingüística Computacional
• NLP – Natural Language Processing
• OSU – The Ohio State University
• POS – Part-of-Speech
• PPP – Presentation, Practice and Production
• SLA – Second Language Acquisition
• SLT – Situational Language Teaching
• TAF – Task Appropriateness Feature
• TBI – Task-Based Instruction
• TSF – Task Strategy Feature
• TF – Transfer Feature
• UG – Universal Grammar
• UM – User Model
APPENDIX B

NILC Lexicon

In NILC’s lexical description the first position is always assigned to part of speech. The other ones are usually consistent, but there are variations among classes. Each item of the list is separated by “;” (semi-colon).

Whenever there are two or more entries for the same item (homonymy) the entries are separated by a # and are ordered according to their frequency in the NILC corpus. The frequency goes from 0 to 36.

eu=<PRON;2G;SI;[RET;}1S;?;?;C;}34;[eu]0;#S;M;SI;N;}?;?;20;[eu]0;>
The number after the canonic form (the last number in the entry) is used to flag if the lexical entry belongs to a locution, e.g. the entry:
“isso=<PRON;2G;SI;[DEM;}2S;?;?;C;}34;[isso]5;>”

belongs to the following locutions:
“por isso que=<LOC;[CONJ;[SUBORD;[CAUS;]]}]1;1;2;[isso]0;>”
“por isso tudo =<LOC;[ADV;[CIR-MOD;]N;} 1;1;2;[isso]0;>”
“por isso =<LOC;[ADV;[CIR-MOD;]N;} 1;0;2;[isso]0; #LOC;[CONJ;[SUBORD; [CAUS;]]}] 1; 0;2;[isso]0;>”
“visto isso =<LOC;[ADV;[CIR-MOD;]N;} 1;0;2;[isso]0;>”

Whenever there is an ”N;” in the middle of the entry, it indicates that the word can be all in capital letters, or no capital letters, or only the first letter is capitalized.

Let’s look at the information in the lexical entries for each part-of-speech.

B.1 Nouns

Examples:
menino=<S;M;SI;N;}?;?;29;[menino]0;>
meninos=<S;M;PL;N;}?;?;30;[menino]0;>
menina=<S;F;SI;N;}?;?;30;[menino]0;>
meninas=<S;F;PL;N;}?;?;29;[menino]0;>
estudante=<S;2G;SI;N;}?;1;30;[estudante]0;>

• POS: S (for substantivo, i.e. noun)
• Gender: M (masculine), F (feminine), 2G (both).
• Number: SI (singular), PL (plural), 2N (both).
• Degree: A (augmentative), D (diminutive), N (null)
• Regency: a list of possible prepositions. I don’t understand the criteria (it is
certainly not based on deverbal nouns)
• Gender generation rule: One number that describes how masculine of feminine
are morphologically formed. I don’t have these rules with me, but I can ask.
• Number generation rule: One number that describes how singular and plural
are formed. Same as above\textsuperscript{31}.
• Frequency in the corpus: from 0 to 36.
• Canonic form followed by locution flag number.

B.2 Proper Nouns

Examples:
luiz=<NOM;M;SI;LI;33;[luiz]5;>
amaral=<NOM;2G;SI;LI;2;[amaral]0;>
o’connor=<NOM;2G;SI;LX;[0;2;]2;[o’connor]0;>

• POS: NOM.
• Gender: M (masculine), F (feminine), 2G (both),
• Number: SI (singular), PL (plural), 2N (both).
• LX/LI: How to capitalize. LI means that first letter is capitalized and the
rest is not. LX followed by a list shows a mixture of capitalization and non-
capitalization, where the numbers on the list show which letters should be
capitalized (e.g. powerpoint=<NOM;M;SI;LX;[0;5;]2;[powerpoint]0;> = Power-
Point).
• Frequency in the corpus: from 0 to 36.
• Canonic form followed by a locution flag number.

\textsuperscript{31}Notice that these two rules are not consistently annotated in the lexicon. Otherwise the words
in the example would show a number for them.
B.3 Adjectives

Examples:

- aberta=<ADJ;F;SI;N;];?;30;[aberto]0;#V;[[PARTIC;F;SI;N;]29;[abrir]0;>
- abertas=<ADJ;F;PL;N;];?;29;[aberto]0;#V;[[PARTIC;F;PL;N;]28;[abrir]0;>
- gramaticais=<ADJ;2G;PL;N;];1;15;[gramatical]0;>
- grande=<ADJ;2G;SI;N;[de;em;por];?;33;[grande]5;#S;2G;SI;N;];?;32;[grande]5;>

- POS: ADJ
- Gender: M (masculine), F (feminine), 2G (both).
- Number: SI (singular), PL (plural), 2N (both).
- Degree: A (augmentative), D (diminutive), SU (superlative), N (null).
- Regency: a list of possible prepositions. I don’t understand the criteria (complements? adjuncts? – “grande” above does not make sense to me)
- Gender generation rule: One number that describes how masculine of feminine are morphologically formed. I don’t have these rules with me, but I can ask.
- Number generation rule: One number that describes how singular and plural are formed\[32\].
- Frequency in the corpus: from 0 to 36.
- Canonic form followed by locution flag number.

B.4 Adverbs

Examples:

- anualmente=<ADV;[CIR-TEMP;]N;28;[anualmente]0;>
- aonde=<ADV;[INT-LUG;]N;19;[aonde]0;#PREP;C;[a;onde;]3;[aonde]0;>
- apaixonadamente=<ADV;[CIR-MOD;]N;6;[apaixonadamente]0;>

- POS: ADV.
- Type: CIR-LUG (circunstância lugar), CIR-TEMP (circunstância tempo), CIR-MOD (circunstância modo), (negação), DUV (dúvida), INT(intensidade), AFIR (afirmação), INT-TEMP (interrogativo de tempo), INT-MOD (interrogativo de modo), (interrogativo de causa), INT-LUG (interrogativo de lugar).
- Degree: A (augmentative), D (diminutive), SU (superlative), N (null).

\[32\]Same evaluation as for nouns.
• Frequency in the corpus: from 0 to 36.

• Canonic form followed by a locution flag number.

B.5 Pronouns

Examples:
eu=<PRON;2G;SI;[RET;]1S;?;?;C;[ ]34;[eu]0>
onde=<PRON;2G;2N;[INDE;REL;]N;?;?;C;[2;onde]5>
se=<CONJ;[SUBORD;[COND;INTE;]]35;[se]0;#PRON;2G;2N;[OBL-AT;REFL;]
3S;?;?;C;[35;[se]0;#SIGL;LC;2;[se]0;>

• POS: PRON

• Gender: M (masculine), F (feminine), 2G (both).

• Number: SI (singular), PL (plural), 2N (both).

• Type: RET (pessoal reto – subject pronoun), OBL-TO (oblíquo tônico – tonic oblique), OBL-AT (oblíquo átomo – atonic oblique), DEM (demonstrativo – demonstrative), TRAT (tratamento – treatment), INDE (indefinido – indefinite), POSS (possessivo – possessive), REL (relativo – relative), REFL (reflexivo – reflexive).

• Person: 1S, 2S, 3S, 1P, 2P, 3P, N (null – no value for person).

• Gender generation rule: One number that describes how masculine of feminine are morphologically formed. I don’t have these rules with me, but I can ask.

• Number generation rule: One number that describes how singular and plural are formed. Same as above.

• Contraction: C = contracted

• Frequency in the corpus: from 0 to 36.

• Base form followed by locution flag number.

B.6 Prepositions

Examples:
para=<PREP;35;[para]0;#PREF;2;[para]0;>
pro=<PREP;C;[para;o;]2;[pro]5;>
pra=<PREP;C;[para;a;]30;[pro]0;>
• POS: PREP

• Contraction: C = contracted. Notice that when the preposition is not contracted the slot simply disappears and the frequency (next category) comes in its place.

• the base forms (pair of words) between brackets followed by frequency: [para;o]2.

• Canonic form followed by frequency: [pro]. The problem of using this canonic form is that it does not refer to the basic form, but to the contracted masculine form.

B.7 Articles

Example:
\[
a = \langle \text{ART}; F; SI; DE; ?; ?; 35; [o]0; \rangle
\]
\[
o = \langle \text{ART}; M; SI; DE; 1; 11; 35; [o]0; \rangle
\]

• POS: ART

• Gender: M (masculine), F (feminine)

• Number: SI (singular), PL (plural)

• Type: DE (definido – definite), I (indefinido – indefinite).

• Gender generation rule: One number that describes how masculine or feminine are morphologically formed.

• Number generation rule: One number that describes how singular and plural are morphologically formed.

• Frequency.

• Canonic form followed by locution flag number.

B.8 Conjunctions

Example:
\[
como = \langle \text{CONJ}; [\text{SUBORD}; [\text{CAUS}; \text{CONFOR}; \text{COMP;}]] 35; [como]5; \rangle
\]
\[
e = \langle \text{CONJ}; [\text{COORD}; [\text{ADIT}; \text{ADVE;}]] 35; [e]5; \rangle
\]

• POS: CONJ

• Type: SUBORD (subordinativa – subordinate), COORD (coordinative – coordinate).
- Sub-type of subordinative: CAUS (causal – causal), CONFORM (conformativa – conformative), COMP (comparativa – comparative), CONC (concessiva – concessive), PROPOR (proporcional – proportional), TEMP (temporal), CONS (consecutiva – consecutive), FIN (final), INTE (integrante – integrating), COND (condicional, conditional).

- Sub-type of coordinate: ADIT (aditiva – additive), ADV (adversativa – adversative), ALTER (alternativa – alternative), CONCL (conclusiva – conclusive), EXPL (explicativa – explicative).

- Frequency (follows sub-type)

- Canonic form followed by locution flag number: [como]5.

B.9 Verbs

The conjugated forms do not bring information about predication(valency). It inherits it from the base form. The bare infinitive is derived from the 1st and 3rd persons of the personal infinitive, and this fact is not explicitly indicated. The participle is followed by gender and number markers.

Examples:

- abalaria=<V;[[FUT-PRET;ELE;FUT-PRET;EU;]N;]9;[abalar]0;>
- abalar=<V;[TD;BI;INT;PRONOM;][FUT-SUBJ;ELE;FUT-SUBJ;EU;INF-PESS;ELE;INF-PESS;EU;]N;[a;contra;de;]25;[abalar]0;>
- ser=<V;[AUX;LIG;TI;][INF-PESS;ELE;INF-PESS;EU;]N;[com;de;para;por;]35;[ser]5;#S;M;SI;N;ℶ;1;30;[ser]5;>
- sendo=<V;[[GERUN;]N;]34;[ser]0;>
- sentada=<V;[[PARTIC;F;SI;N;]20;[sentar]0;#ADJ;F;SI;N;ℶ;?;2;[sentado]0;>

- POS: V.

- Predicação(valency): INT (intransitivo – intransitive), TD (transitivo direto – direct transitive), TI (transitivo indireto – indirect transitive), BI (bitransitivo – ditransitive), LIG (ligação – linking), AUX (auxiliar – auxiliary), PRONOM (pronominal).


180

- Person: the persons follow the verb form/tense. EU (1s), TU (2s), ELE (3s), NOS (1p), VOS (2p), ELES (3p).

- "N"

- Regency: List of prepositions (it can be empty).

- Frequency: 0 – 36.

- Base form followed by locution flag number.

### B.10 Locutions

Examples:

- à base de=<LOC;[PREP;1;1;2;[base]0;>  
  a lápis=<LOC;[ADV;[CIR-MOD;]N;]1;0;2;[lápis]0;>  
  a luz solar=<LOC;[ADJ;F;SI;N;][]0;0;1;1;2;[luz]0;>  
  à medida que=<LOC;[CONJ;[SUBORD;[PROPOR;]]]1;1;2;[medida]0;>  
  aeroporto internacional charles de gaulle=<LOC;[NOM;M;SI;LX;[0;10;24;35;]]2;2;2;[charles]0;>  

- POS: LOC.

- Type: ADV (adverbial), PREP (prepositional), ADJ (adjectival), CONJ (conjunction-like), NOM (substantiva – nominal)

- Number of words before the canonical word.

- Number of words after the canonical word.

- Canonical word.

- Frequency followed by locution flag.

---

33the definition of canonical seems to be in terms of content and not in terms of the syntactic notion of head.
B.11 Numerals

\[\text{um}=\langle\text{ART};\text{M};\text{SI};\text{I};1;3;35;[\text{um}]0;\#\text{NUM};\text{M};\text{SI};\text{CAR};3;?;32;[\text{um}]0;\rangle\]
\[\text{primeiros}=\langle\text{ADJ};\text{M};\text{PL};\text{N};[\text{primeiro}]0;\#\text{NUM};\text{M};\text{PL};\text{ORD};1;1;31;[\text{primeiros}]0;\rangle\]
\[\text{duplo}=\langle\text{ADJ};\text{M};\text{SI};\text{N};[\text{duplo}]0;\#\text{NUM};\text{M};\text{SI};\text{MUL};1;1;3;[\text{duplo}]0;\rangle\]
\[\text{milésima}=\langle\text{NUM};\text{F};\text{SI};\text{FRA-ORD};1;1;2;[\text{milésimo}]0;\rangle\]

- POS: NUM
- Gender: M (masculine), F (feminine), 2G (both)
- Number: SI (singular), PL (plural), 2N (both)
- Type: CAR (cardinal), ORD (ordinal), MUL (multiplicativo – multiplying), FRA (fracionário – fractional), FRA-ORD (fracionário ordinal), couldn’t find any example (coletivo – collective)
- Gender generation rule: One number that describes how masculine or feminine are morphologically formed.
- Number generation rule: One number that describes how singular and plural are morphologically formed.
- Frequency.
- Canonic form followed by locution flag number.

B.12 Interjections

Examples:
\[\text{oi}=\langle\text{INTERJ};2;[\text{oi}]0;\rangle\]
\[\text{atenção}=\langle\text{S};\text{F};\text{SI};\text{N};[\text{atenção}]5;\#\text{INTERJ};2;[\text{atenção}]5;\rangle\]

- POS: INTERJ.
- Frequency: 0 – 36.
- Canonic form followed by locution flag number.

B.13 Acronyms

Examples:
\[\text{abnt}=\langle\text{SIGL};\text{LC};2;[\text{abnt}]0;\rangle\]
\[\text{autocad}=\langle\text{SIGL};\text{LX};[\text{autocad}]0;\rangle\]
• **POS: SIGL.**

• **LX/LC/LI:** How to capitalize. LI means that first letter is capitalized and the rest is not. LX followed by a list shows a mixture of capitalization and non-capitalization, where the numbers on the list show which letters should be capitalized (see example above). LC means that all letters should be capitalized (e.g. ABNT).

• Frequency: 0 – 36.

• Canonic form followed by locution flag number.

### B.14 Prefixes

Examples:

auto=\<PREF;2;\{auto\}0;\#S;M;SI;N;\?;\?;1;20;\{auto\}0;>  
pan=\<PREF;2;\{pan\}0;\#S;M;SI;N;\?;\?;2;\{pan\}0;\#SIGL;LC;2;\{pan\}0;>

• **POS:** PREFIX

• Frequency: 0 – 36.

• Canonic form followed by locution flag number.

### B.15 Abreviations

Examples:

adj.=\<ABREV;M;SI;LN;2;\{adj\}0;>  
cia.=\<ABREV;F;SI;LX;0;2;\{cia\}0;>

• **POS:** ABREV.

• Gender: M (masculine), F (feminine), 2G (both)

• Number: SI (singular), PL (plural), 2N (both).

• **LX/LC/LI:** How to capitalize. LI means that first letter is capitalized and the rest is not. LX followed by a list shows a mixture of capitalization and non-capitalization, where the numbers on the list show which letters should be capitalized (see example above). LC means that all letters should be capitalized (e.g. ABNT). LN means that the word can be all capitalized, or with no capitalized letter, or just the first letter capitalized.

• Frequency: 0 – 36.

• Canonic form followed by locution flag number.