

Contents

Abstract	VIII
List of Tables	X
List of Figures	XII
List of Acronyms	XVI
Introduction	1
1. Second Language Acquisition	6
1.1. Introduction	6
1.2. Second and Third Language Acquisition	6
1.2.1. Second Language Acquisition	7
1.2.2. Specific Properties of Third Language Acquisition	8
1.3. Language Transfer	10
1.4. Language Awareness	11
1.5. Summary	12
2. State of the Art in Related Areas of Language Teaching and Learning	13
2.1. Introduction	13
2.2. Plurilingual Language Teaching and Learning	13
2.2.1. Intercommunicabilité Romane	15
2.2.2. Galatea and Galanet	16
2.2.3. EuroComRom	17
2.2.4. EuRom 4	18
2.3. Computer-Assisted Language Learning	19
2.3.1. CALL	19
2.3.2. ICALL	20

2.3.3.	Use of Animation in Language Teaching	22
2.3.4.	Dictionary Look-Up Tools	27
2.4.	Research Questions for this Thesis	30
2.5.	Summary	31
3.	CALL Design Methodology	32
3.1.	Introduction	32
3.2.	CALL Design	32
3.3.	Hubbard’s Methodological Framework	33
3.4.	Colpaert’s RBRO Model	35
3.5.	ESPRIT Design Methodology	38
3.5.1.	Comparing Methodologies and Assessing their Suitability for ES- PRIT	38
3.5.2.	Analysis and Design Phases of ESPRIT	40
3.6.	Summary	43
4.	General Properties of the ESPRIT Approach	45
4.1.	Introduction	45
4.2.	Languages	45
4.3.	Target Learner Group and Learning Method	47
4.4.	Reusing Existing Software and Language Content Resources	51
4.4.1.	Bilingual Word Lists	51
4.4.2.	Verb Conjugator	52
4.4.3.	Romance Lexicon	52
4.4.4.	Wikipedia Articles	53
4.4.5.	POS Taggers	54
4.4.6.	JFrost Lemmatiser	55
4.5.	Usability and Software Ergonomics	58
4.6.	Linguistic Levels and Modalities	59
4.6.1.	Morphosyntax	60
4.6.2.	Syntax	61
4.6.3.	Lexicon	61
4.6.4.	Semantics	62

4.7. Software Architecture	62
4.7.1. Flash	63
4.7.2. Data Storage and Data Processing	68
4.7.3. XML	68
4.7.4. Natural Language Processing	70
4.8. Summary	71
5. A Toolsuite for Plurilingual CALL Applications	72
5.1. Introduction	72
5.2. Graphical User Interface	73
5.3. Multilingual and Plurilingual Language Resources	76
5.3.1. Multilingual Lexicon	76
5.3.1.1. XML Version	77
5.3.1.2. MySQL Version	79
5.3.2. Full-Form Verb Lists	80
5.3.3. Multilingual Verb Lexicon	81
5.3.4. Plurilingual Lexicon	82
5.4. Dictionary Tools	83
5.4.1. Multilingual Dictionary Tool	84
5.4.1.1. Modules	86
5.4.1.2. Language Processing	88
5.4.2. Plurilingual Dictionary Tool	92
5.4.2.1. Techniques and Resources for the Dynamic Detection of Similar Words	95
5.4.2.2. Combining Sound Correspondences with the Levenshtein String Similarity Measure	96
5.5. Multilingual Concordancer	97
5.6. Plurilingual Input Analysis and Feedback Module	100
5.6.1. Language Processing	101
5.6.1.1. Step 1 – Retrieving Lexical Information	102
5.6.1.2. Step 2 – Morphosyntactic Input Analysis and Error Recognition	102

5.6.1.3. Related Work	110
5.6.2. Graphical User Interface	113
5.6.3. Plurilingual Lexicon Interface Component	115
5.6.4. Multilingual Verb Lexicon Interface Component	116
5.7. Animated Grammar Presentations	117
5.7.1. Interactive Elements	122
5.7.2. Options for Customisation	125
5.8. Authoring Tool for Learning Materials	125
5.8.1. Modules	126
5.9. Summary	129
6. Testing and Implementation	131
6.1. Introduction	131
6.2. Local Testing and Implementation on a Remote Web Server	131
6.3. Diacritical Characters	133
6.4. Handling Inconsistent Tagger Output	136
6.5. Data Communication between Server and Browser	136
6.6. Summary	137
7. Evaluation	139
7.1. Introduction	139
7.2. Evaluation Platforms	139
7.3. Formative Evaluation	145
7.3.1. Your Languages	145
7.3.2. Plurilingual Learning	146
7.3.3. Language-Learning Software	148
7.3.4. Animated Grammar Presentations	148
7.3.5. Multilingual Dictionary Tool	153
7.4. Summative Evaluation	156
7.4.1. Pre-Questionnaire	157
7.4.2. Text Tools	159
7.4.3. Sentence Structures	161
7.4.4. Graphical User Interface	164

7.4.5. Plurilingual Input Analysis and Feedback Module	166
7.4.5.1. Analysable Learner Input	167
7.4.5.2. Unanalysable Learner Input	169
7.5. Summary	172
8. Conclusions and Further Work	173
8.1. Migrating ESPRIT Tools to Other Languages	176
8.2. Migration to Mobile Devices	177
8.3. Creating Browser Extensions	178
Appendices	180
A. URLs	180
B. Evaluation	182
B.1. Formative Evaluation	182
B.1.1. Your Languages	182
B.1.2. Plurilingual Learning	183
B.1.3. Language Learning Software	186
B.1.4. Animated Grammar Presentations	190
B.1.5. Multilingual Dictionary Tool	197
B.2. Summative Evaluation	204
B.2.1. Pre-Questionnaire	204
B.2.2. Text Tools	209
B.2.3. Sentence Structures	212
B.2.4. Graphical User Interface	214
Bibliography	217

Introduction

Research in plurilingual teaching and learning of Romance languages has shown that a combined approach to teaching Romance languages is very promising (cf. Bär, 2004; Klein, 2004). It can exploit the similarities between these languages in many ways in order to teach them contrastively and to raise the language awareness of the learner (cf. Hawkins, 1984). Within the range of this Ph.D. dissertation, it refers to the explicit knowledge about properties and processes in the languages involved and their conscious perception in language learning and language use.

To date, several European projects have been devoted to plurilingual teaching and learning of Romance languages, mainly focusing on acquiring reading competence. These projects only produced ‘static’ materials for classroom teaching or computer-based self-learning, without flexible and dynamic techniques from Natural Language Processing (NLP) or Artificial Intelligence (AI), which would support an interactive learning environment with user input and dynamically generated system feedback. To date, there exist – to the best of my knowledge – no plurilingual ICALL (Intelligent Computer-Assisted Language Learning) systems for plurilingual teaching and learning of foreign languages.

The research goal of my Ph.D. dissertation was the design, development, implementation and evaluation of a plurilingual ICALL software system ESPRIT¹ for French, Spanish and Italian, aimed at advanced learners. I investigated how techniques from NLP can enhance the plurilingual teaching and learning of these languages.

Generally the definitions and the use of the terms *multilingual* and *plurilingual* vary considerably. In the context of this Ph.D. dissertation, *plurilingual* means that grammatical and lexical properties of the languages involved are tightly linked to each other, showing a high degree of similarity in form and function. Here plurilingual relates to languages of the same language family, such as Romance languages. *Multilingual*, in

¹ESPRIT: ESPañol, fRançais, ITaliano

contrast, refers to the simple fact that language content and processing is available for different languages, for example word lists in French, Italian and Spanish based on the same topics or web sites with several language versions.

ESPRIT targets native English and German native speakers who are already at an advanced level in at least one of the Romance languages involved. These learners are expected to be familiar with general lexical and grammatical properties of this language (e.g. word classes, sentence patterns, use of prepositions).² Equivalent properties of the other languages are taught through comparison.

The research questions addressed in this Ph.D. dissertation build upon the general research findings in plurilingual teaching and learning of Romance languages, CALL (Computer-Assisted Language Learning) and ICALL, and the use of animation in language teaching.

Existing materials for plurilingual learning of Romance languages almost exclusively focus on receptive skills and lack any kind of intelligent automatic analysis of learner input as well as flexible and dynamic feedback to support interactive learning. Problems often cited for existing CALL and ICALL software are restricted input and simple feedback (Vandevanter Faltin, 2003: 27ff.). The lack of appropriate evaluation often leads to technically driven development. Limited data sources and a low degree of interactivity can hinder the learner to actively engage in a broad range of language-learning activities.

I developed a plurilingual input analysis and feedback module to provide flexible and precise feedback. I used flexible database technologies and a strict separation between storing and displaying/processing language data so that the same set of language data can easily be reused in different language learning activities. In order to dynamically present grammatical properties and processes and to support a high degree of interactivity, I created custom-made animated grammar presentations and developed an authoring tool for animated text. As much as possible I reused existing language data collection and processing resources. To avoid technically driven development, different components of ESPRIT were continually assessed by language learners.

This thesis is structured as follows:

²Existing research projects in plurilingual learning (such as EUROCOMROM and EUROM 4) are based on a similar lexical and grammatical level of knowledge of learners.

Chapter 1 provides general definitions of second language acquisition (SLA) and outlines the differences between SLA and third language acquisition (TLA). This chapter also introduces the notions of language transfer and language awareness which are central issues to the concept of plurilingual learning.

Chapter 2 describes previous and current research in the areas which are of most importance to this Ph.D. research, namely plurilingual teaching and learning, CALL and ICALL, the use of animation in language teaching and NLP-based dictionary look-up tools. In this chapter I identify a number of shortcomings in previous research projects. This helped to formulate the research questions informing my Ph.D. research.

Chapter 3 relates the design and development process of ESPRIT to two widely recognised CALL design methodologies, Hubbard's Methodological Framework (Hubbard, 1996) and Colpaert's Research-Based Research-Oriented (RBRO) design model (Colpaert, 2004). The chapter identifies the common ground between these methodologies and the design approach adopted for ESPRIT and describes special characteristics of the ESPRIT design approach.

Chapter 4 provides detailed information about the general properties and principles of ESPRIT. The chapter describes the target group and learning method of ESPRIT and gives an overview of existing language tools and resources which were successfully integrated into ESPRIT. It also examines important usability issues and software ergonomics relevant to CALL applications. The chapter describes the linguistic levels and modalities of ESPRIT and provides detailed information about the software architecture used for ESPRIT.

Chapter 5 details the components developed for ESPRIT. This includes a flexible web-based graphical user interface (GUI), different types of language tools and resources, tailor-made animated grammar presentations and an authoring tool for the creation of slide-based learning materials with animated text. Language tools developed for ESPRIT are multilingual and plurilingual dictionary tools, a multilingual concordancer, a plurilingual input analysis and feedback module, a plurilingual lexicon interface component and a multilingual verb lexicon interface

component. Language resources comprise a multilingual lexicon, full-form verb lists, a multilingual verb lexicon and a plurilingual lexicon.

Chapter 6 provides information on the testing of ESPRIT tools in a local server environment and the subsequent implementation of ESPRIT tools on a remote web server. This chapter describes a number of problems which were encountered during testing and implementation stages and the solutions to these problems.

Chapter 7 presents the results of the formative and summative evaluation processes. The chapter also provides a detailed description of a fast and reliable web-based database-driven evaluation platform which can easily be adapted to other evaluation projects.

Chapter 8 summarises this Ph.D. dissertation and outlines possible language-learning scenarios in which ESPRIT tools may be used in ways which differ from the specific context of this Ph.D. dissertation.

Background and Motivation

After having learned French in secondary school for seven years, I started to learn Spanish and Italian in the early and mid 90's, respectively. Right from the beginning, I was aware of a large number of similarities between these languages leading to positive and negative language transfer (cf. Odlin, 1989). *Positive transfer* exists if words or grammatical structures in two languages have a similar form and function (e.g. *casa* in both Spanish and Italian means *house*), whereas in *negative transfer* – often called *false friends* – there is only a similarity of form leading to the (false) conclusion that the function would be the same: *fermer* in French means *to close*, whereas *fermare* in Italian means *to stop*.

Although the similarities between Romance languages – which are for the most part due to their common (Latin) root – have been described extensively in contrastive linguistics for decades (e.g. Bodmer, 1944), a broader interest in research on plurilingual teaching and learning only emerged in the 1990s. Foreign language teaching in secondary schools and at universities, however, has been largely unaffected by this research. Language students at both levels only occasionally get the opportunity to learn similar languages simultaneously in a plurilingual setting.

Plurilingual teaching is potentially highly effective (cf. Bär, 2004), yet plurilingual teaching and learning material is quite hard to obtain. In my Ph.D. research, I aimed to design, develop, implement and evaluate software to support plurilingual teaching and learning of Romance languages which helps language learners to optimally exploit their existing knowledge in any one Romance language.