Modeling second language learners' interlanguage and its variability:

A computer-based dynamic assessment approach to distinguishing between errors and mistakes

Sylvie Thouësny

Licence d'allemand (Université de Franche-Comté)
B.Sc. in Applied Computational Linguistics (Dublin City University)

Thesis submitted for the degree of Doctor of Philosophy
Ph.D.

Dublin City University
School of Applied Language and Intercultural Studies

Under the supervision of Dr. Françoise Blin

September 2011

Declaration

I hereby certify that this material, which I now submit for assessment on the programme of study leading to the award of Ph.D. is entirely my own work, that I have exercised reasonable care to ensure that the work is original, and does not to the best of my knowledge breach any law of copyright, and has not been taken from the work of others save and to the extent that such work has been cited and acknowledged within the text of my work.

ID No.: 52146855

Date:____________________________

Signed:__________________________

Abstract

Despite a long history, interlanguage variability research is a debatable topic as most paradigms do not distinguish between competence and performance. While interlanguage performance has been proven to be variable, determining whether interlanguage competence is exposed to random and/or systematic variations is complex, given the fact that distinction between competence-dependent errors and performance-related mistakes should be established to best represent the interlanguage competence.

This thesis suggests a dynamic assessment model grounded in sociocultural theory to distinguish between errors and mistakes in texts written by learners of French, to then investigate the extent to which interlanguage competence varies across time, text types, and students. The key outcomes include:

1. An expanded model based on dynamic assessment principles to distinguish between errors and mistakes, which also provides the structure to create and observe learners' zone of proximal development;

2. A method to increase the accuracy of the part-of-speech tagging procedure whose reliability correlates with the number of incorrect words contained in learners' texts;

3. A sociocultural insight into interlanguage variability research. Results demonstrate that interlanguage competence is as variable as performance. The main finding shows that knowledge over time is subject to not only systematic, but also unsystematic variations.

Keywords: sociocultural theory, zone of proximal development, interlanguage variability, knowledge modeling, error, mistake, dynamic assessment.

To my late grandmother

Acknowledgements

First and foremost, I would like to take this opportunity to recognise the time and effort devoted by Dr. Françoise Blin who insightfully supervised this doctoral research. From one stage to the next, she not only spent tremendous time reading the successive drafts of this thesis, but also acted as a truly compassionate person when I was no longer in the mood to do research. Without her support and encouragement over the five past years, this thesis would have never taken shape.

I would also like to express my gratitude to Mat Schulze, Peter Wood and Bjanka Pokorny for providing me with a copy of their unpublished manuscript about balanced complexity measure. Special thanks to Mat for offering me the tools to compute the coefficient and for taking the time to discuss it with me. I also acknowledge Mr. Foley for accepting the time consuming task of second rater on part-of-speech tagging, and equally thank Detmar Meurers, Marcus Dickinson and Jennifer Foster for their helpful perspicacious comments and suggestions on the method applied to improve the tagging process.

Furthermore, I wish to show my appreciation to Marta Antón, Trude Heift, Katsunori Kotani, Matthew Poehner, Larry Selinker, Peter Smagorinsky, Judith Szerdahelyi, and Elaine Tarone for benevolently sharing some of their publications with me. I would also like to thank Sharon Crout and Lori Bailey, editorial assistant and director, respectively, of the editorial office of the American Journal of Orthopsychiatry, for providing me with a copy of Luria's article originally entitled "An objective approach to the study of the abnormal child", published in 1961.

A special mention should also be made to the staff and fellow Ph.D. students in the School of Applied Language and Intercultural studies at Dublin City University for their discussions on many occasions, as well as for the students who took part in this study. More particularly, I would like to express my gratitude to Annette Simon who kindly accepted to provide comments on an earlier version of this thesis, and equally thank Cathy Fowley for her continuous comfort and understanding of what it truly means to do a doctorate.

My thanks are equally addressed to my examiners Trude Heift and Monica Ward for not only making my viva voce a truthfully challenging and enjoyable experience, but also offering valuable criticisms of the thesis and providing constructive ideas.

I would also like to give credit to my family in general, who stood by my side in my writing task, but in particular to my two daughters Karine and Fany for their constant support and encouragement and for making my life easier. My gratitude is also extended to Katherine, my daughter in-law, for her valuable proofreading of the final draft of this thesis.

And last but certainly not least, I will never ever thank enough my husband Raphaël for all he did for me, and more specifically for *always* being there for me!

Table of Contents

List of Tables

List of Figures

List of Formulae

List of publications

Articles in refereed journals

Thouësny, S. (2010). Investigating learner variability: The impact of task type on language learners' errors and mistakes. *CALICO Journal.* 28 (1), pp. 21-34.

Chapters in refereed volumes

Thouësny, S. (2011). Dynamically assessing written language: to what extent do learners of French language accept mediation? In S. Thouësny & L. Bradley (Eds.), *Second language teaching and learning with technology: views of emergent researchers* (pp. 169-188). Dublin: Research-publishing. net.

Thouësny, S., & Blin, F. (2011). Modeling language learners' knowledge: What information can be inferred from learners' free written texts? In M. Levy, F. Blin, C. Bradin Siskin, & O. Takeuchi (Eds.), *WorldCALL: International Perspectives on Computer-Assisted Language Learning* (pp. 114-127). New York: Routledge.

Proceedings

Thouësny, S. (2010). Assessing Second Language Learners' Written Texts: An Interventionist and Interactionist Approach to Dynamic Assessment. In *Proceedings of World Conference on Educational Multimedia, Hypermedia and Telecommunications (EDMEDIA), 28-JUN-10 - 02-JUL-10, Toronto, Canada, 2010.* (pp. 3517-3522). Chesapeake, VA: AACE.

Thouësny, S., & Blin, F. (2009). Modeling language learner knowledge state: what are language students' free written productions telling us? In T. Koyama (Ed.), *Proceedings of the WorldCALL 2008 Conference, CALL Bridges the World, August 5 - 8, 2008* (pp. 64-67). Fukuoka, Japan: The Japan Association for Language Education and Technology.

List of presentations

Invited talks

Thouësny, S. (2009). "*Part-of-speech tagger: improving the reliability when processing language learners' language*". Invited talk. National Centre for Language Technology (NCLT Seminar Series 2008/2009), 24-JUN-09, Dublin City University, Dublin, Ireland.

Conferences

Thouësny, S. (2010). "*Assessing Second Language Learners' Written Texts: An Interventionist and Interactionist Approach to Dynamic Assessment*". Paper presentation. ED-MEDIA 2010 World Conference on Educational Multimedia, Hypermedia & Telecommunications, 28-JUN-10 - 02-JUL-10, Toronto, Canada.

Thouësny, S. (2010). "*Dynamically assessing written language: To what extent do learners of French language accept mediation?*". Paper presentation. CALICO 2010 Conference, Enhancing Language Learning: Research, Innovation, and Evaluation in CALL, 08-JUN-10 - 12-JUN-10, Amherst College, Amherst, MA, US.

Thouësny, S. (2010). "Assessing written language of learners of French: a computer-based dynamic assessment approach to distinguishing between errors and mistakes". Paper presentation. ADEFFI Annual Postgraduate Symposium, 10-APR-10, Trinity College, Dublin, Ireland.

Thouësny, S. (2009). "*Investigating Learner Variability: The Impact of Task Type on Language Learner Errors and Mistakes*". Paper presentation. CALICO 2009 Conference, Language Learning in the Era of Ubiquitous Computing, 10-MAR-09 - 14-MAR-09, Arizona State University, Tempe, AZ, US.

Thouësny, S. & Blin, F. (2008). "*Modeling language learners' knowledge state: what are language students' free written productions telling us?*". Paper presentation. WorldCALL 2008 Conference, CALL BRIDGES THE WORLD, 05-AUG-08 - 08-AUG-08, Fukuoka, Japan.

Blin, F. & Thouësny, S. (2007). "*Project-based Language Learning in Virtual Learning Environments: What Can Learners' Errors Tell Us about Their Language Development?*". Paper presentation. CALICO 2007 Conference, The Many (Inter)Faces of CALL, 22-MAY-07 - 26-MAY-07, Texas State University, San Marcos, Texas, US.

Research seminars, workshops, posters

Thouësny, S. (2011). "*Inferring language learners' knowledge from their performance: An extended theoretical model*". School of Applied Language and Intercultural Studies Postgraduate Showcase, 11-FEB-11, Dublin, Ireland.

Thouësny, S. (2010). "*Distinguishing between competence-dependent errors and performance-related mistakes in language learners' written texts*". Poster presentation. Humanities and Social Science Research Day, 26-MAY-10, Dublin City University, Dublin, Ireland.

Thouësny, S. (2009). "*Increasing the Reliability of a Part-of-speech Tagging Tool for Use with Learner language*". Workshop presentation. Automatic Analysis of Learner Language (AALL'09): From a Better Understanding of Annotation Needs to the Development and Standardisation of Annotation Schemes, 10-MAR-09 - 11-MAR-09, Arizona State University, Tempe, AZ, US.

Thouësny, S. (2008). "*Discriminating errors from mistakes in free written productions to access language learners' knowledge state*". Research Seminar. 14-MAR-08, Dublin City University, Dublin, Ireland.

Thouësny, S. (2007). "*Specifications for a language learner model for adaptive feedback in second language learning*". Poster presentation. EUROCALL 2007 Conference, Mastering multimedia: teaching language through technology, 05-SEP-07 - 08-SEP-07, University of Ulster, Coleraine, Northern Ireland, UK.

Chapter 1. Introduction

With a European policy[1] encouraging the learning of at least two foreign languages, intercultural competencies and language skills have become more important than ever. Assessing language learners' language, also called interlanguage (Selinker 1972), since it contains features of both native and target languages, is indispensable in terms of grading and achievement predictability. It is equally important for teachers and learners, as information about learners' linguistic strengths and weaknesses allows for personalised actions tailored to each individual's needs. Such learners' particulars generally constitute the basis of any language learner model. Typically integrated into intelligent computer-assisted language learning applications, learner models enable a representation of the learner knowledge, which is generally used to adapt teaching strategies and support learning activities.

However, most representations of learners' knowledge depend on observations drawn from learners' performance, assuming that an incorrect grammatical or lexical word refers by default to a lack of competence, as opposed to a possible gap in performance, errors and mistakes (Corder 1967), respectively. In effect, grading and predictability with regard to learning accomplishments may be compromised due to non-germane variations in a learner's knowledge estimation. While the relationship between learners' interlanguage variability and factors, such as task type, text form, time pressure, task complexity, or planning, to name but a few, is well documented in second language acquisition research, and more specifically in the psycholinguistic and sociolinguistic domains, the impact of the distinction between errors and mistakes on the variability of a learner's interlanguage has not been significantly addressed.

---

1. Multilingualism: an asset for Europe and a shared commitment. Brussels, 18.9.2008. COM(2008) 566 final. Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions. Online from: <http://ec.europa.eu/education/languages/pdf/com/2008_0566_en.pdf> [Accessed 18 September 2010].

This research is located at the centre of three discrete, yet interconnected domains: learner interlanguage variability, learner knowledge modeling, and learner language assessment. The work presented in this thesis seeks to (a) model the knowledge of intermediate learners of French from their unrestricted written language by differentiating between competence-dependent errors and performance-related mistakes, and (b) investigate the extent to which interlanguage competence varies across time, across text types, and across students.

1.1. Research Background

Sometimes learners write correct grammatical and lexical forms, sometimes they do not. Words or groups of words are commonly judged as ill-formed when deviating from the established and accepted rules of the language one attempts to learn, in other words the target language. As a language learner myself, I have experienced all sorts of teaching and learning methods, such as pen pal, short and long immersion periods in the host country, one to one courses, evening courses, distance learning, adult courses as well as university language courses. One of my early experiences will remain in my memory forever. In my second year of German in France, the lecturer handed back our German to French translation assignment. As only one error was made on the first page, I was rather proud of myself. I soon realised however, that there was little to be proud of. The lecturer had got so irritated after the first incorrect form made that she decided to stop correcting my work forever. Fifteen years later, I have forgotten which incorrect form not only denied me the opportunity to improve my language competence by receiving feedback from my teacher, but also left me completely unmotivated. As Ferris (2004) noted,

> [s]tudents are likely to attend to and appreciate feedback on their errors, and this may motivate them both to make corrections and to work harder on improving their writing. The lack of such feedback may lead to anxiety or

> resentment, which could decrease motivation and lower confidence in their teachers (Ferris 2004 p.56).

Differentiating between incorrect forms in terms of importance has become the focus of my interest, even if the idea of making a distinction between errors and mistakes was not transparent at the beginning. The final year project of my Bachelor of Computational Linguistics gave me the opportunity to develop a web-based computer-aided application for the provision of corrective feedback by language teachers. As my implementation progressed, research questions such as how to conceive a learner model to best predict learners' knowledge started to emerge. Assessing learners' performance to unveil their competence became a priority.

How should I assess students' written language? Assessing a learner's performance may be achieved through the means of techniques such as summative, formative, or dynamic methods. Summative assessment, also referred to as *assessment of learning*, is designed to measure a learner's achievement at a specific point in time (Perie et al. 2009). Test results show what learners are able to perform at a "single occasion" (Mitchell 1992 p.20), which does not necessarily demonstrate what they know. Formative assessment, also characterised as *assessment for learning*, intends to inform an instructional decision and to support learners' attempts to learn (Stiggins 2008). Heritage et al. (2009) point out that assessing learners in such a way implies a constant data elicitation from teaching activities to identify the gap between what learners should achieve, i.e., the desired learning outcome, and what they are actually able to achieve, i.e., the current level of attainment. Dynamic assessment, on the other hand, provides learners with adapted feedback depending on their level of development (Poehner 2008). Although some researchers, such as Leung (2007), find similarities between formative and dynamic assessment, in the sense that the two provide interactive feedback to guide and assist learners in improving their performance, both methods are in fact distinct in essence.

Poehner and Lantolf (2005) maintain that dynamic assessment cannot be simply compared to formative assessment, and state further that while the former has its root in the Vygotsky's theory of mind, the latter is based on teachers' intuition with regard to classroom practices (p.260). Depending on the amount of feedback provided to learners, dynamic assessment provides information on actual and potential development.

What does the literature report about interlanguage competence variability? The literature provides mixed and sometimes confusing results. The main reasons are twofold. Firstly, the term interlanguage conveys different representations depending on the perspective from which it is investigated. It is either a synonym of interlanguage performance or interlanguage competence. For example, while a sociolinguistic approach does not distinguish between competence and performance to investigate the variability of a learner's usage of language (e.g., Tarone 2000), the psycholinguistic perspective generally acknowledges the competence-performance dichotomy (e.g., Rothman 2007). Secondly, although the distinction between competence and performance may be accepted when investigating interlanguage variability, the term *interlanguage* is often associated with the meaning of performance, rather than competence.

Competence, from a generative perspective, is seen as a stable underlying structure, as opposed to performance that reflects variation in the language use (Chomsky 1965). Some researchers, such as Gregg (1989, 1990), disagree about the idea of variable competence (Ellis 1985a). On the one hand, competence is seen as a stable feature but it cannot be directly accessed, and therefore it is difficult not to say impossible to investigate. On the other hand, competence may be variable, but since it cannot be accessed directly, researchers look at interlanguage performance and consequently generally consider incorrect forms as competence-dependent errors by default. As noted by Ellis (1994b), "whereas the linguistic approach ignores variability, the sociolinguistic and the psycholinguistic approaches try to describe and explain it" (p.120). However, neither approaches differentiate between performance-related mistakes and competence-dependent errors to investigate the extent of interlanguage variability. To date, findings on interlanguage variability are left incomplete as interlanguage performance is generally *the* object of investigation.

Although interlanguage variability research, generally undertaken from a sociolinguistic or psycholinguistic approach, is not a common domain of enquiry for sociocultural researchers, I followed Vygotsky's theory, and more precisely his concept of the zone of proximal development to investigate interlanguage competence and its variability. The concept of the zone of proximal development generally allows a focus on what individuals can do with the assistance of other more capable individuals, as opposed to the analysis of learners' independent performance with respect to correct and incorrect forms. The zone of proximal development is pictured by Lantolf and Aljaafreh (1995) as "a window into the person's future mental growth" (p.619). It is precisely from this point of view that this research intends to distinguish between errors and mistakes in learners' interlanguage performance. Observing the next level of maturation in the learners' zone of proximal development should indicate whether the incorrect grammatical feature to be analysed is in fact, a mistake (close to independent performance) or an error (further away from independent performance). As such, a sociocultural perspective provides a new angle of analysis and interpretation of subjects shared by most theories in second language acquisition, such as the "human capacity to learn languages other than the mother tongue", or "individual variability that human beings exhibit in learning [these languages]" (Ortega 2005 p.318).

Variability in interlanguage is also a key concept for sociocultural theory as learner interlanguage may develop through interactions with others. Sociocultural theory does not share the same view in terms of language cognition as other theories. According to Ortega (2007), "Vygotskian sociocultural theory of SLA stands apart from [...] other theories [...] in that language cognition is viewed as neither a linguistic nor a psychological faculty of mind. Instead, SLA sociocultural theorists posit that cognition can be best understood as a social faculty" (p.229). Lantolf (2009) further stresses that the social "environment is not a factor in development – it is the very source of development" (p.365). The differences between sociocultural theory and other mainstream theories in second language acquisition are sometimes referred to as "epistemological tensions" or "paradigm wars" (Ortega 2005 p.322). Despite these strong descriptors, Ortega (2005) stresses the importance of "exploring diverse frameworks for ways of knowing and of generating knowledge about L2 learning" (p.323). For example, some researchers such as Lapkin, Swain and Smith (2002) have successfully combined together cognitive-interactionist and sociocultural theories. This thesis does not aim at resolving the epistemological tensions between mainstream and sociocultural theories, but rather at benefiting from sociocultural research methods to broaden our knowledge of learner interlanguage variability.

The learners' knowledge may have varying designations depending on one's approach to second language acquisition. Sociolinguistic and psycholinguistic perspectives acknowledge the terms interlanguage competence and performance, whereas sociocultural theory refers to actual and potential development as well as interlanguage development. In order to avoid conflicting meaning between sociolinguistic, psycholinguistic, and sociocultural theories, it may be relevant to put forward the definition of the key words in use within this thesis. The term *interlanguage competence* represents the knowledge second language learners have about the language they are learning. *Interlanguage performance* refers to the learners' use of this knowledge when attempting to express meaning in the target language. *Actual development* designates the current knowledge of a learner, what he or she is able to perform today without assistance, and *potential development* represents what a learner will know in a more or less distant future, i.e., what he or she is able to perform today with assistance.

Interlanguage development, from a sociocultural perspective, is the outcome of "how mediational means are appropriated by the individual as a result of dialogical interaction with other individuals" (Aljaafreh and Lantolf 1994 p.476). Accordingly, it includes the observation of not only actual but also potential development. In this dissertation, *interlanguage development* relates to the learner's progress with respect to knowledge (interlanguage competence as opposed to interlanguage performance) as observed at different points in time.

While assistance in sociocultural theory is generally associated with human mediation, the role of the more knowledgeable source may as well be undertaken by other means such as computers (Grigorenko 2009). Consequently, the context for a computer-based environment was created to help collect data on learners' actual and potential development. Learners of French were encouraged to participate in a pilot study to determine the extent of their interlanguage competence, as opposed to interlanguage performance. The analysis of the preliminary findings gave an insight into interlanguage competence variability. These preliminary findings needed to be further investigated, synchronically and diachronically. Learners of French were asked to participate anew in a dynamic-based assessment study so as to determine whether and how interlanguage competence varied at a single point in time between students (synchronic inter-learner analysis), longitudinally between students (diachronic inter-learner analysis), and longitudinally in-between students (diachronic intra-learner analysis).

1.2. Scope of the thesis

Through the lens of a sociocultural perspective, this thesis investigates the extent to which interlanguage competence may be variable. It shows how dynamic assessment, whose core principles and procedures derive from Vygotsky's zone of proximal development method, can help distinguish between both competence-dependent errors and performance-related mistakes. Strongly influenced by researchers, such as Amaral and Meurers (2007, 2008), Heift (2004, 2008), Poehner (2005, 2008), Selinker (1972, 1992), and Tarone (1988, 2007), a learner interlanguage corpus has been created to address the following research questions:

> **Question 1.** Can the learners' zone of proximal development, i.e., their actual and potential development, and the distance in-between, be represented and observed so that errors and mistakes can be distinguished?
>
> **Question 2.** Are interlanguage competence and performance variable across students, time, and text types?
>
> **Question 3.** Does the modeling of the learners' zone of proximal development provide further insight into their interlanguage development?

This thesis models the interlanguage performance of fourteen learners of French as a foreign language in order to reveal their current underlying linguistic knowledge and ability to use this knowledge in a written context. In compliance with Dublin City University Research Ethics regulations, a low-risk project application was submitted and accepted. Students who participated in this study were all volunteers and signed a consent form. Participants were studying French at an intermediate level. Texts written by the participants were collected throughout two academic semesters. The focus on written, rather than spoken language is motivated by two circumstances. Firstly, a transcription of a learner's speech would not indicate whether learners experienced difficulties in certain areas such as orthography or number agreement given the fact that the plural mark in French is generally a silent feature. Furthermore, learners' spoken productions may contain certain characteristics such as filled pauses to express hesitations, which could interfere with the error-mistake distinction due to uncertainties in spoken language. Beattie (1983) noted that verbal expression in speech, and more precisely in spontaneous speech, "is riddled with hesitations and silences" (p.33).

Secondly, the aim is to demonstrate that assessing language learners' linguistic performance at the end of a course does not identify the level of their knowledge. The learners' knowledge is determined by exploring correct as well as incorrect instances of a specific grammatical or lexical feature after ruling out mistakes learners could have written by inadvertence. Learners' incorrect forms are categorised according to their level of development, which is observable "at the level of overt independent performance and at the level where performance is mediated by someone else" (Lantolf and Thorne 2007 p.212).

1.3. Outline of the thesis

After defining the concept of interlanguage as an umbrella term for both learners' competence and performance, chapter two highlights the theoretical perspectives concerned with the investigation of learner interlanguage variability and interlanguage modeling. In particular, it outlines divergent views on whether a difference should be made between competence and performance, and focuses on difficulties in overcoming the error-mistake distinction to represent what learners know.

The next chapter addresses the challenges in distinguishing between competence-dependent errors and performance-related mistakes and proposes dynamic assessment as an appropriate means to distinguish between both types of incorrect forms. After pointing out that dynamic assessment is firmly grounded in sociocultural theory, and providing an overview of Vygotsky's related concepts, such as the zone of proximal development, the double stimulation method and the concept of mediation, limitations as well as potentialities to apply such a model are discussed.

The complications that may arise in operationalising dynamic assessment as a means to distinguishing between errors and mistakes are then explored in the fourth chapter. Four main challenges to conduct an interlanguage variability study are identified: the realisation of the dynamic assessment error-mistake distinction, the application of dynamic assessment principles in asynchronous context and free written documents, the identification of reliable psychometric measures to evaluate learners' interlanguage knowledge, and the creation of an annotated corpus. The chapter then considers the implications as well as the methodological requirements to address the challenges mentioned above. In particular, it lists the following two requirements: the creation of the learners' zone of proximal development, and the representation of this zone of proximal development. The chapter concludes by the presentation of the research design.

The fifth chapter presents the various tools implemented for the collection and annotation of the data. More specifically, it proposes and discusses a method to minimise incorrect part-of-speech tags automatically applied to the corpus, and describes the tools that were used by either the participants or this researcher. After presenting the data organisation, the chapter explains the data analysis. In particular, it demonstrates how the learners' zone of proximal development can be represented so that errors and mistakes can be distinguished. Additionally, it highlights the various metrics used to calculate the lexical and syntactical complexity of the learners' interlanguage and their language accuracy so that learner knowledge can be represented and analysed.

The results of these analyses are reported in the following two chapters. Chapter six details the results obtained after running the psychometric measures on the data of the fourteen participants at a single point in time, i.e., a synchronic inter-learner analysis. Chapter seven details the results obtained when exploring learner interlanguage longitudinally. Principally, it analyses the data from two participants over a short period of time, i.e., a diachronic inter-learner analysis, and explores the data from one participant across different text types over a period of two academic semesters, i.e., a diachronic intra-learner analysis.

Chapter eight summarises the findings of this thesis by revisiting each of the research questions. It establishes that interlanguage competence is subject to free variability, thus invalidating the exclusive approach of stability advanced by linguistic theory. Before discussing the implications of the results in the domain of second language teaching and learning, the concluding chapter examines the limitations of the study with a view towards future research.

Chapter 2. Interlanguage variability and knowledge modeling: a literature review

Learner written interlanguage is one source of data to analyse and model learner knowledge. This, however, is not without complications as one cannot be sure whether learners' incorrect forms are due to competence-dependent errors or performance-related mistakes. The aim of this chapter is to explore past and recent research in interlanguage variability so that issues and challenges are identified, and to determine whether a distinction between both competence and performance can be made when investigating the variability of a learner's interlanguage.

The chapter commences with a brief overview of the interlanguage hypothesis and defines both interlanguage competence and interlanguage performance. After providing a typology of the terms associated with variability, the second section explores the relationship between interlanguage and variability, as found in the literature. In addition, it considers various methods to distinguish between competence-dependent errors and performance-related mistakes. The third section presents and discusses modeling techniques to infer learners' knowledge, as well as representative models in the domain of second language learning. Finally, the chapter concludes with a discussion on the shortcomings of the current approaches to the investigation of learners' interlanguage.

2.1. Interlanguage knowledge: towards a representation

Stetsenko (2010) claims that defining knowledge "the ways through which we know the world, including concepts, conceptual understanding and thinking" (p.69) is a rather challenging task. Learner knowledge has been characterised by a profusion of different qualifiers throughout the years. For example, Reif and Allen (1992) use the following adjectives when referring to knowledge: general, definitional, entailed, supplementary, ancillary, formal, case-specific, declarative, procedural, and conceptual. The type of knowledge that underlies a learner's ability to understand the rules and structures of a foreign language, as well as to judge the grammaticality of any sentences is known as linguistic knowledge. Learners' second language knowledge was first said to be solely influenced by their native language. For example, Lado (1957) claimed that an attentive analysis of learners' errors shows patterns of similarities between both learners' target language and mother tongue. Later, Richards (1974) challenged Lado's (1957) contrastive analysis by listing a variety of errors observed in learners' language, which could not have resulted from a native language transfer, such as a "faulty generalisation" or an "incomplete application of rules" (p.174). The learner's language produced somewhere in-between the native and target languages was termed "interlanguage" by Selinker in 1969.

2.1.1. The interlanguage hypothesis

Selinker (1969) initially used the concept of interlanguage to refer to Israeli students' attempts of producing English:

> An 'interlanguage' may be linguistically described using as data the observable output resulting from a speaker's attempt to produce a foreign norm, i.e., both his 'errors' and 'nonerrors'. It is assumed that such behaviour is highly structured. In comprehensive language transfer work, it seems to me that recognition of the existence of an interlanguage cannot be avoided and that it must be dealt with as a system, not as an isolated collection of errors (Selinker 1969 reprinted in Selinker 1988 p.117).

In Selinker's (1974) paper, interlanguage is defined as a "separate linguistic system" (p.35). A linguistic system is "a grammar, which underlies the use of language, including comprehension and production. Native-speaker grammars are constrained by built-in universal linguistic principles, known as [Chomsky's (1965)] Universal Grammar", and "non-native grammars [are] referred to as *interlanguage grammars*" (White 2003 p.1, emphasis in original).

Selinker's (1972) interlanguage hypothesis suggests that the language produced by learners of foreign or second languages is an independent interim linguistic system with its own set of rules. This underlying grammar not only includes features of the learner's native language and of the language to be learned, but also contains elements which tend to occur only in interlanguages, whose origin is neither the native nor the target language (Gass and Selinker 2008 p.14). Considering that learners' attempts to produce utterances of the language they are learning do not always conform to the native speaker norms of the target language, Selinker (1974) proposed that "one would be completely justified in hypothesising, perhaps even *compelled* to hypothesise, the existence of a separate linguistic system" (p.35, emphasis in original).

In addition to the concept of interlanguage, which is widely recognised among researchers, learner language is also called "approximative system" (Nemser 1971) or "transitional competence" (Corder 1967). Perceived sometimes as synonyms in the literature, Corder (1981) argued that each of these alternatives "draws attention to different aspects of the phenomenon" (pp.66-67). For example, Nemser's (1974) approximative system is directional in the sense that the learner approximates the target language in stages that are separate and distinct. Corder (1981), on the other hand, borrows Chomsky's (1965) notion of competence and puts the stress on the idea that language knowledge is "constantly developing" (Corder 1981 p.67). His notion of transitional competence represents what is being acquired by the learners, that is, "the underlying knowledge of the language to date" (Corder 1981 p.10). While Nemser, Corder and Selinker's models conceptualise the learner's language as a mental grammar in its own right, neither Corder's (1981) transitional competence hypothesis nor

Nemser's (1974 p.55) approximate system hypothesis clearly consider the existence of fossilised linguistic features (Selinker 1992 p.225). Although the term fossilisation is often acknowledged as lacking a unified definition (Han 2004), it is generally understood as an incomplete or faulty language learning process; "a permanent cessation of [...] learning before the learner attains [target language] norms at all levels of linguistic structure" (Selinker and Lamendella 1981 p.217). Rothman et al. (2010) further state that, as opposed to first languages, second languages are characterised by different degrees of fossilisation, which may occur at all proficiency levels (p.45).

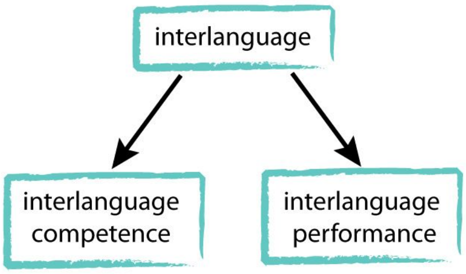
Since Selinker's (1969) publication, the meaning of interlanguage in the literature has generally oscillated between the learner's production of the target language and the learner's interim grammar system. For example, Gass and Selinker (2001) identify interlanguage as the learner's performance: "[t]he language produced by a nonnative speaker of a language (i.e., a learner's output)" (p.455), whereas Ellis (2008a) argues that linguistic competence is "what researchers are talking about when [referring] to 'interlanguage'" (p.17). Researchers have long debated whether investigations on interlanguage should support the generative distinction between competence and performance. The disagreement is mainly driven by the fact that the term competence is often used in the literature in a broader sense than Chomsky's (1965) restrictive definition (DeKeyser 2009 p.119).

2.1.2. Competence and performance

The competence-performance distinction is generally associated with Chomsky's (1965) view of language, where he hypothesises the existence of a universal grammar underlying all languages. This universal grammar corresponds to an in-born capacity of the human brain for learning a first language. Chomsky (1965) concurrently made a distinction between language competence and language performance. Performance does not reflect competence as the latter will be affected "by such grammatically irrelevant conditions as memory limitations, distractions, shifts of attention and interest, and errors (random or characteristic)" (Chomsky 1965 p.3). In Chomsky's (1965) view, competence is characterised by what an individual knows about his or her first language, that is, an intuitive mental representation of linguistic rules, where a native speaker would be able to pinpoint whether a construction is ungrammatical, but not necessarily be able to explain the reason of its non-grammaticality. In other words, linguistic competence refers to the learners' tacit grammatical knowledge of their mother tongue. Performance, on the other hand, is described as what an individual does with this grammar knowledge in concrete situations.

From the second language acquisition perspective, Gregg (1989) strongly supported the application of the Chomskyan theory to L2 research and suggested linguistic competence rather than linguistic performance as a specific domain of enquiry. Critiquing Tarone's (1979) claim that "the aim of applied linguistic research is to describe the way people talk when they are not being systematically observed" (p.181), Gregg (1989) contended that the main issue "with this concept of language is that it fails to distinguish between competence and performance" (p.18). More recently, publications show that researchers are still debating whether a distinction between competence and performance should be made. For example, while Rothman (2007) maintains that the distinction is necessary to explain second language acquisition process (p.609), Verspoor, Lowie and Van Dijk (2008) emphasise the idea that dynamic system theory does not distinguish between competence and performance to investigate interlanguage development. Whether both competence and performance are merged together or juxtaposed, they are nonetheless closely related, since interlanguage competence cannot be accessed straightaway. As stated by Lakshmanan and Selinker (2001), "information about the nature of interlanguage competence can only be derived indirectly, through an examination of interlanguage performance data" (p.393). The concept of interlanguage may, therefore, be regarded as an umbrella term under which interlanguage competence is juxtaposed with interlanguage performance, as illustrated in Figure 2.1.

Figure 2.1. Interlanguage: an umbrella term



Interlanguage performance, also called “meaningful performance” as opposed to “performance of drills” (Selinker 1974 p.32), refers to the learners’ use of their knowledge when attempting to express meaning in their foreign or second language. Interlanguage competence represents the knowledge second language learners have about the actual linguistic system of the target language they are learning. Derived from Chomsky’s (1965) theory, this knowledge is by definition intuitive and grammatical.

2.1.3. Intuitive knowledge

While intuitive may refer to an innate capacity in the case of first language learners, intuitive may also carry the meaning of implicit knowledge. For example, Ellis (1994a) describes implicit knowledge as intuitive “in the sense that the learner is unlikely to be aware of having ever learnt it and is probably unaware of its existence” (p.85). He further adds that there are two types of implicit knowledge: formulaic, i.e., “ready-made chunks of language”, and rule-based knowledge, i.e., “generalised and abstract structures which have been internalised” (Ellis 1994b pp.355-356). As opposed to implicit knowledge, explicit knowledge is characterised as “available to the learner as an [sic] conscious representation" (Ellis 1994a p.84). Explicit knowledge is verbalised, accessible through controlled exercises, declarative and may become procedural (Ellis 2004 pp.235-238).

Anderson (1996) differentiates between declarative and procedural knowledge in such a way that "declarative knowledge is factual knowledge that people can report or describe, whereas procedural knowledge is knowledge people can only manifest in their performance" (p.18). The *knowing that* and *knowing how* distinction has existed for many years (Ryle 1946). For example, knowing that present tense verbs in English take the inflectional morpheme *-s* in the third person is declarative knowledge, and knowing how to apply such a grammatical convention in spontaneous language use without thinking about this is procedural knowledge (DeKeyser 1998 pp.48-49).

Declarative knowledge becomes procedural knowledge first through conscious effort, and then by means of less and less intentional mental exertion (DeKeyser 1998). Therefore, consistent practice may lead to an increased automaticity in terms of learners' ability to use their factual knowledge. For Anderson (2000), when an activity does not demand any further attention nor disrupts a multi-task situation, the knowledge can be situated in the "autonomous stage", which he illustrates as follows:

> When people learn to drive, initially they need to pay all their attention to driving and they are unable to maintain a conversation. As they become more practiced, they are able to maintain a conversation while driving. The driving skill becomes so automatic that it seems to require no attention at all, at least when driving conditions are not demanding (Anderson 2000 pp.325-326).

DeKeyser (2003) further points out that declarative knowledge which has become fully automatised through practice might be considered as equivalent to implicit knowledge considering that "learners eventually lose their awareness of the rules" (p.329). He outlines that after the complete automatisation of the rules, procedural knowledge functions as “implicitly acquired knowledge”, that is, “knowledge without awareness” (DeKeyser 2003 p.329). Opposed to this idea, Bialystok (1994) declares that what changes through practice is how the knowledge is accessed, not how the knowledge is represented. Whether explicit knowledge is automatically performed and “effortlessly” accessed, it still remains explicit, as opposed to implicit (Bialystok 1994 p.567).

2.1.4. Grammatical knowledge

Douglas’ (2000) framework of “language knowledge” contains four subsets of knowledge. The first subset, textual knowledge, contains the knowledge of cohesion and rhetorical organisation. The second subset, functional knowledge, includes the knowledge of manipulative arguments, heuristic – how to explain advantages and disadvantages – and imaginative functions. The third subset, sociolinguistic knowledge, covers the knowledge of dialects, registers, idiomatic expressions and cultural references. The fourth subset, grammatical knowledge, regroups the knowledge of vocabulary, morphology, syntax, and phonology. Douglas’ (2000) representation of language knowledge is mainly based on Bachman and Palmer’s (1990, 1996) model, which divides language knowledge into two subcategories: organisational and pragmatic knowledge. On the one hand, organisational knowledge comprises (a) the abilities to build and comprehend correct sentences in terms of grammatical aspects, i.e., “vocabulary, syntax, phonology, and graphology”, and (b) textual competence, in other words, “knowledge of cohesion and knowledge of rhetorical or conversational organisation” (Bachman and Palmer 1996 pp.67-68). On the other hand, pragmatic knowledge is concerned with the relationship between the learner and the context in which the learner attempts to communicate (Bachman and Palmer 1996).

Douglas’ (2000) grammatical model of knowledge implies considerations of not only the morphological and syntactical aspects of learners’ interlanguage, but also the lexical as well as the phonological facets of learners' language production. In his study of lexical significance, Wilkins (1972) pointed out that "without grammar very little can be conveyed, without vocabulary nothing can be conveyed" (p.111). Additionally, while phonology is generally understood as the study of the organisation and use of sounds in natural languages, some researchers include this area of linguistics to investigate learners' knowledge in written language. For example, L'Haire (2007) employs the phoneticisation method, which transposes a word into its phonetic transcription, to detect a learner's phonetic approximation in written documents. Despite the appealing side of this method, he declares that the phoneticisation method is "not totally satisfying" (L'Haire 2007 p.155).

In summary, the learners' interlanguage competence relates to the proceduralised grammatical knowledge they have about the structure of their second or foreign language, whether it is implicit or explicit. Grammatical knowledge implies that the learner knows how to apply the grammatical code in language use with regard to morphological, syntactical, and lexical aspects of the language. On the contrary, the learners' interlanguage performance designates the learners' attempts to produce language when communicating through the target language. The learner's interlanguage performance, therefore, includes not only correct, but also incorrect grammatical structures and lexicons. As outlined by Littlewood (1984), "[i]t is well known to teachers that learners are often inconsistent in their performance. In one activity, for example, they may give the impression that they have mastered a particular rule, but a moment later, they may apply it wrongly for no apparent reason" (p.81). Since second or foreign languages are acquired or learned, and not inherited, variability in a learner's performance is not only unavoidable, but also "a normal phenomenon in [learners'] second language" (Littlewood 1984 p.81).

2.2. Interlanguage variability: theoretical positions

Learners' interlanguage variability has been widely researched and has principally focused on the relationship between variations and factors, such as task type (Bygate 1999), text form and function (Tarone and Parrish 1988), time pressure (Hulstijn 1989), task complexity (Robinson 2001), planning (Yuan and Ellis 2003), language style (Bell 1984), or transfer of learning (James 2007), to name but a few. Some researchers disagree about types of variations. For example, while Preston (1996) claims that unsystematic variation is likely to be nonexistent, and even be a "chimera" (p.25), Labov (1971) argues that when looking at language in use, one is expected to find systematic and unsystematic variations (p.451). Additionally, other researchers disagree about whether interlanguage should be concerned with variability or not. Yet, a wealth of research has demonstrated variability in interlanguage (e.g., Heift 2008, 2010a, Herschensohn and Arteaga 2009, Howard 2006, Hulstijn 1989, James 2007, Rothman et al. 2010, Tarone 1985, 1988, Tarone and Parrish 1988).

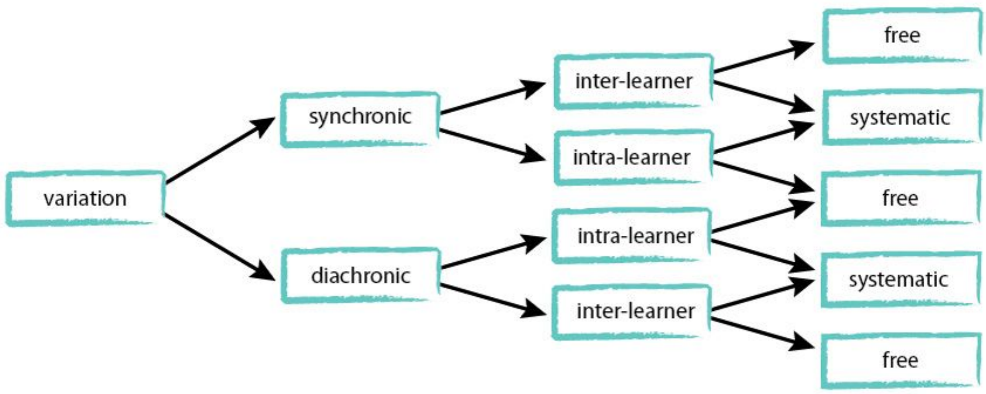
Before discussing the main theoretical positions on interlanguage variability, the subsequent section defines the terminology associated with interlanguage variation.

2.2.1. Typology of variation

According to Ellis (2008b), the concept of variation can be divided into horizontal and vertical variations. However, the usage of these terms may occasionally be ambiguous as one can find at least two different meanings for each of them in the literature. For example, Corder (1981) refers to horizontal when comparing different dialects or registers, whereas the use of vertical refers to learners' shifting from one linguistic form to another (pp.91-92). Adamson (2009) uses this terminology associated with this definition when providing examples of early studies on interlanguage variations. On the other hand, vertical variations may refer to changes over time, i.e., diachronic variations, and horizontal variations may be linked to changes occurring at a single point in time, i.e., synchronous variations (Ellis 1989 p.22). To avoid any confusion between both interpretations, the terms diachronic and synchronic are preferred here when referring to changes over time and at a single point in time, respectively.

The typology presented in this section is based on Ellis' (2008b) typology of variation. Interlanguage variation, as illustrated in Figure 2.2 below, may be investigated from a synchronic and/or diachronic point of view.

Figure 2.2. A typology of variation (adapted from Ellis 2008b p.129)



Diachronic and synchronic variations can be subdivided into inter- and intra-learner variations. While inter-learner variation designates dissimilarities across learners, intra-learner variation refers to language differences found within one individual (James 2007). Inter- and intra-learner variations can in turn be subdivided into systematic and free variations. Systematic refers to fully explainable and predictable variations occurring in the learner's language. For example, VanPatten (1992) found that the English grammatical morpheme *ing* was generally acquired before the inflectional morpheme 3rd-person singular *s*. Sequences in interlanguage development are considered as systematic, given the fact that these interlanguages "follow rules and patterns that change over the course of L2 development, but do so in patterned ways" (Pica 2005 p.265). Free variations, also called unsystematic variations, describe random linguistic alternatives unpredictable by rules (Ellis 2008b).

2.2.2. Interlanguage variability: competence or performance

Early variationist models, such as Labov's (1972) variable rules, were of a great influence on the work of researchers investigating variations in second language acquisition; this is exemplified, for instance, in the work of Dickerson (1975), Ellis (1985b) and Tarone (1985). A variable rule "typically makes the following kind of statement: rule x (e.g., contraction of the copula) is used with probability X in social context Y given that the linguistic context is Z" (Pienemann 2007 p.43). This approach intends to demonstrate that internal and external constraints have an effect on variation. While internal constraints refer to linguistic factors, external constraints designate factors that impact the performance, such as social class or age (Romaine 2003 p.411).

Approaches to investigating variability in learners' interlanguage are commonly split into either a linguistic, psycholinguistic, or sociolinguistic perspective (Ellis 1994b). Firstly, researchers who attempt to apply the Chomskyan model to L2 tend to see learners' competence as a fixed system, and variation as an exclusive feature of performance (Ellis 1994b). For example, Gregg (1990) denies any perspectives which would acknowledge the fact that the learners' competence itself might be variable, and maintains that theories of second language acquisition should not investigate variability. According to Romaine (2003), there is evidence that competence might be variable (p.430). For example, Ellis' (1984) variable competence model considers the variability that is obvious in a learner's interlanguage performance and hypothesises that this variability reveals a competence variable in itself (p.159). Opposed to this belief, Gregg (1990) maintains that he is not against "the idea of an homogeneous competence that changes over time" (p.368). However, he argues that the variable competence model not only does not provide any concrete explanation of what the learner's knowledge might be, but also "explicitly erases the line between linguistic knowledge and linguistic output" (Gregg 1990 pp.377-378). Therefore, interlanguage competence is qualified as a stable underlying structure, in other words invariable. Variability, on the other hand, is ascribed to interlanguage performance. To this researcher's knowledge, there are no studies – from a generative viewpoint – that intend to demonstrate the stability of learners' interlanguage competence. There are, however, attempts to distinguish between competence and performance. For example, Lyons (1972) suggested that all "vocal reflexes" (e.g., sneezing or coughing), "non-linguistic voice-quality" (e.g., dialectical variation), and "verbal, prosodic and paralinguistic features" (e.g., variations occurring under stress) were removed from the learner's language use (pp.51-52). As a result, researchers ought to have access to invariable data, which should reflect the learner's competence (Ellis 1994b p.119). Ellis (2008b) further highlights that the data considered within this trend of research mainly consists of judgement data, i.e., "speakers' intuitions regarding what they think is correct in the L2 rather than actual instances of language use" (p.118).

Psycholinguistic researchers are, according to Ellis (1994b), concerned "with identifying the internal mechanisms responsible for variable performance" (p.120). From Tarone's (2000) point of view, the aim is to explain "how an interlanguage grammar gets acquired over time" (p.186). Depending on these researchers' interpretation, the focus is either on interlanguage performance or on interlanguage competence. For example, Ellis (2005) explores interlanguage performance to investigate language accuracy on planned and unplanned activities. He finds that an unpressured online planned task gives language learners the opportunity to generate more accurate language output, as opposed to a pressured unplanned activity (Ellis 2005). Other studies investigate the structure of learners' second language, in other words, the interlanguage competence. For example, Maritxalar et al.'s (1997) model of second language learners' interlanguage aims to compare an ideal language grammatical structure with real learners' interlanguage performance in order to extract learners' knowledge from work undertaken. The model is based on the analysis of a corpus written by learners of the Basque language and represents ideal levels of knowledge learners should have about the structure of their second language. The levels considered are high and upper intermediate. One limitation of such a method is that the real learner's knowledge is estimated by juxtaposing his or her interlanguage performance with an interlanguage model, which is, in fact, an average of several learners' interlanguage performance. It may be the case that the real learner's interlanguage performance may match the model. However, comparing the interlanguage performance of the real learner to the ideal learner's interlanguage performance will not give information about the real learner's knowledge, as the performance may not reflect a true representation of the learner's knowledge. Furthermore, describing all levels of language structure during the learners' learning process may represent a complicated endeavour.

Sociolinguistic approaches are concerned with the study of language use influenced by societal and cultural factors. Wardhaugh (2006) distinguishes between *micro-sociolinguistics* and *macro-sociolinguistics*, depending on the emphasis put on the research, that is, either linguistic or social, respectively. Tarone (2000) points out that sociolinguistic studies "demonstrate the impact of social factors on interlanguage use at a single point in time and do not show that those social factors affect the acquisition of specific linguistic features of [interlanguage]" (p.185, emphasis removed). For example, Van Compernolle and Williams (2009) seek to analyse how features such as stylistic variations are realised in discourses of learners of French and non-learners (native speakers), both engaged in synchronous computer-mediated communication. To do so, they identify variable contexts found in the learner's interlanguage performance and explore the factors that could be responsible for these variations. In general terms, sociolinguistic approaches are more concerned with language use than interlanguage competence.

Other perspectives of research, such as dynamic system theory or sociocultural theory, take both the cognitive and social angles of the learner's development into consideration. One difference that could be noted between both approaches is that dynamic system theory sees the social "as a resource that is available to the learner", whereas sociocultural theory views it "as the source of development" (Lantolf 2007 p.33). For example, De Bot, Lowie and Verspoor (2007) state that in contrast to most approaches to SLA, variation is not considered as "noise", but rather "as an inherent property of changing system" (p.14, emphasis removed). Additionally, Yan and Fischer (2002) highlight the importance of focusing on variability in cognitive development in order to illustrate the dynamic aspect of the learning and development processes. From this viewpoint, variation is considered as "a direct manifestation of dynamic processes" (Yan and Fischer 2002 p.143). However, variations are observed in the interlanguage performance, not the interlanguage competence. Similar to dynamic system theory, sociocultural researchers do not make any differences between competence and performance. When the term competence is used in the literature related to sociocultural theory, the idea behind the concept refers more to the dictionary's definition, that is, having the necessary ability or knowledge to perform something successfully. Sociocultural researchers do not even talk about learners' interlanguage competence and interlanguage performance. However, they do investigate learners' interlanguage development. According to Lantolf and Aljaafreh (1995), investigating second language development from a sociocultural perspective means that the "learner's interlanguage is not simply determined by the relative accuracy of linguistic performance, but, crucially, is a function of the frequency and quality of regulation (i.e., help) negotiated between novice and expert" (p.620).

The major trends of interlanguage variability research show that the concept of interlanguage, even if acknowledged as being an interim system of second language learners, is often investigated by considering only the interlanguage performance viewpoint. One reason for this relates to the difficulty of extracting what learners know from what they perform. Therefore, researchers and teachers alike frequently assume that what is correctly performed is what learners know, and conversely, what is incorrectly performed is what learners do not know. It is, as noted by Amaral and Meurers (2008), "usually taken for granted that linguistic errors are caused solely by a lack of linguistic competence" (p.328), although not all incorrect forms are competence-dependent errors (Corder 1967).

2.2.3. Competence-dependent errors and performance-related mistakes

According to James (1998), the learner's comprehension of a language corresponds to the domain theory of knowledge, whereas error analysts explore the learner's "linguistic ignorance" (p.62). The analysis of error is defined as a "type of linguistic analysis that focuses on the errors learners made" (Gass and Selinker 2008 p.102). Gass and Selinker (2008) further point out that errors, considered as "red flags", reflect a learner's knowledge state (p.102). Corder's (1967) Error Analysis framework was an alternative to Lado's (1957) Contrastive Analysis Hypothesis, which was based on the assumption that incorrect patterns in second or foreign language productions were mostly due to a transfer from learners' native language. While Contrastive Analysis considers errors as a failure in terms of language acquisition, Error Analysis views errors as evidence of the learners' developing interlanguage (Ellis and Barkhuizen 2005).

In his seminal article entitled "The significance of learners' errors", Corder (1967) argues that the analysis of errors is central to investigating the learners' acquisition process. He makes key points, such as (a) errors are evidence of a built-in syllabus and they are systematic, (b) errors, which reveal the learners' interlanguage, develop in the learner's transitional competence, (c) errors are significant in the sense that they provide evidence to teachers of what still needs to be taught, evidence to researchers of how language is learnt, and evidence to learners when testing their hypothesis about the language they are learning, and (d) errors should be differentiated from mistakes (Corder 1981 pp.8-11).

In the process of distinguishing between both errors and mistakes, the dictionary (Concise Oxford English Dictionary) is of little use in this matter since the first definition of the term error is mistake. Making a distinction between both terms does not come as a natural process to everybody. For example, Vann, Meyer and Lorenz (1984) report one of their participant-teacher's comments as “an error is an error” (p.433). There are plenty of examples in the literature where researchers or teachers maintain the distinction, and just as many examples where they use both error and mistake interchangeably (e.g., Darus and Subramaniam 2009, Gao 2009). Corder (1981) claims that it is essential to draw “a distinction between those errors which are the product of [...] chance circumstances and those which reveal [the learner's] underlying knowledge”, performance mistakes and competence errors, respectively (p.10).

In 1985, Ellis claimed that the distinction between errors and mistakes was “probably unworkable in practice” (1985c p.68). More recently, he points out that, although there exist various proposals to distinguish between errors and mistakes,

> none of [them] are easy to implement in practice. The distinction between an ‘error’ and a ‘mistake’ is nothing like as clear-cut as Corder made out. The gravity of an error is to a very considerable extent a matter of personal opinion (Ellis 2009 p.6).

Errors and mistakes can be distinguished by differentiating between intentionality and non-intentionality (Taylor 1986), or between systematicity and non-systematicity (Corder 1981). Another approach consists of determining the amount of assistance learners required to correct themselves (Ellis 1997, James 1998).

For example, Taylor (1986) suggests to explore “the writer’s semantic and structural intentions” to determine whether an incorrect form[2] is an error or a mistake (p.154). According to his proposal, if learners intend to express themselves rather vaguely in terms of semantic intentions, the incorrect forms occurring in the text are likely to be insignificant. From such a perspective, to distinguish between errors and mistakes would imply not only a high degree of subjectivity from the corrector, but also a high fluctuating threshold between intentionality and non-intentionality when classifying learners’ intentions as vague or not vague.

In Corder’s (1967) view, competence-dependent errors, or simply put, errors are systematic incorrect forms, in the sense that they are evidence of the learner’s use of an underlying system during the learning process. Errors reflect the learner’s interlanguage competence. By contrast, mistakes are incorrect forms, whose systematicity cannot be explicitly described. Also referred to as performance-related errors, mistakes are mainly the result of accidental slips of the tongue, physical or psychological conditions, such as tiredness or specific emotional states.

Keeping open the option that an error might not be on all occasions systematic, and that a mistake might not be constantly unsystematic, the error-mistake distinction can also be approached from the learners’ degree of ability to correct themselves (Ellis 1997, James 1998). Ellis (1997) points out that the differentiation between both concepts could be made not only by checking the “consistency of learners’ performance”, but also by asking “learners to try to correct their own deviant utterances” (Ellis 1997 p.17). If learners are capable of correcting themselves, their incorrect forms are regarded as mistakes. Conversely, if they are not able to self-edit their own performance, their incorrect forms are considered as errors. James (1998) offers a practical classification to

2. It is important to note here that Taylor refers to mistakes as the generic term, and distinguishes between errors and slips instead of errors and mistakes as the terminology used within this thesis.

distinguish between errors and mistakes: if an incorrect form, merely pointed out to the learner, can be corrected by the learner himself or herself without other sources of assistance, it is assumed to be a mistake. In contrast, if the learner is unable to correct himself or herself when the incorrect form is signalled, it means that "the [incorrect] form the learner used was the one intended, and that it is an error" (James 1998 p.78). In James' (1998) own words, a mistake "can only[3] be corrected by their agent if their deviance is pointed out to him or her", whereas an error "cannot be self-corrected until further relevant (to that error) input (implicit or explicit) has been provided and converted into intake by the learner" (p.83). By contrast to the intentionality versus non-intentionality, and systematicity versus non-systematicity methods, James' (1998) proposition seems to be more realistic in practice, and therefore is the one used in this thesis. Figure 2.3 illustrates an incorrect form being either a mistake or an error depending on the implicitness or explicitness of the assistance required in order for learners to propose a correct alternative[4].

Figure 2.3. Distinguishing between mistakes and errors: from implicit to explicit assistance

The more implicit the assistance, the greater the chance an incorrect form may be a mistake. Conversely, there is a stronger possibility of an incorrect form being an error

3. It should be added that if learners can correct incorrect forms without pointing them out, then these incorrect forms are also considered as mistakes.

4. The term *alternative* in this context refers to a learner's correction when self-editing his or her text, i.e., when proposing a replacement to his or her initial incorrect form, which can be successful or unsuccessful. The term is widely used throughout this thesis and, when in the same context, always refers to the same meaning. The aim is to differentiate between corrections made by students and this researcher. A learner's correction is mainly referred to as a learner's alternative.

with more explicit assistance. The distance between implicit and explicit assistance determines the transition zone in which an incorrect form goes from a mistake to an error. The distance between what a learner can do with and without assistance from someone else is very much in line with Vygotsky's (1978) concept of the zone of proximal development. It is widely used amongst sociocultural researchers to determine the learner's level of cognitive development. Yet, distinguishing between an error and a mistake by exploring the level of assistance learners require has to this researcher's knowledge, not been exploited to differentiate what learners know from what they perform.

Concretely representing the learner's interlanguage competence, in other words, modeling the learner's grammatical knowledge without their mistakes, so that one can explore the variability of both types of interlanguage, is still tentative. Liu (2006) points out that "the uncertainty relationship between students' responses [...] and students' competence in learning targets make student modeling a challenging task" (p.287). Heift and Schulze (2007) further point out that the quality of a learner model is determined by considering how "performance-related errors, that is, mistakes or lapses, as opposed to genuine competence-dependent errors" are handled (p.176).

2.3. Knowledge modeling: techniques to handle errors and mistakes

A learner model or student model[5] may be defined as "a representation of a learner's current understanding constructed by a computer-based educational environment according to the learner's actions in that environment" (Bull et al. 2006 p.8). In the case of language learning, learner models usually enable intelligent computer-assisted language learning applications not only to assess students' performance, record their progress and provide them with personalised feedback, but also to adapt tutoring

5. The synonyms (or near synonyms) learner model and student model are used here interchangeably.

strategies depending on learners' own progress (Heift and Schulze 2007 p.172). In addition, learner models can enable the observation, recording, analysis, and even inference of reasons for an ill-formed word, thus potentially providing some insight into the students' language knowledge (Heift and Schulze 2007 p.172).

2.3.1. Modeling techniques to infer learners' knowledge

A learner model is generally described as one core component of intelligent tutoring systems (ITS), adaptive educational systems (AES), adaptive hypermedia systems (AH), adaptive Web systems (AWS), intelligent interactive learning systems (IILS), intelligent computer-assisted language learning (ICALL), to name but a few. The number of components of a tutoring system, as found in the literature, is to some degree variable. For example, Beck, Stern and Haugsjaa (1996) list five major components: the student model, the pedagogical module, the domain knowledge, the communication module, and the expert model. The main constituents, however, are the student model, the teacher model, the expert model, and the user interface (Bernsen and Dybkjær 2008 p.367, Freedman et al. 2000 p.15, Heift and Schulze 2007 p.175, Samuelis 2007 p.81).

Because learners are not physically and mentally the same, tutoring applications in most cases attempt to meet the learning needs of each individual. Elsom-Cook (1993) points out that "[o]ne of the central assumptions underlying research on Intelligent Tutoring Systems is that a teaching interaction should adapt to individual differences between learners" (p.227). Results obtained from the analysis of the learners' input enable the system to not only retrieve information about the learners' idiosyncrasies but also make assumptions about their current knowledge, which is, according to Baker, Corbett and Aleven (2008), "a fundamental part of intelligent tutoring systems" (p.406). Learner modeling often refers to "the process of gathering relevant information in order to identify and represent the knowledge state of the student" (Mitrovic 1997 p.7). Elsom-Cook (1993) identifies alternative purposes to learner modeling. In addition to determining the learners' current knowledge (diagnostic model), learner models may also provide adaptive corrective feedback (corrective model), extend the learners' knowledge (elaborative model), change the teaching strategies (strategic model), anticipate an action effect on learners (predictive model), and assess the learners' level of achievement (evaluative model) (Elsom-Cook 1993 p.228, Heift and Schulze 2007 p.172).

Adaptive student modeling systems may draw inferences about what a learner knows or does not know through the means of different methods. Amongst the modeling techniques found in the literature, the most common ones seem to point in the direction of the overlay technique, the bug library, the probabilistic approach using for instance Bayesian networks, the knowledge tracing method, and the constraint-based modeling technique. This list is certainly not exhaustive, but it provides nevertheless an overview of common methods employed in the modeling of the learner's knowledge. For a more comprehensive view of modeling techniques, the reader is invited to consult other publications on this subject, such as Heift and Schulze (2007) or Conati, McCoy and Paliouras (2007).

According to Brusilovsky and Millán (2007), the *overlay model* is the "most popular form of a structural knowledge model" (p.7). It is assumed that a learner's knowledge is a subset of the expert's knowledge defined within the model, in other words, an overlay on the domain knowledge the learner is supposed to know. The domain knowledge encloses a set of knowledge elements, which represent the expertise knowledge. Knowledge elements, or "fragments of the domain knowledge", have been given various labels, such as "concepts, knowledge items, topics, knowledge elements, learning objectives, learning outcomes" (Brusilovsky and Millán 2007 p.18), "concepts, principles, subskills" (Ohlsson 1994 p.168), or "Knowledge Units, or KUs" (Michaud and McCoy 2003 p.97). Brusilovsky and Millán (2007) recommend the term "concept" as it best represents fragments of not only conceptual, but also procedural knowledge (p.18). The main aspect of the overlay model, is that it characterises the learners' knowledge state in terms of correct knowledge exclusively. In the most basic functioning of the overlay model, the knowledge values are restricted to "known" or "not known" (Brusilovsky and Millán 2007 p.20). An improvement has been made with the introduction of the weighted overlay model. For example, a weight of *one* would indicate that the concept has been mastered, a weight of *half of one* would denote that the concept is partially mastered, whereas a weight of *minus one* would reflect the learner's ignorance (VanLehn 1988 p.62). Within this approach, only the learner's missing conceptions are identified and stored into the overlay model, as opposed to the learner's misunderstandings or misconceptions which are ignored by the system.

In order to capture misconceptions, *the bug library* is a technique commonly used with the overlay model. To infer the learners' knowledge, the learners' performance is compared with a library of expected errors, i.e., "a list of their (recurring) errors" (Ohlsson 1994 p.169). Other denominations, such as "fault model", are sometimes encountered in the literature (e.g., Nguyen and Do 2008). According to Holt et al. (1994), a bug library or catalogue is characterised by a "fixed collection of bugs and misconceptions" (p.9). The aim of this library is to draw inferences about the learners' errors and provide explanations for their faulty performance. A buggy rule, also called mal-rule (Payne and Squibb 1990, Sleeman 1984), describes an error produced by a learner. A library of buggy rules generally contains a considerable amount of errors made by a group of learners. However, the main drawbacks of such a method are that bug libraries do not transfer between populations of students, and that the "error descriptions are very expensive to create because they are built on empirical analyses of errors previously encountered" (Heift and Schulze 2007 p.180). Furthermore, Payne and Squibb

(1990) demonstrate that the instability of mal-rules in terms of frequency of use makes the process of distinguishing between an error and a mistake[6] unfeasible (p.445).

Probabilistic learner modeling seeks to utilise prior information to infer learners' competencies. For example, Conati et al. (2002) use Bayesian networks to infer the students' knowledge (of Newtonian physics), and their most probable following sequences of actions on the basis of prior learners' assessments. Wei et al. (2005) use Bayesian network to assume what the learners know about the different concepts to be learned, and how well they know them. CIMEL, the intelligent tutoring system designed to support the instruction of object-oriented design and programming, is a three-layered model. It intends to (a) identify learners' problem solving strategy, (b) document the learners' past knowledge and (c) deduce the learners' reasons for making an error. These authors affirm that the learner model "performs a diagnosis based on the history of the student's performance to determine the reasons for the student's errors and where there are gaps in his or her knowledge" (Wei et al. 2005, not paginated, section "CIMEL ITS ARCHITECTURE"). The probable reasons for a learner to produce an incorrect form are identified as resulting from either a typo, a lapse in memory, a misunderstanding, or a misconception. For example, after committing a similar error type three times, the learner model categorises the incorrect form as a potential misunderstanding. Whether more occurrences of this type of error should arise, the learner model would label them as misconceptions. This distinction is questionable in the sense that the boundaries between the different categories are not stated explicitly. It is not known when an incorrect form finishes being a typo and starts becoming a lapse in memory, and when this lapse in

6. Payne and Squibb's (1990) use the terms "slip" and "mistake" to designate what has been termed in this thesis a mistake and an error, respectively (see discussion on error and mistake in section 2.2.3. on page 27). To these authors, a slip denotes "a consequence of the performance system", whereas a mistake indicates "short-comings in competence" (Payne and Squibb 1990 p.465).

memory finishes and starts becoming a probable misunderstanding to finally be considered as a misconception.

Knowledge tracing is a process-based approach which was introduced by Anderson in 1983 (republished in 1996). It is a technique used to estimate what learners currently know or have learned, given observations of what they attempted to perform (Beck and Sison 2006). For example, Corbett and Anderson's (1995) cognitive tutoring system, the ACT Programming Tutor (APT), is used for the instruction of the Lisp, Pascal and Prolog programming languages. The learner's procedural knowledge in terms of programming skills is traced, in the sense that each learner's actions and steps are compared to the rules contained in the "ideal student model" (Corbett and Anderson 1995 p.256). If the student's input matches the model entry, the system validates the learner's action, otherwise the system invalidates it until the learner enters a correct answer. To map performance to competence, knowledge tracing assumes two learning parameters, i.e., known and not known, and two performance parameters, i.e., guesses and slips (Corbett et al. 2000). For each correct answer learners entered into the system, there is a probability that they did not have the knowledge to do so, and for each incorrect input they submitted for correction, there is a probability that the learners actually knew the answer. The two main drawbacks of knowledge tracing are known as *identifiability* and *model degeneracy* (Baker et al. 2008). Identifiability refers to the issue that "models with equally good statistical fit to performance data may make very different predictions about a student's knowledge state", and model degeneracy characterises paradoxes, such as indicating a high probability that the learner becomes weaker even "after three correct answers in a row" (Baker et al. 2008 p.407).

The *constraint-based modeling technique*, suggested by Ohlsson (1992, 1994 mentioned in Mitrovic et al. 2002 p.243), is an approach to build a learner model based on learners' incorrect forms (mistakes and errors are, in this paper, used interchangeably). Examples of intelligent tutoring systems using constraint-based modeling technique are exemplified in (a) the SQL-tutor, an intelligent tutoring system for SQL programming[7] (Mitrovic 1997), and (b) the CAPIT system, an application which instructs the punctuation and capitalisation rules in the English language (Mayo et al. 2000). For instance, the domain knowledge in the SQL-tutor is represented as a set of constraints which intends to reproduce the conceptual syntax commonly used to access databases. According to Mitrovic and Ohlsson (1999), "knowledge about a domain can be represented by constraints on correct solutions in that domain" (p.239). Constraint-based modeling is a product-centred approach in the sense that the process by which the learner came to the solution is not taken into account, only the solution state is of importance. As long as learners do not succeed in realising the task, "they are free to perform whatever actions they please" (Martin and Mitrovic 2000 p.384). A constraint, as explained by Kodaganallur, Weitz and Rosenthal (2005), is described as a set of conditions that must be met in order for the learner to have his or her solution accepted by the application. Not all conditions are applicable to all problems contained in the tutoring system. For this reason, a constraint is delineated by a "relevance condition" and a "satisfaction condition" (Mitrovic 1997 p.8). The learner's solution is compared to a predetermined ideal solution using pattern matching technique, where an expression including variables is compared to an expression containing no variables (Mitrovic 1997 p.14). A constraint is considered as violated if and only if its relevance condition is true and its satisfaction condition is false. In other words, a violated constraint indicates that the learner entered an inconsistent solution. The representation of a learner's error in the SQL-tutoring system includes the idea that the probable lack of knowledge could also be a potential mistake (Martin 1999), yet without making any distinction between both types of incorrect form. Furthermore, as noted by Heift and Schulze (2007), constraint-based

7. SQL (Structured Query Language) refers to the language used when accessing and interrogating database systems.

modeling technique "can successfully rely on patterns because of the rather limited domain of a querying language such as SQL [...]. However, for a natural language this is a much more exhaustive and difficult task" (p.184).

The main drawback for some of these previously mentioned methods lies in the fact that modeling techniques infer learners' competence by comparing interlanguage performance to a pre-established set of expert knowledge. This does not identify the type of incorrect form, but rather highlights a discrepancy between the expert knowledge and the learner's performance. This discrepancy could equally refer to a mistake or an error. Other methods base the distinction between competence-dependent errors and performance-related mistakes on the idea of systematicity, which may not be sufficient to reveal a learners' competence (Gregg 1990 p.369). Alternative methods merely assume that a linguistic incorrect form is by default a lack of competence. Although inferring knowledge from performance will always, or probably always, be an assumption, there are researchers in second language learning who intend to overcome performance-related mistakes in order to better represent the learners' knowledge.

2.3.2. Learner models in language learning

In the domain of language learning, there are numerous examples of intelligent tutoring systems (e.g., Amaral and Meurers 2006, Bailin 1990, Bull et al. 1995, Cui and Bull 2005, Dickinson and Herring 2008, Hamel 2008, Heift 2001, 2010b, Heift and Nicholson 2001, Jia 2009, Kang and Maciejewski 2000, McCoy et al. 1996, Nagata 2002, Pérez et al. 2000, Shaalan 2005, Twidale et al. 1992, Virvou et al. 2000). Amongst them, a few applications emphasise the idea that surface description errors are not sufficient to infer the learner's knowledge. For example, (a) McCoy et al.'s (1996) ICICLE, an Interactive Computer Identification and Correction of Language Errors, (b) Amaral and Meurers' (2006, 2008) TAGARELA, a Teaching Aid for Grammatical Awareness, Recognition and Enhancement of Linguistic Abilities, and (c) Heift's (2001) E-Tutor do not assume that an incorrect form is a lack of knowledge by default, but rather attempt to hypothesise their causes.

ICICLE is an intelligent tutoring system designed to support the learning of English as a second language in the case of learners whose native language is ASL (American Sign Language) (McCoy and Masterman 1997, McCoy et al. 1996, Michaud and McCoy 1998). One challenge addressed by the system, when analysing learners' written texts and providing feedback, relates to the identification of the "relevance" of the learners' incorrect forms (Michaud et al. 2000 p.95). The concept of relevance is defined by these authors as the idea of identifying errors that are *relevant* to the learners, which implies to exclude incorrect forms such as (a) "careless mistakes" and (b) errors that "are far beyond the [learner]'s ability to grasp at this level of understanding" (Michaud et al. 2000 pp.95-96). More specifically, the system attempts to exclude not only incorrect forms that could have occurred even though the learner's level of mastery is proved to be high, but also incorrect forms that require far too many explanations for the learner's current level of ability (Michaud et al. 2000 p.96). The learner model aims to capture the learner's grammatical knowledge through the means of rules that characterise standard English structures, as well as bug rules which represent common errors made by hearing impaired learners (Michaud and McCoy 2003, 2006). In such a representation, the learner's knowledge is compared to the expert's knowledge and is considered as a subset of this expert's knowledge. To determine whether a grammatical aspect is acquired, the system draws its prediction on learners' past achievements, if any, otherwise on stereotypical[8] information. In addition, the system compares the number of times constructs of morphology and syntax, called "knowledge units", appear correctly against the number of times the same knowledge units have been attempted (Michaud and

8. A learner's stereotypical representation intends to accommodate all learners' states of knowledge by evaluating an interval, which would be delimitated by minimum and maximum averages of learners' knowledge (Elsom-Cook 1993).

McCoy 2003 p.97). Each grammatical construct is identified as “acquired”, “not acquired” or “ZPD” (Michaud and McCoy 2003 p.97). The labels *acquired* and *not acquired* indicate that the constructs are written always correctly or always incorrectly, respectively. The label *ZPD*, borrowed from Vygotsky’s (1978) zone of proximal development theory (Michaud and McCoy 2003 p.96), designates that the knowledge unit is sometimes correct and other times incorrect, which denotes a variable interlanguage performance. Naming the performance variability of the knowledge constructs as *ZPD* is somehow misleading, since learners’ *relevant errors*, as opposed to *careless mistakes*, are identified through systematicity, and not levels of assistance as expected by the theoretical framework of the zone of proximal development. Furthermore, if a learner’s knowledge unit is consistently incorrect (*relevant errors* that are *not acquired*), it means that a knowledge unit that is sometimes correct, other times, incorrect (labeled as *ZPD)* represents a mistake for the ICICLE model. This, however, is not clearly mentioned.

From a perspective other than systematicity and non-systematicity to distinguish between competence-dependent errors and performance-related mistakes, Amaral and Meurers’ (2008) consider “the fact that the task being performed can play a significant role in determining the students’ production” (p.328). Amaral and Meurers’ (2006) TAGARELA system is a web-based intelligent computer-assisted language learning application designed to assist the teaching as well as the learning of the Portuguese language. Described as an intelligent electronic workbook, the system provides six activity types: listening comprehension, reading comprehension, picture description, fill in the blank, rephrasing, and vocabulary (Amaral and Meurers 2009 p.581). The learners’ input may be of the form of either a word, a phrase or a sentence. The provision of feedback is a fully automated process, and it is provided in accordance with the learners’ needs depending on the appropriateness of their semantic and grammatical input. The student model is one of the six modules included in this system (the other being the web interface, the analysis manager, the feedback manager, the expert module, and the instruction model). Amaral and Meurers (2008) claim that “in order to guarantee valid interpretations of student performance it is not enough to keep track of a student’s production; it is vital to have at least information about the environment where it occurs” (p.328). They identify the fact that the task has an impact on the learners’ interlanguage performance and is habitually not acknowledged (Amaral and Meurers 2007 p.342). As a result, they include the submodules “task appropriateness” and “task strategies” to the existing sources of knowledge in their 2006 version of the system (Amaral and Meurers 2008 pp.333-334). The extended 2009 learner model takes into account learners’ personal information (e.g., age, mother tongue), language competence (linguistic properties, task appropriateness, and task strategies) and native language transfer (lexical and structural L1 transfer error) (p.582). The feedback is then determined depending on the probable origin of the error, which is referred to as a “lack of a specific linguistic ability”, a “lack of a strategic ability”, or an “insufficient mastery of a specific linguistic ability” (Amaral and Meurers 2008 pp.332-333). A lack of linguistic ability, for instance, means that the learner does not know how to make an agreement between words in gender and number; a lack of a strategic ability refers to the learner’s inability to perform a specific task due to lack of other strategies that could be implicated during the realisation of the main task itself; an insufficient mastery of a specific linguistic ability designates that the learner is able to demonstrate linguistic knowledge in simple context, e.g., fill in the blank, but is unable to do so when dealing with a more complicated task, e.g., reading comprehension task. The authors assume the learners’ linguistic knowledge through linguistic ability to use the language in context. The learners’ knowledge is inferred depending on tasks and strategies, and not merely on learners’ performance. However, even if a linguistic incorrect form is either due to a lack of strategic ability or an insufficient mastery of linguistic ability, the idea that an incorrect form could be the result of a slip of the keyboard or a specific emotional state is not envisaged in their model.

Based on other criteria than systematicity, or tasks and strategies, Heift (2003) infers the learner's knowledge based on performance history. E-Tutor, originally named "German Tutor", is an intelligent language tutoring system which not only offers online language exercises and chapters of textbook for the German language, it also analyses learners' sentences (Heift 2010b p.444). More specifically, it detects learners' incorrect forms, provides learners with feedback depending on parameters such as progress level, and keeps track of their performance in terms of grammatical aspects such as agreement or verb participle (Heift 2008, Heift and Schulze 2003). The learner is allowed to enter free input to a certain extent, given that vocabulary and grammatical structures used in the exercises are domain-restricted. The learner model represents "the current skill level of the student across different grammatical constructs and vocabulary" called "the *grammar* nodes" (Heift 2001 p.5, emphasis in original). When a grammatical incorrect form is detected by the system, the type of feedback provided to the learner is determined in accordance with his or her proficiency level, which is based on previous performance. The proficiency levels are either beginner, intermediate or advanced level. The learners' performance history is maintained by collecting evidence, either positive or negative, from their answers in order to assign scores to each language aspect taken into consideration by the system (Heift 2003). Scores, going up and down, are also weighted depending on parameters such as pedagogical objectives. As a result, a learner may be rated as a beginner in one particular aspect and as proficient in another (Heift and Nicholson 2001). If the learner is considered as a beginner then the type of feedback he or she receives is explicit. Conversely, if the learner is referred to as an advanced learner, the feedback provided tends to be less detailed (Heift 2003). Despite the fact that the E-Tutor system determines the level of implicitness/explicitness of the assistance required to help learners correct themselves, the error-mistake distinction is not informed. The reason for this is that the levels of assistance represent the result of the estimation of the learners' knowledge, and not the means to infer the learners' knowledge. Being rated as a beginner and being provided with explicit assistance does not signify that the learner wrote competence errors.

All theoretical positions, modeling techniques, as well as representative models discussed above represent valuable contributions to the understanding of the interlanguage variability phenomenon, as well as the learners' knowledge modeling domain. Although some researchers propose ways of representing the learners' interlanguage knowledge, yet without clearly distinguishing between errors and mistakes, it is frequently assumed that what learners produce incorrectly is, by default, evidence of their level of competence. Even if the understanding of mapping performance to knowledge estimation is improving, the difficulties of inferring the learners' knowledge from observations of interlanguage performance is still not resolved (Beck 2007). Therefore, interlanguage variability research is not fully attended to as the focus should also be on interlanguage competence.

2.4. Summary and conclusion

The chapter began with an introduction of the interlanguage hypothesis as a framework to define learners' grammatical knowledge, as well as their performance. Interlanguage is defined as an umbrella term under which interlanguage competence is juxtaposed with interlanguage performance. While interlanguage competence refers to the learners' proceduralised knowledge of a second or foreign language in terms of grammatical structures and lexicons, the learners' interlanguage performance relates to the learners' language productions.

In addition, the chapter showed that such a representation of a learner's interlanguage does not have unanimity amongst researchers in interlanguage variability research.

Depending on the theoretical framework, variability is investigated from either the learners' performance or the learners' competence. However, even amongst proponents of the competence-performance dichotomy, the general tendency is to analyse the learners' performance without extracting *what they know* from *what they do*. Yet, some researchers have proposed ways of distinguishing between competence-dependent errors and performance-related mistakes.

It was also demonstrated that there is a large amount of research dedicated to modeling learners' knowledge. However, none of these studies clearly address the issue of distinguishing between errors and mistakes either. Although there are some attempts to hypothesise the causes of an incorrect form, again it is often taken for granted that performance reflects competence.

In order to represent the learners' competence more comprehensively – so that interlanguage variability research can be given a new insight and so that the learners' knowledge can be represented without incorrect forms such as chance circumstances – a distinction between errors and mistakes is necessary. In line with James' (1998) method, it was argued that distinguishing between an error and a mistake could be realised by exploring the level of assistance learners require to correct themselves. This is the approach that will be adopted. The subsequent chapter will propose dynamic assessment, rooted in sociocultural theory, as an appropriate framework to distinguish between errors and mistakes.

Chapter 3. Toward a dynamic assessment to distinguish between errors and mistakes

The previous chapter outlined the void in interlanguage variability research with reference to the difficulty of representing the learners' interlanguage competence. This chapter proposes dynamic assessment as a suitable model to identify the learners' grammatical knowledge, in other words, to distinguish between competence-dependent errors and performance-related mistakes.

Dynamic assessment (DA) is commonly defined as an approach which integrates both teaching and assessment activities at the same time, and has its roots in sociocultural theory. After providing a concise overview on Vygotsky and his central concepts relevant to dynamic assessment, the second section presents both interventionist and interactionist approaches to dynamic assessment. In particular, it discusses the error-mistake distinction in light of both approaches. The third section explores the limitations and potentialities of applying such a model to the error-mistake distinction.

3.1. Dynamic assessment: a sociocultural framework

Lantolf and Poehner (2004) define dynamic assessment as an approach that

> integrates assessment and instruction into a seamless, unifed activity aimed at promoting learner development through appropriate forms of mediation that are sensitive to the individual's [...] current abilities. In essence, DA is a procedure for simultaneously assessing and promoting development that takes account of the individual's [...] zone of proximal development (Lantolf and Poehner 2004 p.50).

Vygotsky in Russia and Feuerstein in Israel both developed an alternative to traditional forms of IQ testing. Although the terminology used by both researchers is slightly different, their concepts are nevertheless very much alike[9] (Lantolf and Poehner 2009). Whilst Feuerstein considered linking assessment with intervention as a means to measure one's ability to benefit from mediation, Vygotsky investigated the potential development of an individual through the analysis of his or her independent and mediated performance (Sternberg and Grigorenko 2002). According to Poehner and Lantolf (2005), dynamic assessment, as opposed to other types of assessment, is not "based on teachers' intuitive classroom practice", but rather it is rooted in Vygotsky's developmental theory (p.260).

3.1.1. Vygotsky: a brief overview

Followed by a series of Russian battles in 1917 that destroy the Tsarist autocracy, Marxism led to a new way of living and working in Russia. The foundation of Marxism not only stresses socialism and collectivism but also places a great emphasis on history as a means to fully understand any culture or society (MIA)[10]. Strongly influenced by Marx's philosophy, Russian researcher Lev Semenovich Vygotsky was a prominent psychologist. Called once the *Mozart of Psychology* by Toulmin (1978), Vygotsky died in 1934 at the young age of thirty-eight, leaving behind him extensive materials (Luria 1935). Vygotsky's work is acknowledged to be partitioned into three different periods, (a) from responses to a stimulus to social behaviourism, (b) from social behaviourism to psychological materialism, and (c) from psychological materialism to cultural-historical theory (Veresov 2005). Cultural-historical theory is widely referred to as sociocultural (or hyphenated as socio-cultural) theory within language learning and teaching literature

---

9. According to Lantolf and Poehner (2009), "Feuerstein attributes various features of his work to the influence of his mentor, André Rey, who was a colleague of Piaget and no doubt acquainted with Vygotsky's research" (p151).

10. The Marxists Internet Archive (MIA) is an all volunteer, non-profit public library, accessible from http://marx.org.

(Poehner 2008). Sociocultural is a term that conveys different meanings depending on the research communities. Lantolf (2004) specifies that

> despite the label 'sociocultural' the theory is not a theory of the social or of the cultural aspects of human existence [...]. It is, rather [...] a theory of mind [...] that recognises the central role that social relationship and culturally constructed artefacts play in organising uniquely human forms of thinking (Lantolf 2004 pp.30-31).

Amongst the most invoked concepts of Vygotsky's theory of mind, the notion of the zone of proximal development (ZPD) is an essential part of dynamic assessment. Lantolf and Poehner (2009) consider DA as "the process of actualising the ZPD" (p.150).

3.1.2. The concept of the zone of proximal development

Vygotsky (1978) defines the concept of the zone of proximal development as "*the distance between the actual developmental level as determined by independent problem solving and the level of potential development as determined through problem solving under adult guidance or in collaboration with more capable peers*" (p.86, emphasis in original). Expressed in a different way, the zone of proximal development illustrates "those functions that have not yet matured but are in the process of maturation, functions that will mature tomorrow but are currently in an embryonic state" (Vygotsky 1978 p.86).

According to Kinginger (2002), the concept of the zone of proximal development is often considered as incomplete, thus leaving open its theoretical interpretation. In this regard, Wells (1999) avers that this concept "is the only aspect of Vygotsky [...] that most teachers have ever heard of and, as a result, it is not infrequently cited to justify forms of teaching that seem quite incompatible with the theory as a whole" (p.313). Chaiklin (2003) further points out that Vygotsky's ZPD should not be considered as the interpretations commonly found in the literature (p.39). According to Chaiklin (2003), common interpretations include the idea that the concept (a) applies “to learning all kind of subject matter”, (b) considers that “learning is dependent on interventions by a more competent other”, and (c) describes the “property of the learner that permits the best and easiest learning” (p.41). Opposed to these interpretations, Chaiklin (2003) asserts that the concept of the zone of proximal development is only intended for the analysis of the child development up to adolescence. In addition, he suggests the word scaffolding as a more relevant terminology than ZPD for the frequently inappropriate use of this concept in teaching practices (Chaiklin 2003 p.57). With regard to child development, Vygotsky’s conceptualisation of new methods of exploring mental processes need not be restricted to young people. According to Poehner (2008), the concept of the zone of proximal development may be extended to the study of adult learners of second or foreign languages (p.33). With regard to scaffolding, it is generally understood as the assistance provided by “an adult or expert” to “somebody who is less adult or less expert” during an instruction phase (Wood et al. 1976 p.89). While Holton and Clarke (2006) embrace the idea that scaffolding is rooted in a sociocultural framework, Lambert and Clyde (2003) strongly argue that “[t]he concept of scaffolding is not Vygotsky’s”, and assert that the use of both terms, i.e., scaffolding and zone of proximal development as synonyms denotes “ignorance not only about Vygotsky’s work [...] but about contemporary theories of cognition as well” (p.75). They add that scaffolding does not define “a process of cognitive transition from an actual to a potential level of ability”, but instead “appears to be more about consolidating current levels of expertise and gaining a basic level of mastery” (Lambert and Clyde 2003 p.76). As a result, scaffolding would not be an appropriate method to distinguish between competence-dependent errors and performance-related mistakes, since what needs to be known is the learners’ actual level of development. If a grammatical feature has matured and has been incorrectly written, then it is more likely to be a mistake.

The belief that learners' interlanguage competence may be extracted from their interlanguage performance through the observation of their amount of assistance is supported by the work of Aljaafreh and Lantolf (1994). These authors demonstrate the significance of interaction when considering the effects of negotiated feedback on second language learners' development. Their findings establish that different learners may have different ZPDs for the same incorrect forms, which implies that learners will require different levels of assistance in order to produce a correct replacement. As a result, they claim that they cannot "assume that the error represents the same problem for each learner, because the learners each produce it from a different location in the ZPD" (Aljaafreh and Lantolf 1994 p.474). From this point of view, an incorrect form may be a lack of knowledge for one learner, whereas it may represent a mistake for another. While learners requiring implicit feedback are close to independent performance, learners necessitating explicit feedback are further away from it. The incorrect forms "for which implicit strategic feedback proved to be effective are considered to be high in ZPD [...] since the learner is close to independent performance, while those that require explicit feedback are said to be low in the ZPD [...], because the learner is further away from producing the correct form without help" (Aljaafreh and Lantolf 1994 p.471). Independent performance outlines the learners' actual development, i.e., what they have internalised. In comparison, mediated performance indicates that the learner's interlanguage is still in the process of development (Lantolf and Poehner 2009). Therefore, the amount of feedback required to correct the incorrect form indicates its location within the learner's ZPD. One key principle associated with the concept of the zone of proximal development refers to the notion of mediation.

3.1.3. Mediation: tools and signs, and interactions

Mediation can be understood from two different angles: mediation as the notion of use of tools and signs, and mediation as the notion of interactions. Firstly, sociocultural research aims to understand one's cognitive functioning and how it is related to cultural and historical events. Human beings do not directly connect with the social and physical world, they only interact with the world through intermediaries (Wertsch 2007). According to Vygotsky (1997b), whose focus was essentially directed to language, human activities are mediated through auxiliary means referred to as *tools and signs*. Tools and signs are created by humans and society to fulfil a specific need. Vygotsky (1978) considered tools "as a means of mastering nature", and signs, more specifically, linguistic signs "as a means of social intercourse" (p.53). The tools in the case of work activities, whose functions are to help the individual operate on, and improve the external world, are technical tools (e.g., hammers or computers) (Lantolf and Poehner 2009 p.138). They are those one can touch and manipulate, and are often considered as artefacts "in the sense of historical artefacts that continue to exist after the humans who used them have disappeared" (Wertsch 1998 pp.30-31). The tools in regards of thinking activities, whose functions are to modify and improve the psychological world, relates to symbolic tools or signs (Lantolf and Poehner 2009 p.138).

Kozulin (2003) claims that the other face of mediation is human mediation. Vygotsky (1997b) specified that mediation can be achieved through the help of another more knowledgeable person (p.161). Mediated activities can be understood as a means to direct someone's attention to the object of consideration through interactions. Interactions, according to Vygotsky (1978), are an essential phase of the learners' cognitive development. Learning and development, and more specifically the development of higher mental functions[11], occurs initially as a result of interactions between people on the *interpsychological* level and then on the *intrapsychological* level (Vygotsky 1978 p.57). Interactions are materialised through the use of tools and signs to

11. Challenged with "the existing fact that newborn infants already possess certain mental functions", Vygotsky introduced, as a mean to differentiate between inborn cognitive functions and human mental functions social in origins, the concept of *higher mental functions* (Subbotsky 1994 para 1).

satisfy a specific need or solve a particular problem. They may include "verbal formulations and appeals" as a signal directed to those susceptible to help solve the task (Vygotsky 1978 p.29). Within the ZPD, interactions between learners and more knowledgeable persons represent a major aspect of the concept; it determines the learners' actual and proximal development. Learners interact with experts as an attempt to fulfil their lack of knowledge. More experienced others are not necessarily human beings; this role can be indeed undertaken by other media such as books or computers (Grigorenko 2009). Computer-mediated tools are now regarded as a source of assistance, "*a medium for learning and not a method for L2 instruction*" (Adair-Hauck et al. 2000 p.272, emphasis in original). With the varying amount of assistance provided through interactions by people or computers, learners regulate themselves by appropriating "the regulatory means employed by others" (Lantolf 2009 p.14).

3.1.4. Assistance: going from other-regulation to self-regulation

From a sociocultural perspective, cognitive development can be understood as the transition from object-regulation to self-regulation, passing through other-regulation. According to Lantolf (2000), individuals "are controlled first by the objects in their environment, then by others in this environment, and finally they gain control over their own social and cognitive activities" (p.6). Expressed in a different way, regulation refers to the manner an individual engages with his or her environment to successfully complete the task. For example, Aljaafreh and Lantolf (1994) determine their transitional "levels of help, or regulation" as other-regulation, partial self-regulation, and self-regulation (pp.470-471). While the first transitional stage (other-regulation) indicates that learners cannot perform without the help of somebody else, the second stage (partial self-regulation) demonstrates the learners' capability of "detecting and correcting their own mistakes without outside feedback" (Aljaafreh and Lantolf 1994 p.470). The third stage (self-regulation) denotes fully automatised skills, which indicates that learners are able to complete the task without relying on the assistance of another person (Aljaafreh and Lantolf 1994).

Providing assistance in the form of corrective feedback to help learners correct their written texts is a debatable topic. Corrective feedback, defined as “any indication to a learner that his or her use of the target language is incorrect” (Lightbown and Spada 1999 p.172), is either considered as harmful or helpful, or somewhere in between to the learner. For example, Truscott (2004), in response to Chandler (2003), reiterates that correcting learners’ grammatical incorrect forms may be ineffective as well as harmful to their fluency and writing skills. He concludes by stating that “the evidence, especially regarding grammar errors, points to a clear conclusion: Correction is a bad idea” (Truscott 2004 p.342). Chandler (2004), on the other hand, argues that error correction is helpful to language learners and increases accuracy in language writing. To support her claim, she lists previous studies such as the one from Ferris and Roberts (2001) whose findings show that the control group, who received no feedback, was outperformed by the group who received feedback. Yet, Hyland and Hyland’s (2006) review on feedback indicates that there are no clear answers as for the effectiveness of feedback on second language learners’ writing. Moreover, Ellis et al. (2009) show that there is no indisputable conclusions with respect to feedback types. While they suggest that “explicit feedback seems more likely to promote the cognitive comparison that aids learning” (p.330), Lantolf and Poehner (2011) argue that

> if the instructional aim is simply to help learners arrive at a correct response, then explicit feedback is certainly an efficient means. However, [...] if the intention is to promote development then process must be foregrounded, as in the ZPD (Lantolf and Poehner 2011 p.17).

Some researchers, such as Nassaji and Swain (2000), or Aljaafreh and Lantolf (1994) provide feedback within the learners’ zone of proximal development. Nassaji and

Swain's (2000) findings demonstrate that "help provided within the ZPD was more effective than help provided randomly", where randomly, in their context, means regardless of the learners' zone of proximal development (p.48). As for Aljaafreh and Lantolf (1994), they propose a regulatory scale characterised in terms of gradual feedback moving from implicit to explicit and dispensed to learners through interventions and interactions (p.471). Each step within the regulatory scale[12], listed as zero to twelve, describes the level of interaction that must be undertaken between a tutor and a learner in order for the learner to complete the task. For example, tutors at level zero do not intervene; learners are asked to correct themselves independently. At level twelve, tutors provide additional explanations in the event learners still do not understand the solution that was given to them at level ten. In between, the tutors' assistance ranges from indicating "something may be wrong in a segment" up to providing "clues to help the learner arrive at the correct form" (Aljaafreh and Lantolf 1994 p.471). Providing learners with a means of solving a problem, i.e., including a stimulus-means between the object and the response is also referred to as the double stimulation method (Vygotsky 1978 p.74).

3.1.5. The double stimulation method

Vygotsky's (1978) double stimulation method is equally called "the functional method of double stimulation" (e.g., Vygotsky 1978 p.74), or other designations such as the "instrumental method", the "experimental-genetic method" or the "historical-genetic method" (Engeström 2007 p.364). It is defined as follows:

> The task facing the child in the experimental context is, as a rule, beyond his present capabilities and cannot be solved by existing skills. In such cases a neutral object is placed near the child, and frequently we are able to observe

12. Aljaafreh and Lantolf's (1994) regulatory scale is given in full in Appendix B on page 282.

> how the neutral stimulus is drawn into the situation and takes on the function of a sign. Thus, the child actively incorporates these neutral objects into the task of problem solving. We might say that when difficulties arise, neutral stimuli take on the function of a sign and from that point on the operation's structure assumes an essentially different character (Vygotsky 1978 p.74).

According to Vygotsky (1978), double stimulation seeks to delineate the *history* (from beginning to end) of the developmental process by placing learners in a situation that is assumed to be above their natural capabilities (p.74). To do so, the individual "is put in a structured situation where a problem exists [...] and is provided with active guidance towards the construction of a new means to the end of a solution to the problem" (Van der Veer and Valsiner 1991 p.169, cited in Engeström 2007 p.364). A close examination of the learners' behaviour should help "shed light on the ways that newly acquired [...] concepts or skills become stable or 'fossilised'" (Portes et al. 1993 p.6). Guidance, whether in the form of teachers' interventions or in the form of interactions between teachers and learners, aims at providing learners with help to complete the task as well as promoting their cognitive development. Whilst an approach to dynamic assessment implies the teacher's participation (intervention), it is equally important to note the involvement of the learner during this process (interaction) (Van der Aalsvoort and Lidz 2002).

3.2. Dynamic assessment: concepts and principles

The term dynamic assessment originated in research investigating children's abnormal behaviours (Mathews 1961) and is nowadays predominantly applied in areas such as learning disabilities or adults' language impairments (e.g., Keane 1987, Moore-Brown et al. 2006, Navarro and Calero 2009, Samuels 2000). However, other practitioners have started to widen the use of dynamic assessment practices to second language assessment and pedagogy (e.g., Ableeva 2008, Erben et al. 2008, Gibbons 2003, Guthke et al. 1986, Kozulin and Garb 2002, Lantolf and Poehner 2004, Oskoz 2005, Peña and Gillam 2000, 2000, Poehner 2005, 2007, 2008, Poehner and Lantolf 2005, Schneider and Ganschow 2000).

Dynamic assessment is sometimes described as an umbrella term regrouping a myriad of different approaches and methods (Elliott 2003). Key terms, such as dynamic assessment, interactive assessment, dynamic testing or learning potential assessment, to name but a few, alternately act as synonyms or distinct words (Haywood and Lidz 2007). For example, Grigorenko and Sternberg (1998) consider dynamic testing and dynamic assessment as having different meanings. They add that "the goal of dynamic assessment is to evaluate, to intervene, and to change. The goal of dynamic testing, however, is much more modest: It is to see whether and how the subject will change if an opportunity is provided" (Grigorenko and Sternberg 1998 p.76). According to Poehner (2008), assessing dynamically "posits a qualitatively different way of thinking about assessment from how it is traditionally understood by classroom teachers and researchers" (p.1).

Dynamic assessment is commonly described according to the type of "mediated assistance" provided to learners in order for them to attain their goal (Lantolf and Poehner 2004 p.54). For example, Daniel (1997) distinguishes two groups with different intervention processes: the first approach provides "standard interventions" and the second one "nonstandardised interventions" (p.1041). While the former relates to the use of measures to determine the amount of prompts learners require to be able to provide a correct alternative, the latter refers to Feuerstein et al.'s (1979) idea of associating intervention with assessment (Daniel 1997 p.1041). Lantolf and Poehner (2004) refer to both approaches as "interventionist" and "interactionist", respectively (p.54). During an interventionist approach, teachers are "not free to respond to learners' needs [...] but must instead follow a highly scripted approach to mediation in which all prompts, hints, and leading questions have been arranged in a hierarchical manner" (Poehner 2008 pp.44-45). By contrast, mediated assistance between learners and teachers during an interactionist orientation is negotiated rather than established in advance, which is in line with Vygotsky's concept of the zone of proximal development (Lantolf and Poehner 2004 p.58).

3.2.1. Interventionist approaches to dynamic assessment

Sternberg and Grigorenko (2002) discern two subcategories within the interventionist approach: the "sandwich" and the "cake" formats (p.27). In the *sandwich* format, "the instruction is given all at once[13] between the pretest and the posttest", whereas in the *cake* format, "the instruction is given in graded layers after each test item, as needed" (Sternberg and Grigorenko 2002 p.28).

Mediation within the *sandwich* structure can be either individualised or group directed. In the case of a group, the instruction is not personalised, it is identical for all examinees. The assistance provided to learners tends to be more explicit when provided to individuals, and more implicit when administered to group settings (Sternberg 2005). For example, although Budoff and his colleagues (Budoff 1987, Budoff and Friedman 1964, Corman and Budoff 1973) used intelligence tests when assessing institutionalised teenagers, they were among the first researchers to experiment with the *sandwich* format, also called the test-intervene-retest method. In such a context, learners were allowed to be trained and retested after being first tested through standardised intellectual assessment. The post-test is used to (a) compare the learner's assisted performance with the unassisted performance and (b) analyse how the learner has improved (Grigorenko 2009). The learning potential is estimated as an improvement in performance due to

13. Lidz (1991) states that post-tests can be administered one day or even one month after training (p22).

instruction between the pre- and post-tests. However, investigating an improvement in terms of learning potential in a learner's second performance does not identify the nature of the incorrect form in the first performance. The incorrect form may have been a mistake that was not repeated during the post-test. This result would have been considered as an improvement. However, there is no conclusive evidence to support this statement.

Within a *cake* format, learners are given questions one by one. If they answer the first question correctly, then the second question is given. Otherwise, they are provided with graded assistance, "like layers of icing on a cake", designed to make the solution more obvious each time (Sternberg 2005 p.25). The information is then used by the teacher to determine "how many and what kinds of hints the examinee needs in order to solve the item" (Sternberg 2005 p.25). Prevailing models of cake format to dynamic assessment are typically linked with names such as Brown and Ferrara (1985), Carlson and Wiedl (1979), or more recently Glaspey and Stoel-Gammon (2007). Brown and Ferrara (1985), for instance, base their graduated prompt method on Vygotsky's zone of proximal development. Mostly situated in the context of mathematics, as well as reading and spelling, Brown and colleagues utilise either human or computer interaction as mediators (Haywood and Lidz 2007). The assistance provided within this approach relies on a series of predetermined standardised prompts designed to assist the learner until the task is finally achieved. As for most interventionist approaches, the help provided to the learner is organised from more implicit to more explicit (Poehner 2008). The learner's attainment is then calculated by counting the amount of hints required by the learner to provide correct constructs. More precisely, the learner's learning potential is estimated by measuring the amount of instruction needed in order to master the specific task, and by analysing how far this knowledge can be transferred to new contexts (Grigorenko 2009). Carlson and Wiedl's (1979) method, known as the *testing-the-limits* approach, was developed using multiple choice questions designed for testing abstract reasoning in children. The approach attempts to respond to *what if* questions of how learners "would perform if, for example, provided with more time", seeking, thus, an enhancement of learners' performance (Haywood and Lidz 2007 p.325). Unsatisfactory test scores are here not interpreted as evidence of learners' deficient abilities, but rather as different learners' intellectual and educational experiences (Sternberg and Grigorenko 2002). More recently, Glaspey (2008) introduces the Glaspey Dynamic Assessment of Phonology (GDAP), formerly known as the *scaffolding scale stimulability* approach to dynamic assessment (Glaspey and Stoel-Gammon 2007). Glaspey and Stoel-Gammon (2007) explain that the assessment of stimulability typically takes place after a static assessment, the aim being to modify linguistic complexities in order to stimulate the learner's production. By merging the stimulability approach with dynamic assessment, Glaspey (2008) combines a graduated prompt approach with the testing/training unity paradigm.

Grigorenko (2009) points out that interventionist approaches to dynamic assessment are generally suitably used in computerised systems. For example, Erben, Ban, and Summers (2008) have undertaken dynamic assessment with multiple choice questions up until now. The cake format in particular, is commonly used in computer-based dynamic assessment to evaluate and quantify changes in performance through psychometric measure (Embretson 2000, Kalyuga and Sweller 2005). Although the assistance provided to learners arranged from implicit to explicit could enlighten the error-mistake distinction, one limitation to consider within the cake format is that it uses standardised assistance for each learner. Standardised assistance implies that every learner is given comparable procedures (De Beer 2006). Prompts defined in advance and identical for everybody may work well for multiple choice questions, but may not be appropriate for non-restricted written texts. Students would be unlikely to write the same words and sentences and thus, unlikely to produce the exact same incorrect forms in the same context.

3.2.2. Interactionist approaches to dynamic assessment

Feuerstein, Rand and Hoffman (1979) developed the Learning Potential Assessment Device (LPAD) to assess one's learning potential by providing learning opportunities through active guidance. The role of the mediator is to "observe how a student approaches a problem and to explicate and remediate the difficulties experienced by a student" (Grigorenko 2009 p.117). The learning potential assessment device, a battery of instruments designed to allow teachers' observations and interventions, is rooted in R. Feuerstein and S. Feuerstein's (1999) concept of Mediated Learning Experience (MLE). The mediated learning experience relates to the interactive relationship between the learner and the teacher. R. Feuerstein and S. Feuerstein (1999) identify attributes which characterise the quality of an interaction: "intentionality and reciprocity", "mediation of transcendence", and "mediation of meaning" (p.15). Firstly, *intentionality* and *reciprocity* are defined as (a) the "intention to mediate to the mediatee" and (b) the means "to turn an implicit intention into an explicit, volitional, and conscious act" (Feuerstein and Feuerstein 1999 p.17). Poehner (2008) points out that, while it seems evident that a mediator would in fact intend to mediate, *intentionality* remains an important factor for the following reason: it contrasts with incidental learning in the sense that the concept of mediated learning experience "is focused on the child's cognitive development through guiding him as he participates in various activities" (p.57). *Reciprocity* refers to the idea that the learner's participation is viewed as an active co-construction of knowledge between both the learner and the teacher (Poehner 2008 p.58). Secondly, *mediation of transcendence* is described as the "individuals' ability to re-contextualise their learning and apply it to new, more demanding problems" (Poehner 2007 p.325). Thirdly, *mediation of meaning* is defined by R. Feuerstein and S. Feuerstein

(1999) as the significance of objects, actions or events in one's environment that could not be fully understood without proper mediation. The learning potential assessment device, in which the components of the mediated learning experience are concretised, is a new way of considering cognitive development (Feuerstein et al. 1998). Indeed, the focal point moves from "what an individual is able to do (at a given moment in time) to what the individual *can become able to do* in immediate time frame and in subsequent, future interactions" (Feuerstein et al. 1998 p.98, emphasis in original).

A more recent example of interactionist approach to dynamic assessment is illustrated by Poehner's (2005) work in the context of L2 learners of French. Focusing on spoken language, advanced undergraduate students met their teacher outside the classroom in order to improve their oral communication. After independently producing a past-tense narrative in the target language, learners were asked to replicate the task with the help of an examiner during the sessions. The aim of the individualised tutoring was to address the learners' issues identified during the assessment in a highly and flexible way. After twelve sessions of mediated assistance focusing on past tense and aspect – two sessions a week – the learners were "re-administered the original independent and mediated narration tasks so that any development during the enrichment programme could be observed" (Poehner and Lantolf 2005 p.246). Pohner's (2005) findings illustrate what he calls a "double-sided coin metaphor", that is, that dynamic assessment "at once assesses and promotes development" (p.313). Within such a method, learners are provided with assistance, which is negotiable and tailored to their individual needs. With respect to the error-mistake distinction, a mistake would be automatically detected with the first level of interaction, as teachers and learners alike would know immediately that the incorrect form was not intended.

With regard to written language, Antón (2009) reports on the implementation of a dynamic assessment approach, which could be categorised as interactionist. Second language learners of Spanish were given twenty minutes to write an essay without assistance. The correction of their text was performed under the supervision of the teacher. Firstly, students self-edited their written text after which they were then given a dictionary and a grammar manual. Finally, they were invited to interact with the examiner about concerns they might have about the composition of the text. According to the author, learners have the opportunity to revise what they think they do not know (Antón 2009). However, it is not clear whether learners have the opportunity to revise the incorrect forms they are unaware of. If they do not know something is wrong, it is more likely that they will never ask about it, and therefore miss opportunities to interact and correct themselves. Skinner and Madden (2010) further point out that even if learners believe they need help, it is not certain that they will ask for it (p.21).

So far, the discussion on interventionist and interactionist approaches to dynamic assessment shows that not all methods are appropriate to help distinguish between errors and mistakes. An interventionist method is more conducive to quantitative analysis, counting prompts and hints to measure the learner's ability to provide a correct alternative. Although this approach is commonly computerisable, the assistance is often pre-established in advance and not adapted to individuals. An interactionist method, on the other hand, is qualitatively oriented. It assesses and promotes the learner's cognitive development following Vygotsky's concept of the zone of proximal development, as illustrated above with Poehner's (2005) work. However, adapting Poehner's (2005) method to help distinguish between an error and a mistake would not be appropriate since he explored learners' potential learning after six weeks time, whereas the aim of this thesis is to first reveal the learner's actual knowledge, and then investigate its variability. Furthermore, the application of such a method seems to point towards the analysis of the learners' spoken, rather than written language. Table 3.1 summarises the main differences between both interventionist and interactionist approaches.

Table 3.1. Interventionist and interactionist approaches to dynamic assessment

Interventionist	**Interactionist**
• Quantitative analysis: psychometrical measures.	• Qualitative analysis: ZPD.
• Computer-based application: assessment conducted with large numbers of individuals simultaneously (multiple choice questions).	• Face-to-face: small numbers of students.
• Written and spoken language.	• Spoken language.
• Mediation between learners and teachers established in advance; • Assistance not tailored to learners' responsivity; • Hints ranging from implicit to explicit • Standardised mediation.	• Mediation between learners and teachers negotiated; • Mediation tailored to learners' responsivity.
• Individual or group settings.	• Individual.

3.3. Distinguishing between competence errors and performance mistakes

Despite the fact that not all approaches to dynamic assessment are in line with the concept of the zone of proximal development, DA may nevertheless constitute an adequate framework to elicit information about a learner's competence in written language. Dynamic assessment not only gives information about learners' future development, but also enlightens what learners are able to perform today independently. Therefore, learners' incorrect forms for which there is no potential development, as they have been already internalised, should be highlighted.

3.3.1. Dynamic assessment: limitations and potentialities

Although many researchers have advocated the use of dynamic assessment as a means to assessing learners' spoken or written language (Daniel 1997, Erben et al. 2008, Grigorenko and Sternberg 1998, Lantolf and Poehner 2004, Poehner and Lantolf 2005), DA has not received much attention from researchers from either psychological or educational fields (Sternberg and Grigorenko 2002). Sternberg and Grigorenko (2002) advance three probable reasons for such a relative reserve (pp.30-31). Firstly, very little has been published on the reliability and validity of dynamic assessment. Secondly, the paucity of details in describing the methods applied makes replication almost impossible. Thirdly, educators and psychologists seem to ignore dynamic assessment because of an insufficient familiarity with the assessment model they have in mind. In addition, De Beer (2006) outlines the time issue with regard to its application in real educational settings. Indeed, when teachers may have classes of up to one hundred learners, putting dynamic assessment into practice can be considered quite a challenge (Poehner 2007). Furthermore, dynamic assessment whose primary purpose is to “interpret” rather than “measure” has “not been accepted with open arms by members of the testing community” (Lantolf and Poehner 2004 p.66). Although there are some studies on interactionist approaches directed to spoken language, very few examples are published in the context of written language, and even fewer are computer-based. For example, Ebadi (2010)[14] analyses and develops a one-to-one interactionist approach using written and spoken prompts in various collaboration tools to explore learners’ grammatical structure in their writings. To this researcher’s knowledge, there are no other examples of computer-based dynamic assessment grounded in an interactionist approach and directed to non-restricted written language.

Even lacking examples of computer-based interactionist approaches in the case of free written texts in L2, merging both interventionist and interactionist theoretical principles should help explore the learners’ ZPD in the context of documents written as a response to an open task. For example, simultaneously assessing a large number of students through computer-based applications (interventionist) with mediation tailored to individuals and negotiable in the sense that learners can accept or ignore the assistance (interactionist) should enable the creation and observation of the learners’ ZPD. This will

14. Stated in a private email correspondence.

allow the monitoring of their actual and potential development. In view of such potentialities between both interventionist and interactionist approaches, James' (1998) practical classification of incorrect forms in terms of errors and mistakes can be operationalised in line with dynamic assessment principles.

3.3.2. Operationalising the DA-based error-mistake distinction

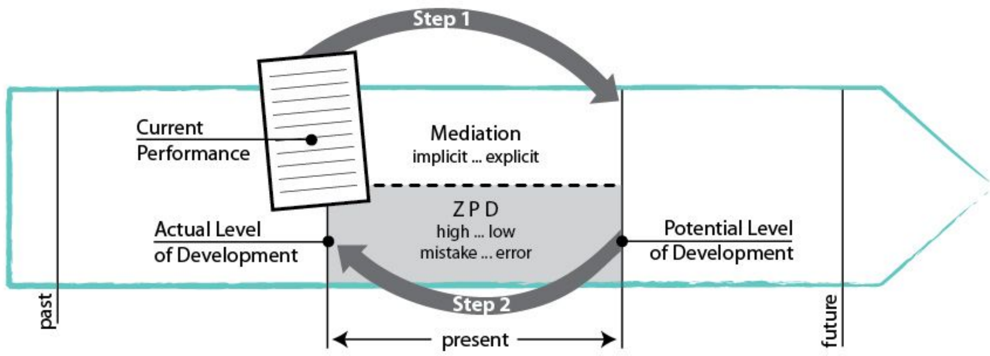
On the one hand, an incorrect form that necessitates implicit feedback is considered to be high in a learner's zone of proximal development, whereas an incorrect form that demands more explicit explanations is deemed to be low in a learner's zone of proximal development (Aljaafreh and Lantolf 1994 p.471). On the other hand, an incorrect form that necessitates implicit feedback is considered to be more likely a mistake, whereas an incorrect form that requires more explicit assistance to be corrected is characterised as an error (James 1998 p.83). Therefore, the distinction between errors and mistakes is linked to the representation of the learners' zone of proximal development. In other words, positioning each incorrect form in the ZPD, i.e., determining the learners' potential development – their future –, will help establish whether the incorrect form to be analysed is in fact an error or a mistake.

According to Valsiner (2001), there are two different ways of predicting a learner's future achievement: the "past-to-present" and the "present-to-future" models (p.86). Within a past-to-present model, the past leads to the present stage of functioning, and development refers to "a 'sequence of stages' that a person is assumed to pass through on the way to the final stage" (Lantolf and Poehner 2004 p.52). For example, traditional summative assessment is designed to measure the learner's performance at a specific point in time, conventionally at the end of a course or an academic period (Perie et al. 2009). Despite the fact that summative assessment is concerned with what has been learned, results obtained from such assessment "are often used to make decisions about individuals' futures (e.g., admittance to a program, promotion to a higher level of study, or conferral of a degree or certification)" (Poehner 2007 p.323). Valsiner (2001) further reports that "the dynamic *changes* of the past that have led to the present can also explain the future" which he defines as a *"post factum"* prediction (Valsiner 2001 p.86, emphasis in original). In that case, the future is acknowledged as being the continuance of past performances which may contain performance-related mistakes.

On the other hand, present-to-future models focus on the "future-in-the-making" (Valsiner 2001 p.86). They predict the learners' future attainments from what they can achieve in collaboration with other more knowledgeable persons (Poehner and Lantolf 2005). Valsiner (2001) points out that such models "give researchers a focus on the processes of emergence – or construction – of novelty" (p.86). A present-to-future model "sees ability not as a stable trait but as a malleable feature of the individual and emergent from the activities in which the individual participates" (Lantolf and Poehner 2004 p.53). In other words, present-to-future models, which underpin dynamic assessment, enable the prediction of what learners will be able to accomplish in a proximal time through guided assistance provided by more expert persons.

While past-to-present models tend to infer the learners' future achievements from past and present performances, present-to-future models intend to predict learners' future from their level of potential development, i.e., from the amount of assistance required to complete the task. However, neither model infers the learners' current knowledge. Therefore, I propose an expanded version of the present-to-future model. Its aim is to reveal what learners currently know, as opposed to what they will know. Starting from the learners' independent performance, the prerequisite to distinguish between errors and mistakes is to determine the placement of the incorrect forms into the learners' ZPD. In contrast with the present-to-future model, the expanded model proposed in this thesis looks ahead at the learners' level of potential development, and then looks back at their independent performance to infer their knowledge at time present (Figure 3.1).

Figure 3.1. Expanded model



Learners' written texts produced at time *present* correspond to their current independent performance. Step 1 refers to the creation of the zone of proximal development, which is estimated for each incorrect form included in the learners' documents. The learners' potential level of development reflects the learners' achievements realised with mediation going from implicit to explicit, that is, what they are able to perform with assistance at time *present*.

In step 2, the zone of proximal development is observed. If learners require no more assistance than merely pointing the ill-formed words out to edit them, then the incorrect forms are high in the ZPD. When the learners are nearly at the point of completing the task by themselves, the actual and potential development of these incorrect forms are close to one another. The distance between both levels is almost nonexistent. By contrast, if learners are in no way able to provide correct alternatives with implicit feedback, the incorrect forms are considered low in the ZPD. The distance between both the levels of actual and potential development is now noticeable. Incorrect forms that are low in the learners' ZPD are considered as errors, whereas those that are high in the ZPD are regarded as mistakes. Mistakes may then be extracted from the learners' current performance to unveil their actual level of development, i.e., their present knowledge.

The expanded model presented above constitutes the context necessary to create and observe the learner's ZPD which is necessary to operationalise the error-mistake distinction proposed by James (1998). While the other two models (past-to-present and present-to-future) aim at predicting the learners' future attainments, the proposed model reflects the steps to unveil their *present knowledge* based on the observation of their *potential development*. Once the ZPD is created and observed, the error-mistake distinction can be made. Given the fact that incorrect forms highly placed in the ZPD will be considered as being *very close* to independent performance and consequently, will more likely be performance-related mistakes, it is worth mentioning that these incorrect forms are still pointed out to learners, thus implying a certain amount of assistance.

3.4. Summary and conclusion

The chapter started with a brief overview of Vygotsky's sociocultural theory as well as a discussion of the principal concepts underpinning dynamic assessment. The focus was on the concept of the zone of proximal development in which dynamic assessment is firmly grounded. Linked to the concept of the ZPD, notions such as mediation, regulation, and double stimulation were also discussed .

Interventionist and interactionist approaches to dynamic assessment, generally categorised depending on the type of assistance offered to learners during the assessment were then presented. To a large extent, assistance, either in the form of teachers' interventions or in the form of interactions between teachers and learners, aims to promote learners' cognitive development.

Embracing DA as a means to distinguish between errors and mistakes resulted in the necessity of adopting principles from both interventionist and interactionist perspectives. While the former is well adapted for a computerised approach, the latter is more in line with the concept of the ZPD. A representation of the error-mistake distinction was explored, and an expanded version of Valsiner's present-to-future model was proposed.

The next chapter will explore the methodological issues that need to be addressed in order to first operationalise this model and then investigate the variability of a learner's interlanguage competence. Adopting dynamic assessment as a means to distinguish between errors and mistakes involves the application of the model proposed within this chapter. In practice, one challenging task of this doctoral research is to create and represent the learner's zone of proximal development, so that it can be observed.

Chapter 4. Interlanguage variability and dynamic assessment: methodology

Performance is variable, whereas the extent to which competence is cannot be thoroughly elucidated. Some light can be shed on the subject by differentiating between competence-dependent errors and performance-related mistakes. The previous chapter proposed and discussed a dynamic assessment based model to better reflect the learners' interlanguage competence. This chapter identifies the methodological issues and requirements to conduct this doctoral research.

The first section starts by exploring the challenges of operationalising dynamic assessment as a means to distinguish between errors and mistakes. Consideration of the implications as well as the methodological choices to address these challenges are then explored. Finally the presentation of the context of this study and the research design are highlighted.

4.1. Challenges in applying a dynamic assessment based error-mistake distinction on interlanguage variability research

The key challenge of the current research is to operationalise the expanded model proposed in the previous chapter, i.e., to operationalise the error-mistake distinction. More specifically, the design of the context to realise the learners' zone of proximal development and the representation of this zone are crucial elements within the model. Without them, the zone cannot be observed and therefore the distinction between errors and mistakes cannot be determined. Having proposed dynamic assessment as an appropriate environment to apply the distinction, the different challenging tasks are based on the process of (a) operationalising the expanded model, and in particular representing the learners' zone of proximal development, (b) providing assistance tailored to individuals in open-ended essays through the means of a computer, (c) finding appropriate psychometric measures to represent the learners' current knowledge, and (d) creating and annotating a collection of learners' written texts.

4.1.1. Operationalising the expanded model

The aim of this model is to concretise the learners' zone of proximal development for each incorrect form found in a written text. Depending on its location, i.e., high or low in the ZPD, the incorrect form will likely be a mistake or an error, respectively. The concept of the zone of proximal development is a *theoretical* framework and, according to Kozulin and Garb (2002), no "standardised procedure for the ZPD assessment" was originally produced (p.113). Vygotsky's (1998) most comprehensive explanations could be as follows:

> We show the child how such a problem must be solved and watch to see if he can do the problem by imitating the demonstration. Or we begin to solve the problem and ask the child to finish it. Or we propose that the child solve the problem that is beyond his mental age by cooperating with another, more developed child or, finally, we explain to the child the principle of solving the problem, ask leading questions, analyse the problem for him, etc. (Vygotsky 1998 p.202).

To this researcher's knowledge, there is no systematic method to represent the learner's zone of proximal development concretely or graphically. Vygotsky (1987) labelled the ZPD with numbers designating mental age differences between the levels of actual and potential development. For example, a child's zone expressed with the number "4" designates that he or she was able to perform a task designed for a child 4 year older with the assistance of an adult (Vygotsky 1987 p.209). No other examples of such a method were found in the literature in general, and more specifically in the L2 learning and sociocultural context.

As previously outlined, Lantolf and Poehner (2009) claim that the representation of the ZPD can be achieved through the means of dynamic assessment (p.150). Kozulin and Garb (2002), for instance, follow the test-intervene-retest method to investigate the "feasibility of the development and implementation of a dynamic assessment procedure" (p.112). After administering static tests reflecting text-comprehension tasks to students, the teacher reviews and mediates each test item in an interactive way. The post-test evaluates how much the students have benefited from the mediation. To operationalise the learners' learning potential, Kozulin and Garb (2002) developed the Learning Potential Score (LPS), which reflects "both gain made by the student from pre- to post-test and an absolute achievement score at the post test" (p.121). According to the authors, the LPS score determines whether the learner has a high or low learning potential (Kozulin and Garb 2002). For instance, a student who obtained a low score at the pre-test and a high score at the post-test is considered as having a "very high LPS" , whereas a student with two identical scores at both tests is considered as having a "very low LPS" (Kozulin and Garb 2002 p.121). Although the authors indicate that dynamic assessment may be developed and implemented in the context of English as a foreign language, their method would be extremely difficult to duplicate due to a paucity of information concerning the procedure, and particularly on how the mediation process was conducted on each item.

4.1.2. Providing assistance in the case of free written language

Dynamically assessing texts implies correcting learners' incorrect forms and providing them with different levels of assistance. Although computer-based assessment technology has spread geographically (horizontally) and throughout language courses to become an essential component of any study (vertically) (Chapelle 2009), teachers' use of technology is more related to administrative purposes (Judson 2006, Palak and Walls 2009, Windschitl and Sahl 2002). As a result, the provision of corrective feedback is generally "written on drafts or given orally" (Parr and Timperley 2010 p.68). Yet, there have been some methods for marking and returning electronically-submitted written documents. A common manner to proceed is to print the assignment, correct it in a traditional fashion with a pen, and return it to the student in person. Another method is to use some mainstream applications to transmit feedback. For example, most people are familiar with word processors and their revision tools including in-line editing. Word processors have however a certain number of disadvantages. On the one hand, the same program, even sometimes the same version, must be used by both learners and teachers, otherwise a change of formatting could incur the risk of loosing annotations. On the other hand, changes are directly added in the original text and unless it is converted into a read-only document, corrections could be modified or deleted by the learner without the knowledge of the teacher. Szerdahelyi (2010) suggests other means for correcting asynchronous texts using for instance web technology such as discussion boards or wiki spaces. Despite these options, there is no allowance for several levels of assistance. Computer-based authoring options exist over the Internet to help teachers assess learners' language; these tools are designed to develop tests mostly based on multiple-choice, true-false, fill-ins or short answer questions (Douglas and Hegelheimer 2007 p.118). The idea of using the web to provide learners with correction and assistance is attractive as it enables learners to correct their texts at their own pace without installing specific applications on their computer. Moreover, students who do not own a computer could also access their corrections by using any computing facilities. Additionally, learners and teachers do not have to share the same physical space to interact, assistance being provided remotely. According to Hwang, Ang, and Francesco (2002), some learners tend to avoid face to face feedback to limit possible embarrassments, thus preferring the electronic scenario. Another advantage is that the data is already computerised which facilitates the collection of learners' texts as well as the creation of the corpus. However, using online access to assess learners' texts as well as to retrieve their corrections is not without ethical concerns, especially with regard to data protection.

4.1.3. Identifying appropriate psychometric instruments to measure knowledge

Thorne (2005) argues that Vygotsky's sociocultural theory differs ontologically from other traditional views of psychological theories of mind that prioritise the individual. The ontology is different in the sense that "it is an ontology of the individual as a socially constructed being", as opposed to an "ontology of the autonomous individual" (Lantolf 2009 p.365). Consequently, since both trends of research do not share the same belief about the individual, traditional concepts of measurement may not be appropriate to sociocultural theory and therefore, not suitable to dynamic assessment (Poehner 2008). Opposed to the idea of dynamic assessment as a means to measure ability, Snow (1990) argues that DA faces major issues with regard to psychometric measurements. He further states that without associating the term assessment to "the concept of measurement [...]; the term is meaningless" (Snow 1990, not paginated, para 13). While Sternberg and Grigorenko (2002) report that researchers' endeavours to quantify learners' emergent ability have not been consistent (p.30), Kaniel (2000) claims that on the account of the "dynamic and qualitative nature of the assessment – intervention, the complex mechanisms of validity and reliability cannot automatically be adopted" (p.669). Indeed, the issue of reliability and validity within dynamic assessment research is a recurrent topic in the literature (Poehner 2005 p.116). This suggests that dynamic assessment does not provide any method to measure the learners' potential and actual development psychometrically. Consequently, the search for psychometric instruments outside the context of dynamic assessment and sociocultural theory is a necessary step to be able to represent learners' grammatical knowledge.

Measuring learners' knowledge

A learner's grammatical knowledge, as defined in Section 2.1.4 on page 19, includes vocabulary as well as syntactical aspects of the language. Measuring the learner's knowledge implies looking at errors (not mistakes) as well as correct constructs. Indeed, analysing learners' errors does not indicate what learners know, it merely informs of what learners could not produce correctly. In order to obtain a fuller picture of the learners' current knowledge, correct instances of the target language should also be taken into account. A common method for estimating the extent to which a learner has acquired a linguistic aspect refers to the "obligatory occasion analysis" (Ellis 1994b p.74). Obligatory occasions are specific linguistic features whose obligatory presence in a learner's performance is conditioned by the target language. For example, there are two obligatory occasions for the past tense -*ed* in (a) *my sister visited us yesterday* and (b) *my father *arrive yesterday*, one being correct (visited), the other being incorrect (arrive) (Ellis 1994b p.74). The obligatory occasion method is defined as a means to examine "how accurately learners use specific linguistic (usually grammatical) features" (Ellis and Barkhuizen 2005 p.73). Accuracy is thus determined by counting all obligatory occasions of one grammatical construct in a learner's production, as well as all correct instances of the same construct (Ellis and Barkhuizen 2005 p.83). First developed in the domain of native language (Brown 1973), the obligatory occasion method of analysis has been borrowed by L2 researchers to study the order of morpheme acquisition, either longitudinally or cross-sectionally (e.g., Hakuta 1974 and Dulay and Burt 1973, respectively). For example, Hakuta (1974) investigated over a period of forty weeks the presence, absence and order of grammatical morphemes of a Japanese girl learning English. Morpheme studies, abandoned in the 1980's, were criticised for investigating small sets of morphemes and for grouping them disparately (Ellis and Barkhuizen 2005 p.78). The obligatory occasion analysis, however, has not been totally abandoned. In his investigation of past tenses in the case of learners of French, Poehner (2005) undertakes a similar approach without explicitly naming it. While dynamically assessing his students, he counts “the total number of verbs used [... and] whether these verbs were correctly formed and appropriately used” (p.152).

Other perspectives consider a learner’s interlanguage as a set of multiple components as well as a result of specific conditions. For example, Ellis (2003) states that

> [t]he extent to which [learners] call on their lexical knowledge, their rule-based system, or a combination of the two is influenced by a number of task design features [...] and implementational conditions [...], and in turn influences what aspect of performance, i.e., fluency, complexity, and accuracy is prioritised (Ellis 2003 p.287).

Language complexity, accuracy and fluency

The notions of complexity, accuracy and fluency (CAF) in a learner’s language use and acquisition have been investigated with a large diversity of measurement tools (Pallotti 2009). While fluency is mostly concerned with “the capacity to produce speech at normal rate and without interruption”, accuracy deals with error frequency, and complexity determines the advanced level of the language (Skehan 2009 p.510). According to Housen and Kuiken (2009), complexity and accuracy are “linked to the current state of the learner’s [...] interlanguage knowledge” (p.462). Accuracy can be estimated with the help of the obligatory occasion analysis (described above), but this measure fails to take into account the vocabulary aspect. The accuracy measure can be complemented by the percentage of error-free clauses or error per hundred words (Ellis and Barkhuizen 2005 p.139).

Complexity relates to “the range of forms that surface in language production and the degree of sophistication of such forms” (Ortega 2003 p.492). Lu (2009) lists several syntactic complexity measures, such as length of production (e.g., mean length of sentences) and sentence complexity (e.g., mean number of clauses per sentence). Length-based measures used to determine the syntactic complexity of learners' language have been demonstrated to explain insufficiently how the complexification has been achieved (Wolfe-Quintero et al. 1998 p.15). Alternately, Norris and Ortega (2009) demonstrate that a length-based measurement is "a good measure of overall or general complexity because such a broader measure might be able to capture large-scale or long-term variation that would be missed by finer-grained, more specific metrics" (p.568). According to Housen and Kuiken (2009), methodological issues with regard to complexity and accuracy analysis may be addressed with a refinement of the measurement instruments, as well as a better definition of the grammatical features to be analysed (p.470). To overcome some of these methodological issues, Norris and Ortega (2009) propose to avoid redundancy in terms of metrics, and suggest an approach that includes three dimensions: the overall complexity (e.g., mean length of T-unit[15]), the complexity by subordination (e.g., mean number of clauses per T-unit), and the complexity by phrasal elaboration (e.g., mean length of clauses) (p.574). Alternatively, Schulze, Wood and Pokorny (forthcoming) propose to analyse textual complexity with "*one* integrated and balanced measure of complexity", which includes four different vectors based on lexical and grammatical complexity and sophistication measures (not paginated, para 2, emphasis in original).

Manually processing analysis of lexical and syntactic complexity, or counting by hand incorrect and correct occurrences to determine their accuracy can become very tedious and time consuming, especially with regard to large collection of learners' texts. Lu (2008) claims that there is a "need for computational tools that can automate the process

15. According to Hunt (1965), a T-unit can contain one single clause (like simple sentences) or multiple clauses (like complex sentences). He added that he could not call these units "minimal sentences" as the word sentence already carried different meanings. Therefore, he proposed a fresh neutral term for these minimal terminable units: "T-unit" (Hunt 1965 p.37).

with high accuracy" (p.153). To address this need, he released Synlex[16], a web-based L2 lexical and syntactic complexity analyser for the English language during summer 2010. To this researcher's knowledge, there is no equivalent for the French language. Any measurement of a learner's language accuracy and complexity starts with specific information contained in a corpus such as part-of-speech. The next section outlines the challenge in constituting an annotated corpus.

4.1.4. Creation and annotation of learners' corpora

While data collection was, hitherto, a time consuming and tedious work, computers have eased the task in terms of time and workload. Learner corpora, and more specifically interlanguage learner corpora, are defined as

> electronic collections of authentic foreign or second language data. They differ from the data types commonly used by second language acquisition (SLA) and foreign language teaching (FLT) researchers in two major respects: (a) they are computerised [...] and (b) they are big and therefore constitute a much more reliable basis to describe and model learner language than has ever been available before (Granger 2003 p.465).

Creating an annotated corpus to investigate interlanguage variability firstly implies consideration of the type of data to be collected, and then enriching the corpus with appropriate annotations.

Data elicitation and collection

Eliciting data to create an interlanguage learner corpus is not without concerns. For, example, Tarone (1988) highlights a few factors, such as physical settings or time

16. Available online at http://calper.la.psu.edu/tools

allocated for the task. These factors all influence variability in learners' interlanguage, and consequently it is recommended to consider these parameters carefully before data collection (Tarone 1988 pp.120-121). In addition, Ellis and Barkhuizen (2005) recommend making a detailed description of the conditions on how the data is collected, since "the nature of the sample that is collected may influence the nature and distribution of the errors observed" (p.57). The task of data elicitation is generally accompanied by the decision of either collecting the information longitudinally or cross-sectionally. According to Meisel, Clahsen, and Pienemann (1981), a "longitudinal study gives a description of linguistic performance at several points of time, analysing changes which occur between point t_{t+1}", whereas cross-sectional studies "analyse the linguistic performance of a number of L2 learners at a certain point of time" (pp.111-113).

Moreover, the investigation of the learners' interlanguage should imply the analysis of learners' "meaningful performance", where meaningful is representative of real life authentic situations (Selinker 1974 p.32). Tarone (1994) points out that, since the concept of interlanguage was first introduced, researchers were encouraged to use as data "utterances produced by second-language learners when they were trying to communicate meaning in the target language" (p.1717). Similarly, Ellis and Barkhuizen (2005) claim that the learner's use of a second language in natural contexts is considered as "the best data for examining their underlying knowledge systems (competence)" (p.9). One reason behind building and using a corpus of learners' authentic texts is to draw information from it, which implies the annotation and encoding of linguistic features contained within the texts. As noted by Leech (1997),

> [c]orpora are useful only if we can extract knowledge or information from them. The fact is that to extract information from a corpus, we often have to begin by building information in – that is, by adding annotations (Leech 1997 p.4).

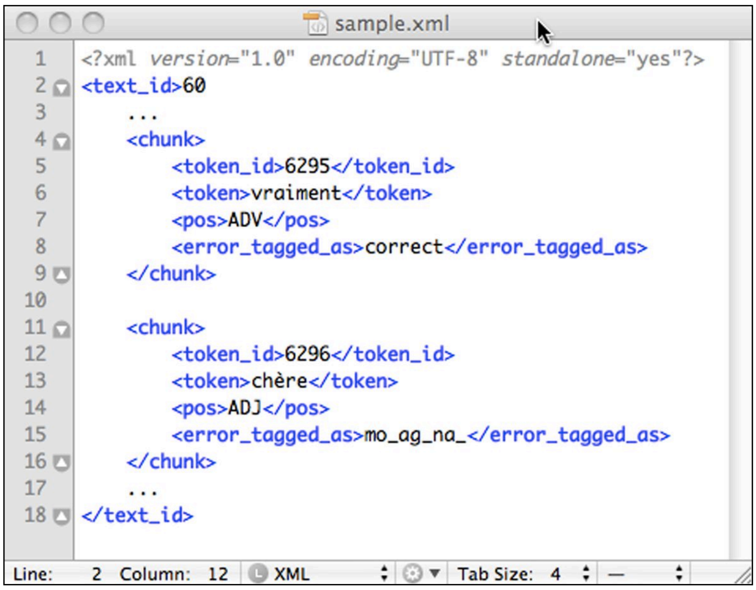
Annotated corpora not only enable machines as well as humans to retrieve information, but also provide a "clear and objective record of analysis that is open to scrutiny and criticism" (McEnery et al. 2006 p.30).

Corpus annotation and encoding

While annotation denotes the practice of making "explicit the linguistic features of a text" (Bowker and Pearson 2002 p.83), encoding refers to "the actual symbolic representation", which is used to describe the linguistic information added to the corpus (Leech 2005, not paginated, para 5). Leech (2005) further points out that the symbolic representation used should be simple and unambiguous. Some researchers, such as Meurers and Wunsch (2010) emphasise the standardised aspect of the encoding process, as standardised forms "can uniformly be accessed in all resources" (p.3).

To annotate and encode any corpus, the main tendency is to use metadata, i.e., "data about data" (Burnard 2005 p.30). Extensible Markup Language (XML) is defined as a "metalanguage, that is, a language used to describe other languages" (TEI 2009 p.xxxi). An example of an annotated corpus in XML format is illustrated in Figure 4.1 below.

Figure 4.1. Excerpt of an annotated corpus in XML format



While the annotations may be added manually, computerised data facilitates semi- and fully automatic annotations. For example, semi-automatic annotation tools such as an error editor "enable researchers to introduce linguistic annotation interactively" (Granger 2002 p.16).

Error annotation

Language learner corpora are particularly "useful when they are error-tagged" (Granger 2003 p.465). Learners' incorrect forms can affect a whole sentence or a single word, and even inside a word, a single morpheme. Ferris (2010) points out that the literature does not clearly indicate whether research designs should only be concerned with treatable error types or should also consider untreatable categories. According to Ferris (2010), treatable refers to errors easy to describe, i.e., "errors that can be addressed by reminders of a clear and succinct rule", and untreatable designates errors that interfere with meaning at semantic and syntactic levels, i.e., "lexical issues or sentence structure" (p.196). One approach in error categorisation is to tag the incorrect forms according to "their nature (grammatical, lexical, etc.)" rather than "their source (interlingual, intralingual, etc.)", as the error interpretation would be less subjective (Granger 2002 p.19). For example, Granger's (2003) error classification is organised per domain and subdomain. Domains referring to general levels (e.g., morphology, grammar, lexis or syntax) are divided in subdomains or error categories such as word order or gender (Granger 2003 pp.467-468).

One issue in identifying and annotating learners' incorrect forms automatically refers to the use of spelling and grammar checkers not targeting L2 errors; they generally fail to detect learners' ill-formed words. Granger (2003) reports a success rate of 25% to 35% when correcting second language learners' written texts with non-targeted L2 spellcheckers (p.466). Several studies tested the efficacy of checkers and demonstrated that learners' interlanguage could not be automatically corrected. For example, Rimrott and Heift (2005) analyse incorrect forms performed by learners of German to determine the type of errors that can be successfully detected. Their findings show that not only were 20% of the incorrect forms "accidental typographical mistakes"[17], but they also demonstrate that generic spellcheckers – designed to correct these error types in texts written by native speakers – could only detect half of them (Rimrott and Heift 2005 p.17). To overcome such limitations, researchers have developed tools targeted at nonnative writers. For example, L'Haire's (2007) FipsOrtho is a spell-checker designed for learners of the French language focusing on error types such as phonetic spelling, morphology or missing apostrophe (p.138). The author states that, although the findings are promising, incorrect forms such as lexical and agreement error types were not systematically detected and had to be manually annotated. Ndiaye and Vandeventer-Faltin (2003), who have also developed a spell-checker tailored to learners of French, add that further empirical analysis of the spell-checker with corpora of learners' authentic language are necessary to improve the error detection results. As a consequence, learner corpora are generally annotated manually with the assistance of an error editor, and then analysed through a computer-aided error analysis system.

The concept of computer-aided error annotation and analysis was introduced by Dagneaux, Denness and Granger (1998) to generate lists and counts of specific error types "in their context and alongside instances of non-errors" (p.173). For example, the French Interlanguage Database (FRIDA) corpus, based on Dagneaux, Denness and Granger's (1998) system, not only includes error tags, but also probable corrections to these incorrect forms (Granger 2003). According to Granger (2003), an error annotation should allow for "informativeness and manageability", "flexibility", "reusability", and "consistency" (p.467). An *informative* error annotation refers to enough information on

17. According to Rimrott and Heift (2005), performance mistakes are incorrect forms that contain single or multiple letter violations (addition, omission, substitution, and transposition) and word boundary violations (p.23).

the learner's incorrect forms, but not too much in order for the information to be still *manageable*. *Reusable* has to be understood in the sense that the error categories should be general enough to be reused by other languages. *Flexible* designates that the error-tag process should be easily added or deleted by the annotators, and finally *consistent* in the sense that error tagging principles should be applied consistently by all annotators. Leech (2005) further points out that annotations should also be separable from the raw corpus without loosing any information about the original writings.

Part-of-speech tagging

Part-of-speech tagging is also considered as a valuable source of information in annotated corpora (Leech 2005). The part-of-speech tagging process is defined as the action of assigning the most reliable part-of-speech tag (e.g., noun, verb, adjective, adverb, ...) to not only each word in a learner's text, but also each punctuation mark, symbol or abbreviation, in other words, each token in the input (Jurafsky and Martin 2000). One issue with the tagging procedure however is that one token may be eligible for more than one part-of-speech tag. Several techniques exist to disambiguate a word class in order to apply the part-of-speech tag adapted to the situation. Since Greene and Rubin's (1971[18]) TAGGIT, which was initially used to tag the Brown Corpus[19], taggers in terms of accuracy have sufficiently improved to be worthy of attention. In general terms, taggers can be classified into four main groups depending on the technique they use to disambiguate linguistic units: linguistic, machine-learning, statistical, and hybrid models (Màrquez et al. 2000). For example, within the *linguistic approach*, linguists manually formalise the grammar knowledge as a set of descriptive rules or constraints that characterise the syntactic properties of a given corpus. This set of logical rules, capable

18. Mentioned in Abney 1997 p.119.

19. The Brown Corpus is a collection of American English texts compiled and printed by Kučera and Francis at the Brown University in 1961 (Pradhan et al. 2008).

of generating an infinite number of possible sentences along with their structural description, is then used to assign appropriate description to each sentence in the corpus (e.g., Greene and Rubin 1971, Karlsson 1995). Within a *machine-learning approach*, the aim is "to automatically induce a model for some domain, given some data from the domain" (Jurafsky and Martin 2000 p.118). There are a few taggers for the English language employing machine-learning techniques, such as the Markov model (e.g., Cutting et al. 1992), transformation-based error-driven learning (e.g., Brill 1995), or decision trees (e.g., Black et al. 1992, Màrquez et al. 2000, Schmid 1994). *Statistical methods* have focused on probabilities using stochastic algorithms to "pick the most-likely tag" for each word in a text (Jurafsky and Martin 2000 p.303). Examples of stochastic taggers are illustrated in the work of Marshall (1983) or Faaß et al. (2009). Finally, other taggers, called *hybrids* use a combination of two or more techniques, such as rule-based and machine learning methods in the case of Tlili-Guiassa's (2006) tagger used for the Arabic language.

In general terms, taggers achieve promising scores in terms of tagging accuracy. A typical success rate evolves above 95% (Granger 2002 p.18). However, as noted by Díaz-Negrillo, Meurers, Valera, and Wunsch (2010), "[w]hen applied to a new genre of text, taggers perform worse than they applied to the genre they were developed for" (p.4). Since taggers are generally trained with texts written by native speakers, tagging a text performed by a second language learner might be problematic. Indeed, the tagger will encounter potentially ambiguous words due to misspellings, foreign words, proper nouns, invented words or other unknown words. As stated by Granger (2003), the particularity of learner language is that it reflects "a high rate of misuse, i.e., orthographic, lexical, and grammatical errors" (p.466). As a result, "[i]n order to make up for this degradation of performance, post-correction steps are usually added to modify tags that are systematically wrongly assigned" (Díaz-Negrillo et al. 2010 p.4).

Thus far, the above discussion not only underscores the importance of electronic collections of authentic language learners' data, but also focuses on the reliability aspect of the annotation process to analyse learners' language. While interlanguage learner corpora must be enriched with annotations to provide the opportunities for a wide range of analysis, the up-to-date techniques with regard to reliability are promising but still far from efficient. The error annotation process for instance, cannot be fully automated, since generic and spell- and grammar checkers not targeting L2 errors perform rather poorly in terms of error detection. As a result, error annotation is often a computer-assisted process. Furthermore, texts written by second or foreign language learners may be difficult to annotate with regard to part-of-speech tags, as language learners' incorrect forms may considerably diminish the tagging accuracy of any part-of-speech tagger. To overcome these issues, post editing is often necessary.

The difficulties of applying reliable tags in an interlanguage learner corpus are critical, and must be taken into serious consideration. Indeed, all findings derived from the analysis of an annotated learner corpus may be compromised if the annotations are not reliable.

4.2. Implications for this thesis

The current research proposes dynamic assessment as a means to distinguish between competence-dependent errors and performance-related mistakes when approaching the study of interlanguage variability. While dynamic assessment normally focuses on emergent abilities, the aim of this study is to reveal the learners' actual development, and to then analyse certain aspects of it. The previous discussion not only emphasised the need for a computerised learner corpus annotated with reliable tags in order to analyse and measure the learners' actual knowledge, but also outlined the void in dynamic assessment with regard to reliable psychometric measurement instruments.

4.2.1. Research questions

As stated in the opening chapter, this thesis seeks to investigate learners' interlanguage variability from a new angle. In particular, it suggests the use of sociocultural theory and dynamic assessment to differentiate between competence-dependent errors and performance-related mistakes, and addresses the following research questions:

> **Question 1.** Can the learners' zone of proximal development, i.e., their actual and potential development, and the distance in-between, be represented and observed so that errors and mistakes can be distinguished?
>
> **Question 2.** Are interlanguage competence and performance variable across students, time, and text types?
>
> **Question 3.** Does the modeling of the learners' zone of proximal development provide further insight into their interlanguage development?

Thus far, it has been maintained that the analysis of learners' interlanguage competence could not be undertaken without differentiating between errors and mistakes. While interventionist and interactionist approaches to dynamic assessment provide a suitable framework to distinguish between competence-dependent errors and performance-related mistakes, they do not offer appropriate methods to answer all aforementioned questions. Consequently, there are specific methodological requirements that need to be taken into account.

> **Requirement 1.** *Creation of the learners' zone of proximal development.* The context and tools enabling interactions between a novice and a more expert person when dynamically assessing written texts need to be considered, so that the zone between learners' actual and potential development can be created. More specifically, applications that will assist

in both the provision of multi-layered levels of assistance and the collection of learners' answers to this assistance, need to be designed and implemented in order to fulfil the conditions necessary to realise a learner's ZPD**.**

Requirement 2. *Representation of the zone of proximal development.* Once the zone of proximal development exists for each incorrect form encountered in learners' texts, the ZPD needs to be concretely represented, so that a difference between competence-dependent errors and performance-related mistakes can be operated. In particular, the collection, annotation and storage of information about assistance required to correct incorrect forms will enable (a) the representation of the zone of proximal development, (b) the identification of both errors and mistakes, and (c) the observation of the learners' interlanguage competence as well as their potential abilities to perform.

The first requirement represents a challenge in terms of design and implementation of tools to create the context necessary to realise the learners' zone of proximal development. The second requirement suggests the creation of an annotated corpus which will contain information on each incorrect form such as error type or the amount of assistance required for learners to propose correct replacements. A few methodological choices, such as error classification, levels of assistance, and corpus annotation needed to be addressed in order to conduct this doctoral research.

4.2.2. Error classification

The error taxonomy used for the current study is adapted from (a) Mackey, Gass and McDonough (2000) with regard to morphosyntactic and lexical errors, (b) L'Haire (2007) with reference to syntax and punctuation, and (c) Granger (2003) with respect to grammar and typography. The present classification includes the following main categories: selection, syntax, morphosyntax, misspelling and typography. The selection error category indicates an incorrect selection in terms of lexemes or grammatical aspects, such as the use of an inappropriate tense or incorrect word class. The syntactic error category relates to the syntax of the sentence, including omission or addition of words. The morphosyntactic error category is characterised by an incorrect, misplaced or missing morpheme in a semantically correct word. The misspelling error category refers to lexemes that are semantically correct but incorrectly spelt. Finally, the typographic error category is linked to typography and punctuation.

The error classification is hierarchically arranged from main error categories to error sub-categories and error types. Table 4.1 below lists all error types used in this research along with their category distribution and their corresponding codes. The first column refers to the error main category, e.g., *[mo]* stands for *morphosyntactic error category*. The second column corresponds to sub-categories of the error main category, for example, *[ag]* refers to *agreement*. As for the last column, it designates the error type, a sub-sub-category of the main category, e.g., *[na]* denotes a *noun adjective* error type. As a result, an *[mo_ag_na_]* error sequence refers to a morphosyntactic error as the main category, an agreement error as a sub-category, and more specifically, as a noun adjective agreement error as specific error type.

In the morphosyntactic category, a distinction is made between *determinant noun agreement (mo_ag_dn)* and *gender agreement (mo_ag_ge)* to identify an incorrect agreement with respect to number from an incorrect agreement in terms of gender. For example, in the segment **cette exemple est pertinente (this example is relevant - correct form: cet exemple est pertinent),* the mismatch between the noun and its determinant is identified as an incorrect gender agreement, not as an incorrect determinant noun agreement, as the student used the feminine form for the determinant as well as the adjective, which signals that he or she did not know the gender of the noun *exemple* (masculine).

Table 4.1. Error category, sub-category, and type

Error category		**Error sub-category**		**Error type**	
[se]	selection	[vo]	vocabulary	[mu]	nonstandard word or expression
				[lw]	inappropriate choice
				[re]	inappropriate register
				[sn]	not understandable
		[gr]	grammar	[cw]	inappropriate connection word choice
				[rv]	reflexive verb
				[ar]	incorrect article type
				[pr]	incorrect preposition
				[mo]	incorrect mood
				[te]	incorrect tense
				[vo]	incorrect voice
				[cl]	incorrect word class
[sy]	syntactic			[ad]	word addition
				[om]	word omission
				[wo]	word order or incorrect syntax
				[un]	syntax not understandable
[mo]	morphosyntactic	[ag]	agreement	[dn]	determinant noun agreement
				[ge]	gender agreement
				[na]	noun adjective agreement
				[pa]	pronoun antecedent agreement
				[pp]	past participle agreement
				[sv]	subject verb agreement
		[fo]	form	[pl]	incorrect plural form
				[co]	incorrect conjugation form
				[wf]	incorrect word formation
[sp]	spelling			[hy]	hyphen
				[ms]	misspelling
				[ac]	incorrect or missing accent
[ty]	typography			[ca]	capitalisation
				[ab]	abbreviation
				[mp]	missing punctuation
				[wp]	inappropriate punctuation
				[es]	missing or inappropriate space

Additionally, a differentiation is made between *subject verb agreement (mo_ag_sv)* and incorrect *conjugation form (mo_fo_co)*. The former identifies a mismatch between the subject and the verb in terms of number such as: *les profs *pose des questions (the teachers *asks questions - correct form: posent)*, in which the number for the subject (plural) does not correspond to the one for the verb (singular). The latter, i.e., *conjugation form (mo_fo_co)*, recognises an incorrect formation in terms of verb conjugation: e.g., *vous *faisez* (you do) instead of *vous faites*. The error type *mo_fo_wf*, on the other hand, signals an incorrect word formation not related to verbs, such as in the case of adverb formation. In French, adverbs are usually formed by adding the suffixe *ent* to the feminine form of the adjective: e.g., *naturelle + ment = naturellement (natural - naturally)*. The incorrect word *naturelment (naturally) reflects an incorrect word formation as the rule to build the adverb is not respected.

In the typography category, the *missing or inappropriate space (ty_es)* relates to the rules governing the usage of spaces (blank areas) occurring, for instance, between words or near punctuation marks. While no space would be placed between a word and an exclamation mark in English (e.g., Hello!), one space would be required in French (e.g., Salut !).

The classification (Table 4.1 above) is used as a guideline to label as many ill-formed words as possible in a learner's performance. It is important to note that the aim of the error annotation is for research purposes only and not for teaching, otherwise, fewer incorrect forms would have been marked. As suggested by Van der Linden (1993), students should not be overwhelmed with too many corrections. However, it was decided to assign one unique error type per word or group of words, thus defining a level of precedence over types of errors in the event of ambiguity – from high to low – selection, syntactic, morphosyntactic, misspelling, and typographic error categories. Precedence was originally given to meaning and communication over form to follow the teacher's priorities in terms of pedagogical objectives.

For example, the following sentence illustrates the case of an ambiguous error type in respect of its source:

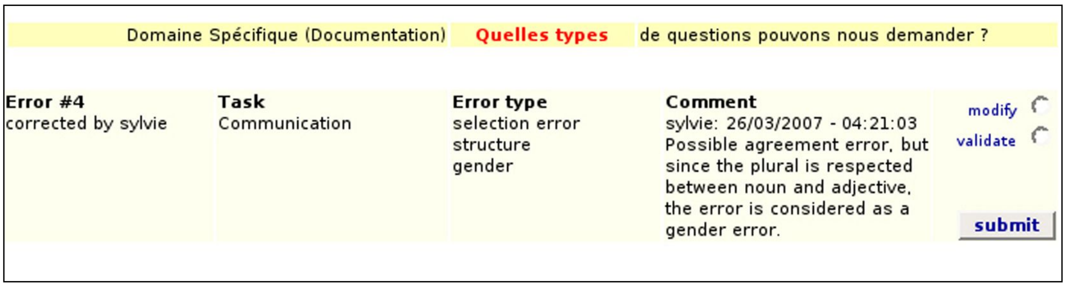
[1]. ...le projet pour moi *été une bonne expérience...

(...the project was a good experience...)

The error type could be a word omission (sy_om_, the auxiliary is missing). However, since the learner who produced this sentence consistently wrote the imperfect tense incorrectly **été* instead of *était*, the other possible error type could be an incorrect tense (se_gr_te_, the compound past tense is more appropriate in context). Since selection (se_) errors take precedence over syntactic (sy_) errors, the error type applied for this particular incorrect form was an incorrect selection in terms of tense (se_gr_te_).

According to Granger (2009), "all researchers highlight the need for detailed documentation on the system and tests to assess inter- and intra-rater reliability" (p.24). The question of inter- and intra-rater reliability in encoding incorrect forms was not directly addressed within this research. However, as I was the only error corrector, my ability to mark texts written by learners of French was checked by a third party, an experienced French lecturer. Indeed, this doctoral research gave me the opportunity to correct students' texts for the first time. The aim was to determine whether I was qualified for the task. To do so, an application to help validate or invalidate my initial corrections was developed especially for this purpose (Figure 4.2). A set of 700 incorrect words, along with their context and error type, was randomly selected. As the experienced teacher agreed on 96.76% of my markings, it was decided that I was qualified to correct all participants' texts. The remaining 3.24% were mostly due to ambiguous errors, which led to the elaboration of the error hierarchy as mentioned above.

Figure 4.2. Application to check this researcher's error annotation



4.2.3. Levels of assistance

The dynamic assessment procedure adopted does not reflect a specific format as discussed in Section 3.2 on page 54. The assistance is given in graded layers on each incorrect form, which tends to point towards the cake format. However, the sandwich structure could also be considered since the assistance is provided all at once. Furthermore, the fact that the mediation is tailored to each individual's needs, and negotiable in the sense that learners can choose to ignore or accept to read their feedback when correcting themselves implies an interactionist approach.

The levels of assistance provided to learners are adapted from Aljaafreh and Lantolf's (1994) regulatory scale (presented in Section 3.1.4 on page 51, and listed in Appendix B on page 282). While these authors specify twelve levels of assistance, only four of them were considered. The main reason for this is that most of the conditions could not be applied to the current research. For example, the level *"tutor indicates that something may be wrong in the segment"* could not be re-used given the fact that each incorrect form is pointed out. The levels of assistance are thus as follows:

> **Level 1.** The highlighted incorrect word, or group of words indicates that something is wrong, no further information is provided.
>
> **Level 2.** The error type is provided for each highlighted incorrect word or group of words, narrowing down the nature of the incorrect form.

Level 3. Detailed explanations about the nature of the incorrect form is given to help the learner find the correct answer, yet without providing it.

Level 4. The correct form is provided.

4.2.4. Error annotation and part-of-speech tagging

The error annotation was first performed using Markin*TM* [20], an existing tool used to assist language teachers in the marking of students' texts. The advantage of using this computer-assisted feedback software was twofold: firstly, the set of error tags developed with the error classification was easily adaptable; and secondly, the output could be exported in *text only* format, which enabled the error types and additional annotations to be computerised and easily processed. However, the main drawback with this application was that the character encoding used, i.e., Latin 1 (charset=iso-8859-1), did not correspond to the character encoding used by the other applications in operation within this study (charset=utf-8). This was a serious issue with regard to character corruption encoding. Furthermore, error annotating texts with Markin offered the option to develop a three-layered level of assistance, whereas a four level was preferred for better placement of the incorrect form within the learner's zone of proximal development. As the existing solution was not satisfying, the design and implementation of an error editor using UTF-8 encoding and a four-level regulatory scale were a prerequisite to conduct the current research.

Additionally, part-of-speech tags were applied for each token of the input through the means of TreeTagger[21], a tool designed to assign the most likely part-of-speech tag along with its lemma[22] information to each token in a text. TreeTagger was selected as it is

20. http://www.cict.co.uk/software/markin/index.htm

21. http://www.ims.uni-stuttgart.de/projekte/corplex/TreeTagger/

22. Dictionary entry form.

language independent[23] and is available online for downloading. The software has been developed and implemented within the "textual corpora and tools for their exploration" project[24] at the University of Stuttgart in Germany. TreeTagger was trained on native English input and tested on data from the Penn Treebank corpus (Marcus et al. 1993). According to Schmid (1994), it achieved a tagging accuracy of 96.34%. However, after the first attempt to tag learners' texts, it was noticed that the tagger performance was conditional on the correctness of the input. As a consequence, a method to improve the tagger accuracy was proposed and applied.

4.3. Research design

During the academic year of 2007/2008 and the first semester of 2008/2009, I was given the responsibility of tutoring classes of students of French at Dublin City University, Ireland. Prior to my first tutoring experience, I started the design and the development of prototypes (Phase 1). My first involvement in a class gave me the opportunity to develop these tools further and to evaluate them in concrete educational settings (Phase 2). Time was then devoted to collecting enough data for interlanguage variability analysis and interpretation (Phase3).

As illustrated in Table 4.2 below, the first phase of the research, from January 2007 to August 2008 was primarily concerned with the development, implementation and beta-testing of tools required to provide learners with correction and assistance. The learners' texts to be error-corrected were collected, then error-annotated and part-of-speech tagged. A pilot study conducted in laboratory settings provided the opportunity to highlight areas that needed improvement. The identification of limitations at this early stage enabled the refinement of not only the tools used to collect the data, but also the

23. TreeTagger has been used to tag German, English, French, Italian, Dutch, Spanish, Bulgarian, Russian, Greek, Portuguese, Chinese and old French texts.

24. TC project: http://www.ims.uni-stuttgart.de/projekte/tc/

corpus annotation with regard to reliability (Phase 2, from September 2008 to June 2009). In particular, the error taxonomy and encoding, the regulatory scale, and the method to improve the tagger accuracy were refined in accordance to the limitations identified during the first phase. After finalising the implementation of all the tools required to conduct the study, tests were conducted in a semi-controlled educational setting. These tests outlined weaknesses in the research design, such as the fact that learners were not systematically reading the assistance provided to them. In order to ensure that the learners' behaviour with regard to assistance use could not impact the error-mistake distinction, it was decided to monitor learners' actions with regard to feedback access. Finally, the opening of the online access to all participants enabled the data to be collected. Then, the data was from September 2009 (Phase 3) onwards annotated, encoded, analysed and finally interpreted.

Table 4.2. Timeline of the research study

Phase 1 **pre-research study** (Jan. 07 to Aug. 08)	• Development of a taxonomy of learners' error types; • Design and implementation of a pilot application to help learners correct themselves with and without assistance; • Identification of a suitable part-of-speech tagger, and improvement of its tagging accuracy; • Evaluation of the pilot application in laboratory settings.
Phase 2 **research study** (Sep. 08 to Jun. 09)	• Design and implementation of an error-annotation tool; • Revision of the tool designed to assist learners' correct themselves with and without assistance; • Implementation of a monitoring system to check learners' access to feedback; • Data collection.
Phase 3 **data analysis** (from Sep. 09)	• Constitution of the electronic corpus; • Annotation and encoding of the corpus; • Analysis of the learners' errors and mistakes; • Interpretation in terms of interlanguage variability.

4.3.1. Context

This study was conducted in a French language class at Dublin City University over the first semester of the 2008/2009 academic calendar. Students who consented to participate in this study had the opportunity to continue the experiment over the second semester of the same academic year. Intermediate learners of French were recruited as they were the only students available. This was not considered as a limitation as intermediate learners usually exhibit larger numbers of error types than beginners or advanced learners. This class was composed of eighty-nine learners of French in their first year, and included twelve sessions of three hours a week. The module, whose aim was to further develop students' skills, was methodically structured around action-reflection cycles in order to stimulate and enhance the development of learner autonomy and language use. The curriculum of this course covered a project-based approach. The main project was to describe one typical week at the university from the student's point of view. The task involved the creation of a wiki. Students engaging in a group were able to contribute or modify the content of the wiki by adding texts, images and sound files.

4.3.2. Participants

From the eighty-nine learners of French enrolled in the language course, nineteen students consented to participate in this research study. Participants were provided with as much information as possible about the aim and context of the study. The procedure for collecting and processing data were described in plain English during an oral presentation and in writing with the plain language statement before the study commenced (Appendix A.1 p.280). Participants were asked to read and sign a consent form, either a paper version (Appendix A.2 p.281) or an electronic version accessible via

Moodle[25], the virtual learning environment currently used at the university. Participation was strictly voluntary and it was ensured that the participants were made aware that if they decided not to take part in the study, it would not at all affect their grades. In line with Dublin City University Research Ethics Committee, all students who participated in this research study were volunteers and signed the electronic consent form, which allows for the collection, analysis, use, and quotation of the data.

In order to protect the privacy and confidentiality of all the participants, the record of demographic data was kept to a minimum. Yet, name and email address were considered to be necessary information given the fact that a correspondence between the participants and this researcher was initiated. Naming people by their first name was judged to be more polite than referring to them with numbers when directly addressing them. However, when presenting or publishing extracts of their texts to highlight certain aspects of them, the participant's anonymity is preserved. Numbers (student #1, student #2, ...) are assigned rather than using learners' real first names or even aliases. Furthermore, online access to the different instruments are, as one might expect, username and password protected. As indicated by Prior, Rogerson and Fairweather (2002), "security is not only a technical issue", if weak, the security issue may easily turn into an ethical concern (p.29).

Almost all participants had English as their first language, except for one student whose native language was Portuguese. All were studying French at an intermediate level and came from different programmes such as applied languages or international business. There were three males and sixteen females[26]. The participants' age ranged from eighteen

25. http://moodle.org

26. No distinction between male and female was done within this research, as gender is considered from a performative point of view by this researcher, which is defined as something one "does" as opposed to something one "has" (Cameron 2005 p.491).

to twenty-four, except for one student who was fifty years old. Table 4.3 below shows an overview of all participants with regard to gender and native language.

Table 4.3. Overview of the participants

Student id	Male/Female	Native language
#2	F	English
#3	F	English
#4	F	English
#5	F	English
#6	F	English
#7	F	English
#8	F	English
#9	F	Portuguese
#10	F	English
#11	M	English
#12	F	English
#14	M	English
#15	F	English
#16	F	English
#17	F	English
#19	M	English
#20	F	English
#21	F	English
#22	F	English

The reader will have observed that there is no student #1, #13, #18. The missing numbers refer to students who sent texts for submission, but never signed the consent form even after the study was completed. As students could change their mind without being asked about their reasons, their refusal of signing the consent form was acknowledged as their volition of leaving the study. Therefore, the data corresponding to these three students (demographic and texts) was removed from the study and the database, as specified in the low-risk project document sent to the research ethics committee. From the nineteen students who signed a consent form, fourteen of them sent at least one text for correction during the first semester, the remaining five participants never submitted a text. While three of the fourteen active participants continued to submit their written texts for correction during the second semester, the others chose not to pursue after the first semester. It is worth mentioning that most of the participants were Irish students and that Irish students through their educational system, either in primary or secondary schools, do not receive any specific tutoring in linguistic knowledge. To minimise this shortcoming, reference grammar specifically written for the needs of English speakers learning French was used when providing them with corrections. Additionally, as mentioned by Fowley (2011), "it is easy to forget that in Ireland for example, many of the young people whom we routinely call digital natives have only lived online since 2005 or 2006" (p.20). This implies that some students may feel uncomfortable with computers, or more broadly speaking with technology. This statement is equally relevant and must be kept in mind when providing students with the tools necessary to help them participate in this research.

4.3.3. Writing tasks

Participants submitted their texts for correction on a voluntarily basis. Learners had the option to send their written documents via Moodle or as an attachment by e-mail directly to myself. Upon receipt of some feedback, students were asked to correct themselves with and without assistance, and to re-submit their texts along with alternatives to each incorrect form.

Participants could send any text types, including those they had to submit to their lecturer as summative assessment. The various text types submitted were a discussion *forum*, a reflective account of their learning experience named *bilan*, an *email* correspondence, an essay resulting from various written task-based activities named *écrire*, a few *wiki* texts, a *survey*, and a transcription of speeches labelled *parler*. They are described as follows:

> A **forum** discussion was set up on Moodle to encourage learners to express their challenges. Only sparse posts were submitted. Student #2 was one of

them. She posted seven posts during the first month of the academic semester, mostly to keep the members of her team aware of her progress. She soon gave up as no one responded.

Bilan was an obligatory assessment due at the end of the course and graded by their teacher (not this researcher). Learners were asked to reflect on their language learning and outcomes. They were asked to report their experiences, either positive or negative, and how they overcame their problems in doing the various tasks.

Email refers to a private correspondence between student #2 and this researcher throughout the whole academic year, for which she asked to be corrected. Emails are mostly short informal texts.

Ecrire results from written task-based activities learners had to perform. The document was formally corrected and graded by their teacher as continuous assessment. A task completed, for instance, was to organise a trip abroad in a French speaking country.

Wiki reflects a collaborative work between students. The project was to describe one typical week at the university from the student's point of view. The wiki naturally included texts written as a result of learners' collaboration, implying amongst other things a peer-review. However, the texts submitted for correction within this study were texts students had to write individually, which were then added to their wiki group in their personal section.

Survey is a one page document which was submitted by student #9 only. The document helped her collect information from French students in Ireland about differences between the Irish and French educational system.

Parler refers to a written transcript of a speech learners had to record and include into their wiki pages. From an informal conversation with student #11, it is known that he was principally interested in syntax, which was the reason for which he submitted this text type.

The distribution of text types received per learner is shown in Table 4.4 below.

Table 4.4. Distribution of text types received per participant

<table>
<tr><th>Student</th><th>Forum</th><th>Bilan</th><th>Email</th><th>Ecrire</th><th>Wiki</th><th>Survey</th><th>Parler</th></tr>
<tr><td>#2</td><td bgcolor="#c0c0c0"></td><td bgcolor="#c0c0c0"></td><td bgcolor="#c0c0c0"></td><td bgcolor="#c0c0c0"></td><td bgcolor="#c0c0c0"></td><td></td><td bgcolor="#c0c0c0"></td></tr>
<tr><td>#5</td><td></td><td></td><td></td><td></td><td bgcolor="#c0c0c0"></td><td></td><td></td></tr>
<tr><td>#6</td><td></td><td bgcolor="#c0c0c0"></td><td></td><td></td><td bgcolor="#c0c0c0"></td><td></td><td></td></tr>
<tr><td>#7</td><td></td><td bgcolor="#c0c0c0"></td><td></td><td></td><td bgcolor="#c0c0c0"></td><td></td><td></td></tr>
<tr><td>#8</td><td></td><td></td><td></td><td></td><td bgcolor="#c0c0c0"></td><td></td><td></td></tr>
<tr><td>#9</td><td></td><td></td><td></td><td></td><td bgcolor="#c0c0c0"></td><td bgcolor="#c0c0c0"></td><td></td></tr>
<tr><td>#10</td><td></td><td bgcolor="#c0c0c0"></td><td></td><td></td><td bgcolor="#c0c0c0"></td><td></td><td></td></tr>
<tr><td>#11</td><td></td><td></td><td></td><td></td><td bgcolor="#c0c0c0"></td><td></td><td bgcolor="#c0c0c0"></td></tr>
<tr><td>#12</td><td></td><td></td><td></td><td></td><td bgcolor="#c0c0c0"></td><td></td><td></td></tr>
<tr><td>#14</td><td></td><td></td><td></td><td></td><td bgcolor="#c0c0c0"></td><td></td><td></td></tr>
<tr><td>#15</td><td></td><td></td><td></td><td></td><td bgcolor="#c0c0c0"></td><td></td><td></td></tr>
<tr><td>#16</td><td></td><td bgcolor="#c0c0c0"></td><td></td><td></td><td bgcolor="#c0c0c0"></td><td></td><td></td></tr>
<tr><td>#17</td><td></td><td bgcolor="#c0c0c0"></td><td></td><td bgcolor="#c0c0c0"></td><td bgcolor="#c0c0c0"></td><td></td><td></td></tr>
<tr><td>#19</td><td></td><td></td><td></td><td></td><td bgcolor="#c0c0c0"></td><td></td><td></td></tr>
</table>

While most of the students submitted wiki and bilan texts, student #2 for instance sent for correction text types such as forum, bilan, email, écrire, wiki, and parler.

4.4. Summary and conclusion

The chapter started by exploring the methodological challenges arising from the use of a dynamic assessment framework to distinguish between errors and mistakes in a learner's text. More specifically, it identified complex situations with regard to (a) the operationalisation of the model proposed in this thesis, (b) the design of tools to apply interventions in written context, (c) the identification of psychometric instruments to measure learners' current knowledge, and (d) the creation and annotation of a corpus.

After restating the research questions, the second section considered the methodological requirements that needed to be addressed before the study started. Two major requirements were identified: the creation of the learners' zone of proximal development, and the representation of this zone. In addition, it addressed the various decisions taken with regard to error classification, levels of assistance, and corpus annotation.

After showing the different phases of the research design, the third section outlined the educational settings in which the study took place. It also presented the participants and discussed ethics. Additionally, it outlined the fact that the fourteen participants who submitted at least one text for correction were learners of French at an intermediate level, and that the study ran over a period of two semesters during the 2008/2009 academic calendar. While the majority of the participants submitted texts during the first semester, three students chose to continue the experiment during the second period. The next chapter will detail the instruments to collect, annotate, and analyse the data.

Chapter 5. Construction of instruments and empirical analyses

The previous chapter considered the challenges as well as the methodological choices necessary to conduct this research. In particular, it highlighted the issue in terms of reliability when annotating learner corpora, and more specifically, part-of-speech tagging texts written by language learners. This chapter presents the instruments and methods used to overcome the aforementioned issues in order to answer the research questions outlined in the introductory chapter.

The first section opens with the methodology used to minimise incorrect part-of-speech tags so that reliability can be improved. An overview of the tool's architecture along with a description of the students' and researcher's tools that were designed and implemented to process, collect, store, annotate, and analyse the data, are then presented. Following a summary of the data collected as well as the procedures used to distinguish between errors and mistakes, and the psychometric measures adopted to represent the learners' knowledge, the synchronic and diachronic analyses are outlined.

5.1. Improving the part-of-speech tagging process

Since errors in annotations will result in larger errors in the analysis of grammatical and lexical forms with regard to psychometric instruments, all components in a given corpus must be annotated and encoded with consistent tags. A major challenge in part-of-speech tagging is to identify the unknown and ambiguous words, since unambiguous words can be automatically and accurately processed (Mihalcea 2003). The method proposed to improve the tagging accuracy is a three-step approach, which (a) identifies the lemmas that are unknown for the tagger, (b) checks the part-of-speech tags automatically applied against an extended set of rules based on recurrent tagging errors, and (c) cross-references the part-of-speech tags with the error-annotated corpus.

5.1.1. Identifying unknown lemmas

If the word to be labelled was present in the training corpus (the one used to train the tagger), the lemma is identified, otherwise it is tagged as unknown. In the case of an unknown lemma, the tagging tool computes the most likely part-of-speech tag by extrapolating the best option (Daelemans and Van der Bosch 2005). While an exhaustive training corpus is not feasible, finding ambiguous unknown words in a text not seen before by the tagger is unavoidable. The first improvement therefore aims to reduce the amount of unknown lemmas that are, by definition, not recognised by and ambiguous for the tagger.

Figure 5.1 shows a sample output obtained from TreeTagger after running the command shell *"Apres ca j'ai une pause" (After this, I take a break)*, which contains two incorrect words: **Apres (After)* and **ca (this)*.

Figure 5.1. TreeTagger output

The sentence is tokenised and each token is annotated with its part-of-speech and lemma information. Some of the lemmas are unknown to the tagger. For example, the incorrect word **Apres (After)* is tagged as a proper noun (NAM) and the unrecognised lemma is marked as *<unknown>*. One cause identified for not recognising certain lemmas is due to capital letters. TreeTagger does not apply the proper part-of-speech tag for a few tokens, either placed at the beginning or in the middle of a sentence. For example, Table 5.1 shows the part-of-speech and lemma information provided by the tagger after running the input *Cette semaine, en Allemagne (This week, in Germany)*, and the output obtained after transposing all uppercases into lowercases.

Table 5.1. Lowercase transposition

Input:	**Cette**	**semaine**	**,**	**en**	**Allemagne**
original tagging	*NAM <unknown>	NOM <semaine>	PUN <,>	PRP <en>	NAM <Allemagne>
after lowercase conversion	DET:DEM <ce>	NOM <semaine>	PUN <,>	PRP <en>	*NOM <unknown>

The first token of the sentence *Cette (this)*, starting with an appropriate uppercase letter, is labelled as a proper noun (NAM). The lemma information <*unknown*> indicates that the tagger guessed the part-of-speech tag, which is an incorrect guess in this case. On the other hand, the tag applied to the last token *Allemagne (Germany)* – starting with an appropriate uppercase letter – is recognised by the tagger and properly tagged. The last row shows the results obtained with the same tokens transposed into lowercases. While the lemma information and part-of-speech tag provided for the token *cette* are now correct (<ce>, DET:DEM, demonstrative determiner), the token *Allemagne,* converted into lowercases, is no longer recognised as a proper noun. The tagger estimated its part of speech as being most likely a noun (NOM). Converting the original input into lowercases provides accurate pos-of-speech and lemma information for most tokens. However, transposing the entire text into lowercases may be at the cost of valuable information and thus lowering the tagger performance. The algorithm, used to reduce the amount of unidentified words, checks each unknown lemma in the original input and looks for lemma information in the lowercase version. The following record, displayed in Figure 5.2, lists a sample of tokens whose unknown lemmas were identified with the lowercase conversion method.

Figure 5.2. Examples of unknown lemmas resolved with the lowercase transposition

token	lemma_1 pos-tagging from treetagger	lemma_2 after uc to lc transposition	pos_1 pos-tagging from treetagger	pos_2 after uc to lc transposition	man_pos_tagged_1 human #1
J'	unknown	je	NAM	PRO:PER	PRO:PER
J'	unknown	je	ADJ	PRO:PER	PRO:PER
Il	unknown	il	NOM	PRO:PER	PRO:PER
Fermer	unknown	fermer	NAM	VER:infi	VER:infi
Mettez	unknown	mettre	NAM	VER:pres	VER:impe
Calme	unknown	calme	NAM	ADJ	ADJ
Une	unknown	un	ABR	DET:ART	DET:ART
Logement	unknown	logement	NAM	NOM	NOM
Clubs	unknown	club	NAM	NOM	NOM

The column *token* enumerates the tokens in their original format. Columns *lemma_1* and *pos_1* show the lemma information and the part-of-speech tags obtained after running TreeTagger on the original input. Columns *lemma_2* and *pos_2* list the lemma information and the part-of-speech tags retrieved from the text when converted into small letters. The last column *man_pos_tagged_1* refers to the hand-annotated tags, which are used as an indication to check how the tagger performs with new procedures. For example, the first token *J'-(I)*, correctly spelt and not recognised by TreeTagger, was tagged as a proper noun (NAM), which is incorrect. The lowercase conversion allows TreeTagger to recognise the token. As a result, the lemma information *je* is provided and the most likely part-of-speech tag is estimated, i.e., personal pronoun (PRO:PER), which corresponds to the hand-annotated tag.

If the lemma information is still unavailable, the token is checked against an additional word bank (created for the purpose of this study) of unambiguous abbreviations, adjectives, adverbs, proper nouns, compound nouns or English words commonly used in the participants' written documents. This word bank is a list of learners' *unambiguous* tokens with lemma and part-of-speech information that were not initially recognised by TreeTagger (see Figure 5.3 for a few examples).

Figure 5.3. Extract of the word bank

token	pos	lemma
music	NOM	musique
ca	PRO:DEM	ça
equipement	NOM	équipement
intruction	NOM	instruction
informathique	NOM	informatique
comfortable	ADJ	confortable

If the token with an unknown lemma has an entry in the word bank, the part-of-speech and lemma information is automatically updated. Otherwise, if the token has no entry in the word bank, it may be manually added into the database, see Figure 5.4.

Figure 5.4. Insertion of new entries into the word bank

intruction	NOM	NOM	unknown instruction
vetements	NOM	NOM	unknown vêtement
comfortable	ADJ	ADJ	unknown confortable
peut-etre	VER:infi	ADV	unknown peut-être
parceque	VER:subp	KON	unknown parce que
francais	VER:impf	ADJ	unknown
			submit

The token and its part-of-speech as originally estimated by TreeTagger are given in columns 1 and 2. The user (the person who feeds TreeTagger with the input) chooses in the drop down list the part-of-speech tag that corresponds to the entry. If the user provides the lemma in column 4, the token is considered as unambiguous and stored into the database. Otherwise, the application interprets an empty lemma box as the token being ambiguous, which in turn implies that the token can be labelled with more than one tag.

When a token is incorrect and ambiguous, for instance **francais (French - adjective or noun)*, the part-of-speech tag can be disambiguated with the surrounding context of the token. The next step shows how a set of additional hand-written rules based on the observation of inconsistencies in the original tagging format can improve the tag accuracy.

5.1.2. Rule-based part-of-speech tagging

The rule-based part-of-speech tag review is based on the BRILL tagging method, which first assigns default initial tags, i.e., the most frequent tag for that word, and then applies replacement rules based on the analysis of lexicon, morphological and contextual rules (Brill 1995). The method adopted here is to assign the most probable tag to each token with TreeTagger, and then to apply replacement rules depending on prior and/or subsequent part-of-speech tags, lemma information and tokens themselves.

TreeTagger performance is checked against a set of 45 additional rules written by this researcher in order to exclude consistent incorrect part-of-speech tags. The replacement rules are described under the form of phonological rules including regular expressions. Regular expressions[27] are used to show the search pattern, whereas phonological rules are generally outlined as

Rule #1. A->B/X__Y, where A becomes B between X and Y;

Rule #2. A->B/X__, where A becomes B after X;

Rule #3. A->B/__Y, where A becomes B before Y.

The whole set of additional rules written to improve the tagger accuracy is classified into three categories. The first one includes rules that refine the original tag set of TreeTagger by adding and subtracting part-of-speech tags. The second category contains rules that look at specific part-of-speech tags and update them depending on prior and/or

---

27. Appendix C.1 on page 283 provides a brief overview of the symbols and syntax commonly used in regular expressions.

subsequent tokens, part-of-speech and lemma information. The third group relates to rules that capture issues involving specific tokens (see Appendix C.2 on page 283 for the complete set of rules, Appendix C.3 on page 285 for TreeTagger tag set, and Appendix C.4 on page 285 for the updated tag set). Table 5.2 below lists some of the rules for the three categories.

Table 5.2. Classification of the hand-written rules and examples

<table>
<tr><th>Classification</th><th>Examples</th></tr>
<tr><td>Category 1:
Adding or subtracting part-of-speech tags.</td><td>Rule that refines the numeral (NUM) part-of-speech tag into adjective (ADJ), numeral adjective (ADJ:NUM):
Rule #2. ^(NUM)$->^(ADJ)$/__
if token=[^(.*\d.*)] AND if token=(i(è|e|é)re*s*)$
ELSE
Rule #4. ^(NUM)$->^(ADJ:NUM)$/__
if token=[^(.*\d.*)] AND if token!=^(c|m|x|i|l)$</td></tr>
<tr><td>Category 2:
Updating existing tags depending on context.</td><td>Rule that replaces adjective (ADJ) tags by noun (NOM) tags when situated between a determiner (DET) and a verb (VER):
Rule #19. ^(ADJ)$->^(NOM)$/^(DET)__^(VER)</td></tr>
<tr><td>Category 3:
Capturing issues involving specific tokens.</td><td>Rule that changes the part-of-speech of the token pour, always misidentified as a conjunction (KON), into a noun (NOM) or a preposition (PRP) depending on context:
Rule #17. token->^(NOM)$/^(DET)__
if token=^(pour)$ AND if pos!=^(PRP)$
ELSE
Rule #20. token->^(PRP)$/__
if token=^(pour)$ AND if pos!=^(PRP)$</td></tr>
</table>

For example, refining the tag set by adding part-of-speech tags is illustrated with the case of the part-of-speech NUM (numeral) in the first category. TreeTagger classifies any cardinal or ordinal adjectives written either in figures or in letters as a numeral without any distinction. Whereas a cardinal numeral adjective indicates that the adjective is a definite number, an ordinal adjective refers to an ordering relation of elements. The following example (Line 2) shows that the cardinal number *trois (three)*, the ordinal number *première (first)* and the cardinal number *500* written in figures are all tagged as numeral, which is correct but not specific enough, especially with a view to analysing possible agreement errors:

[2]. J'ai écrit trois(NUM) fois la première(NUM) partie de 500(NUM) mots.

(I wrote three times the first part of 500 words.)

Distinguishing between these instances of numeral tags is relevant since (a) an ordinal adjective can be compared to an adjective that qualifies a noun in terms of number and gender agreement, and (b) a cardinal adjective written in letters is most of the time invariable but susceptible to be incorrectly written. Rules #2 and #4 below help make a difference between numeral in digit (NUM), numeral adjective (ADJ:NUM), and other numeral considered as an adjective (ADJ):

```
Rule #2. ^(NUM)$->^(ADJ)$/__
if token=[^(.*\d.*)] AND if token=(i(è|e|é)re*s*)$

ELSE

Rule #4. ^(NUM)$->^(ADJ:NUM)$/__
if token=[^(.*\d.*)] AND if token!=^(c|m|x|i|l)$
```

If the token, if tagged as a numeral (NUM), is not a digit and ends with a combination of i+e+è+re+s (e.g., prem-ier, prem-ière, prem-iers, ..., (*first*)), the part-of-speech tag is replaced by the adjective tag (ADJ). Otherwise, the token, if not a digit nor a Roman numeral, is tagged as a numeral adjective (ADJ:NUM). The previous example displayed in Line 2 above is now tagged as:

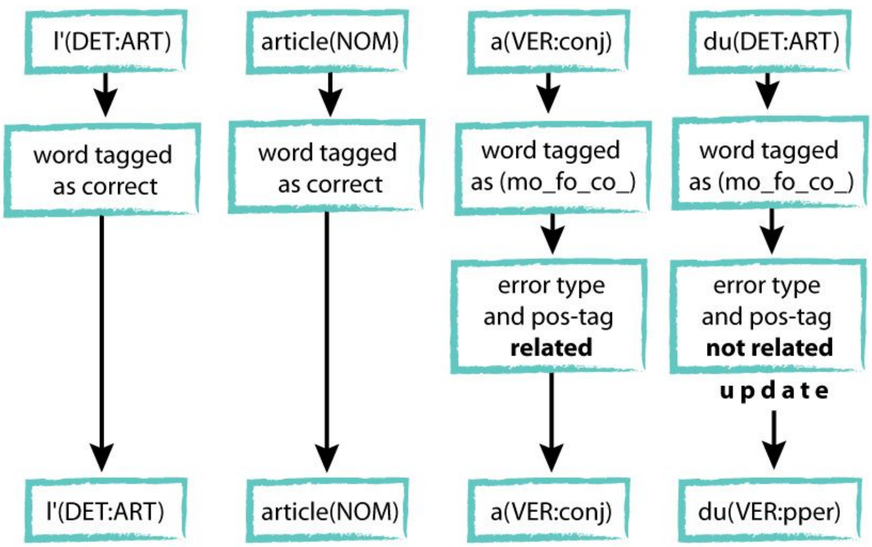
[3]. j'ai écrit trois(ADJ:NUM) fois la première(ADJ) partie de 500(NUM) mots.

After checking each token against the set of rules based on recurrent tagging errors, the next step involves the cross-reference of the part-of-speech tagged corpus with the error annotated corpus.

5.1.3. Cross-reference between error annotation and part-of-speech tag

If the token was also error annotated, the part-of-speech tag applied to this token is compared to the error type tag. In the event of incompatible tags, the part-of-speech tag is updated in accord with the error type. For example, Figure 5.5 illustrates the procedure with the following sentence: *l'article *a du*... (the paper had to...)*. The tokens *l'* and *article*, not marked as incorrect by the corrector, are not further processed by the system; the initial tagging is kept and stored as is. Token *a*, on the other hand, is enclosed with an error tag. Since the error type (mo_fo_co_ - incorrect conjugation form) and the part-of-speech tag (VER:conj - conjugated verb) are related, the initial tagging is not modified. However, the part-of-speech and error type of the last token *du* are not related. The token is tagged as a determiner (DET:ART) whereas the error type involves an incorrect formation in terms of a verb (mo_fo_co_). The initial tag is therefore updated in accord with the error type. In this case, the system replaced the part-of-speech tag DET:ART (determiner) by VER:pper (past participle).

Figure 5.5. Cross-reference between part-of-speech tagged and error-annotated corpora



If the error type involves an incorrect conjugation form (mo_fo_co_), and the part-of-speech is different from a conjugated verb (VER:conj) or a past participle (VER:pper) tag, the following rules are applied.

Rule #9:

[^(VER:pper|VER:conj)]->^(VER:pper)$/^(VER:conj)$ ^(ADV)?$__

ELSE Rule #7:

[^(VER:pper|VER:conj)]->^(VER:conj)$/[^(VER:conj)]__

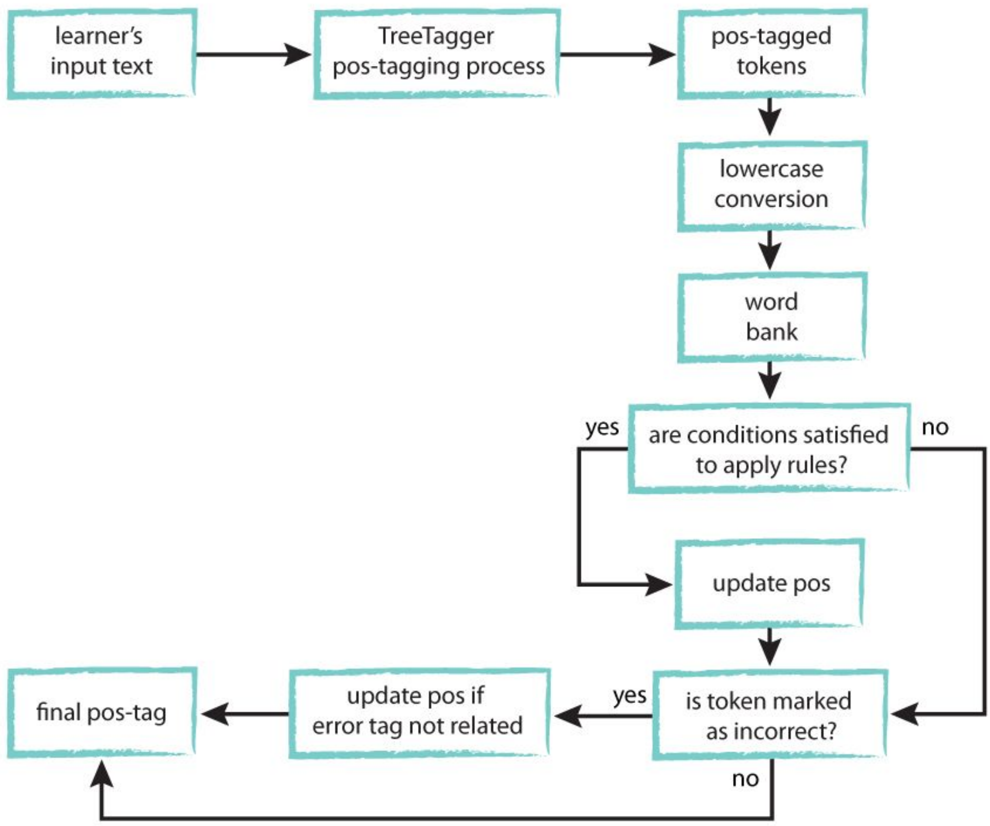
ELSE Rule #27:

[^(VER:pper|VER:conj)]->^(VER:pper)$/^(VER:inf)$__;

Rule #9 checks first whether the current tag is preceded by a conjugated verb (VER:conj) followed by zero or one adverb (ADV). If it does then the tag should be a past participle (VER:pper). Also, if the current tag does not follow any conjugated verb (rule #7), then the tag itself should be a conjugated verb (VER:conj). Otherwise, if the tag follows an infinitive verb (rule #27), then it is probable that the token is a past participle (VER:pper). A full account of the rules written for the cross-referencing between part-of-speech tagged and error-annotated corpora is given in Appendix C.5 on page 286.

The whole process to improve the part-of-speech tagging accuracy can be summarised as follows (Figure 5.6 below). Firstly, the learner's raw input is fed into TreeTagger, then information on unknown lemmas is searched within the lowercase version and the word bank. The next step implies checking each token against a set of rules intended to reduce consistent tagging errors. If the rules can be applied, the current tag is replaced, otherwise the tag is left unchanged. Finally, each token if marked as incorrect in the error-annotated corpus is (a) compared to the error tag, (b) updated if they are not related, and (c) stored along with the part-of-speech, lemma information as well as the error type.

Figure 5.6. Process to improve the tagging accuracy



5.1.4. *Inter-rater agreement analysis*

To evaluate the tagging process accuracy of TreeTagger, the output produced before and after improvements is compared to a hand-annotated tagged corpus. The baseline strategy simply consisted of hand-tagging a sample, token by token using the exact same part-of-speech tag set that is given with TreeTagger. The *Petit Larousse* and *Petit Robert* dictionaries[28] were used as references when hesitating between different tags. A set of 10,108 tokens corresponding to 14 students' written texts, in which 8,424 words and 1,440 lexical and grammatical errors were encountered, was manually tagged by this researcher. The mean error rate per hundred words is equal to 17.11%. To compare

28. Le Petit Larousse Illustré 2007 and Le Petit Robert 2009 are both French monolingual dictionaries.

TreeTagger with the gold standard data, the percentage of agreement excluding chance between human and machine is estimated through the means of Krippendorff's (1970) alpha and Cohen's (1960) kappa coefficients, – two measures using distinct ways of computing the expected agreement (Di Eugenio and Glass 2004) –, so that the scores give a partial view of the tagging accuracy. In addition, the tagging accuracy is evaluated with recall, precision and F-measure, three widely used metrics in part-of-speech tagging to estimate the effectiveness of a system (Voutilainen 2003). Each coefficient, i.e., alpha, kappa, recall, precision and F-measure is further detailed in Appendices C.6, C.7 and C.8 starting on page 287). To ensure high quality in tagging, this researcher's reliability was also checked against the tagging of another human; a former teacher in France, who also has a passion for French grammar. The corpus used is much more modest than the one mentioned above; it includes 2,022 tokens in total.

5.1.5. Human versus human

The inter-human rater reliability was measured with the kappa coefficient. The percentage of agreement excluding chance between both raters is equal to 95.7% (Table 5.3).

Table 5.3. Inter-rater reliability: human versus human

		Value	Asymp. Std. Error[a]	Approx. T[b]	Approx. Sig.
Measure of Agreement	Kappa	.957	.005	146.202	.000
N of Valid Cases		2004			

a. Not assuming the null hypothesis.
b. Using the asymptotic standard error assuming the null hypothesis.

This result was considered as slightly low. However, the cause of the disagreement can be easily explained. The tag set included a part-of-speech tag named *POS* in the event raters could not apply a specific tag to a particular token. For example, the token **Que-ce-que (what is it)*, not properly tokenised, was not manually assigned with a part-of-speech tag as it is impossible to identify the right one. Rater #2 applied the unspecified tag (POS) much more frequently than rater #1, in particular when the tokens were English words. Although there were a few English words in the corpus, they had to be tagged as any other tokens. Figure 5.7 shows a few examples of those tokens tagged with the *POS* tag.

Figure 5.7. Unspecified part-of-speech tag

token_id	token	man_pos_tagged_1 human #1	man_pos_tagged_2 human #2
1175	hui	ADV	POS
1448	Que-ce-que	POS	POS
9625	Home	NOM	POS
9626	and	KON	POS
9627	Away	ADV	POS

Given the fact that most discrepancies between both raters occurred because of English words and the frequent use of the *POS* tag by rater #2, the result obtained with the kappa coefficient was deemed to be satisfactory.

5.1.6. Machine versus human

Table 5.4 below summarises all results obtained from the various agreement analysis discussed above.

Table 5.4. Summarising the results of the agreement analysis

	Alpha	**Kappa**	**Precision**	**Recall**	**F-measure**
Before improvement	89.62%	89.60%	78.43%	77.67%	78.03%
After improvement	97.60%	97.60%	97.21%	96.01%	96.61%

The table clearly demonstrates an improvement. For example, Krippendorff's alpha coefficient rose from 89.62% ($\alpha = .8962$) to 97.60% ($\alpha = .9760$). Craggs and Wood (2005) point out that "it is impossible to prescribe a scale against which all coding schemes can be judged", which suggests that the threshold at which agreement measures are considered sufficient is to be determined in a subjective way (p.293). The results obtained are certainly encouraging. With the alpha, kappa, and precision results placed above the upper bound at which taggers tend to not surpass (Vanroose 2001), i.e., 97%, it was decided that the part-of-speech tagging was reliable enough to draw conclusions from the annotated corpus.

5.2. Overview of the tools' architecture

The majority of tools implemented for the purpose of this research are installed on a private server[29]. They are accessible via the Internet, and are username and password protected. The other tools are installed on a local server, which is accessible from a local area network (LAN), which are also username and password protected. All tools used by the participants are web-based applications. They do not require the installation of any extensions, additional software or allocated disk space on a learner's personal computer (client), as everything is stored on the server. The tools run in most existing browsers. The applications were implemented using PHP[30], a scripting language generally used in the development of dynamic web pages, and MySQL[31], an open source database which runs on several platforms including Debian GNU/Linux[32], the open source operating system used to run the part-of-speech tagger. Additionally, TinyMCE[33], a Javascript WYSIWYG editor, was used when implementing the error corrector application.

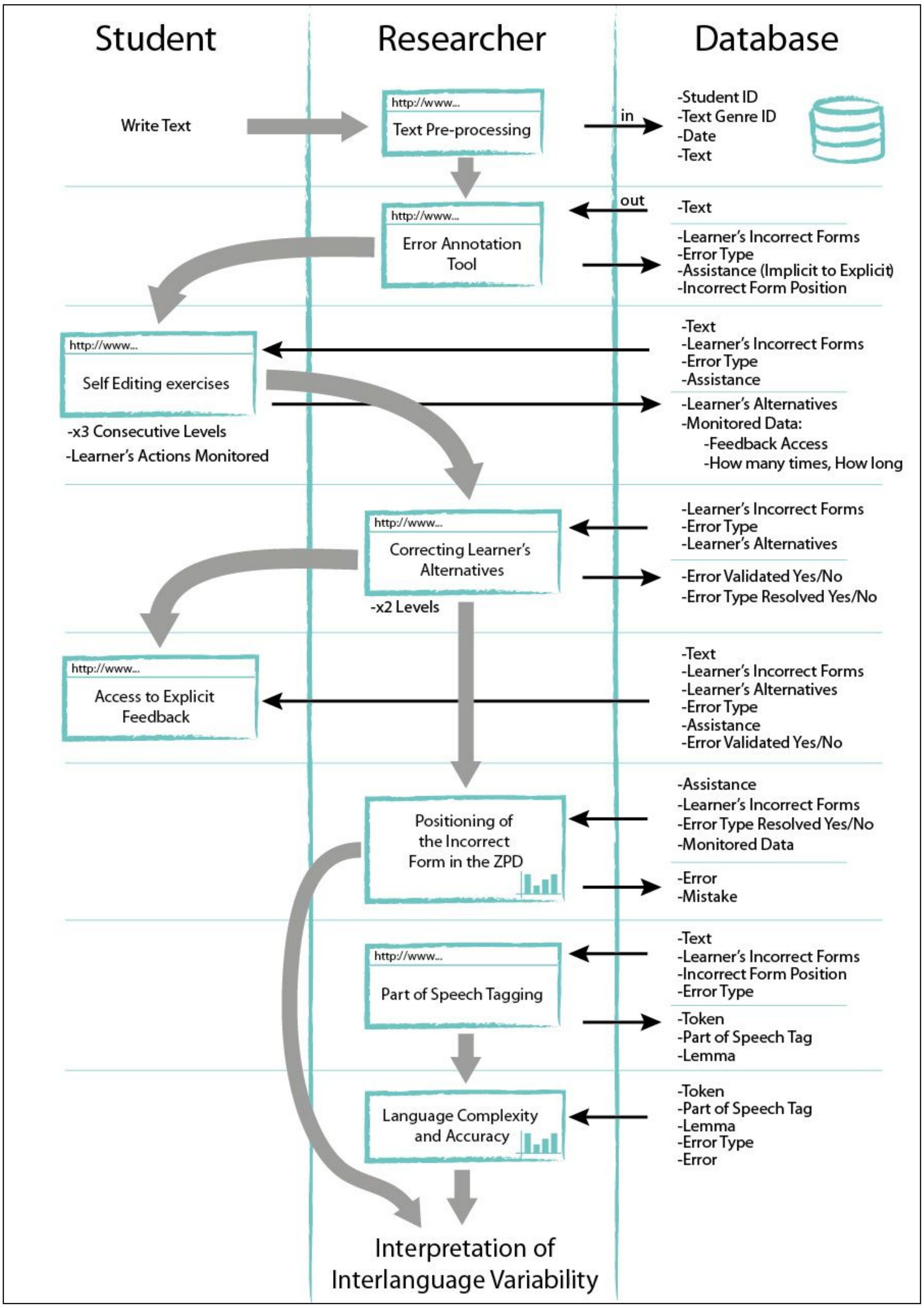
The inter-dependencies between the various tools in action and the data movement going in and out of the database to collect and analyse the data are illustrated in Figure 5.8 below. The large arrows in grey identify the next step in the process, whereas the thin

---

29. http://www.servage.net
30. http://www.php.net
31. http://www.mysql.com
32. http://www.debian.org
33. http://tinymce.moxiecode.com

arrows in black indicate the information going in the database while running the current tool, or going out of the database to enable the current tool to run.

Figure 5.8. Architecture



The process starts with the participants' texts received in an electronic format. The texts are pre-processed before being entered into the database. Once the pre-processed input is stored, it is then re-used in the error-annotation tool, so that each incorrect form can be flagged. After error-annotating the text, information such as incorrect form, error type, and level of assistance is sent into the database, which is then exported to the learners' self-editing tool. After participants proposed alternatives to their incorrect forms, all learners' replacements were checked for correctness. While information regarding level of assistance and feedback access was then used to position learners' incorrect forms in their zone of proximal development, learners' texts were also part-of-speech tagged to help represent their language accuracy and complexity. A sample text in the full life-cycle of the system is illustrated in Appendix D.1 on page 290.

5.3. Students' tools

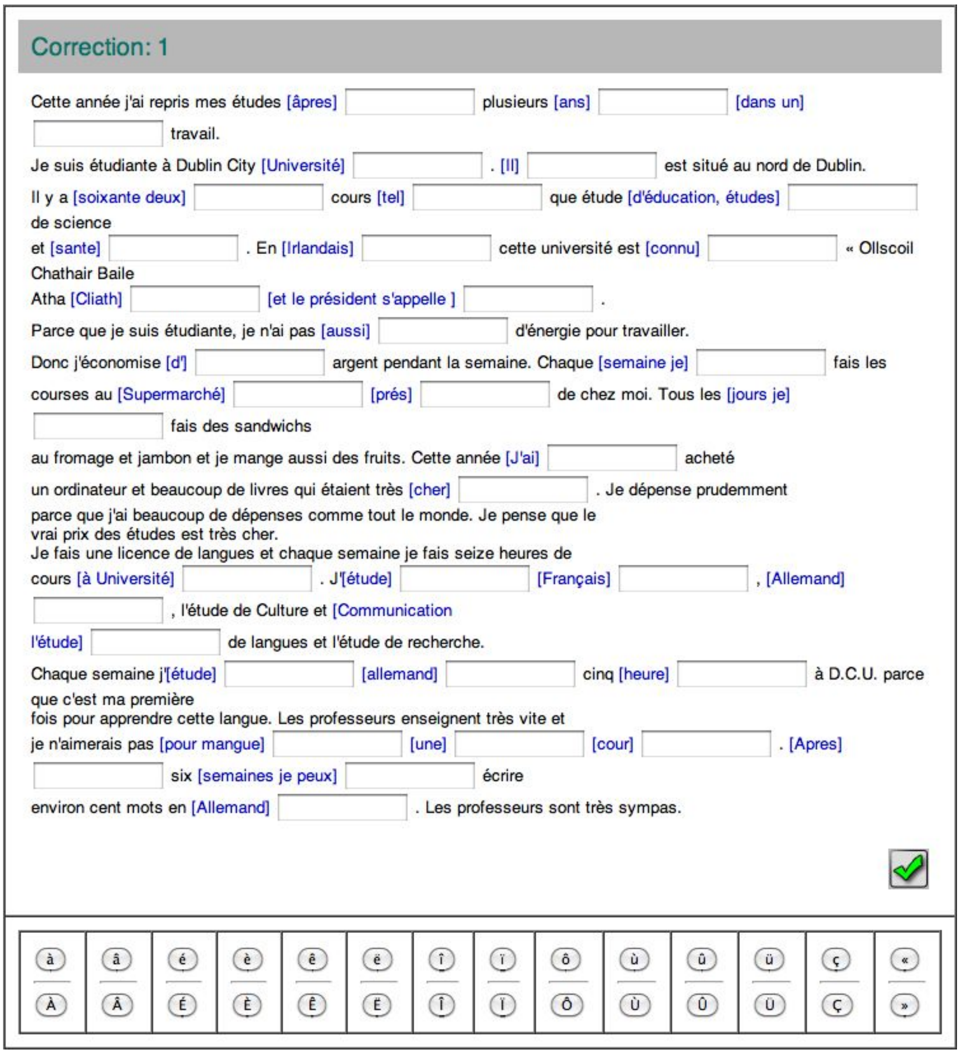
The tools assisted in (a) the provision of learners' alternatives to their incorrect forms, and (b) the possibility for them to consult all of their data after their task was completed.

5.3.1. Self-editing exercises

Learners were invited to review each of their documents via a web-based application and to provide alternatives to all parts of the text that were marked as incorrect by the corrector. While it is commonly assumed that if learners are able to produce a correct alternative with implicit assistance, which designates that they already have a certain control over their subject (Lantolf 2009 p.360), it could also be the case that producing a valid answer after the first level of assistance may result from a learner's guess. For this reason, learners were asked to correct themselves three times consecutively and independently of what they may have proposed at previous levels of assistance. The very first self-editing exercise was performed under supervision mainly in the event of technical issues, and also for final oral recommendations before starting. Learners were advised not to use external help such as the assistance of a third party or dictionaries, the aim being to estimate what they knew at a particular point in time with controlled feedback. Although it was not suggested to spend too much time on each incorrect form, learners were allowed to take as long as they needed to complete the task. The other self-editing exercises in the case of multiple participations were performed independently without supervision. Learners knew however that they could contact me in the event of unexpected technical issues.

Figure 5.9 below shows a screen capture of the web-based application used by participants to self-correct their texts at level one.

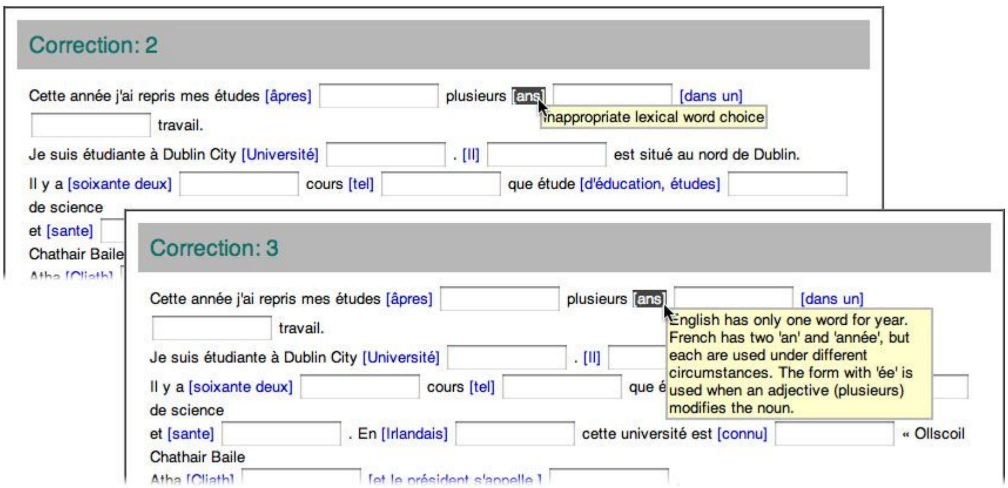
Figure 5.9. Self-editing exercise: first level



For the first level, learners had to correct all sequences marked as incorrect without any indication about the error type. Incorrect forms (blue on screen) are enclosed within squared brackets to make them more salient to learners. An input text box is displayed beside each incorrect form on the right hand side. If students knew or believed they knew correct replacements, they entered alternatives in the spaces provided for that purpose, otherwise they were left blank. After completing the first level of correction, learners had to continue the task by clicking on the check button. They could access the second level of the exercise upon submitting the alternatives proposed at level one.

For the second level of correction, the learners were provided with comments on the error type such as "inappropriate lexical word choice" (Figure 5.10). The comments were visible with a mouse roll-over action on the incorrect forms. Participants were aware that they had to correct all highlighted incorrect forms once again, independently of the fact that they were correct or not at the previous level. For the third and final level of correction, students were provided with detailed information about their incorrect forms. Once more, learners were asked to supply an alternative in accordance with the additional comments provided. Figure 5.10 illustrates the self-editing exercises at levels two and three.

Figure 5.10. Self-editing exercises: second and third levels



5.3.2. Data access

After correcting themselves, students were thanked for their participation and were offered the possibility to access not only correct alternatives to their ill-formed words that were proposed by the corrector (Figure 5.11), but also all the alternatives they wrote at any of the levels (Figure 5.12).

Figure 5.11. Correct alternative provided: fourth level

Figure 5.12. Information made available to students after correcting themselves

Correction 1

incorrect form	feedback	my correction	teacher's correction	Alternative validated?
[âpres]	-	àpres	après	pending
[ans]	-	année	années	pending
[dans un]	-	au	de travail	pending

Figure 5.12 above reports the alternatives a learner entered at level one. The *pending* observation under the title *Alternative validated?* indicates that the learner's propositions were not approved as correct yet. The learner may have proposed something different than the corrector, and may wish to know whether this was correct or not. Once the corrector verified the learner's replacement for acceptability, participants were able to check whether their alternatives were appropriate or not. Figure 5.13 shows that the alternatives provided at level one were either validated or invalidated. For instance, the student's alternative *au* (fourth row) suggested as a replacement to the incorrect sequence **dans un* is different from the assessor's correction but nevertheless appropriate in context.

Figure 5.13. Validation report of alternatives provided at level 1

incorrect form	feedback	my correction	teacher's correction	Alternative validated?
[âpres]	-	àpres	après	✗
[ans]	-	année	années	✗
[dans un]	-	au	de travail	✓
[Université]	-	Universitée	étudiante à l'université de DCU / étudiante à	✗

5.4. Researcher’s tools

The researcher’s tools assisted in (a) the pre-processing of learners’ texts received for correction, (b) the error-annotation of these texts, (c) the validation of learners’ alternatives proposed as replacements to their incorrect forms, and (d) the monitoring of learners’ behaviours with regard to feedback access.

5.4.1. Pre-processing the learners’ texts

Before being stored into the database, the documents submitted by the participants were converted into *text only* format, encoded in UTF-8, and pre-checked mostly in terms of special characters not recognised by the tagger tokenising module. For example, curly single quotes (’) were replaced by straight single quotes (') to avoid inconsistencies in part-of-speech tagging such as *c’est (this is)* tagged as [*c’est (adjective)*] instead of [*c' (demonstrative pronoun) est (verb)*]. The form used to submit each text into the database (Figure 5.14) requests the submission date, the student name, and the text type in which the document is best suited (e.g., forum).

Figure 5.14. Database insertion of new texts

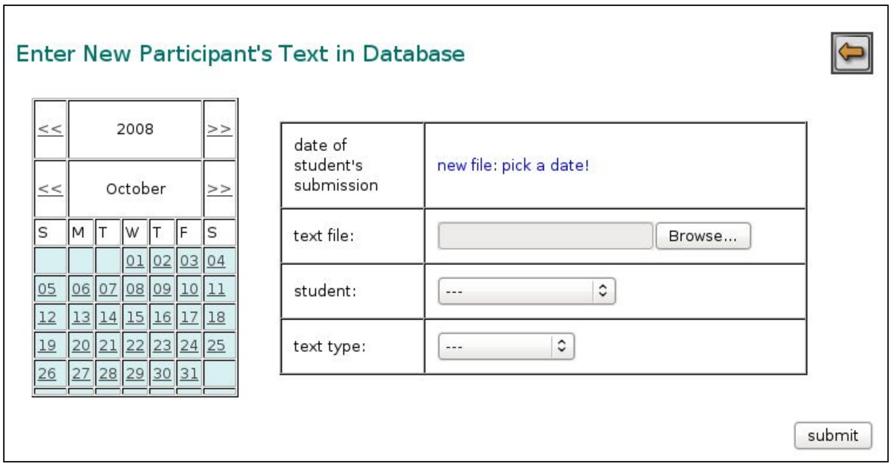
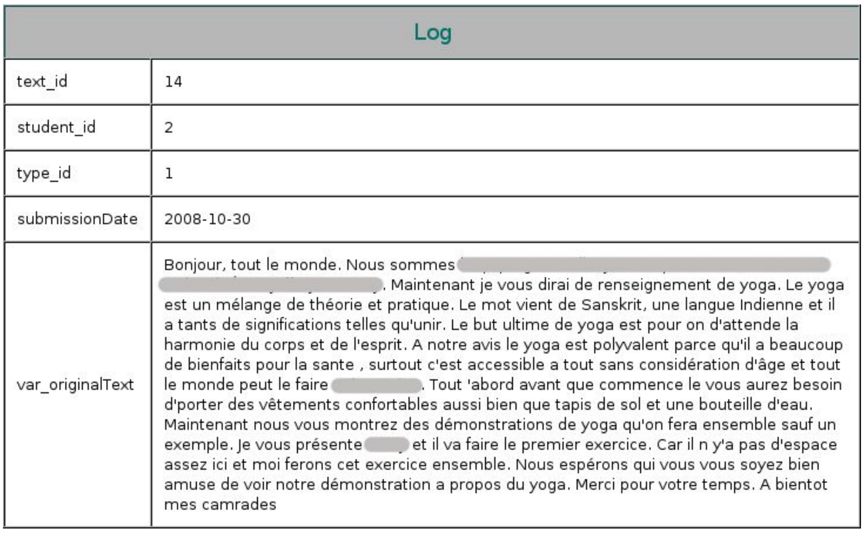


Figure 5.15 below represents the log obtained after inserting a new text into the database. Each database entry is identified with unique numbers as they are faster to process than character strings. For example, text type (type_id in Figure 5.15) is equal to 1 which designates that the text written here was classified under the category *forum*. The text types and their corresponding database identification are given in Appendix D.2 on page 291.

Figure 5.15. Log of data recorded in the database[34]



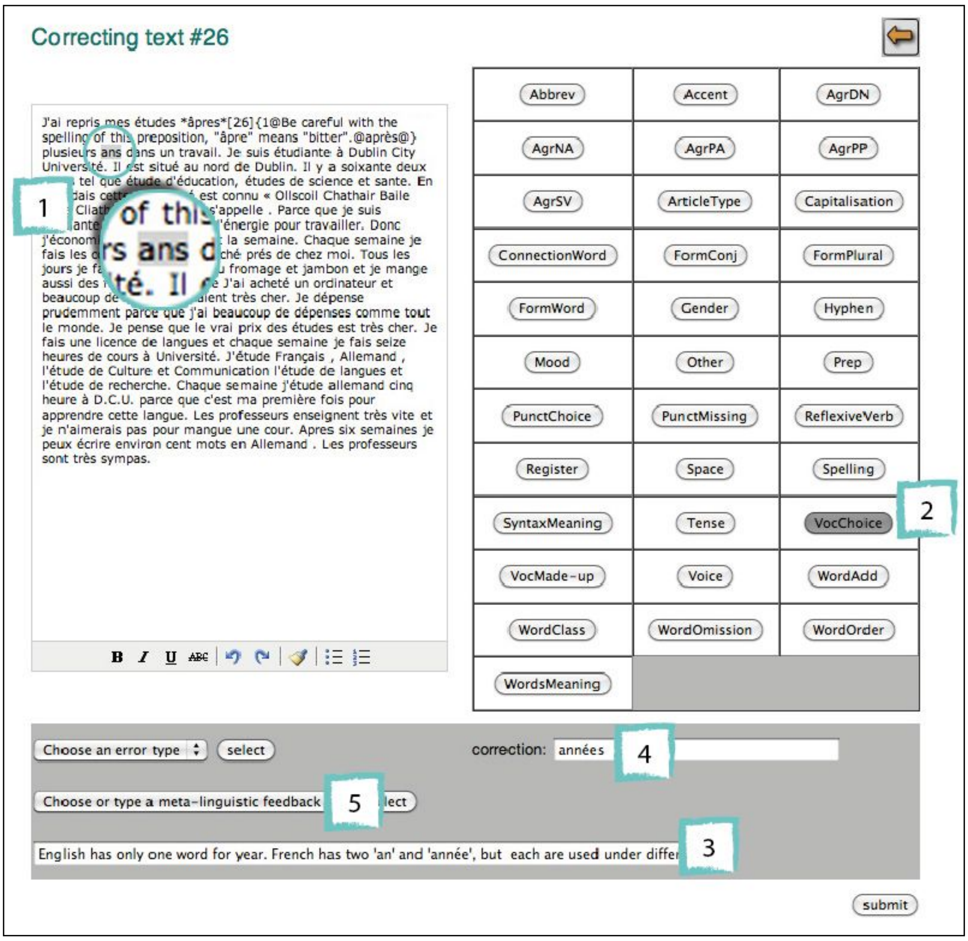
Log	
text_id	14
student_id	2
type_id	1
submissionDate	2008-10-30
var_originalText	Bonjour, tout le monde. Nous sommes [illegible]. Maintenant je vous dirai de renseignement de yoga. Le yoga est un mélange de théorie et pratique. Le mot vient de Sanskrit, une langue Indienne et il a tants de significations telles qu'unir. Le but ultime de yoga est pour on d'attende la harmonie du corps et de l'esprit. A notre avis le yoga est polyvalent parce qu'il a beaucoup de bienfaits pour la sante , surtout c'est accessible a tout sans considération d'âge et tout le monde peut le faire [illegible]. Tout 'abord avant que commence le vous aurez besoin d'porter des vêtements confortables aussi bien que tapis de sol et une bouteille d'eau. Maintenant nous vous montrez des démonstrations de yoga qu'on fera ensemble sauf un exemple. Je vous présente [illegible] et il va faire le premier exercice. Car il n y'a pas d'espace assez ici et moi ferons cet exercice ensemble. Nous espérons qui vous vous soyez bien amuse de voir notre démonstration a propos du yoga. Merci pour votre temps. A bientot mes camrades

34. All marks that could eventually reveal the identity of this participant have been masked in order to preserve her anonymity.

5.4.2. Computer-aided error editor

The functioning of the computer-aided error editor is mirrored in the software Markin (mentioned in Section 4.2.4 on page 92), mostly with regard to the encoding of the meta-linguistic information and the interface disposition. The process to correct learners' texts consists of (a) highlighting the incorrect words or groups of words, (b) choosing a meta-linguistic feedback[35] in the drop down list or writing a personalised feedback, (c) entering an appropriate alternative, and (d) applying the corresponding error type. The procedure to correct students' texts is explained below with Figure 5.16 as visual support.

Figure 5.16. Web-based application to correct learners' texts



35. Meta-linguistic feedback was provided in English with the help of Hawkins and Towell's (2001) grammar book.

For example, the highlighted ill-formed word **ans* in the above figure (area 1) is an inappropriate word choice (se_vo_lw_). Before selecting the *vocChoice* button (area 2), relevant information that may help the learner correct herself is entered in the corresponding field (area 3). In addition, a correct form is provided in area 4. The meta-linguistic feedback may also be chosen from the drop down list (area 5), which contains several annotations that describe recurrent incorrect forms requiring the exact same comment over and over again. Table 5.5 shows examples of meta-linguistic feedback[36] often associated with the inappropriate word choice error type.

Table 5.5. Meta-linguistic feedback associated with an inappropriate word choice

Error type	Description
se_vo_lw_ (incorrect word choice)	English has only one word for year. French has two 'an' and 'année', but each are used under different circumstances. The form with 'ée' is usually used when an adjective modifies the noun.
	The pronoun 'il' cannot usually be used to refer to events or actions. While 'ce' is normally used with 'être', 'cela' and 'ça' are used with other verbs. 'Cela' tends to be used in written French whereas 'ça' is widely used in the spoken language.
	Word for word translation.
	Inappropriate word in context.
	Climatic conditions can be expressed by an impersonal use of 'faire' followed by an adjective or a noun.
	This verb only exists in an impersonal form, which means that it only takes the pronoun 'il' as subject.

Each incorrect form is annotated with the editor tool in the following way:

incorrect form[errorType]{index@meta-linguistic feedback@correct form}

The **incorrect form** sequence corresponds to the first level of assistance in the regulatory scale, where the incorrect form is solely highlighted. The *[errorType]* section coincides with the second level of assistance where only the error type of the incorrect form is issued. An *index* is given in order to make the sequence unique in the input text.

36. See Appendix D.4 on page 291 for the complete list of pre-determined meta-linguistic feedback.

The *{_@meta-linguistic feedback} and {_@correct form}* strings provide the information for the third and fourth level of assistance, respectively. Information inside the curly brackets are separated with the @ sign, which is used afterwards as a splitting character when processing the string. Each incorrect form is then stored in the database along with its text identification, its index in the text, its position in both the original and encoded texts, the error type, the meta-linguistic feedback and the correct form. The position in the text is not only used to retrieve the incorrect form context when needed, it is also necessary to improve the part-of-speech tagging accuracy when cross-referencing the error annotation with the part-of-speech tag. Line 4 below represents a concrete example of annotated incorrect forms. The learner's original writing is highlighted in bold and the information related to the incorrect forms (error type and levels of assistance) is in italics.

> [4]. **J'ai repris mes études *âpres****[26]*[37]*{1@Be careful with the spelling of this preposition, "âpre" means "bitter".@après@}* **plusieurs *ans****[2]{2@English has only one word for year. French has two 'an' and 'année', but each are used under different circumstances. The form with 'ée' is used when an adjective (plusieurs) modifies the noun.@plusieurs années@}* ***dans un***...

5.4.3. Correcting learners' alternatives

After learners were provided with their corrected texts, they had to propose alternatives to their incorrect forms. The validation tool helped determine whether these alternatives were correct or not. It is a semi-automatised process based on a matching method using regular expressions[38]. This procedure enabled more than three-quarters of the alternatives to be marked automatically, the rest being checked manually. The validation of learners'

37. See Appendix D.3 on page 291 for the error types and their corresponding database codes.

38. See Appendix C.1 on page 283 for a short account on regular expressions.

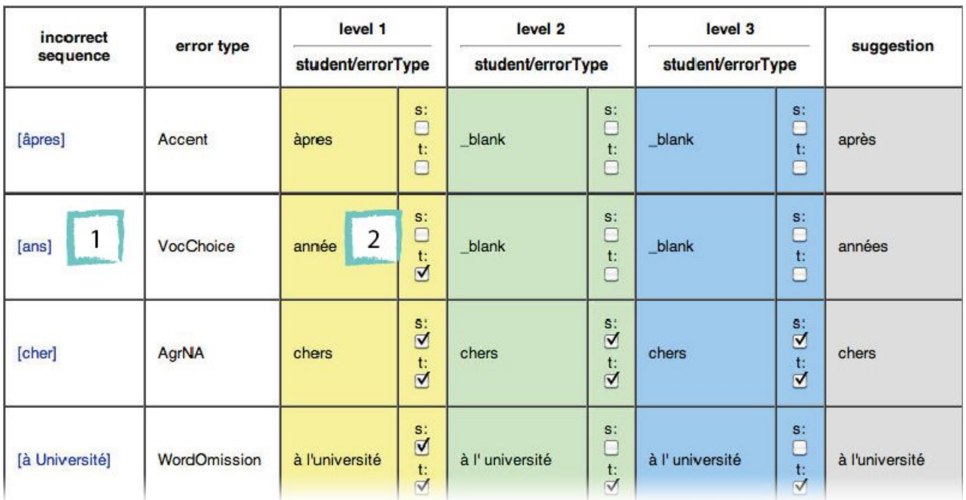
alternatives is conditioned on two criteria. The first criterion determines whether the alternative is well-formed and the second one specifies whether the original error type is resolved.

> **Criterion 1.** *From the students' perspective.* Since students were provided with a report on the alternatives they proposed indicating whether their replacements were correct or not, the alternatives were marked as appropriate, only if the replacements suggested were *perfectly* well-formed. Otherwise, the replacements were invalidated, that is, marked as incorrect.

> **Criterion 2.** *From the research perspective.* If students were able to resolve the original error types, then their corrections were marked as appropriate, even if learners produced other error types when proposing their replacements.

The motivation behind the two levels of validation was justified by the fact that learners could not be provided with the same error validation as the one used for this research. For instance, the incorrect form **ans (years),* displayed in Figure 5.17 below (area 1), is an incorrect choice in terms of vocabulary (VocChoice).

Figure 5.17. Validation process



incorrect sequence	error type	level 1 student/errorType		level 2 student/errorType		level 3 student/errorType		suggestion
[âpres]	Accent	àpres	s: ☐ t: ☐	_blank	s: ☐ t: ☐	_blank	s: ☐ t: ☐	après
[ans] 1	VocChoice	année 2	s: ☐ t: ☑	_blank	s: ☐ t: ☐	_blank	s: ☐ t: ☐	années
[cher]	AgrNA	chers	s: ☑ t: ☑	chers	s: ☑ t: ☑	chers	s: ☑ t: ☑	chers
[à Université]	WordOmission	à l'université	s: ☑ t: ☑	à l' université	s: ☐ t: ☑	à l' université	s: ☐ t: ☑	à l'université

With the alternative *année* proposed by the learner at level 1 – without assistance – (area 2), the student solved the error type. However, by forgetting the -*s* mark plural – *plusieurs *année (several years)* –, the learner created another error type which could be classified as a noun adjective agreement error. While the box below the letter *s (student)*, if checked, identifies the learners' entries as perfectly correct, the box below the letter *t (type)*, if checked, recognises that the error type was resolved. With only one validation method, the dilemma would be to decide between what is acceptable for the student or resolved in terms of error type for this research. Accepting the word **année* as a correct alternative, which is half true since the word has been rectified in accordance to the error type, is not an acceptable option for the student, as the plural mark is missing. However, not validating the alternative as correct would denote that this specific error type was not resolved, which is false since the incorrect word was correctly replaced. Although the incorrect form **année* is marked as inappropriate in terms of replacement for the student, the fact that the error type has been noticed is recorded in the database for further analysis.

5.4.4. Monitoring the learners' actions when self-editing their texts

A learner's alternative or a blank field does not establish whether they are the outcomes of reading the feedback or skipping it. Monitoring the learners' behaviour with regard to feedback access was identified as a necessary step, especially after witnessing that some participants were not consulting the assistance provided at level two and/or level three. This observation is consistent with other studies, in which researchers state that it is not because help is available to learners that they will use it (e.g., Fischer 2007, Heift 2010b).

Accessing feedback does not signify that the learner has read it. However, since learners had to move their cursor over the incorrect forms to display the feedback, and if the message was opened sufficiently long, it was assumed that the learner had the intention of reading the message. To determine whether a feedback was accessed/read, the computer-based application (a) recorded the time at which any feedback pop up was opened and closed, (b) computed the difference between both records, (c) calculated the time required to read the feedback, and (d) evaluated whether the access was sufficiently long for the feedback to be read. The time required to read the feedback was determined depending on the amount of words included in the message and the speed rate at which learners read. While it is commonly assumed that the speed rate at which a native speaker reads is around 250 words per minute, Ziefle (1998) found that reading on a computer screen, as opposed to paper, slowed down the rate to 180 words per minute. Following her findings, the speed rate adopted in this study was set up at 180 words per minute. The time in milliseconds required to read one word is thus equal to 333.33ms (60/180*1000).

From the student perspective, it was considered appropriate to show[39] learners – apart from a few texts written before the implementation of the monitoring system – a detailed account of their actions, so that they could reflect on their correction and whether they needed assistance. Figure 5.18 below shows, for instance, that the assistance for the incorrect word **était (was)* was accessed once during 3.86 seconds. In fact, the feedback was accessed twice; the other access lasted 21 milliseconds. Since the feedback message includes three words, i.e., *subject verb agreement*, the time required to read this message is calculated to be at least 999 milliseconds. As a result, the access that lasted 21 milliseconds was not considered long enough to ensure the reading of the message.

39. The system was not implemented to record the learners' access of these reports. Except from one participant who orally expressed her satisfaction with the detailed information on each incorrect form, whether the other students benefited from it or not is not known.

Figure 5.18. Monitoring system

incorrect form	feedback	my correction	teacher's correction	Alternative validated?	Feedback read? How many times? Average time?
[francais]	misspelling	français	français	✓	✓ 1.46 seconds (1 time)
[les grammaires]	inappropriate word choice	la grammaire	les exercices de grammaire	✗	✓ 1.53 seconds (1 time)
[était]	subject verb agreement	-	étaient	✗	✓ 3.86 seconds (1 time)
[renconterai]	misspelling	recontrerai	rencontrerai	✗	✓ 1.35 seconds (3 times)
[apres]	incorrect or missing accent	après	après	✓	✗ Not read!
[heures!]	missing or inappropriate space	heures !	heures !	✓	✓ 2.64 seconds (1 time)
[bientot]	incorrect or missing accent	bientôt	bientôt !	✓	✗ Not read!

5.5. Data organisation and analysis

The data is stored in a relational[40] as opposed to flat[41] database to efficiently organise the information, and more specifically, to decrease the time required to access the database. The time parameter is a factor that should not be neglected, especially when accessing the database remotely. For example, if the queries take too much time and resources to be completed, this may force the database to crawl and consequently experience a time out error resulting in incomplete data retrieval.

5.5.1. Overview of the data organisation

Figure 5.19 shows an entity-relationship diagram (created with MySQL Workbench[42]) which graphically represents the relational organisation of the data. The database

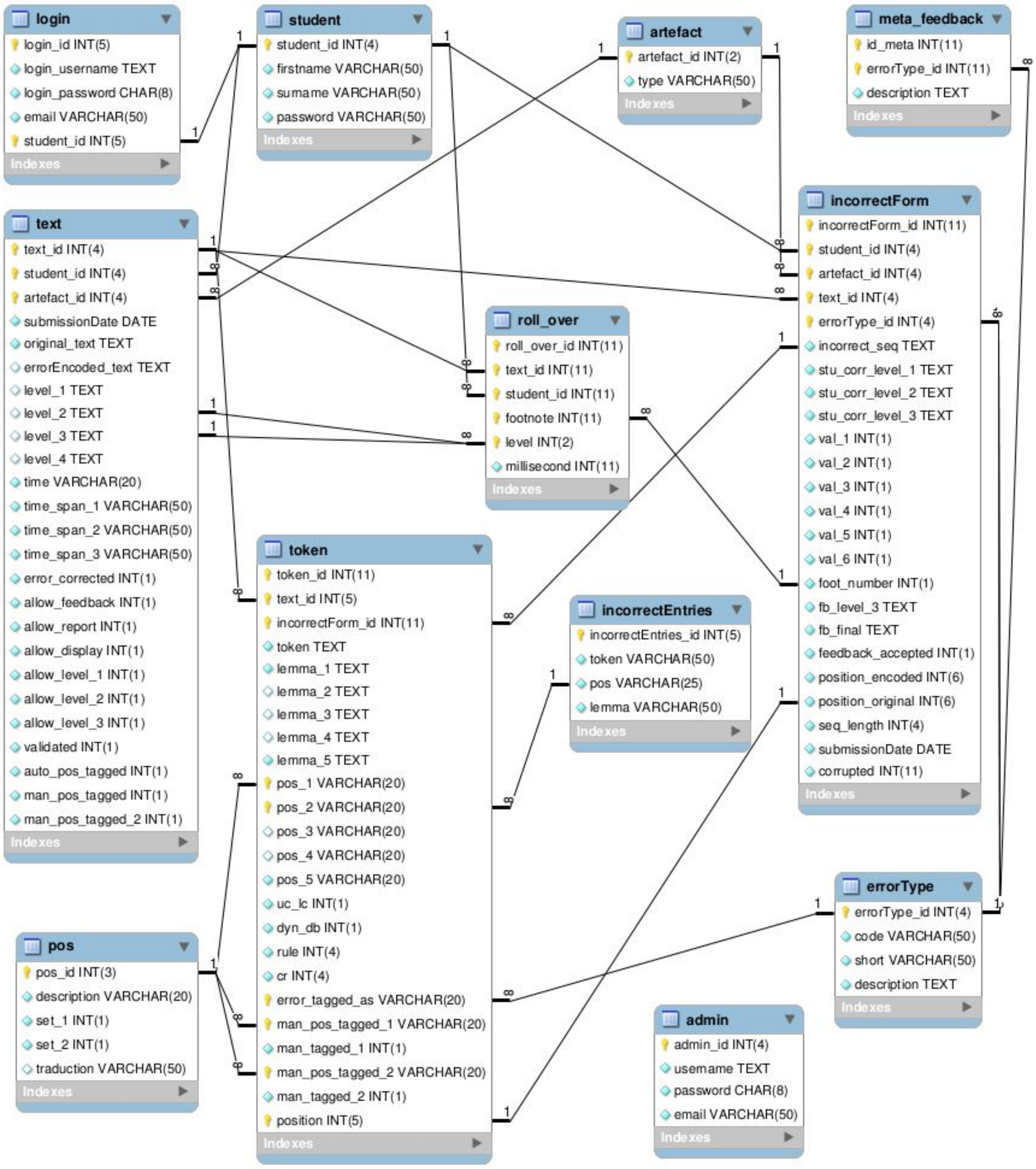
40. A relational database contains multiple tables that are connected to each other with one or more common fields (Pravec 2002 p.101). Codd (1979, 1982), a pioneer in database management, invented the relational model to improve productivity by avoiding redundant data.

41. A flat database stores a collection of data in one table, which means that each record can be saved and displayed in one single file (Pravec 2002 p.101).

42. http://www.mysql.com/products/workbench/

includes twelve tables in total: admin, login, students (student), texts (text), text types (artefact), error types (errorType), incorrect forms (incorrectForm), meta-linguistic feedback (meta_feedback), feedback access (roll_over), tokens (token), unknown part-of-speech lemmas (incorrectEntries), and part-of-speech tags (pos).

Figure 5.19. Entity-relationship diagram



The *token* table, for instance, lists each token in a text and associates each of them with a text identification so that information on that text can be easily retrieved. By following the text identification in the *text* table, further information on that text, such as the student identification or the submission date, can also be retrieved, and so on.

As a relational database cannot be represented in one single flat file, the XML format is therefore more appropriate to display the different elements of the database that are used for analysis. An extract of the fully annotated corpus taken from the database and converted into XML format is displayed below:


```
<?xml version="1.0" encoding ="UTF-8" standalone="yes"?>
<text>
<text_id>88</text_id>
<student_id>2</student_id>
<text_type>email</text_type>
<original_text>Bonjour Sylvia. J'ai fini mes examens en francais , les grammaires était bien, mais
l'écrit était difficile... </original_text>
<errorEncoded_text>Bonjour Sylvia. J'ai fini mes examens en
*francais*[29]{1@Misspelling.@français@} , *les grammaires*[2]{2@Inappropriate word in
context.@les exercices de grammaire@} *était*[22]{3@Verbs agree with their subjects in person and
number.@étaient@} bien, mais l'écrit était difficile...</errorEncoded_text>
...
<chunk>
        <token_id>18943</token_id>
        <token>était</token>
        <error_tagged_as>mo_ag_sv_</error_tagged_as>
        <incorrectForm>
                <incorrectForm_id>2820 </incorrectForm_id>
                <incorrect_seq>était</incorrect_seq>
                <stu_corr_level_1>_blank</stu_corr_level_1>
                <stu_corr_level_2>_blank</stu_corr_level_2>
                <stu_corr_level_3>_blank</stu_corr_level_3>
                <val_student_1>0</val_student_1>
                <val_student_2>0</val_student_2>
                <val_student_3>0</val_student_3>
                <val_errorType_1>0</val_errorType_1>
                <val_errorType_2>0</val_errorType_2>
                <val_errorType_3>0</val_errorType_3>
                <fb_level_2>subject verb agreement</fb_level_2>
                <fb_level_3>Verbs agree with their subjects in person and number.</fb_level_3>
                <fb_level_4>étaient</fb_level_4>
                <feedback_read_count_level_2>1</feedback_read_count_level_2>
                <feedback_read_average_time_l2>3.86</feedback_read_average_time_l2>
                <feedback_read_count_l3>1</feedback_read_count_l3>
                <feedback_read_average_time_l3>3.86</feedback_read_average_time_l3>
                <submissionDate>2009-05-01</submissionDate>
        </incorrectForm>
        <pos_5>VER:conj</pos_5>
        <position>68</position>
</chunk>
<chunk>
        <token_id>18944</token_id>
        <token>bien</token>
        <error_tagged_as>correct</error_tagged_as>
        <pos_5>ADV</pos_5>
        <position>74</position>
</chunk>
<chunk>
        <token_id>18945</token_id>
        <token>,</token>
        <error_tagged_as>correct</error_tagged_as>
        <pos_5>PUN</pos_5>
        <position>78</position>
</chunk>
...
</text>
```

The example above is an extract of text #88. It shows the type of information contained within the corpus. More specifically, it includes the student identification, the text type, the original writing, the error annotated text, and information on each token. Each token is attached to an identification, an error type if appropriate – otherwise the token is marked as *correct* –, a part-of-speech tag, and the numeric position of its occurrence in the text. If the token is assigned an error type, the information also includes the level of assistance required, the alternatives provided by the learner, whether the feedback was read, the correct alternative, and whether the learner's alternatives were validated. In addition, the submission date is also recorded. The annotated text #88 can be found in full in Appendix D.5 on page 297.

5.5.2. Overview of the data collected

Learners' participation with regard to the amount of (a) texts submitted, (b) tokens, (c) words, (d) incorrect forms – which could contains one single word or more –, and (e) incorrect words, is outlined in Table 5.6 below.

The corpus contains 77 texts written by 14 students. In total, it includes 19,870 tokens, in which 2,579 incorrect forms were counted. The fact that most students submitted more than one text for correction implies a relatively high level of motivation to participate in this research. Otherwise, they would have stopped after the first submission. The word count is relatively impressive, especially for student #2 who sent in total 7,264 words for correction. Also interesting is the percentage of incorrect forms learners made, which should enable the observation of a wide range of error types, as intermediate learners may exhibit greater variations in terms of error than beginners.

Table 5.6. Overview of the learners' participation

Student id	Text count	Token count	Incorrect token count	Word count	Incorrect form count	Incorrect word count
#2	50	8,128	1,488 18.31%	7,264	1,033 14.22%	1,459 20.09%
#5	1	420	66 15.71%	356	43 12.08%	59 16.57%
#6	3	1,262	254 20.13%	1,098	158 14.39%	247 22.50%
#7	4	1,052	214 20.34%	936	145 15.49%	210 22.44%
#8	2	865	175 20.23%	738	127 17.21%	167 22.63%
#9	2	527	124 23.53%	470	101 21.49%	119 25.32%
#10	3	1,501	254 16.92%	1,349	156 11.56%	246 18.24%
#11	2	1,280	112 8.75%	1,107	78 7.05%	102 9.21%
#12	1	366	73 19.95%	338	37 10.95%	72 21.30%
#14	1	373	89 23.86%	339	60 17.70%	80 23.60%
#15	1	381	65 17.06%	347	49 14.12%	62 17.87%
#16	2	1,002	178 17.76%	926	96 10.37%	177 19.11%
#17	4	2,288	658 28.76%	2,089	457 21.88%	635 30.40%
#19	1	425	55 12.94%	395	39 9.87%	54 13.67%
Total: 14 students	77	19,870	3,805 (19.15%)	17,752	2,579 (14.53%)	3,689 (20.78%)

The error rate, commonly calculated by dividing the number of error types by the total amount of words in a text is equal to 14.53%. However, I believe that the error rate calculated as such is not a representative measure of the real situation. For example, if a learner wrote a sentence of four words, such as *l'information est *intéressant (the information is interesting),* and one of these words *intéressant* is incorrect due to an inappropriate noun adjective agreement *(correct form: intéressante)*, the error rate is equal to one incorrect form, i.e., one error type, divided by four words. As a result, the error rate percentage equals 25%. If another learner wrote a sentence of 4 words, such as *Intéressante l'information est (Interesting the information is)*, and the whole sentence is marked as an inappropriate syntactical structure with one unique error type, the error rate is also equal to one error type divided by four words. Here again, the result equals 25%. Yet, both results do not carry the same value, as in one case, the whole sentence is incorrect (e.g., incorrect syntax error type), and in the other, there is only one single incorrect word (e.g., incorrect noun adjective agreement). For this reason, not only the error rate is given, but also the incorrect word rate is calculated. Both measures represent the minimum and maximum scores a text can be assigned in terms of error proportion. Table 5.6 above shows a global error rate of 14.53% and a global incorrect word rate that reaches 20.78%. Consequently, the percentage of incorrect forms cannot be estimated to be only 14.53%, but rather in between the minimum and maximum values of 14.53% and 20.78%, respectively. A convenient representation could be to calculate the average of both measures, which means that the percentage of incorrect sequences for the whole corpus would be 17.65%.

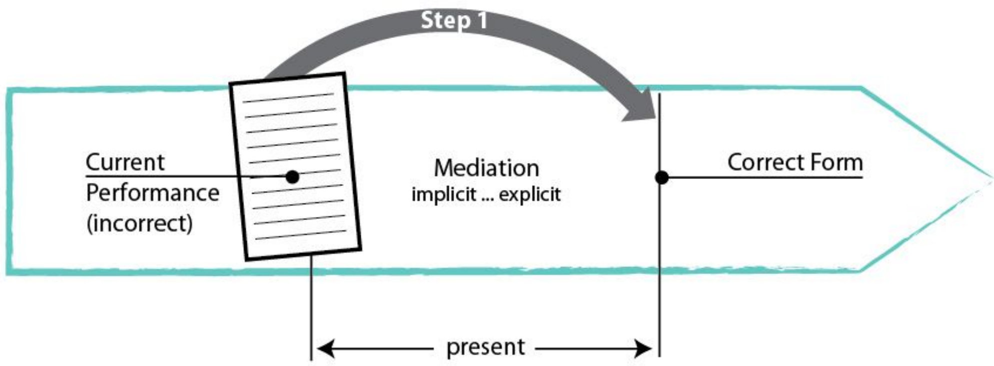
As shown in Figure 5.8 on page 116, there are two major evaluations to be performed on the data collected: (a) the positioning of the incorrect form in the learner's zone of proximal development, so that a distinction can be made between errors and mistakes, and (b) the computing of the learner language complexity and accuracy, so that the learner's knowledge can be represented. The distinction between errors and mistakes and the evaluation of the language complexity were performed through the means of PHP applications. Details on the implementations can be found in Appendices E.1 on page 306 and E.2 on page 313, respectively. The output of both evaluations was arranged in a tabular format so that descriptive and frequency analyses could be performed with the statistical software SPSS[43].

5.5.3. *Distinguishing between errors and mistakes*

If a word did not carry any error type tag, it was assumed that it was something the learner knew. Although the possibility that the word could have been guessed in the first place should not be excluded, the distinction between correct forms and guesses is beyond the scope of this thesis. If the learners produced an incorrect form, it was assumed that it was an error by default. The level of assistance required to correct this incorrect form helped situate it within the learner's zone of proximal development, and thus helped differentiate between errors and mistakes. The fact that learners did not systematically access the feedback to help them correct their texts required an additional procedure to represent the ZPD. In the pilot study (see research design in Section 4.3 on page 93), the first level of assistance on the regulatory scale (incorrect form highlighted) was considered to be different from the second level (error type provided), which was also regarded as dissimilar from the third level (meta-linguistic feedback). Knowing that learners did not systematically access the assistance, level 1 may be considered as equivalent to level 2 or 3 if the feedback is not read. This had to be taken into consideration when positioning the incorrect form in the ZPD. Figure 5.20 represents the first step (out of two) to distinguish between errors and mistakes as presented in Section 3.3.2 on page 64. All words or groups of words marked as incorrect in the current performance are attempted to be corrected by their respective authors with first little assistance (highlight only), and then more and more explicit feedback until providing learners with the correct answers.

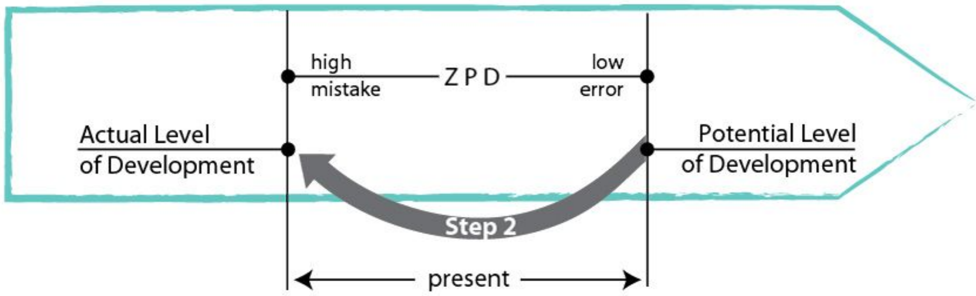
43. SPSS is a registered brand name IBM SPSS Statistics (formerly SPSS Statistics): http://www.spss.com/statistics/

Figure 5.20. Error-mistake distinction model: first step



The second step, illustrated in Figure 5.21, consists of examining the amount of assistance that was necessary to position the incorrect forms in the learners' ZPD. If students used minimum assistance then the features marked as incorrect are nearly at the point of internalisation; the incorrect forms are high in the ZPD and are more likely to be performance-related mistakes. In such a situation the zone of proximal development is almost nonexistent; the level of potential development nearly matches the level of actual development. However, learners' current performance does not reflect learners' level of actual development as the former includes incorrect forms that are considered as correct in the latter. On the other hand, if learners require more than very little assistance to propose correct answers, the zone of proximal development becomes tangible; the grammatical or lexical aspect marked as incorrect is still in development. Consequently, the levels of actual and potential development are now distinct; the more assistance, the wider the zone. If the incorrect forms are low in the ZPD, then they are more likely to be competence-dependent errors. In the event of incorrect forms recognised as errors, then the current performance reflects the level of actual development.

Figure 5.21. Error-mistake distinction model: second step



In order to avoid learners' lucky guesses while self-editing themselves, and since they had to correct their texts three times consecutively, it was decided that learners had to provide a correct form at least twice with minimum assistance for an incorrect form to be considered as a mistake. The conditions for an incorrect form to be considered as a mistake or an error are displayed in Table 5.7 below.

Table 5.7. Error-mistake distinction depending on feedback access and alternatives

Learners' alternatives provided with:				
Scenarios	**Incorrect form highlighted (level 1)**	**Error type provided (level 2)**	**Meta-linguistic feedback provided (level 3)**	**Error or mistake**
#1	0/-	0/-	0/-	error
#2	0/-	0/-	1	error
#3	0/-	1	0/-	error
#4	0/-	1	1(feedback read)	error
#5	0	1 (feedback read)	1(feedback not read)	error
#6	-	1 (feedback read)	1(feedback not read)	mistake
#7	0/-	1 (feedback not read)	1(feedback not read)	mistake
#8	1	0/-	0/-	error
#9	1	0/-	1 (feedback read)	error
#10	1	0/-	1 (feedback not read)	mistake
#11	1	1 (feedback read)	0	error
#12	1	1 (feedback read)	-	mistake
#13	1	1 (feedback not read)	0/-	mistake
#14	1	1	1	mistake

A zero (0) means that the learner entered an incorrect form, a dash (-) designates that the field was left blank, and a one (1) refers to a learner's correct alternative. Scenario #13, for instance, considers that the incorrect form is a mistake on the condition that the learner provides a correct replacement at level one, a correct replacement without accessing the assistance at level two, and an inappropriate replacement or nothing at level three. The fact that the learner can propose two correct alternative forms should exclude the possibility of a lucky guess. Although the order in which learners entered their corrections was not taken into consideration, two correct responses in succession could place the incorrect form higher in the ZPD than a correct/incorrect/correct sequence; the sequence could be a valuable indicator. However, and as mentioned at the beginning of this section, since learners did not access their feedback consistently, all levels of self-editing exercises, with respect to amount of assistance, could potentially be identical with one another. Consequently, no preference was given to sequence order.

Knowing whether learners accessed the assistance or not is a prerequisite to establish their level of potential development. Texts without indication about how learners behaved with the assistance were nevertheless processed. However, a detailed report was made on incorrect forms that could not be definitively distinguished with certainty because of the lack of information with regard to feedback accessed. If all necessary information is available, the errors and mistakes are considered as *confirmed*. However, if the distinction could not be determined due to incomplete information, then errors and mistakes are labeled as *not confirmed*. Additionally, a difference is made between (a) errors (not mistakes) that could be corrected with assistance at levels two and/or three, and (b) errors that could not be self-edited even with meta-linguistic feedback, or errors that could be edited only once without feedback, in which case the alternative could have been a lucky guess. A feature marked as incorrect, where the student was able to provide at least one correct alternative with assistance, is considered as having a higher position in the ZPD than an error for which the learner could not suggest any correct form even with assistance. In summary, an error that can be corrected with assistance is located in between high and low in the ZPD, and everything else (except from mistakes) takes a lower position in the ZPD. Table 5.8 below lists the code applied to each incorrect form depending on whether the distinction between errors and mistakes could be conclusively established or not.

Table 5.8. Incorrect form status and corresponding code

Incorrect form status	**Confirmed**	**Not confirmed**
mistake: high in ZPD	1	3
error: between high and low in ZPD	0	2
other error: low in ZPD	-1	

Now that the placement of the incorrect form in the learner's zone of proximal development is established and that the distinction between errors and mistakes can be operated, thus revealing the learner's current knowledge, the next stage consists of measuring this current knowledge.

5.5.4. Representing learners' current knowledge

Following the discussion in Section 4.1.3 on page 73, the learner's knowledge is estimated through the means of interlanguage complexity, as well as interlanguage accuracy.

Interlanguage complexity

The learners' interlanguage complexity is estimated with Schulze, Wood and Pokorny's (forthcoming) balanced complexity method. More specifically, it is approximated from textual lexical and grammatical complexity by focusing on letter, word form, bigram, and period unit. The indicators adopted are mean word length, text-length adjusted type token ratio, unique bigram ratio, and mean period unit length, respectively. Although the reliability of this multidimensional and organic balanced complexity measure has so far only been tested on German texts, it is firmly grounded on reliable and valid metrics used to measure textual complexity in *all* languages (Schulze et al. forthcoming).

1. Lexical complexity/diversity: text-length adjusted type token ratio. The first vector is computed by dividing the number of different words (types) by the square root of twice the amount of words (Carroll 1964). This measure gives an indication about the lexical variety of a text taking into account the length of the text.

Formula 5.1. Adjusted type token ratio

$$v_1 = \frac{t}{\sqrt{2w}}$$

where t and w designate the number of *type*s and *words*, respectively.

A type is commonly acknowledged as every different form a word can encounter in a text. For example *cat* and *cats* counts for two different types. I argue that the changing lexical diversity would be better viewed by counting different lemmas that occur only once in a text as opposed to counting every unique different form of a word. For example, if a student uses the same verb to express the idea of writing by using the word *écrire (to write)* under different forms such as various tenses, he or she still refers to the same lexeme over and over again. However, by making use of synonyms, the learner shows the extent of his or her lexical diversity. While the first situation only informs about the learner's capability to conjugate the same verb, it does not provide any information about the learner's lexical knowledge as the second situation does. Yet, the type token ratio would be identical for the two examples. Taking advantage of the pos-tagged corpus, the adjusted type token ratio is instead computed by counting the amount of unique lemmas and the amount of words in a text. The *adjusted type token ratio* has to be understood as the *adjusted lemma word ratio*[44]. This adaptation differs from Schulze et al.'s (forthcoming) approach, who focus on text surface items that are directly observable. Counting lemmas implies analysing the text from a morphological perspective in order to identify the dictionary entry of each word. This can be achieved with any lemmatisers or some part-of-speech taggers such as TreeTagger.

2. Lexical sophistication: mean length of a word. The second vector is determined by dividing the number of letters by the number of words in a text. To determine the total amount of letters a text contains, all characters within a word, i.e., a token that is not part-of-speech tagged as a punctuation mark or as a digit, are counted. Longer words are usually associated with the idea of complexity as they tend to include inflectional and derivational affixes.

Formula 5.2. Mean word length

$$v_2 = \frac{l}{w}$$

where l and w designate the number of *letters* and *words*, respectively.

3. Syntactic sophistication: mean period unit length. The third vector is obtained by dividing the number of words by the total amount of sentences marked by periods (exclamation mark, question mark, and full stop). Mean length is understood as a measure to describe the level of sophistication of a text, the longer the sentences, the more sophisticated they are.

44. A token, in this thesis, already carries the meaning of an item that has been part-of-speech tagged. This item could be a punctuation mark, which is not taken into account when computing the type token ratio. For this reason, and in order not to get confused, the term token always refers to a part-of-speech tagged item, unless specified in the context of type/token ratio, where a close synonym would be the term *word*.

Formula 5.3. Mean period unit length

$$v_3 = \frac{w}{p}$$

where *w* and *p* designate the number of *words*

and *sentences marked by periods*, respectively.

Schulze, Wood and Pokorny (forthcoming) state that sentence boundaries are usually accurately marked in texts written by second language learners. As it was observed in the corpus that the final punctuation mark at the very end of the text was sometimes forgotten – especially in emails and forum discussions –, the last token of the text is checked, and in the event of a missing final punctuation mark, the amount of periods is incremented by one in order to virtually count a non-existing full stop.

4. Syntactic complexity/diversity: unique bigram ratio. The fourth vector, based on the adjusted type token ratio, is calculated by counting the number of different bigrams ([word1 word2], [word2 word3], [word3 word4], ...), and by dividing this number by the square root of twice the total amount of bigrams in the text. This measure captures the level of syntactic complexity by looking at word combinations.

Formula 5.4. Unique bigram ratio

$$v_4 = \frac{u}{\sqrt{2b}}$$

where *u* and *b* designate the number of *unique bigrams*

and total amount of *bigrams*, respectively.

Each letter within a word is normalised to avoid identical bigrams to be counted as unique due to, for instance, incorrect or missing accents.

Schulze, Wood, and Pokorny (forthcoming) point out that the four vectors combined together describe the overall complexity of a text, as one could not produce complex sentences without extending their lexical and syntactical complexity and sophistication. The balanced complexity formula is as follows:

Formula 5.5. Balanced complexity

$$C_B = \left(|v_1 - 1| + \left| v_2 - \frac{1}{\sqrt{(2w)}} \right| + \left| v_3 - \frac{1}{\sqrt{(2w)}} \right| + |v_4 - 1| \right) - \left(\max(v_1, v_2, v_3, v_4) - \min(v_1, v_2, v_3, v_4) \right)$$

The four vectors, detailed above, when assembled together offer an overview of the learners' language complexity. Interpreting the balanced complexity measures as absolute scores would be a rather difficult endeavour. However, the result can be used to compare learner knowledge, or shed light on changes over time (Schulze et al. forthcoming).

Interlanguage accuracy

Extending the obligatory occasion analysis to error types, as opposed to morphemes as discussed in Section 4.1.3 on page 73, learner interlanguage accuracy is calculated by dividing the number of correct forms by the total amount of correct and incorrect forms in a text. The number of incorrect instances is calculated by counting words that are error annotated. The number of correct instances, on the other hand, is determined by counting part-of-speech tags that are not marked as incorrect. For example, since adjectives are all susceptible to agree in number and gender with the nouns they modify, estimating the percentage of accuracy in noun adjective agreement in a text denotes counting *noun adjective agreement* error types and *adjective* part-of-speech tags not marked as incorrect. Thus, interlanguage performance accuracy, represented through the ratio of incorrect to correct forms (*IncF:CorF*), is calculated as follows:

Formula 5.6. Interlanguage performance accuracy

$$P = \left(\frac{CorF}{CorF + IncF} \right) \times 100$$

Interlanguage competence accuracy, on the other hand, must include mistakes in the count. It is calculated in four different ways (Formulae 5.7 to 5.10) in order to capture the uncertainty of a few error-mistake distinctions due to incomplete information on feedback access.

Formula 5.7. Correct forms + mistakes

$$K_1 = \left(\frac{CorF + M}{CorF + IncF} \right) \times 100$$

where M equals mistakes

Formula 5.8. Correct forms + mistakes + mistakes not confirmed

$$K_2 = \left(\frac{CorF + M + M_{nc}}{CorF + IncF} \right) \times 100$$

where M and M_{nc} equal mistakes and mistakes not confirmed, respectively.

Formula 5.9. Correct forms + mistakes + errors not confirmed

$$K_3 = \left(\frac{CorF + M + E_{nc}}{CorF + IncF} \right) \times 100$$

where M and E_{nc} equal mistakes and errors not confirmed, respectively.

Formula 5.10. Correct forms + mistakes + errors and mistakes not confirmed

$$K_4 = \left(\frac{CorF + M + E_{nc+}M_{nc}}{CorF + IncF} \right) \times 100$$

where M, E_{nc} and M_{nc} equal mistakes, errors not confirmed, and mistakes not confirmed, respectively.

Table 5.9 below displays the percentages of accuracy in terms of overall performance (P) and overall knowledge (K_i) of two students for whom some of the information on feedback access is incomplete. Student #8, for instance, wrote 690 correct forms and 175 incorrect forms in total. According to the formula above, her percentage of accuracy in performance attains 79.8%. As the error analysis totalled 24 mistakes, her percentage of accuracy in knowledge (K_1) equals 82.5%.

Table 5.9. Percentage of success in performance and knowledge variants

Formulae	**Student #8**	**Student #9**
$P = \left(\frac{CorF}{CorF + IncF}\right) \times 100$	$\frac{690}{690+175} \times 100 = 79.8\%$	$\frac{403}{403+124} \times 100 = 76.5\%$
$K_1 = \left(\frac{CorF + M}{CorF + IncF}\right) \times 100$	$\frac{690+24}{690+175} \times 100 = 82.5\%$	$\frac{403+58}{403+124} \times 100 = 87.5\%$
$K_2 = \left(\frac{CorF + M + M_{nc}}{CorF + IncF}\right) \times 100$	$\frac{690+24+0}{690+175} \times 100 = 82.5\%$	$\frac{403+58+0}{403+124} \times 100 = 87.5\%$
$K_3 = \left(\frac{CorF + M + E_{nc}}{CorF + IncF}\right) \times 100$	$\frac{690+24+6}{690+175} \times 100 = 83.2\%$	$\frac{403+58+0}{403+124} \times 100 = 87.5\%$
$K_4 = \left(\frac{CorF + M + E_{nc+}M_{nc}}{CorF + IncF}\right) \times 100$	$\frac{690+24+6+0}{690+175} \times 100 = 83.2\%$	$\frac{403+58+0+0}{403+124} \times 100 = 87.5\%$

While the four calculations of knowledge (K_1 to K_4) for student #9 are equal to 87.5%, which implies that all mistakes and errors were confirmed, the scores achieved by student #8 vary from 82.5% to 83.2% due to six errors not confirmed. Student #9 illustrates the case that although some of the information on feedback access is not known, the error mistake distinction can on some occasions be nevertheless operated (see Table 5.7 on page 137 for the conditions). When the information on feedback access is fully known, K1, K2, K3 and K4 are all equal; errors and mistakes are confirmed.

Thus far, the language accuracy scores in current performance, actual development, and potential development – for most students – are identified and stored. From these measures, the learners' ability to perform their knowledge can be inferred and represented (Figure 5.22 below). A spider chart is preferred as it shows the size of the gap between 100% success rate and the learners' actual achievements in terms of performance and competence. Furthermore, the spider chart is also described as one "powerful graphical tool for evaluating [...] performance" (Rogers 1995 p.16).

Figure 5.22. Representing learners' overall current performance, actual and potential development, zone of proximal development, and ability to perform knowledge

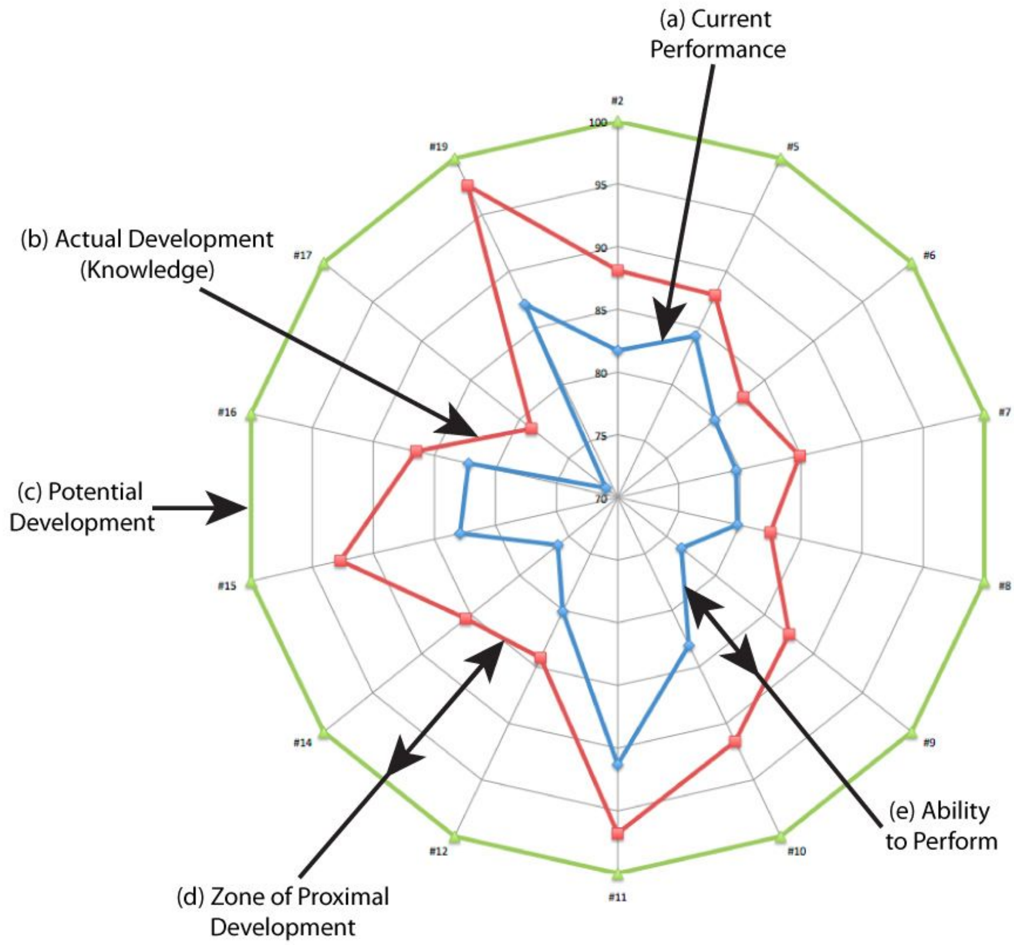


Figure 5.22 above illustrates (a) the learners' overall current performance calculated by counting each correct and incorrect form in the whole corpus, (b) their level of actual development by taking out mistakes from their current performance, (c) their theoretical level of potential development, i.e., what learners would ideally be able to do at the final stage of assistance, where the correction is provided, (d) the distance between the actual

and potential development, that is, their zone of proximal development, and (e) the distance between the current performance and the actual development, that is, their ability to perform knowledge. The figure shows, for instance, that while the performance achieved by student #17 is equal to 71.2%, her actual knowledge reaches 78.8% (language accuracy, all error types taken together). Additionally, the figure illustrates her zone of proximal development which is comprised between 78.8% and 100%, what she is able to hypothetically achieve with the help of the corrector at level 4 in the regulatory scale of assistance. The spider chart further indicates that learner #17 made a relatively high amount of mistakes compared to learner #6. While the former wrote 173 mistakes in 658 incorrect forms (26.29%), the latter produced 37 mistakes in 254 incorrect forms (14.56%). This indicates that learner #17 has a lower ability to perform her knowledge than student #6 whose performance is closer to her competence. The ability to perform knowledge is calculated as follows:

Formula 5.11. Ability to perform knowledge

$$A = \left(\frac{P}{K}\right)$$

where *A, P,* and *K* equal ability, performance, and knowledge, respectively.

An ability score less than 1 indicates that the learner knows better than she can perform, whereas an ability score greater than 1 would have denoted that the learner could perform better than she knows. In such a case, the learner would have guessed correct forms when writing her document independently.

5.6. Empirical analyses: synchronic and diachronic

To investigate the process of cognitive changes during a short span of time, microgenesis has proved to be well adapted (Lantolf 2005). Not only, does microgenesis consider changes in performance over relative short periods of time, where short-time spans may designate minutes, weeks, and even months, it also views "variability as an important phenomenon, rather than as a nuisance to be minimised" (Flynn and Siegler 2007 p.139). The texts as well as the intervals on the timeline – going from October 2008 to April 2009 – at which students submitted their written productions are listed in Table 5.10.

Table 5.10. Timeline submission

	T1: Oct 2008	**T2: Nov 2008**	**T3: Dec 2008**	**T4: Feb 2009**	**T5: Mar 2009**	**T6: Apr 2009**
#2	1 bilan 7 forum 2 email	4 wiki 2 bilan 6 email 2 écrire	2 wiki 1 bilan 6 email 1 parler	3 email 1 écrire	3 email 2 écrire	1 bilan 5 email 1 écrire
#5	-	-	1 wiki	-	-	-
#6	-	-	1 bilan 2 wiki	-	-	-
#7	-	-	1 bilan 3 wiki	-	-	-
#8	-	-	2 wiki	-	-	-
#9	-	-	1 survey 1 wiki	-	-	-
#10	-	-	1 bilan 1 wiki	-	-	1 bilan
#11	-	-	1 parler 1 wiki	-	-	-
#12	-	-	1 wiki	-	-	-
#14	-	-	1 wiki	-	-	-
#15	-	-	1 wiki	-	-	-
#16	-	-	1 bilan 1 wiki	-	-	-
#17	-	-	1 wiki 1 bilan	-	1 écrire	1 bilan
#19	-	-	1 wiki	-	-	-

The three different options of data analysis that emerge from Table 5.10 are a synchronic inter-learner analysis, a diachronic inter-learner analysis, and a diachronic intra-learner analysis.

5.6.1. Synchronic inter-learner analysis

> **A synchronic inter-learner analysis**: examination of the data at a single point in time (December 2008), across all participants (#2, #5, #6, #7, #8, #9, #10, #11, #12, #14, #15, #16, #17, and #19) and two possible text types (wikis and bilans).

The aim of the synchronic inter-learner analysis is to examine the variations in both interlanguage performance and interlanguage competence at one unique single point in time (T3), and to compare the results obtained between learners. Learners' actual and potential development are observed within this one-time snapshot exclusively. This means that the potential development of the learners, i.e., an estimate of their future knowledge after T3, is observed at T3.

5.6.2. Diachronic inter-learner analysis

> **A diachronic inter-learner analysis**: examination of the data of students #10 and #17 over a short time span (from December 2008 to April 2009) and across one unique text type (bilan).

The diachronic inter-learner analysis intends to observe the variations in-between learners' levels of actual and potential development at two distinct points in time (T3 and T6), and to investigate the types of changes that occur between the two snapshots. Because Vygotsky's sociocultural theory of mind offers the possibility of examining the learners' developmental process, the measures obtained at both points (T3 and T6) are juxtaposed to help determine (a) the types of variations between both measures, and (b) whether the potential development observed at T3 can be a prediction of knowledge that will occur at T6.

5.6.3. Diachronic intra-learner analysis

> **A diachronic intra-learner analysis:** examination of the data of student #2 over time (from October 2008 to April 2009) and across text types (forum, bilan, email, écrire, wiki, and parler).

The diachronic intra-learner analysis aims at measuring one student's knowledge over a longer time frame and at several points in time (T1, T2, T3, T4, T5, T6), and to observe the variations in-between text types during this time frame. The actual and potential development of the learner are thus investigated to determine whether the interlanguage competence is subject to random and/or systematic variations, and so as to establish whether the potential development observed at one point in time may be used to predict the knowledge that is observed in the next points.

5.7. Summary and conclusion

The chapter opened with the method proposed to improve the tagging accuracy when processing language learners' written documents. The method is based on three sequential steps: firstly, the identification of unknown lemmas through the means of a lowercase conversion and a word bank; secondly, the reduction of incorrect tags consistently applied with a set of additional rules, and thirdly, the elimination of incorrect tags by cross-referencing the part-of-speech tags with the learners' error-annotated corpus. The section concluded by displaying the results obtained.

The instruments necessary for data collection and annotation were presented. More specifically, the web-based applications utilised by the learners to provide corrections to their incorrect forms were described. The section also showed the tools required by this researcher to pre-process, error-annotate learners' texts, and validate the learners' alternatives provided to their incorrect forms when self-editing their texts. Additionally, it outlined how the learners' actions with regard to feedback access when correcting themselves were monitored.

After presenting the data organisation and an overview of the data collected, the different metrics used to explore learners' interlanguage were illustrated. The first analysis focused on the actualisation of the expanded model proposed in this thesis, so that errors and mistakes can be distinguished. Thereafter, specific measures were described to examine the learners' language accuracy and complexity, so that learner knowledge can be represented and investigated. More specifically, this section showed how the concept of obligatory occasion analysis could be extended to error types to calculate the learners' language accuracy, and it outlined the balanced complexity measure, which is used to represent interlanguage complexity.

The texts collected and annotated provide data for a cross-sectional and longitudinal empirical analysis. Chapters 6 and 7 will present the findings with regard to interlanguage competence variability obtained through synchronic and diachronic investigations, respectively.

Chapter 6. A synchronic inter-learner analysis

The previous chapter presented the tools, as well as the methods to collect, annotate, and analyse the corpus. In particular, it listed and explained the various approaches used in this study to distinguish between a competence-dependent error and a performance-related mistake, and to represent learner knowledge. The means previously described are applied within this chapter in order to answer the research questions listed in the introduction. More specifically, this chapter proposes to analyse interlanguage competence across students at a single point in time.

The chapter starts with the examination of the learners' responses to interventions since they (a) determine the placement of their incorrect forms in the zone of proximal development, and (b) help distinguish between competence-dependent errors and performance-related mistakes. The analysis focuses then on learners' errors and mistakes, and their frequencies depending on the main error categories. The following section explores the variations occurring in actual and potential development between students, and in particular within the learners' zone of proximal development. More specifically, it inspects the errors and mistakes learners produced in selection, syntax and morphosyntax.

6.1. Learners' responses to interventions

As mentioned in Section 4.3.3 on page 98, the data analysed within this chapter includes the wiki and bilan text types written by fourteen participants (#2, #5, #6, #7, #8, #9, #10, #11, #12, #14, #15, #16, #17, and #19) in December 2008 (Table 6.1). Given the fact that the texts written by students #5, #8 and #9 were submitted before the implementation of the monitoring system, the four texts written by these three students are not taken into consideration in the present analysis (this section only), as the information on feedback access is not available.

Table 6.1. Data considered for the cross-sectional analysis

Student id	Text id	Text type	Token count	Incorrect sequences	Incorrect tokens	Incorrect sequence rate	Incorrect token rate	Date
2	65	bilan	485	69	111	14.2%	22.9%	12/2008
	34	wiki	35	7	9	25.7%	25.7%	12/2008
	48	wiki	446	33	48	7.4%	10.8%	12/2008
5	28*	wiki	420	43	66	10.2%	15.7%	12/2008
6	64	bilan	536	74	111	13.8%	20.7%	12/2008
	36	wiki	293	40	69	13.7%	23.5%	12/2008
	37	wiki	433	44	74	10.2%	17.1%	12/2008
7	68	bilan	474	82	131	17.3%	27.6%	12/2008
	35	wiki	145	9	13	6.2%	9.0%	12/2008
	57	wiki	388	46	62	11.9%	16.0%	12/2008
	58	wiki	45	8	9	17.8%	20.0%	12/2008
8	38*	wiki	271	52	83	19.2%	30.6%	12/2008
	39*	wiki	594	75	101	12.6%	17.0%	12/2008
9	42*	wiki	453	79	96	17.4%	21.2%	12/2008
10	63	bilan	471	56	85	11.9%	18.0%	12/2008
	44	wiki	417	51	79	12.2%	18.9%	12/2008
11	45	wiki	661	49	64	7.4%	9.7%	12/2008
12	46	wiki	366	37	73	10.1%	19.9%	12/2008
14	50	wiki	373	60	89	16.1%	23.9%	12/2008
15	51	wiki	381	49	65	12.9%	17.1%	12/2008
16	53	bilan	514	49	106	9.5%	20.6%	12/2008
	52	wiki	488	47	72	9.6%	14.8%	12/2008
17	67	bilan	493	99	133	20.1%	27.0%	12/2008
	55	wiki	413	74	110	17.9%	26.6%	12/2008
19	61	wiki	425	39	55	9.2%	12.9%	12/2008
Total			10,020	1,271	1,914	12.7%	19.1%	12/2008

* Texts for which there is no information on feedback access.

The data for the present analysis (minus students #5, #8, and #9) thus includes 1,022 incorrect sequences in total, for which 1,022 error types and 1,022 meta-linguistic feedback were given to students at levels two and three, respectively. Figures 6.1 and 6.2 below provide an overview of the learners' behaviour with regard to feedback access at both levels two and three. In addition, the diagrams display the amount of alternatives provided (with and without assistance), and whether these alternatives were correct.

Figure 6.1. Overview of feedback access and alternatives provided at L2

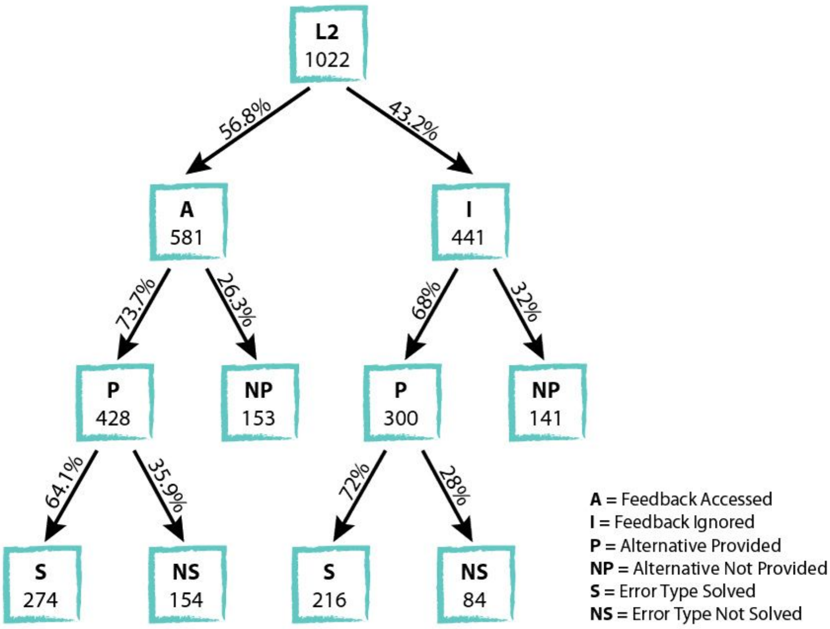
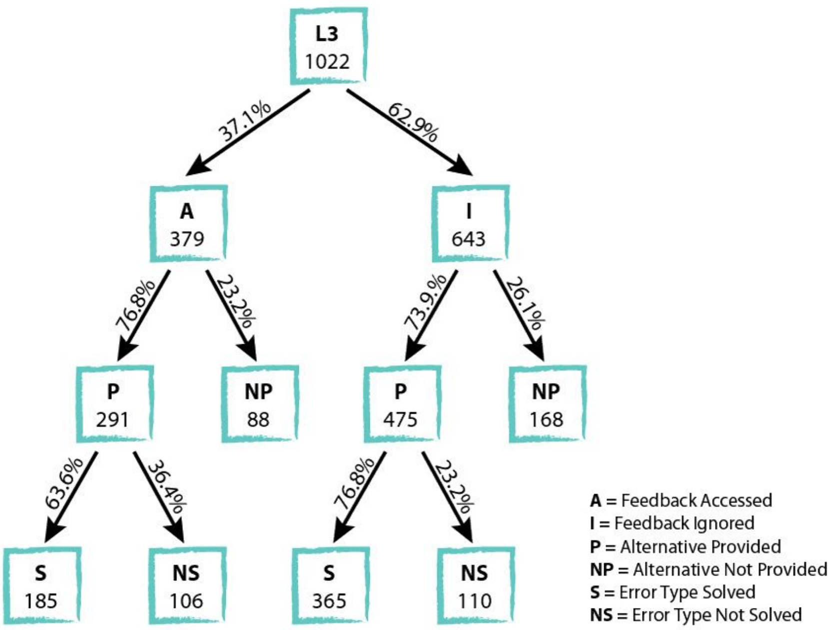


Figure 6.2. Overview of feedback access and alternatives provided at L3



From the above illustrations, one can notice that 53% (N=1084) of the annotations were not accessed (L2: N=441; L3: N=643), which implies that they were not read. Table 6.2 provides the percentage of feedback each student did not access in their wiki/bilan at both levels, and outlines different behaviours with regard to acceptance of assistance.

Table 6.2. Percentage of feedback not accessed per student

Students	**Amount of feedback provided at levels 2 and 3**	**Amount and % of feedback not accessed at levels 2 and 3**	
#2	218	48	22.0%
#6	316	295	93.4%
#7	290	128	44.1%
#10	214	93	43.5%
#11	98	67	68.4%
#12	74	22	29.7%
#14	120	35	29.2%
#15	98	51	52.0%
#16	192	93	48.4%
#17	346	192	55.5%
#19	78	60	76.9%

Student #2 for instance opened 78% of all her feedback (22% were ignored), whereas student #6 did not access 93.4% of them (6.6% were accepted). Mediation is considered as *ignored* when learners did not access the assistance made available to them at the different stages of their correction task. Alternately, accessing the assistance presupposes that learners *accepted* it.

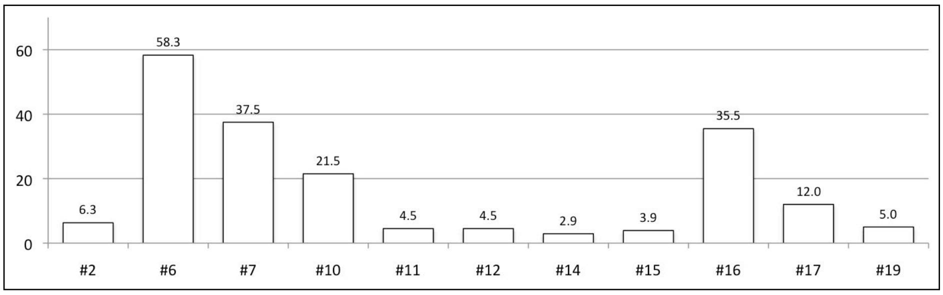
6.1.1. Ignoring the assistance

In the event of learners' refusal to access feedback, that is, when learners did not open the pop-up windows that included the error types (level two in the regulatory scale) or the metalinguistic feedback (level three), learners could either provide alternatives or leave the fields blank. Altogether students did not propose any alternatives to 28.5% (N=309)[45]

45. Feedback ignored: L2: N=441; L3: N=643. Alternatives not provided L2: N=141; L3: N=168 (Figures 6.1 and 6.2 on page 154).

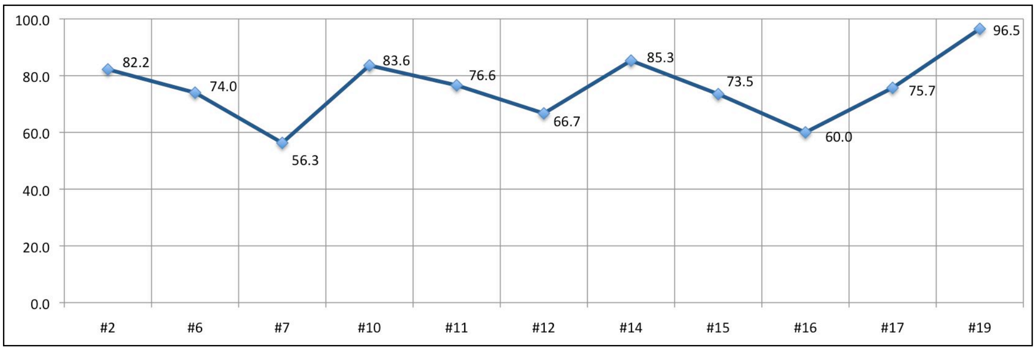
of all incorrect forms for which the feedback was not accessed/read. Breaking down the analysis per student, Figure 6.3 illustrates the percentage of incorrect forms that were not attempted when the assistance was not accessed. While student #6 scored rather weakly in terms of engagement with 58.3% of incorrect forms left without any replacement, the other students generally proposed alternatives to their incorrect forms despite the fact that they did not access the assistance. For example, student #14 did not provide any alternatives to 2.9% of all his incorrect forms for which no feedback was read, which implies that 97.1% of his incorrect forms were attempted, whether successful or not.

Figure 6.3. Percentage of incorrect forms that were not attempted when feedback was not read



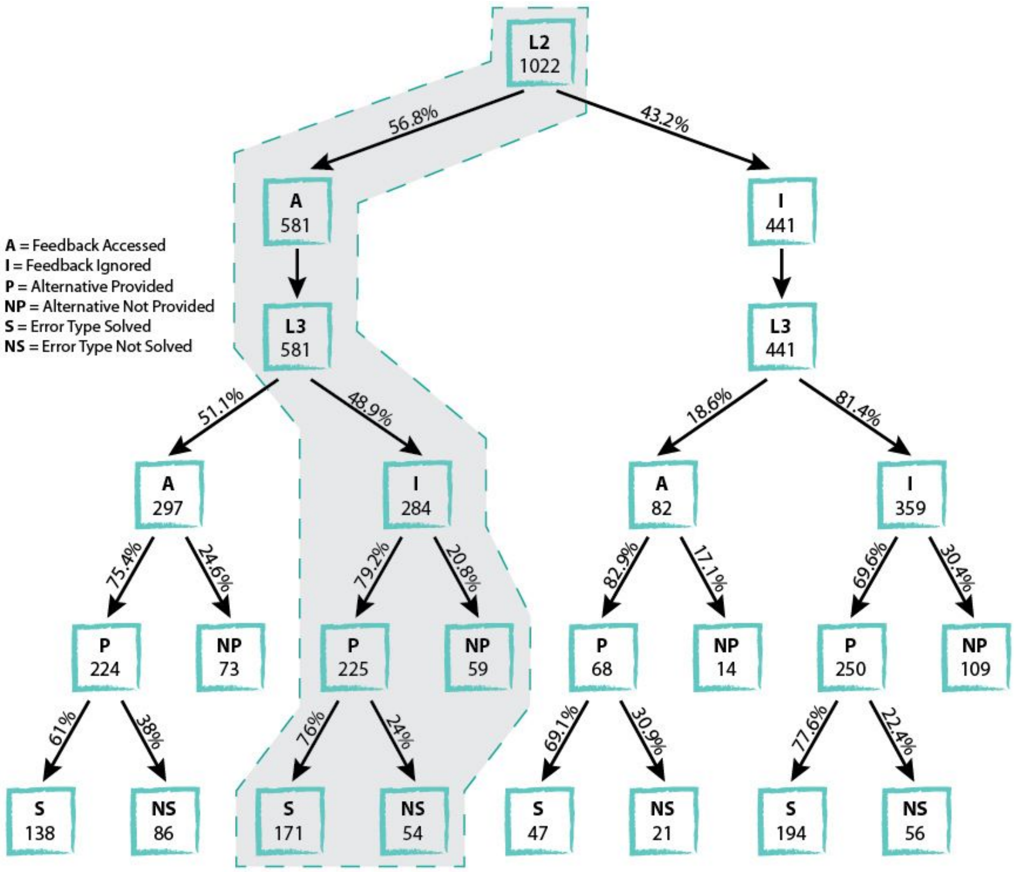
The percentage of correct alternatives when assistance is not accessed is illustrated in Figure 6.4. For example, while student #7 was probably over confident with 56.3% of success when editing her text, student #19 was able to successfully correct 96.5% of his incorrect sequences for which no assistance was required. For most students, the percentage of success in correcting themselves without assistance is higher than 70% with an average of 75.5%. The initial expectation was that students would consistently require assistance to help them correct themselves, and that providing them with feedback was an essential feature of their correction phase. This graphic suggests that (a) learners do not require assistance in a systematic way and (b) if learners do not seek help when proposing a replacement to their incorrect forms, it is most likely because they do not need it.

Figure 6.4. Percentage of correct alternatives without seeking assistance



Another observation that can be made from the data is that students are more inclined to provide alternatives without assistance at level three (N=475) than at level two (N=300). Figure 6.5 below shows (grey area) that students did not access 48.9% of their feedback at L3, for which they required assistance at L2.

Figure 6.5. Alternatives provided at L3 depending on feedback access at L2



One probable explanation for not accessing the assistance and providing an alternative at L3 is that learners did not need to re-access it after reading the feedback at L2. For example, student #12's alternatives below (Table 6.3) illustrate the possibility that she probably did not need feedback at level three after accessing it at level two.

Table 6.3. Alternatives provided by student #12 at L3 without seeking assistance

Student #12	**Correct alternatives at level 3 without assistance**	**Incorrect alternatives at level 3 without assistance**
Situation #1	**At level 2: feedback accessed and correct alternatives provided**	
	• [allemand (*Allemande), *German*] • [Certaines (*certaines), *some*] • [cours, je (*cours je), *lecture, I*] • [nous (*on), *we*] • [Quand (*quand), *When*] • [regardons (*regarder), *see*] • [retourne (*retournerai), *go back*] • [va (*aller), *go*] • [voyage (*Voyage), *voyage*]	• [*en les (*dans les), *in the*]
Situation #2	**At level 2: feedback accessed and incorrect alternatives provided**	
	• [appartement (*appartements), *apartment*]	-
Situation #3	**At level 2: feedback accessed and no alternatives provided**	
	• [cours (*conférences), *lectures*] • [chaque (*ce), *each*]	-
Situation #4	**At level 2: feedback not accessed and correct alternatives provided**	
	-	• [*une (*à une), *one*]
Situation #5	**At level 2: feedback not accessed and no alternatives provided**	
	• [parce que (*que), *because*] • [pour (*de), *for*]	-

In a sequence [word1 (*word2), *word3*], word1 represents the alternative provided by the student at level 3, and word2 refers to the initial incorrect form, followed then by word3, an approximative translation.

Table 6.3 above shows that student #12 entered 16 alternatives at level three without seeking assistance. In addition, it indicates that she was able to correct 10 incorrect sequences at level two with assistance (Situation #1), and that she provided correct alternatives to 9 of them at level three without assistance. The reading of the error type at level two can be considered as a refresher, which could explain why the learner did not feel the need to open the feedback at level three to correct herself. The fact that she could provide a correct alternative at L3 without assistance and that she could not at L2 with assistance (Situations #2 and #3), or without assistance (Situation #5) cannot be clearly justified from the data only. Situation #4 may illustrate the case of a lucky guess as she was able to provide a correct answer at L2 but not at L3, both without assistance.

6.1.2. Accepting the assistance

Accepting the assistance designates that the feedback was opened long enough to be read (see discussion in Section 5.4.4 on page 127). Table 6.4 below shows that over the 960[46] feedback accessed, learners suggested alternatives to 719 (74.9%) incorrect forms.

Table 6.4. Alternatives provided after reading the feedback

Amount of alternatives provided		**Amount of alternatives not provided**	
719	74.9%	241	25.1%

The analysis of whether an alternative was provided after reading the feedback shows varying percentages when splitting the data per error category. Figure 6.6 displays the percentage of all alternatives provided in selection, syntax, morphosyntax, misspelling, and typography, and Figure 6.7 shows the percentage of correct alternatives provided within these categories after reading the feedback at both levels two and three.

46. Feedback accessed at level 2: N=581; at level 3: N=379. Alternatives provided at level 2: N=428; at level 3: N=291 (Figures 6.1 and 6.2 on page 154).

Figure 6.6. Alternatives provided after reading the feedback per error category

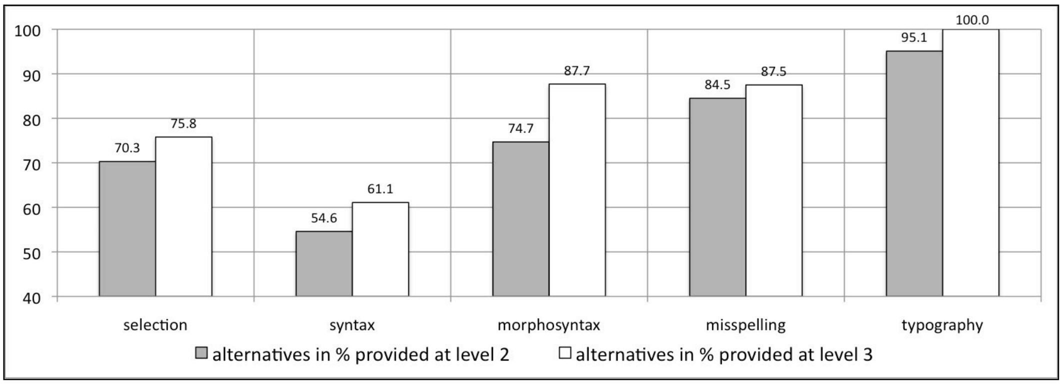
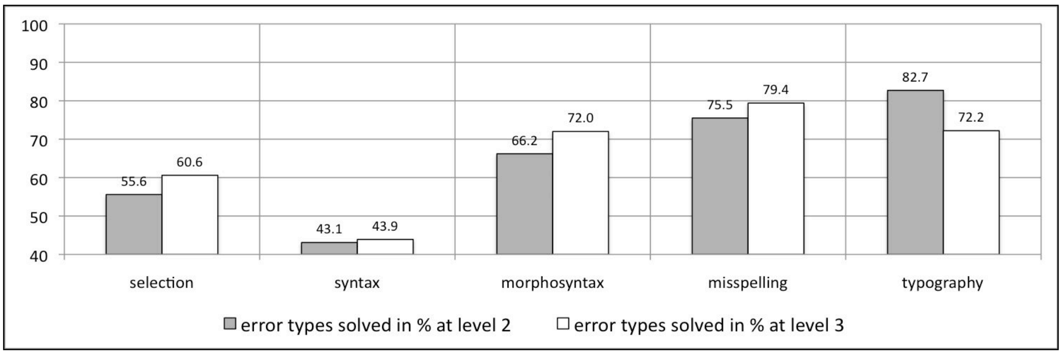


Figure 6.7. Correct alternatives provided after reading the feedback per error category

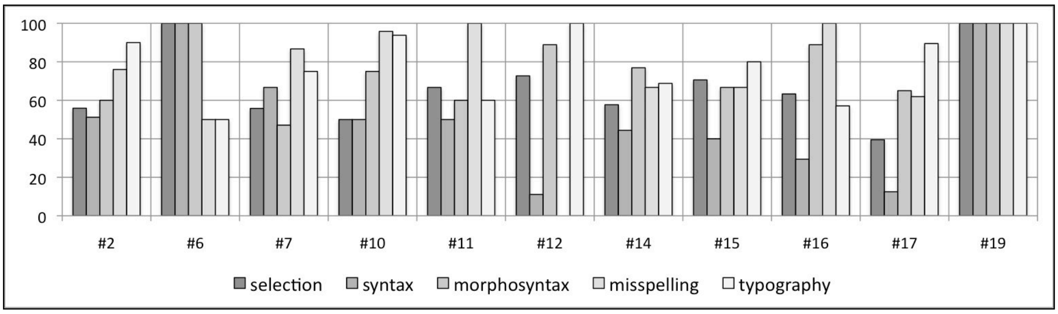


Learners seem to have less difficulties in proposing alternatives to incorrect sequences related to selection, morphosyntax, typography and misspelling. However, when the focus is on syntax, learners tend to be more challenged. While participants provided 95.1% of alternatives at level two in typographic error types, 82.7% of them were resolved. The percentage of accuracy decreases with meta-linguistic feedback, which implies that a level two of assistance would be sufficient for learners to correct typographic errors. Only 55.6% of alternatives to selection error types, 43.1% of alternatives to syntactic error types, and 66.2% of alternatives to morphosyntactic error types were correct after reading the feedback at level two. The figures improve when learners are provided with meta-linguistic feedback, 60.6%, 43.9% and 72.0%, respectively.

The effect of corrective feedback on learner uptake (e.g., Heift 2004, Heift and Rimrott 2008, Lyster and Ranta 1997) or the effect of feedback types in relation to error types (e.g., Lyster 1998) is well documented in the literature. Lyster (1998), for example, demonstrated that “lexical errors favoured the negotiation of form [and] grammatical and phonological errors invited recasts” (p.184). While his findings point to the possibility that L2 learners of French would be able to correct more incorrect lexical choices than incorrect agreements, Figure 6.6 and Figure 6.7 above show that learner uptake and ability to correct themselves are higher in morphosyntactic than selection error categories. With regard to the relationship between error categories and learner uptake, it is rather difficult to determine how the results displayed in both Figure 6.6 and Figure 6.7 compare with findings from other studies as the feedback and/or the error type classification may differ from this research methodology.

In general terms, it may be postulated that learners tend to have more difficulties in providing correct alternatives after accessing the assistance for selection, syntactic and morphosyntactic error types than misspelling and typographic categories. However, it could be argued that the participants’ higher ability to correct themselves in misspelling and typographic categories could be influenced by the content of the feedback itself, and more specifically by its length. Indeed, the feedback for those error types generally includes fewer words than the ones for selection, syntactic and morphosyntactic categories, which implies a more comprehensible metalinguistic feedback. Furthermore, when looking at the distribution of correct alternatives per student (Figure 6.8), the conclusion to be drawn in terms of learners’ ability to correct specific error categories is different.

Figure 6.8. Percentage of correct replacements provided by students at L2/L3 with assistance



Student #19, for instance, does not struggle with any of the error types, since he achieved 100% of correct alternatives in all five error categories after accessing the assistance. Selection, syntax and morphosyntax are not an issue for student #6 either, who performed 100% correctly after reading the feedback. As for student #12, she could not correct herself even with a full explanation about the incorrect forms in the syntax feedback. The synchronic snapshot between learners illustrated in Figure 6.8 above demonstrates that one individual may have issues in one category and skills in development in another. Such strengths or weaknesses, which may not correspond to the group average, can illustrate the case of idiosyncratic interlanguage variability between participants. It is therefore relevant to explore the data of each learner as an individual rather than as a group average.

6.1.3. Negotiating the assistance

While the computer-based application used to collect the data was not originally designed for negotiation, student #14 discussed (literally) the content of one feedback. The sequence marked as *not understandable* in terms of word meaning is as follows:

> [5]. *Je commence chaque jour avec deux Wheatabix *d'être polarisé sur mon jour*. (I always start my day with two Wheatabix in order to be physically and morally fit for the day)*

The participant attempted an unfruitful alternative at level one, i.e., without assistance, **à être polarisé sur mon jour*. Then, the learner accessed the feedback at level two twice with an average time of 15 seconds each, yet without providing any alternatives. At level three, the learner accessed the meta-linguistic annotations three times with an average time of 11 seconds for each reading, and tried to reformulate the sequence differently **d'être en forme*. After the self-editing exercises, the student came to the corrector and said: "I don't think you understood properly what I meant to say". This was an interesting comment, which led to a conversation in French about the meaning of the whole sentence, and the meaning of the word *polariser (polarise)* in this context. The participant added that he was tempted to argue during the self-editing exercises, and wished to write (in English) the reasons for which he thought his formulation was correct. Yet, the learner did not take the liberty to do so, since he said the guidelines were to enter alternatives or to leave the field blank if he had no clue how to edit the incorrect sequence.

Although student #14 sent only 373 words in one text for correction, he respected the guidelines and proposed replacements to almost all of his incorrect forms (only 2.9% of them were left without corrections). His engagement in this study and willingness to correct himself pressed him to ask for additional assistance outside the self-editing exercises. This single example of negotiation retrieved from this research cannot be generalised, but it certainly opens the door to the possibility of integrating learners' interactions as opposed to interventions in computer-based dynamic assessment targeting written language.

6.1.4. Discussion

The aim of this first section was to address the learners' contribution to dynamic assessment and to investigate how learners respond to assistance when correcting their texts written in the French language. How learners responded to feedback is characterised as learner uptake. Lyster (2007) outlines that learner uptake does not merely refer to "what learners claim to have learned from a particular lesson" (p.117). Rather, it is "concerned not only with feedback itself but also with the range of possible learner responses to feedback" (Lyster 2007 p.117). The motivation behind monitoring learners' access to feedback is justified by the fact that knowing whether learners responded to interventions is primordial to help represent their zone of proximal development.

The assistance offered to learners was displayed on request, as opposed to systematically being given with each incorrect form. As Clarebout and Elen (2006) suggested, the use of such tools should presuppose "that learners are good judges of their learning needs" (p.390). It was not only demonstrated that a large portion of annotations designed to help learners correct themselves were never accessed, but also established that learners' refusal to access feedback was not to be interpreted as a definitive refusal of engagement. Rather, a learner's choice of not opening the pop-up windows might be interpreted as a demonstration of self-confidence, as they nevertheless proposed replacements to a large amount of their incorrect forms. Furthermore, considering the fact that the majority of their replacements provided without assistance were correct, this suggests that learners did not need help. However, the reasons for not providing any alternatives when the assistance was not accessed are not so straightforward. Firstly, the learners may have been more interested by the final correction and may have decided to skip some of the incorrect forms along with their feedback so as to attain the final level faster. Secondly, given the fact that a comprehensive error annotation was adopted**,** learners may have felt overwhelmed by the amount of highlighted incorrect forms to be corrected, and thus decided to skip some of them in order to lighten the burden of correcting themselves. Thirdly, learners may have thought that they would not be able to self-edit a particular incorrect form even with feedback, so there was no point of reading the assistance. In addition to focusing on learners' refusal of assistance, this section investigated the extent to which learners accepted and negotiated help. Generally, alternatives were proposed after learners accessed the assistance. Reasons for which the learners left the field blank after reading the feedback might be that they did not understand the message, or that the message was not adapted to their incorrect forms, or simply that they did not know how to correct themselves even after being provided with help. Furthermore, not entering alternatives after reading the feedback has been shown to be a form of negotiation with learner #14. He would have argued the content of the message at the time of the self-editing exercise, but did not do so as the guidelines were to enter an alternatives if he knew one.

Identifying learners' behaviour in terms of access to feedback may assist students and teachers alike in reframing the type of assistance that is required in order for learners to self-edit their incorrect forms in the long term, that is, to help them perform beyond their level of current performance. However, and more importantly for this thesis, it also helps to distinguish between errors and mistakes in learner interlanguage performance.

6.2. Learners' errors and mistakes in interlanguage

As explained earlier, the distinction between errors and mistakes is determined by (a) the alternatives learners proposed in replacement to their incorrect forms, and (b) the level of assistance they required to correct themselves, or more specifically whether they ignored or accepted assistance to provide alternatives. Table 6.5 below lists the text types submitted by each student during the time frame under investigation within this chapter, along with the language level of complexity associated with each text. In particular, it details each vector and provides the overall balanced complexity measure. The motivation behind this is to determine whether wiki and bilan had to be considered as two different text types with regard to lexical and grammatical complexity, or whether they could be merged together and analysed as one, as not all students submitted a bilan.

Table 6.5. Balanced complexity measure for each wiki and bilan

Student id	Text id	Text type	Vector 1	Vector 2	Vector 3	Vector 4	Balanced complexity measure
#2	65	bilan	5.67	4.26	11.55	12.67	23.67
	34	wiki	3.00	4.31	10.67	3.68	11.75
	48	wiki	5.35	4.39	12.24	11.59	23.64
#5	28	wiki	5.81	3.85	11.48	11.75	22.91
#6	64	bilan	5.42	4.50	9.96	12.34	22.31
	36	wiki	5.66	4.59	9.14	10.54	21.89
	37	wiki	6.10	4.13	10.38	12.68	22.66
#7	68	bilan	5.95	4.46	21.35	13.26	26.07
	35	wiki	4.40	4.83	11.18	7.62	19.12
	57	wiki	6.00	4.14	17.35	12.09	24.29
	58	wiki	3.62	5.38	6.50	4.36	14.76
#8	38	wiki	5.27	4.47	6.85	9.57	18.96
	39	wiki	6.31	4.09	9.48	13.86	21.91
#9	42	wiki	6.24	4.02	17.13	12.75	24.96
#10	63	bilan	5.27	4.41	11.56	12.15	23.57
	44	wiki	5.44	4.20	15.42	11.41	23.18
#11	45	wiki	6.82	4.39	17.21	15.38	28.93
#12	46	wiki	4.51	4.13	18.72	10.88	21.58
#14	50	wiki	5.79	4.07	19.76	11.63	23.47
#15	51	wiki	6.34	4.20	18.26	12.09	24.74
#16	53	bilan	5.23	4.21	18.42	13.45	25.05
	52	wiki	5.52	4.13	14.70	12.41	24.12
#17	67	bilan	5.87	4.40	20.41	12.46	25.07
	55	wiki	6.33	4.02	16.43	12.16	24.46
#19	61	wiki	5.09	3.88	14.63	11.83	22.61

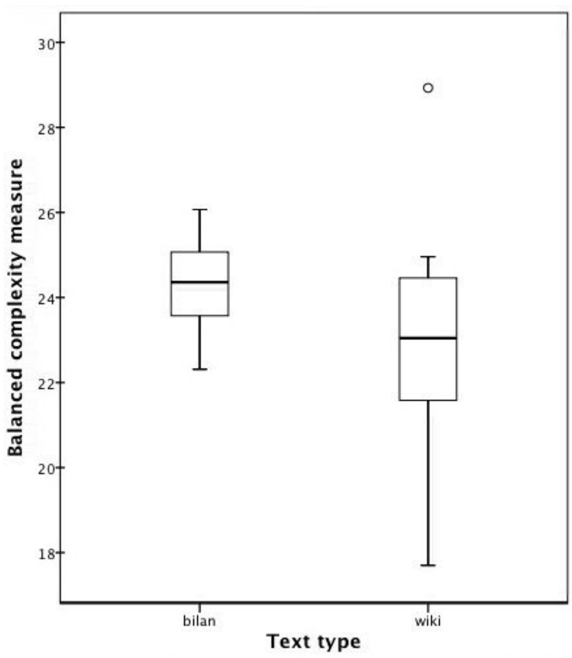
Vector 1: lexical complexity (text-length adjusted type token ratio); **Vector 2**: lexical sophistication (mean length of a word); **Vector 3**: syntactic sophistication (mean period unit length); **Vector 4**: syntactic complexity (unique bigram ratio).

The reader may have remarked that some learners have submitted more than one wiki text. The reason for this is that some participants submitted their wiki in one document (e.g., student #5), and some others sent their wiki texts for correction over a period of several days, in two or even three parts (e.g., student #7).

A repeated measures ANOVA was conducted to assess the effect of text types on balanced complexity measures, so as to decide whether both *wiki* and *bilan* could be considered as one text type. Given the fact that sphericity was not an issue as the repeated measures variable only had two levels, SPSS repeated measures demonstrated that there were no significant differences between the *wiki* and *bilan* groups ($F(1,5)=3.86$, *p*=.107)[47]. With a p>0.05, the mean difference between the two text types is estimated to be not statistically significant. These results suggest that whether learners wrote for their wiki or their bilan, they wrote with an identical lexical and syntactical level of complexity. Consequently, no differences were made between these two text types during this time frame.

Additionally, a boxplot (Figure 6.9 below) illustrates the distribution of the balanced complexity coefficients. Through this graphical representation, the degree of dispersion and skewness in the data can be seen at a glance. The medians, representing the middle values, equal 24.36 (bilan) and 22.73 (wiki). While 75% of the bilan dataset ranges from a coefficient of 22.31 to 25.32, half of the wiki dataset fluctuates between a score of 22.73 and 25.07. The plot indicates that the data spread is to some degree centralised around these values, thus reflecting a relatively homogeneous group with respect to balanced complexity measures.

Figure 6.9. Boxplot of balanced complexity measure per text type



47. The repeated measures ANOVA results are detailed in Appendix E.3 on page 318.

The one-way within subjects ANOVA and boxplot described above show that not only did learners use the same level of language complexity in their wikis and bilans, but they also all performed at a relatively similar level of lexical and syntactical complexity.

6.2.1. Counting errors and mistakes

As detailed in Section 5.5.3 on page 135, each incorrect form was labelled as either an error or a mistake depending on the amount of assistance required to produce a correct alternative, and were then counted. Table 6.6 below, a screen capture of the corpus as seen in SPSS illustrates the information linked to each token necessary to compute learner knowledge. Each token is identified by a number and linked to text and student identifications, so that it can be traced back to its original context. Part-of-speech and error tags are required to calculate language accuracy depending on the status of the incorrect form. As explained in Section 5.5.3, a status of 0, 1, 2, or 3 designates *error confirmed*, *mistake confirmed*, *error not confirmed*, and *mistake not confirmed*, respectively. A status of -1 indicates that the learner's alternative was incorrect. Status 0, on the other hand, signals apart from being a confirmed error that the learner was able to propose a correct alternative with assistance (either at level two or level three), thus outlining a potential development. No status (.) denotes that the token was not marked as incorrect in the first place. The columns P and K1, K2, K3, K4 refer to the learners' performance and knowledge, respectively. As discussed in Section 5.5.4 on page 139, there are four different methods to calculate knowledge depending on whether the error or mistake was confirmed or not. As a result, and depending on the type of calculation, mistakes and errors (confirmed and not confirmed) are either counted as correct (0) or incorrect (1). For example, a confirmed mistake is marked as incorrect (1) in the *performance* column, and correct (0) in the *knowledge* columns. Finally, the *potential development* (PD) column recapitulates all incorrect forms for which learners were capable of providing a correct alternative with assistance, and marked them as 0.

Conversely, incorrect forms learners could not correct, even with assistance, are encoded as 1. Using SPSS descriptive statistics, incorrect forms, errors and mistakes were then counted, so that variations across students' performance and knowledge could be established.

Table 6.6. Data before frequency analysis

token_id	text_id	student_id	pos	error_tagged_as	status	P	K1	K2	K3	K4	PD
7020	38	8	NAM	ty_es_	2	1	1	1	0	0	1
7021	38	8	VER:conj	sy_un_	-1	1	1	1	1	1	1
7022	38	8	VER:conj	sy_un_	-1	1	1	1	1	1	1
7023	38	8	PRP	sy_un_	-1	1	1	1	1	1	1
7024	38	8	PRO:PER	sy_un_	-1	1	1	1	1	1	1
7025	38	8	PRP	correct	.	0	0	0	0	0	0
7026	38	8	PRO:PER	correct	.	0	0	0	0	0	0
7027	38	8	PRP	se_gr_cw_	-1	1	1	1	1	1	1
7028	38	8	NAM	ty_ca_	1	1	0	0	0	0	0
7029	38	8	KON	correct	.	0	0	0	0	0	0
7030	38	8	NAM	ty_ca_	1	1	0	0	0	0	0
7031	38	8	VER:conj	correct	.	0	0	0	0	0	0
7032	38	8	PRP	correct	.	0	0	0	0	0	0
7033	38	8	NOM	sp_ms_	-1	1	1	1	1	1	1
7034	38	8	SENT	correct	.	0	0	0	0	0	0
7035	38	8	SENT	correct	.	0	0	0	0	0	0
7036	65	2	KON	sy_un_	-1	1	1	1	1	1	1
7037	65	2	PRO:PER	sy_un_	-1	1	1	1	1	1	1
7038	65	2	VER:conj	sy_un_	-1	1	1	1	1	1	1
7039	65	2	ADV	sy_un_	-1	1	1	1	1	1	1
7040	65	2	VER:infi	sy_un_	-1	1	1	1	1	1	1
7041	65	2	PRP	se_gr_ar_	0	1	1	1	1	1	0
7042	65	2	DET:ART	se_gr_ar_	0	1	1	1	1	1	0
7043	65	2	NOM	correct	.	0	0	0	0	0	0

P: (errors + mistakes = incorrect forms); **K1**: (mistakes=correct forms); **K2**: (mistakes + mistakes not confirmed = correct forms); **K3**: (mistakes + errors not confirmed = correct forms); **K4**: (mistakes + mistakes not confirmed + errors not confirmed = correct forms).

6.2.2. Juxtaposing performance and knowledge

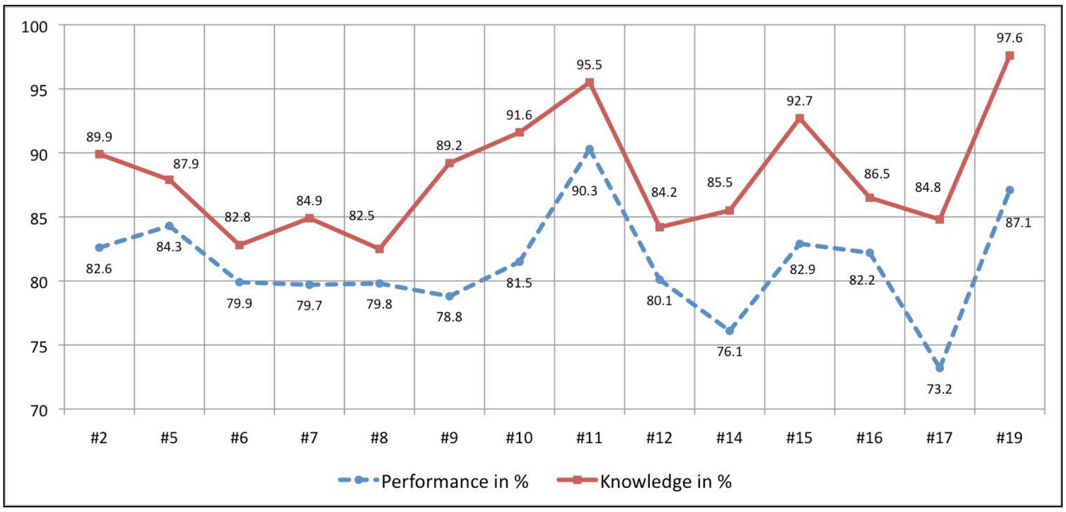
The comparison between learners' performance and knowledge showed disparate results. Table 6.7 below outlines the number of incorrect forms, mistakes and errors, and whether they were confirmed or not. In addition, the above table displays the percentage of mistakes and errors included in learners' incorrect forms. No unconfirmed mistakes were reported in this data sample. Students #5 and #8 however exhibit unconfirmed errors. Given that all incorrect forms were declared as errors by default, these unconfirmed errors were thus considered as competence-dependent errors in the following results.

Table 6.7. Amount of correct and incorrect forms including errors and mistakes

Student	**Total tokens**	**Correct forms**	**Incorrect tokens**		**Errors not confirmed in error total**	**Mistakes not confirmed in error total**	**% of mistakes in incorrect tokens**	**% of errors in incorrect tokens**
			Error	**Mistake**				
#2	966	798	98	70	0	0	41.7%	58.3%
#5	420	354	51	15	15	0	22.7%	77.3%
#6	1262	1008	241	13	0	0	5.1%	94.9%
#7	1052	838	159	55	0	0	25.7%	74.3%
#8	865	690	151	24	6	0	13.7%	86.3%
#9	453	357	49	47	0	0	49.0%	51.0%
#10	888	724	75	89	0	0	54.3%	45.7%
#11	661	597	30	34	0	0	53.1%	46.9%
#12	366	293	58	15	0	0	20.5%	79.5%
#14	373	284	54	35	0	0	39.3%	60.7%
#15	381	316	28	37	0	0	56.9%	43.1%
#16	1002	824	135	43	0	0	24.2%	75.8%
#17	906	663	138	105	0	0	43.2%	56.8%
#19	425	370	10	45	0	0	81.8%	18.2%
Total	10,020	8,116	1,324	580	21	0	30.5%	69.5%

The overall percentage of language accuracy in each learner's interlanguage performance, which is computed by dividing all correct forms by the total amount of correct and incorrect forms contained in learners' texts (CorF/(CorF+IncF)*100), is given in Figure 6.10 below (dashed line). In addition, the figure illustrates the percentage of accuracy in learners' language when considering both errors and mistakes distinctively, thus revealing the learner's knowledge (plain line). This percentage is calculated by dividing the amount of correct forms and mistakes by the total amount of correct and incorrect forms ((CorF+M)/(CorF+IncF)*100).

Figure 6.10. Learners' language accuracy in performance and knowledge



Although students performed at an approximate same level of language complexity, as previously observed, the percentage of accuracy in interlanguage performance and knowledge varies from one student to another, going from 73.2% to 90.3% for performance, and 82.5% to 97.6% for knowledge. The above figure demonstrates that knowledge differs from performance, and that it does not vary in a systematic way. For example, the gap between student #19's performance and knowledge (87.1% and 97.6%) is wider than the one for student #6 (79.9% and 82.8%, respectively).

Without knowing the percentage of mistakes each learner is inclined to produce in his or her written documents, knowledge cannot be intuitionally deduced from performance. While students performed with a relatively homogeneous incorrect form rate (see Table 6.8 below for the percentage details), the percentage of mistakes included within the learners' incorrect forms varies depending on individuals (Figure 6.11).

Table 6.8. Incorrect form rate in learners' texts

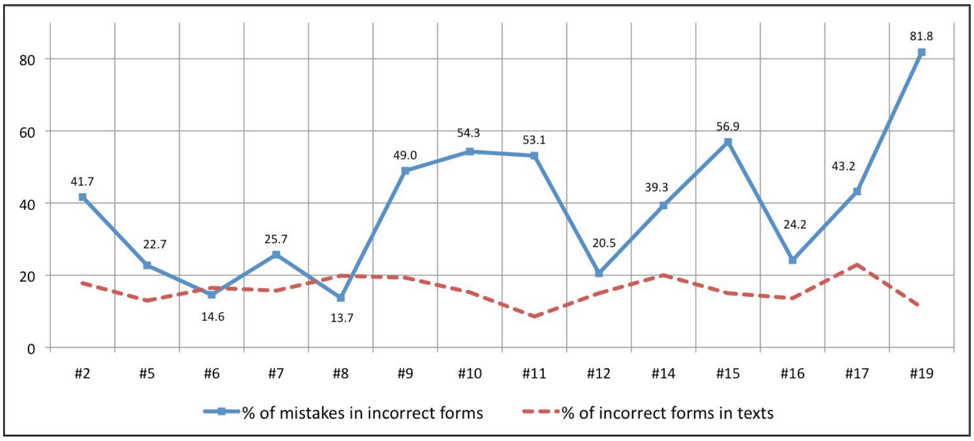
Student	Error type rate[a] in %	Incorrect token rate[b] in %	Average of incorrect form rate[c] in %
#2	15.77	19.80	17.79
#5	10.20	15.70	12.95
#6	12.57	20.43	16.50
#7	13.30	18.15	15.73
#8	15.90	23.80	19.85
#9	17.40	21.20	19.30
#10	12.05	18.45	15.25
#11	7.40	9.70	8.55
#12	10.10	19.90	15.00
#14	16.10	23.90	20.00
#15	12.90	17.10	15.00
#16	9.55	17.70	13.63
#17	19.00	26.80	22.90
#19	9.20	12.90	11.05

a. The error type rate is computed by dividing the total amount of error types, which may contain more than one incorrect token, by the the total amount of tokens.

b. The incorrect token rate is calculated by dividing the number of incorrect tokens by the total amount of tokens.

c. Both the error type rate and incorrect token rate are combined together so as to obtain a balance between the two, which is called average of incorrect form rate.

Figure 6.11. Percentage of (a) mistakes in incorrect forms and (b) incorrect form rate per student



The bottom dashed line in Figure 6.11 above shows the mean percentage of incorrect forms per student, and the plain line indicates the percentage of mistakes that are included in learners' incorrect forms. Some students exhibit high percentages of mistakes suggesting that they were incapable of demonstrating the whole extent of their knowledge such as student #19 with 81.8% of performance-related mistakes in his performance. Some of the other students were able to perform close to their level of knowledge such as student #8 with 13.7% of mistakes. Not only does the percentage of performance-related mistakes vary greatly between students, as mentioned above, it also fluctuates between participants' texts. Table 6.9 below lists the percentages of mistakes included in incorrect forms for each student who submitted more than one text.

Table 6.9. Percentage of mistakes in incorrect forms per student's text

Students	**Text identification**	**Number of incorrect forms**	**Number of errors**	**Number of mistakes**	**Percentage of mistakes in incorrect forms**
2	34	9	1	8	88.9%
	48	48	23	25	52.1%
	65	111	74	37	33.3%
	Total	168	98	70	41.7%
6	36	69	58	11	15.9%
	37	74	61	13	17.6%
	64	111	98	13	11.7%
	Total	254	217	37	14.6%
7	35	13	13	0	0.0%
	57	62	35	27	43.5%
	58	8	5	3	37.5%
	68	131	106	25	19.1%
	Total	214	212	55	25.7%
8	38	74	51	23	31.1%
	39	101	100	1	1.0%
	Total	175	151	24	13.7%
10	44	79	27	52	65.8%
	63	85	48	37	43.5%
	Total	164	75	89	54.3%
16	52	72	47	25	34.7%
	53	106	88	18	17.0%
	Total	178	135	43	24.2%
17	55	110	57	53	48.2%
	67	133	81	52	39.1%
	Total	243	138	105	43.2%

Student #7, for instance, sent four documents for correction. In one of them, she made no mistakes, thus performing at the level of her knowledge. The percentage of mistakes in her other documents however varies from 19.1% to 43.5%. These results emphasise the variable aspect of interlanguage performance, demonstrating that performance-related mistakes are likely to be random, as no systematic patterns could be established between learners and their texts

6.2.3. Errors and mistakes per error category

Splitting learners' incorrect forms per error category shows that the largest number of ill-formed tokens is generally associated with selection and syntactic error types (shaded areas in Table 6.10). For example, student #6 produced 32 incorrect sequences in syntax which include 96 incorrect tokens. In addition, she wrote 61 incorrect sequences in selection which contain 85 incorrect tokens, and between 16 and 31 incorrect tokens in morphosyntax, misspelling and typography.

Table 6.10. Incorrect token and incorrect sequence frequency per error type category

Student	**Selection**		**Syntax**		**Morphosyntax**		**Misspelling**		**Typography**	
	token	**seq.**	**token**	**seq.**	**token**	**seq.**	**token**	**seq.**	**token**	**seq.**
#2	27	24	65	29	21	19	20	20	35	17
#5	21	15	10	3	11	9	14	12	10	4
#6	85	61	96	32	26	26	31	30	16	9
#7	69	59	74	36	27	26	11	9	33	15
#8	60	43	31	13	16	16	37	36	31	19
#9	27	24	13	8	23	23	11	11	22	13
#10	30	23	30	14	27	27	26	23	51	20
#11	27	22	10	6	9	9	4	3	14	9
#12	26	20	31	6	6	5	0	0	10	6
#14	28	20	13	6	11	10	10	9	27	15
#15	23	20	11	7	6	6	7	7	18	9
#16	58	46	76	19	13	12	1	1	30	18
#17	66	56	67	29	30	30	29	28	51	30
#19	10	8	15	9	6	6	8	8	16	8

Splitting learners' mistakes per error category reveals that students did not produce the same percentage of mistakes in all error types, except for student #19 who made more than 80% of performance-related mistakes in almost all categories (Table 6.11 below). In addition, Table 6.11 shows that students were more inclined to produce mistakes in spelling and typographic error types (shaded areas). For example, student #2 produced 31 mistakes in a total of 35 incorrect forms in typography (31/35=0.886), which designates that 88.6% of her incorrect forms in this error category were mistakes.

Table 6.11. Percentage of mistakes in learners' incorrect forms per error type category

St.	**Mistakes in incorrect selection**		**Mistakes in incorrect syntax**		**Mistakes in incorrect morphosyntax**		**Mistakes in incorrect spelling**		**Mistakes in incorrect typography**		**Mistakes in all incorrect forms**	
#2	6/27	22.2%	12/65	18.5%	10/21	47.6%	11/20	55.0%	31/35	88.6%	70/168	41.7%
#5	7/21	33.3%	0/10	0%	4/11	36.4%	4/14	28.6%	0/10	0%	15/66	22.7%
#6	5/85	5.9%	4/96	4.2%	3/26	11.5%	19/31	61.3%	6/16	37.5%	37/254	14.6%
#7	11/69	15.9%	14/74	18.9%	5/27	18.5%	3/11	27.3%	22/33	66.7%	55/214	25.7%
#8	2/60	3.3%	2/31	6.5%	2/16	14.3%	11/37	29.7%	7/31	22.6%	24/175	13.7%
#9	13/27	48.1%	3/13	23.1%	12/23	52.2%	5/11	45.5%	14/22	63.6%	47/96	49.0%
#10	7/30	23.3%	5/30	16.7%	15/27	55.6%	16/26	61.5%	46/51	90.2%	89/164	54.3%
#11	17/27	63.0%	4/10	40.0%	4/9	44.4%	4/4	100%	5/14	35.7%	34/64	53.1%
#12	7/26	26.9%	0/31	0%	3/6	50.0%	-	-	5/10	50.0%	15/73	20.5%
#14	5/28	17.9%	6/13	46.2%	4/11	36.4%	5/10	50.0%	15/27	55.6%	35/89	39.3%
#15	10/23	43.5%	4/11	36.4%	3/6	50.0%	4/7	57.1%	16/18	88.9%	37/65	56.9%
#16	22/58	37.9%	2/76	2.6%	8/13	61.5%	0/1	0%	11/30	36.7%	43/178	24.2%
#17	25/66	37.9%	4/67	6.0%	16/30	53.3%	20/29	69%	40/51	78.4%	105/243	43.2%
#19	4/10	40.0%	15/15	100%	5/6	83.3%	7/8	87.5%	14/16	87.5%	45/55	81.8%

By contrast, Table 6.12 lists the percentages of learners' errors per main category. The shaded areas represent the categories for which learners exhibit the highest percentage. Syntax and selection are the two categories in which learners made most of their errors followed by the morphosyntax. For example, student #2 wrote 53 errors in 65 incorrect forms in syntax (53/65=0.815) denoting that 81.5% of her incorrect forms were errors in this category.

Table 6.12. Percentage of errors in learners' incorrect forms per error type category

St.	**Errors in incorrect selection**		**Errors in incorrect syntax**		**Errors in incorrect morphosyntax**		**Errors in incorrect spelling**		**Errors in incorrect typography**		**Errors in all incorrect forms**	
#2	21/27	77.8%	53/65	81.5%	11/21	52.4%	9/20	45.0%	4/35	11.4%	98/168	58.3%
#5	14/21	66.7%	10/10	100%	7/11	63.6%	10/14	71.4%	10/10	100%	51/66	77.3%
#6	80/85	94.1%	92/96	95.8%	23/26	88.5%	12/31	38.7%	10/16	62.5%	217/254	85.4%
#7	58/69	84.1%	60/74	81.1%	22/27	81.5%	8/11	72.7%	11/33	33.3%	159/214	74.3%
#8	58/60	96.7%	29/31	93.5%	14/16	87.5%	26/37	70.3%	24/31	77.4%	151/175	86.3%
#9	14/27	51.9%	10/13	76.9%	11/23	47.8%	6/11	54.5%	8/22	36.4%	49/96	51.0%
#10	23/30	76.7%	25/30	83.3%	12/27	44.4%	10/26	38.5%	5/51	9.8%	75/164	45.7%
#11	10/27	37.0%	6/10	60.0%	5/9	55.6%	0/4	0.0%	9/14	64.3%	30/64	46.9%
#12	19/26	73.1%	31/31	100%	3/6	50.0%	-	-	5/10	50.0%	58/73	79.5%
#14	23/28	82.1%	7/13	53.8%	7/11	63.6%	5/10	50.0%	12/27	44.4%	54/89	60.7%
#15	13/23	56.5%	7/11	63.6%	3/6	50.0%	3/7	42.9%	2/18	11.1%	28/65	43.1%
#16	36/58	62.1%	74/76	97.4%	5/13	38.5%	1/1	100%	19/30	63.3%	135/178	75.8%
#17	41/66	62.1%	63/67	94.0%	14/30	46.7%	9/29	31.0%	11/51	21.6%	138/243	56.8%
#19	6/10	60.0%	0/15	0.0%	1/6	16.7%	1/8	12.5%	2/16	12.5%	10/55	18.2%

Coupling the percentage of errors with the various vectors of the balanced complexity measures (see Table 6.5 above on page 166) may provide additional useful information to help situate a learner's knowledge within the student group. For example, students #14 and #19 seemed to experience fewer difficulties in syntax than the other participants. The analysis of the syntactical sophistication (vector three of the balanced complexity measure) indicates that student #14 used the highest level of language sophistication in his document (v3=19.8), whereas student #19 scored less than the group average (v3=14.6, average v3=14.9), thus implying a less sophisticated written language. As a result, since student #14 used the most sophisticated language with regard to syntax, he could be considered as a more advanced learner than student #19, despite the fact that the latter made no syntactical errors in his text.

On the other hand, students #2, #5, and #6, for instance, not only exhibited a high percentage of errors in their syntax, but also used a more simple language than the group average. With a coefficient of 9.8, student #6 is noticeably less advanced with regard to syntax sophistication than the other participants within this group.

6.2.4. *Discussion*

This section demonstrated that, although participants performed at a relatively similar level of language complexity, the data supported the idea of variable performance between participants. While many factors such as planning time or task types have been proven to be accountable for performance variability (Section 2.2 on page 21), past and current research has not determined the impact of mistakes on the variable aspect of interlanguage with precision. Varying percentages of mistakes ranging from 13.7% to 81.8% were observed in learners' interlanguage performance. The high percentage of mistakes written in some learners' texts strengthens the idea that not all incorrect forms in a text are to be considered as competence-dependent errors, as generally assumed in the literature (Section 2.3 on page 31). This finding supports the idea that interlanguage performance on its own cannot be used as such to analyse interlanguage competence.

Performance, in most cases, does not reflect the learners' knowledge; it only attests to their usage of the language. It could be assumed that the action of taking out learners' mistakes from their performance could smoothen the student group disparity with regard to language accuracy (they all performed at a similar level of language complexity and they all had a comparable mean error rate). If the student disparity in performance was to be mitigated after the action of removing mistakes from their performance, this would suggest that students with low scores made the highest amount of mistakes. Conversely, students with high scores in language accuracy would make less mistakes. However, the knowledge curve was as irregular as the performance curve. The fact that knowledge is as variable as performance between learners indicates that students with high accuracy scores may also write a high amount of performance-related mistakes, and students with low scores may perform at the level of their knowledge, i.e., without mistakes. As a result, the inter-learner knowledge variability is unsystematic as the percentage of mistakes cannot be predicted. Mistakes are not only a factor that impact learners' interlanguage performance (lowering down language accuracy scores), it also affects the inter-learner knowledge variability as it varies between learners.

In addition, it was established that learners' incorrect forms in selection, syntax and morphosyntax were often associated with a lack of knowledge as opposed to typography or misspellings. It was also demonstrated that learner performance could not be mapped to learner knowledge due to unsystematic performance-related mistakes. While finding the number of mistakes with an average approach could work for a few students[48], this section showed the necessity of analysing individual learners, as a high degree of variability in mistake production was observed between and within participants.

6.3. Variations in actual and potential development

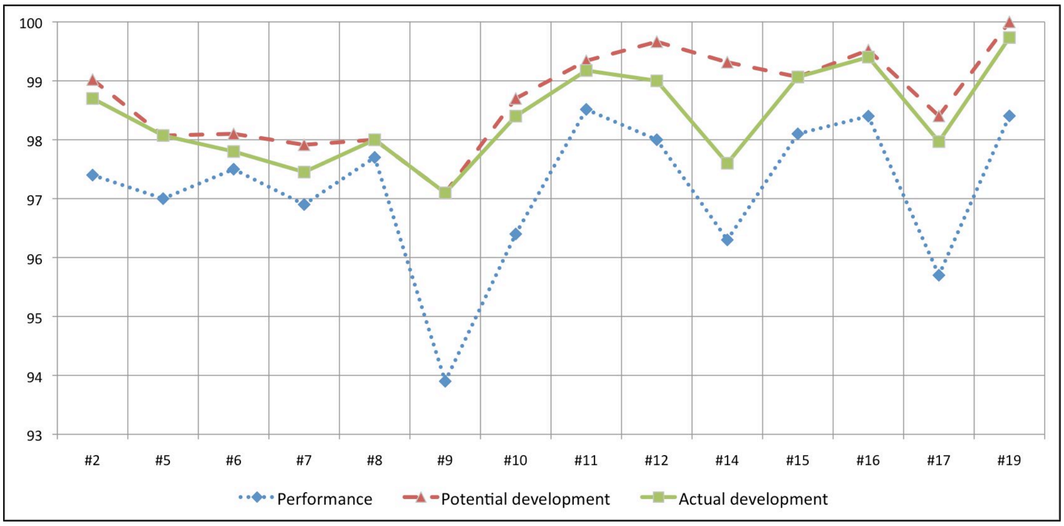
Following the discussion on interlanguage competence addressed in Chapter 2, the types of errors considered within this section are consequently the morphosyntactic, selection, and syntactic error categories.

6.3.1. Morphosyntactic error category

Figure 6.12 below represents the overall results learners attain in morphosyntactic error category. The scores considered reflect the learners' language accuracy in performance, actual and potential development. As a reminder, language accuracy is calculated in (a) performance by dividing all correct forms by all correct and incorrect forms, (b) actual development by considering mistakes as correct forms, and (c) potential development by counting incorrect forms for which learners were able to provide appropriate alternatives with assistance and without being provided with the correct answer.

48. An example of a calculation to estimate the number of mistakes is given in Appendix F.1 on page 319.

Figure 6.12. Language accuracy in performance, actual and potential development: morphosyntactic error types



The dotted line refers to the learners' performance, the plain line illustrates their actual development, and the dashed line shows the extent of their potential development. Frequencies and percentages are detailed in Appendix F.5 on page 322.

As demonstrated in Section 6.2, learners did not have the same ability to perform their knowledge; some students made more mistakes than others. For example, students #2 and #6 exhibit a similar percentage in terms of language accuracy in performance (97.4% and 97.5%, respectively), whereas their levels of actual development demonstrate that student #2 was slightly ahead of student #6 with 98.7% as opposed to 97.8%. On the other hand, students #11 and #14, exhibiting different scores in performance (98.5% and 96.3%), had an identical score in potential development (99.3%), indicating a similar level of cognitive development in morphosyntax. Table 6.13 below details learners' language accuracy per morphosyntactic error type. In addition to listing the performance scores, the table also provides the actual and potential development for each student.

Table 6.13. Percentage of language accuracy in morphosyntactic error types

Students		**dn**	**ge**	**na**	**pa**	**pp**	**sv**	**co**	**pl**	**wf**
#2	Performance	99.0	97.9	89.4	100	94.1	100	96.5	100	100
	Actual Dev.	100	99.0	**93.6**	100	94.1	100	96.5	100	100
	Potential Dev.		100	**93.6**		100		97.7		
#5	Performance	100	93.8	69.2	98.1	-	97.1	97.1	100	100
	Actual Dev.	100	100	**84.6**	**98.1**	-	**97.1**	**97.1**	100	100
	Potential Dev.			**84.6**	**98.1**	-	**97.1**	**97.1**		
#6	Performance	98.1	99.0	89.1	100	66.7	100	100	100	100
	Actual Dev.	**99.0**	**99.0**	89.1	100	66.7	100	100	100	100
	Potential Dev.	**99.0**	**99.0**	90.6		72.2				
#7	Performance	100	98.0	85.4	95.5	69.2	97.8	97.8	100	100
	Actual Dev.	100	100	85.4	95.5	69.2	**97.8**	**98.9**	100	100
	Potential Dev.			90.2	96.6	73.1	**97.8**	**98.9**		
#8	Performance	100	97.6	90.9	100	12.5	97.6	98.8	100	99.4
	Actual Dev.	100	**97.6**	**93.9**	100	**12.5**	**97.6**	100	100	**99.4**
	Potential Dev.		**97.6**	**93.9**		**12.5**	**97.6**			**99.4**
#9	Performance	96.0	98.0	76.5	100	100	97.6	85.4	100	100
	Actual Dev.	**98.0**	100	**94.1**	100	100	100	**89.6**	100	100
	Potential Dev.	**98.0**		**94.1**				**89.6**		
#10	Performance	100	93.5	78.7	100	92.3	95.9	95.9	100	99.6
	Actual Dev.	100	**98.9**	**89.4**	100	**96.2**	**98.6**	98.6	100	**99.6**
	Potential Dev.		**98.9**	**89.4**		**96.2**	**98.6**	100		**99.6**
#11	Performance	100	100	95.0	95.5	50.0	100	96.4	100	100
	Actual Dev.	100	100	100	**97.0**	100	100	**96.4**	100	100
	Potential Dev.				**97.0**			**96.4**		
#12	Performance	100	97.6	100	100	-	100	93.3	100	98.6
	Actual Dev.	100	97.6	100	100	-	100	100	100	100
	Potential Dev.		100			-				
#14	Performance	100	97.4	85.7	100	100	96.2	96.2	100	100
	Actual Dev.	100	97.4	**90.5**	100	100	100	96.2	100	100
	Potential Dev.		100	**90.5**				100		
#15	Performance	100	97.4	92.6	94.1	80	100	100	100	100
	Actual Dev.	100	100	**96.3**	**94.1**	100	100	100	100	100
	Potential Dev.			**96.3**	**94.1**					
#16	Performance	97.0	99.0	97.5	96.3	92.9	100	97.8	100	99.2
	Actual Dev.	100	100	100	97.5	**92.9**	100	100	100	**99.2**
	Potential Dev.				98.8	**92.9**				**99.2**
#17	Performance	100	95.3	73.2	100	72.7	98.7	100	100	99.4
	Actual Dev.	100	100	**85.4**	100	95.5	**98.7**	100	100	99.4
	Potential Dev.			**85.4**		100	**98.7**			100
#19	Performance	100	100	75.0	100	-	100	97.6	100	100
	Actual Dev.	100	100	87.5	100	-	100	100	100	100
	Potential Dev.			100		-				

dn: determiner noun agreement; **ge**: gender agreement; **na**: noun adjective agreement; **pa**: pronoun antecedent agreement; **pp**: past participle agreement; **sv**: subject verb agreement; **co**: incorrect conjugation form; **pl**: incorrect plural form; **wf**: incorrect word formation.

The dash (-) signifies that learners #5, #12, and #19 did not produce any past participle in their documents. Figures in bold indicate that the level of actual development equals their potential development, inferring that learners could not propose any correct alternative even with assistance. Shaded areas denote a difference between both measures, implying that learners could perform – with assistance – beyond their current capabilities. When the actual development reaches 100%, the *Potential Dev.* field is left blank as the incorrect features are fully internalised, and that the zone is nonexistent. Several aspects can be observed from the above table, such as:

- *Learners perform at the level of their knowledge:* performance equals actual development (e.g., student #7 in noun adjective agreement - *na*).

- *Learners perform below their capabilities:* actual development is greater than performance. They were able to correct themselves with minimum assistance denoting that their incorrect forms were mistakes (e.g., student #19 in noun adjective agreement - *na*).

- *Learners have a zone of proximal development:* potential development is greater than actual development. Learners were able to improve their language accuracy with assistance suggesting that their incorrect forms are in the process of being internalised (e.g., student #10 in incorrect conjugation form - *co*).

- *Learners do not have a zone of proximal development:* potential development equals actual development. Learners could not correct themselves without being provided with the correct answer which is given after the self-editing exercises. The correct answer corresponds to the fourth level of the regulatory scale (e.g., student #2 in noun adjective agreement - *na*).

The figures obtained for student #16 in pronoun antecedent agreement (pa), for instance, illustrates a gradual development from performance to potential development . She wrote 77 pronouns marked as *correct* and 3 pronouns marked as *pronoun antecedent agreement*. The three incorrect forms written by learner #16 are listed along with their context below in Lines 6 to 8.

> [6]. *Il est important d'étudier ... pour les examens partiels parce qu'****il*** *aide vraiment pour le résultat général (It is important to study for continuous assessment as they improve the final grade for sure) - correct form: ils*

> [7]. *le rôle de la France dans la guerre et l'impact qu'****il*** *avait sur la France (France within war, and the impact of it on France) - correct form: elle*

> [8]. *J'ai aimé faire cette partie parce qu'****il*** *était intéressante (I loved doing this part as it was really interesting) - correct form: elle*

Student #16's percentage of performance accuracy in pronoun antecedent agreement, as seen in Table 6.13, is equal to 96.3% (CorF/CorF+IncF = 77/(77+3)). This student was able to correct one noun antecedent agreement with minimum assistance (Line 6) which was categorised as a mistake. The percentage of accuracy in knowledge (actual development) is consequently equal to 97.5% (78/(78+2)). Furthermore, although she could not edit Line 7 at level 2 (she read the feedback four times with an average duration of 2.54 seconds and entered **ils*, an incorrect replacement), she proposed *elle*, a correct alternative, at level 3 after reading the meta-linguistic feedback once during 7.9 seconds[49]. Her zone of proximal development in terms of pronoun antecedent agreement is thus equal to 98.8% (79/(79+1)). The last incorrect form (Line 8) was written incorrectly at level one, and she did not enter any replacements at level two nor

49. See Appendix F.2 on page 319 for details in feedback access and replacements.

three despite the fact that she read the feedback once at level two for 2.09 seconds. Student #16's development with regard to noun antecedent agreement is displayed in Table 6.14 below.

Table 6.14. Student #16's development in pronoun antecedent agreement

Current performance	**Actual development**	**Potential development**
96.3%	97.5%	98.8%
student #16 wrote her text independently	student #16 entered correct alternatives without assistance	student #16 entered correct alternatives with assistance

This example demonstrates that the learner's actual and proximal capabilities are not far above her current performance. Calculating and storing each learner's scores offers the possibility to compare all learners' actual and proximal development in one specific error type. Table 6.15 below for instance displays each learner's score obtained in noun adjective agreement.

Table 6.15. Learners' development in percentage in noun adjective agreement (na)

Students	**Current performance**	**Actual development**	**Potential development**
#2	89.4	93.6	93.6
#5	69.2	84.6	84.6
#6	89.1	89.1	90.6
#7	85.4	85.4	90.2
#8	90.9	93.9	93.9
#9	76.5	94.1	94.1
#10	78.7	89.4	89.4
#11	95.0	100	
#12	100	100	
#14	85.7	90.5	90.5
#15	92.6	96.3	96.3
#16	97.5	100	
#17	73.2	85.4	85.4
#19	75.0	87.5	100

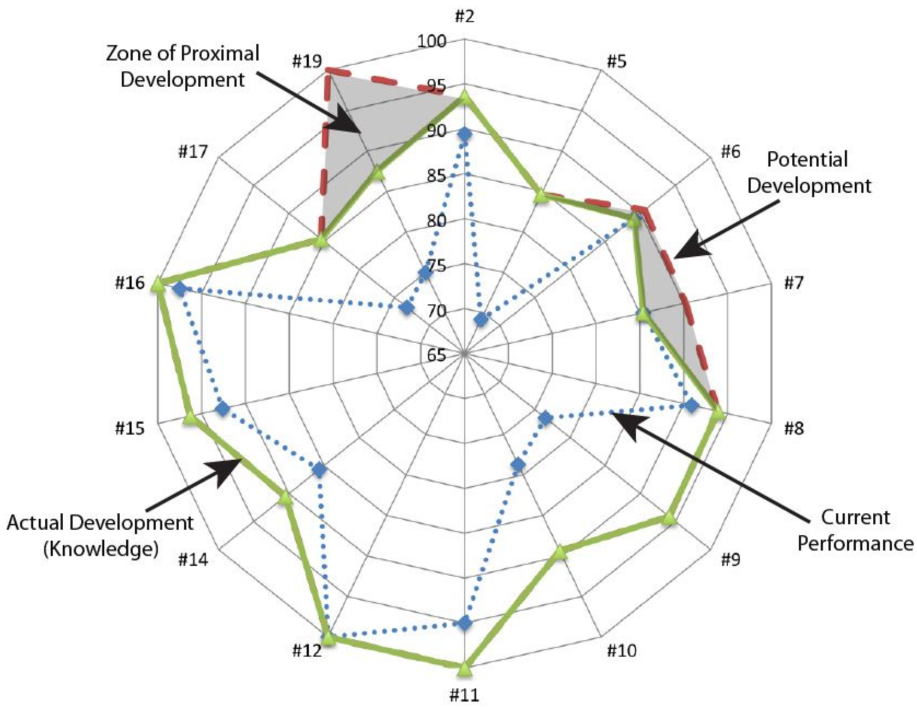
When comparing performance scores in noun adjective agreement, Table 6.15 shows that, for instance, student #19 scored less (75%) than student #6 who achieved 89.1%. Such a *poor* result would naturally lead to a smaller grade. Furthermore, particular attention would be given by any teacher to student #19 so that he could at least catch up with the class average. The above table also illustrates that the representation of the learners' zone of proximal development reflects a different student distribution in terms of needs than the one established with the performance scores. Student #19 and #6, whose performances are quite distant (75% and 89.1%), have in fact a very similar actual level of development (87.5% and 89.1%). While student #6 could barely improve her performance with assistance, which demonstrates that there is a small distance between her actual and potential level of development, student #19 fully corrected himself with assistance indicating a potential development of 100%. Such results imply that learner #19 had practically internalised the noun adjective agreement feature, although his original score did not imply such an expertise.

Student #2, on the other hand, could not edit her incorrect noun adjective agreements any better with assistance than she did without. Her levels of actual and potential development are identical (93.6%). Such a result may imply that the assistance she received (when accessed) was not appropriate to her needs. Even if her performance score was greater than the one of student #19 (89.4%>75%), her development is less advanced reaching only 93.6% as opposed to 100 % for student #19. The two incorrect forms student #2 could not correct with assistance are listed in Lines 9 and 10.

> [9]. *J'aime être *tout seul* (I love to be on my own) - correct form: toute seule*. Learner's corrections proposed at levels 1, 2, and 3: *toute *seul*.
>
> [10]. **mon [ma] *troisièmes année (my third year) - correct form: troisième*. Corrections proposed at levels 1, 2, and 3: *trois (three)*.

In Line 9, the incorrect sequence includes two adjectives that qualify the author herself (feminine singular). Student #2 never updated the second adjective *seul* even after reading the feedback at level two and three once for a duration of 2.13 seconds and 9.4 seconds, respectively. In Line 10, she provided each time the word *trois* as a replacement instead of removing the *-s* plural mark on the adjective (she only accessed the feedback at level two once during 1.67 seconds). In this example, no variations could be observed between her actual and potential development. Like learner #2, students #5, #8, #9, #10, #14, #15, and #17 did not improve their corrections (as listed in Table 6.15 above). Figure 6.13 below graphically represents the variations encountered between learners' actual and potential development in noun adjective agreement.

Figure 6.13. Learners' ZPD in noun adjective agreement

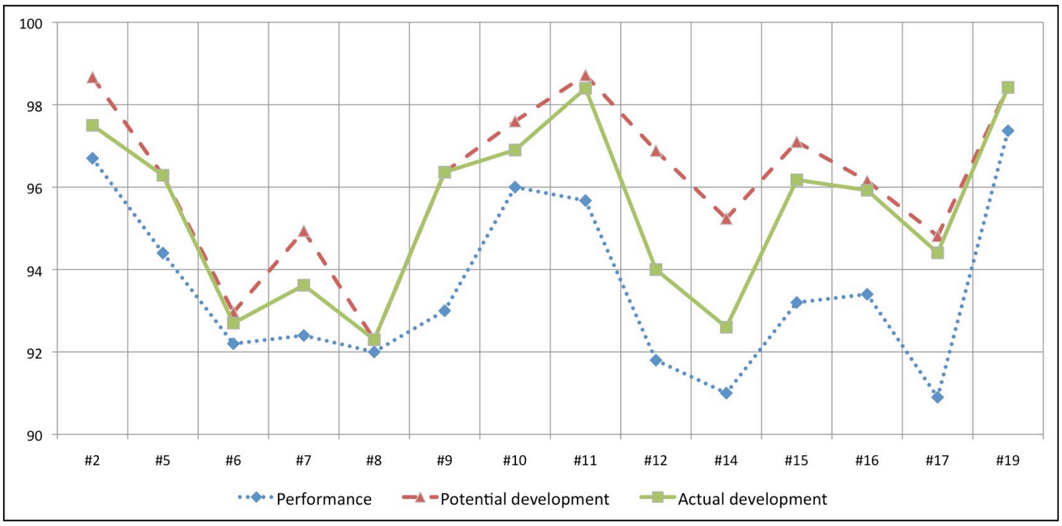


The inner concentric form (dotted line) represents the learners' current performance, the plain line refers to the learners' actual development (knowledge), the dashed line illustrates the learners' potential development, and the shaded areas symbolise the learners' zone of proximal development. The spider web illustrates variances between current performance and actual development, i.e., it demonstrates that not all students were able to perform at the level of their knowledge. It also confirms that their ZPD is not systematically represented. Their actual and potential development are often merged together. Given the fact that the previous discussion on assistance access showed that a large proportion of feedback was unread (53%), distinguishing between both actual and potential levels of development is not always feasible, as the difference between what a learner could do with assistance is not definitely known for each incorrect form. As a result, the zone of proximal development may be nonexistent if the learners' intervention failed to create it, or simply unknown if the learners' behaviour was not recorded.

6.3.2. Selection error category

The analysis of the selection error category displays similar findings as the morphosyntactic error category. Figure 6.14 below illustrates learners' scores obtained in performance, actual and potential development with regard to language accuracy in the selection category (see Appendix F.3 on page 320 for a full account of the frequencies and percentages).

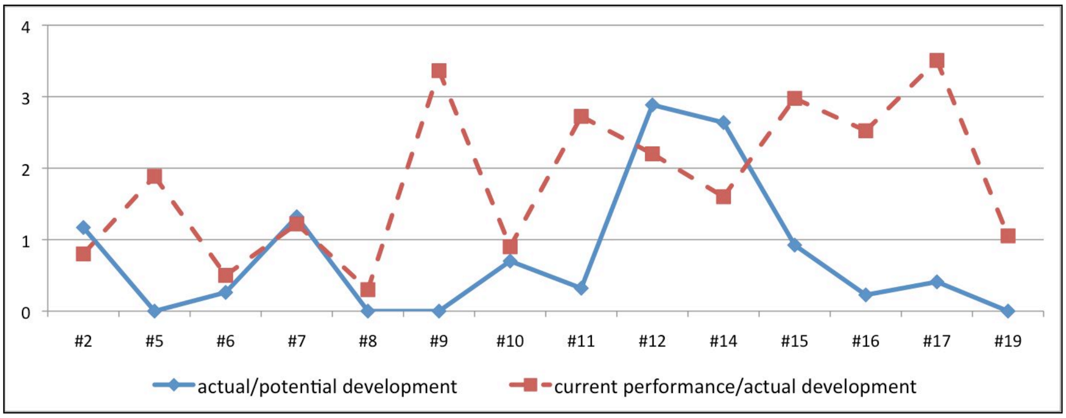
Figure 6.14. Language accuracy in performance, actual and potential development: selection error types



The selection error category reveals more variances between current performance and actual development, and between actual development and potential development, than in the morphosyntactic category. Some learners, such as student #17 may have a relatively large gap between current performance and actual development (90.9% and 94.4%), and a small distance between actual and potential development (94.4% and 94.8%). By contrast, other learners, such as student #6 may perform at the level of their knowledge (92.2% and 92.7%) and show little potential development (92.7% and 93%). Or others such as student #7 may display gradual improvements between all three measures (92.4%, 93.6% and 94.9%).

The gap between the current performance and actual development (e.g., student #9: 96.4%-93%=3.4%), and the distance between actual and potential development (e.g., student #12: 96.9%-94%=2.9%) do not appear to correlate (Figure 6.15); no patterns predictable by rules could be determined.

Figure 6.15. Gap between performance and actual development, and distance between actual and potential development, in percentage



Breaking up the data per error type within the selection category, Table 6.16 lists each learner's score obtained in performance, actual and potential development with regard to language accuracy.

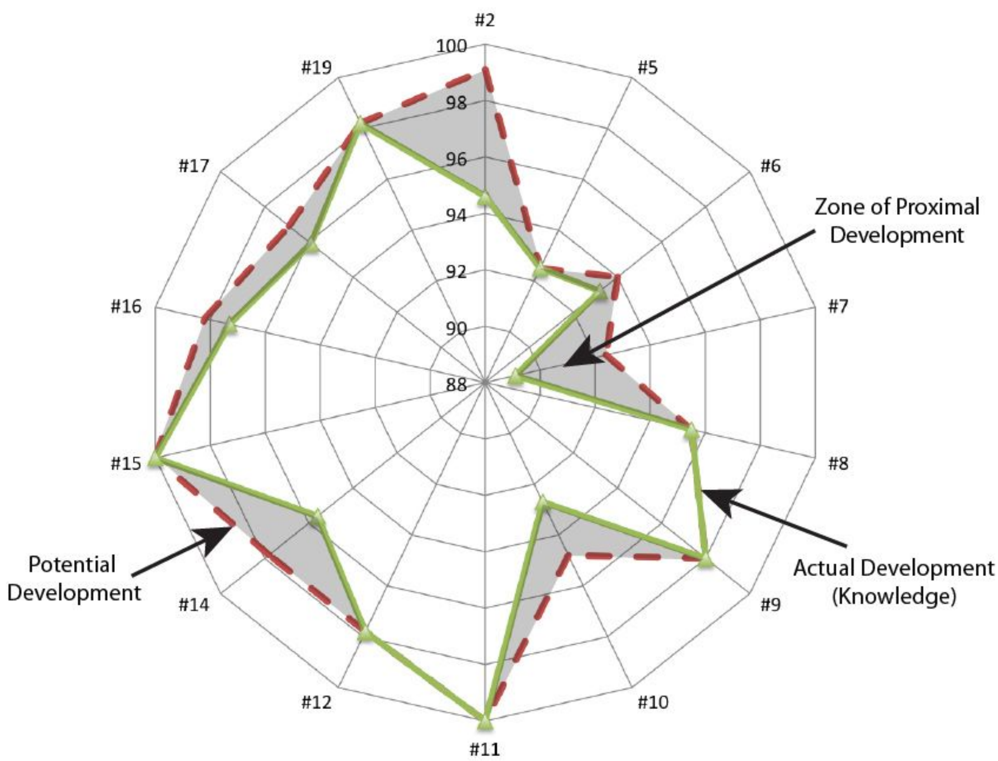
Table 6.16. Percentage of language accuracy in selection error types

St.		**mu**	**lw**	**re**	**sn**	**cw**	**rv**	**ar**	**pr**	**mo**	**te**	**vo**	**cl**
#2	P	99.6	98.9	100	100	100	100	93.7	92.6	100	98.2	100	100
	A	100	98.9	100	100	100	100	94.6	95.7	100	100	100	100
	D		99.1					99.1	97.9				
#5	P	98.0	97.1	100	100	100	100	87.5	95.0	97.1	100	100	100
	A	100	**97.8**	100	100	100	100	**92.5**	**97.5**	**97.1**	100	100	100
	D		**97.8**					**92.5**	**97.5**	**97.1**			
#6	P	99.0	94.5	100	97.2	100	100	93.2	85.4	99.1	98.2	100	98.7
	A	**99.0**	**95.0**	100	**97.2**	100	100	93.2	87.5	100	**98.2**	100	98.7
	D	**99.0**	**95.0**		**97.2**			94.0	89.6		**98.2**		99.0
#7	P	99.2	96.7	100	100	96.7	100	88.2	86.5	96.8	96.8	100	99.2
	A	100	**97.5**	100	100	**96.7**	100	89.1	88.5	96.8	**96.8**	100	100
	D		**97.5**			**96.7**		92.4	90.4	98.9	**96.8**		
8#	P	99.5	94.3	99.3	96.3	100	98.8	95.5	86.7	100	98.8	100	98.9
	A	**99.5**	**94.3**	**99.3**	**96.3**	100	**98.8**	**95.5**	**89.3**	100	**98.8**	100	**98.9**
	D	**99.5**	**94.3**	**99.3**	**96.3**		**98.8**	**95.5**	**89.3**		**98.8**		**98.9**
#9	P	99.1	96.2	100	100	91.4	97.6	96.1	77.3	100	100	100	100
	A	100	**98.1**	100	100	**94.3**	**97.6**	**98.0**	**88.6**	100	100	100	100
	D		**98.1**			**94.3**	**97.6**	**98.0**	**88.6**				
#10	P	100	99.4	100	100	100	97.3	91.7	94.0	97.3	97.3	98.6	100
	A	100	**99.7**	100	100	100	100	92.7	**95.5**	**97.3**	**97.3**	98.6	100
	D		**99.7**					94.8	**95.5**	**97.3**	**97.3**		
#11	P	100	97.6	100	100	98.2	100	100	82.4	100	100	100	99.5
	A	100	98.8	100	100	100	100	100	**95.6**	100	100	100	100
	D		99.6						**95.6**				
#12	P	98.7	97.2	100	100	92.3	100	97.8	81.6	93.3	96.6	100	100
	A	**98.7**	97.2	100	100	100	100	**97.8**	87.8	**96.7**	96.6	100	100
	D	**98.7**	98.1					**97.8**	95.9	**96.7**	100		
#14	P	100	95.9	100	97.8	86.4	96.2	95.6	88.4	96.2	100	100	100
	A	100	96.7	100	**97.8**	**95.5**	96.2	95.6	**90.7**	96.2	100	100	100
	D		99.2		**97.8**	**95.5**	100	97.8	**90.7**	100			
#15	P	94.6	96.5	100	100	100	96.6	97.7	94.3	87.5	100	100	100
	A	94.6	**98.6**	100	100	100	100	100	**97.1**	96.9	100	100	100
	D	96.4	**98.6**						**97.1**	100			
#16	P	99.6	96.0	100	99.2	98.6	100	94.7	90.7	100	94.6	100	99.6
	A	**99.6**	**96.8**	100	**99.2**	**98.6**	100	97.3	**99.2**	100	94.6	100	**99.6**
	D	**99.6**	**96.8**		**99.2**	**98.6**		98.2	**99.2**		95.7		**99.6**
#17	P	100	97.2	100	100	94.8	100	92.8	75.9	98.7	96.3	100	99.5
	A	100	**97.6**	100	100	**94.8**	100	95.9	88.5	**98.7**	**98.8**	100	**99.5**
	D		**97.6**			**94.8**		96.9	90.8	**98.7**	**98.8**		**99.5**
#19	P	100	99.3	100	100	92.3	97.6	96.4	100	100	100	100	100
	A	100	100	100	100	**92.3**	100	**98.2**	100	100	100	100	100
	D					**92.3**		**98.2**					

mu: nonstandard word; **lw**: inappropriate lexical choice; **re**: inappropriate register; **sn**: not understandable; **cw**: inappropriate connection word; **rv**: reflexive verb; **ar**: incorrect article type; **pr**: incorrect preposition; **mo**: incorrect mood; **te**: incorrect tense; **vo**: incorrect voice; **cl**: incorrect word class.

The figures in the shaded areas denote a zone of proximal development, whereas the figures in bold indicates an identical level of actual and potential development. Additionally, the terms *performance*, *actual development*, and *potential development* are abbreviated to P, A, and D, respectively. While the figures in the above table indicate that the zone of the proximal development is most of the time nonexistent in several categories (e.g., student #5 in inappropriate lexical choice (lw)), the incorrect article type (ar) however displays differences between actual and potential development in most students' scores. Figure 6.16 illustrates the extent of learners' capabilities with and without assistance in this error type, i.e., incorrect article type.

Figure 6.16. Learners' zone of proximal development in incorrect article type

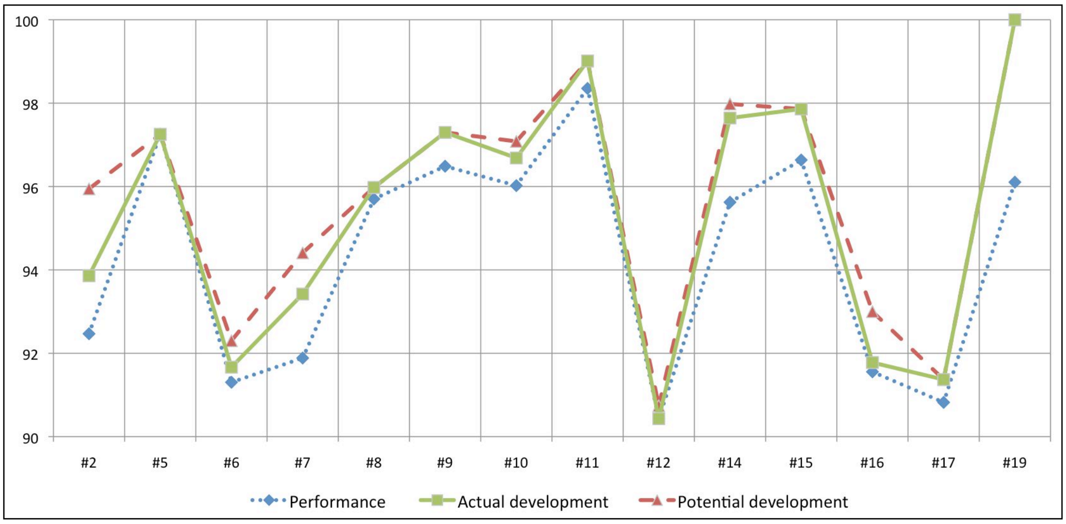


Given the fact that learners performed at an identical level of language complexity (Section 6.2), the above figure confirms the presence of random variability between learners, where high scores in knowledge does not necessarily lead to high potential development or vice versa.

6.3.3. Syntactic error category

The syntactic error category has been shown (Section 6.2.3) to be the area where students experienced difficulties with regard to language accuracy. Within this category, learners produced the majority of their incorrect forms and most of them were errors (not mistakes), indicating a lack in their knowledge (Figure 6.17). Frequencies and percentages are detailed in Appendix F.4 on page 321.

Figure 6.17. Language accuracy in performance, actual and potential development: syntactic error types



The syntactic error category unveils variances between learners, but does not vary as much as the morphosyntactic and selection error types. While most students slightly improved their performance with minimum assistance (e.g., students #6 and #11 with a gap of 0.66% between current performance and actual knowledge), they also show a short not to say nonexistent distance between actual and potential levels of development. The following tabular entries (Table 6.17) consider each syntactic error type and include word addition (ad), word omission (om), incorrect word order (wo), and not understandable syntax (un) error types.

Table 6.17. Percentage of language accuracy in syntactic error types

Students		ad	om	wo	un
#2	Performance	98.6	95.9	99.8	97.8
	Actual Dev.	98.6	96.8	100	**98.2**
	Potential Dev.	99.0	98.6		**98.2**
#5	Performance	100	100	100	97.3
	Actual Dev.	100	100	100	**97.3**
	Potential Dev.				**97.3**
#6	Performance	98.5	98.1	98.2	95.8
	Actual Dev.	**98.5**	98.5	**98.2**	95.8
	Potential Dev.	**98.5**	98.8	**98.2**	96.2
#7	Performance	99.1	97.2	99.1	96.1
	Actual Dev.	99.5	98.0	**99.4**	96.1
	Potential Dev.	99.8	98.4	**99.4**	96.6
#8	Performance	99.3	98.6	100	97.7
	Actual Dev.	**99.3**	**98.9**	100	**97.7**
	Potential Dev.	**99.3**	**98.9**		**97.7**
#9	Performance	99.7	98.9	99.4	98.3
	Actual Dev.	**99.7**	**99.2**	100	**98.3**
	Potential Dev.	**99.7**	**99.2**		**98.3**
#10	Performance	99.2	98.4	98.9	99.5
	Actual Dev.	**99.5**	**98.4**	99.3	**99.5**
	Potential Dev.	**99.5**	**98.4**	99.7	**99.5**
#11	Performance	99.5	99.7	100	99.2
	Actual Dev.	**99.5**	100	100	**99.5**
	Potential Dev.	**99.5**			**99.5**
#12	Performance	100	99.7	100	90.7
	Actual Dev.	100	99.7	100	**90.7**
	Potential Dev.		100		**90.7**
#14	Performance	99.3	99.0	100	97.3
	Actual Dev.	99.3	**99.7**	100	**98.6**
	Potential Dev.	99.7	**99.7**		**98.6**
#15	Performance	99.4	99.7	99.4	98.1
	Actual Dev.	**99.4**	**99.7**	100	**98.8**
	Potential Dev.	**99.4**	**99.7**		**98.8**
#16	Performance	99.8	98.9	99.4	93.2
	Actual Dev.	**99.8**	98.9	99.6	93.2
	Potential Dev.	**99.8**	99.2	100	93.9
#17	Performance	98.8	97.6	99.4	94.4
	Actual Dev.	**98.8**	**97.9**	**99.7**	**94.4**
	Potential Dev.	**98.8**	**97.9**	**99.7**	**94.4**
#19	Performance	99.5	96.6	100	100
	Actual Dev.	100	100	100	100
	Potential Dev.				

Again the shaded areas indicates a difference between actual and potential development, as opposed to the bold figures which signify that there is none. Table 6.17 above reveals that learners had more difficulties in correcting themselves even with assistance in syntactic error types than other categories, as most of their performance results equal their actual and potential development. For example, learner #6 could not perform any better in word addition (ad) and word order (wo) with and without assistance (performance, actual and potential development=98.5% and 98.2%, respectively).

6.3.4. Discussion

This section not only focused on what learners could do without assistance (actual development), but also on what learners could do with assistance (potential development). The various scores obtained from the morphosyntactic, selection and syntactic error categories point to the fact that some learners benefited from the assistance in one error category, but did not necessarily in the other categories. Table 6.18 below summarises the error types for which each learner's[50] zone of proximal development was observable, implying a difference between actual and potential development scores. A tick (✅) indicates that learners attain 100% of success in actual development with regard to language accuracy. The question mark in past participle agreement (mo_ag_pp_) refers to the fact that learners did not produce any past participle verb in their written texts. As a result, one cannot be sure whether the learners would have improved their knowledge with feedback in this error type category. The empty cells designate that learners did not perform any better with or without assistance. The shaded areas signal that learners extended their capabilities from the feedback dispensed at levels two and/or three.

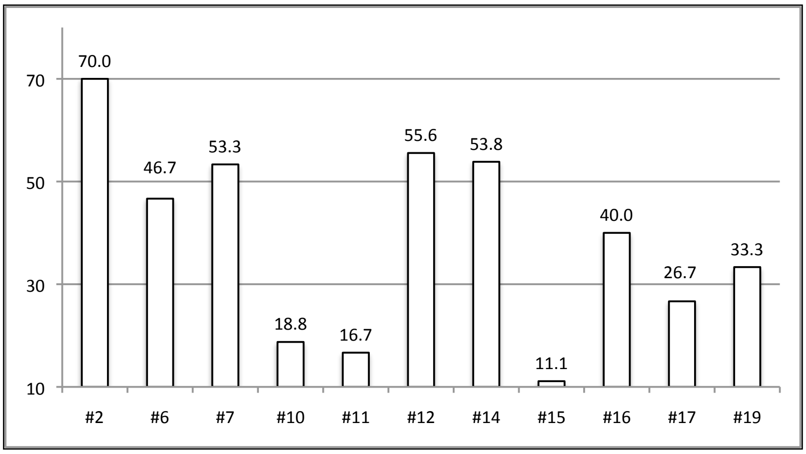
50. Table 6.18 does not include learners #5, #8 and #9 as the figures on display represent the zone of proximal development. Recall that the zone of proximal development is the distance between the two measures calculated with and without assistance. Since only one of the two measures required is known for these three students, the distance cannot be estimated and visualised.

Table 6.18. Observable zones of proximal development per student and error type

Error types	#2	#6	#7	#10	#11	#12	#14	#15	#16	#17	#19
mo_ag_dn_	✓		✓	✓	✓	✓	✓	✓	✓	✓	✓
mo_ag_ge_			✓		✓			✓	✓	✓	✓
mo_ag_na_					✓	✓			✓		
mo_ag_pa_	✓	✓		✓		✓	✓			✓	✓
mo_ag_pp_					✓	?	✓	✓			?
mo_ag_sv_	✓	✓			✓	✓	✓	✓	✓		✓
mo_fo_co_		✓				✓		✓	✓	✓	✓
mo_fo_wf_	✓	✓	✓		✓	✓	✓	✓			✓
se_vo_mu_	✓		✓	✓	✓		✓			✓	✓
se_vo_lw_											✓
se_vo_re_	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
se_vo_sn_	✓		✓	✓	✓	✓		✓		✓	✓
se_gr_cw_	✓	✓		✓	✓	✓		✓			
se_gr_rv_	✓	✓	✓	✓	✓	✓		✓	✓	✓	✓
se_gr_ar_					✓			✓			
se_gr_pr_											✓
se_gr_mo_	✓	✓			✓				✓		✓
se_gr_te_	✓				✓		✓	✓			✓
se_gr_vo_	✓	✓	✓		✓	✓	✓	✓	✓	✓	✓
se_gr_cl_	✓		✓	✓	✓	✓	✓	✓			✓
sy_ad_						✓					✓
sy_om_					✓						✓
sy_wo_	✓				✓	✓	✓	✓			✓
sy_un_											✓
Total error types for which there is a ZPD	7 out of 10	7 out of 15	8 out of 15	3 out of 16	1 out of 6	5 out of 9	7 out of 13	1 out of 9	6 out of 15	4 out of 15	1 out of 3

The overall trends that emerge from Table 6.18 reflects an unsystematic variation between learners and error types, as an unsystematic variation was equally observed between current performance and actual developmental levels in Tables 6.13, 6.16 and 6.17 above. Learners' developmental features may be improved with assistance in one error type (shaded areas in Table 6.18) and not in another (blank areas in Table 6.18). For example, learner #2 was able to improve her performance with assistance in incorrect past participle agreement (mo_ag_pp_), incorrect vocabulary choice (se_vo_lw_), or word omission (sy_om_). By contrast, she could not increase her level of language accuracy in error types such as noun adjective agreement (mo_ag_na_), incorrect verb conjugation (mo_fo_co_), and not understandable syntax (sy_un_). Student #6, on the other hand, displays a potential development in noun adjective agreement (mo_ag_na_) and not understandable syntax (sy_un_) error types. The meta-linguistic feedback was perhaps not adapted to learner #2's specific needs, or learner #2's level of self-confidence could also be too high with regard to the self-editing task. The discussion on learners' responses to interventions (Section 6.1) made clear that learners were mostly correcting themselves without reading the assistance. Such behaviour mainly gives information on actual development, as opposed to potential development, where some level of assistance is required to establish the latter. Figure 6.18 below illustrates the percentage of error types per student for which the zone of proximal development could be represented.

Figure 6.18. Percentage of error types per student for which the potential development could be observed



Student #2, for instance, refined her language accuracy with assistance in seven error types out of ten that needed improvement. This denotes that this learner's potential development could be observed in 70% of all the error types for which she had to improve herself. Student #15, on the other hand, showed a developmental level in only one error type out of nine. In other words, her zones of potential development could be observed for 11.1% of the error types for which she had weaknesses.

In summary, the section has shown that representing learners' potential development is not as straightforward as estimating the learners' actual development, as the level of assistance to correct themselves is a decisive factor that *must* be acknowledged. Without knowing it, only the actual level of development can be estimated. If learners accessed their feedback to correct themselves, the extent of their potential development could be determined. By contrast, if learners did not access their feedback, only their actual development could be established. With regard to variability, it is not because students took full advantage of the dynamic assessment in one particular error type that they behaved in a similar manner in all error types. Had they done so, it could be advanced that specific error types or metalinguistic feedback were more appropriate than others to create the ZPD. However, since learners acted variably in terms of feedback access in error types, levels of potential development between learners point to random variations, where random designates not predictable by rules. Furthermore, differences in learners' potential development were observed between learners with identical levels of actual development, which supports Aljaafreh and Lantolf's (1994) claim: one cannot presume that "any two learners who attain identical scores on a test are necessarily at the same stage in their interlanguage growth" (p.473).

6.4. Summary and conclusion

The chapter started by outlining the learners' responses to interventions, as mediation is a key concept in the zone of proximal development, which is in itself the core of the model proposed in this thesis to distinguish between errors and mistakes. The first section highlighted the fact that learners mostly did not access the feedback provided to them, but that this information was not to be considered as a refusal of engagement, as most learners provided alternatives which were generally correct.

The chapter continued with an analysis of learners' errors and mistakes in interlanguage. While some learners may perform very close to their knowledge in one error type, they may in other situations display a high percentage of mistakes in their productions, which indicates that learners knew more than they were able to write. Furthermore, it was also demonstrated that learners were more inclined to produce mistakes in spelling and typographic error types rather than in syntactic, selection and morphosyntactic categories.

Within the third section, it was shown that the learners' ZPD could not always be represented as learners mainly proposed alternatives without consulting the assistance. As a result, their actual and potential levels of development could not be distinguished, and this is consistent for most of the error types investigated within this section, i.e., selection, syntactic and morphosyntactic error types. Additionally, findings not only corroborated that interlanguage performance was variable, but also demonstrated that interlanguage competence exhibited variability between learners due to learners' variable degree of ability to perform their knowledge.

Consequently, and considering the variable aspect of interlanguage competence in a synchronic analysis, determining the ZPD of each learner is a necessary indicator if one intends to analyse learners' development and with it their future attainments. The following chapter will explore the interlanguage of three participants over a period of two academic semesters in order to not only investigate the differences in learners' actual development over time, but also observe the movements of the incorrect types in the learners' zone of proximal development.

Chapter 7. A diachronic inter-learner and intra-learner analysis

The aim of the previous chapter was to analyse inter-learner variability at a single point in time across equivalent text types with regard to lexical and syntactical complexity. In particular, it addressed the issue of learners' responses to interventions, and mainly focused on learners' errors and mistakes and on their actual and potential development. This chapter investigates interlanguage variability diachronically. More specifically, it analyses the inter- and intra-learner differential levels of actual and potential development over time.

Variability in interlanguage is analysed first with two students' texts written at two distinct points in time (inter-learner analysis), and then with a various range of text types written by one student over a longer period of time (intra-learner analysis). The investigation includes the observation of the learners' responses to interventions, their levels of language sophistication and complexity, as well as the variations in their actual and potential development.

7.1. Diachronic inter-learner analysis

As mentioned in Section 5.6 on page 147, the diachronic inter-learner analysis considers the texts written by students #10 and #17 over a short time span. The motivation behind selecting these two students is justified by the fact that they were available; only a few participants continued the experiment over the second semester. Table 7.1 recapitulates the data for each student with respect to total amount of tokens, incorrect tokens and sequences, as well as incorrect form rates. While student #10 submitted two texts at time T3 (December 2008) and one text at time T6 (April 2009), student #17 sent in addition one text for correction at time T5 (March 2009). The average of incorrect form rates range from 11.3% to 16.1% for student #10, and from 20.3% to 30.2% for student #17.

Table 7.1. Diachronic inter-learner analysis: overall data

Stu.	Time	Text id	Text type	Token count	Incorrect sequence	Incorrect token	Incorrect sequence rate	Incorrect token rate	Average of incorrect form rate
#10	T3	#44	bilan	471	51	85	10.8%	18.0%	14.4%
	T3	#63	wiki	417	55	79	13.2%	18.9%	16.1%
	T6	#89	bilan	613	49	90	8.0%	14.7%	11.3%
#17	T3	#55	bilan	493	74	133	15.0%	27.0%	21.0%
	T3	#67	wiki	413	98	110	23.7%	26.6%	25.2%
	T5	#81	écrire	739	113	187	15.3%	25.3%	20.3%
	T6	#90	bilan	643	161	228	25.0%	35.5%	30.2%

As previously discussed in Section 6.2 on page 165, wiki and bilan were not considered as significantly different with the one-way within subjects (or repeated measures) ANOVA test, as learners used the same language sophistication and complexity for both text types. Consequently, wiki and bilan were grouped together for the following analysis.

7.1.1. Learners' responses to interventions

The observation of learners' behaviour with respect to feedback access (Table 7.2 below) revealed a regular pattern in both students: they accessed more feedback at level L2 (error type) than at level L3 (meta-linguistic feedback), and they accessed less feedback at T6 than T3. For example, student #10 accessed the feedback for 67.9% of all her incorrect forms at L2/T3 and 44.3% at L3/T3, and she opened 57.1% of her feedback at L2/T6 and 38.8% at L3/T6.

Moreover, Table 7.2 below shows that student #17 did not access any of her feedback at T5 (feedback accessed=0%). Since she did not provide any alternatives either[51] at level two or three, each incorrect form within this text (written at T5) would be automatically

51. One probable reason could relate to a technical issue, where the system may have failed to record the data. Yet, the database marked the different steps as successfully attempted. Another reason might be that she simply skipped both levels to attain quicker the corrections proposed at level four.

considered as errors[52]. It may be the case that they all are errors, but there is a reasonable doubt that they may equally have been mistakes, or at least some of them. Consequently, it was decided that integrating this text into the analysis would distort the findings. As a result, the text written by student #17 at T5 was disregarded from the diachronic inter-learner analysis.

Table 7.2. Feedback access per student, time period, and level of assistance

Students	**Time**	**Levels of assistance**	**Total feedback**	**Feedback not accessed**		**Feedback accessed**	
				Frequency	**Percent**	**Frequency**	**Percent**
#10	T3	L2	106	34	32.1%	72	67.9%
		L3	106	59	55.7%	47	44.3%
	T6	L2	49	21	42.9%	28	57.1%
		L3	49	30	61.2%	19	38.8%
#17	T3	L2	172	71	41.3%	101	58.7%
		L3	172	120	69.8%	52	30.2%
	T5	L2	113	113	100%	0	0%
		L3	113	113	100%	0	0%
	T6	L2	161	120	74.5%	41	25.5%
		L3	161	131	81.4%	30	18.6%

Figures 7.1 and 7.2 below provide an overview of both learners' behaviour with regard to feedback access at times three and six, both levels of assistance, two and three are grouped together.

52. The error-mistake distinction requires at least two alternatives to determine whether the incorrect form is due to a lack of knowledge or a gap in performance.

Figure 7.1. Overview of feedback access and alternatives provided: student #10

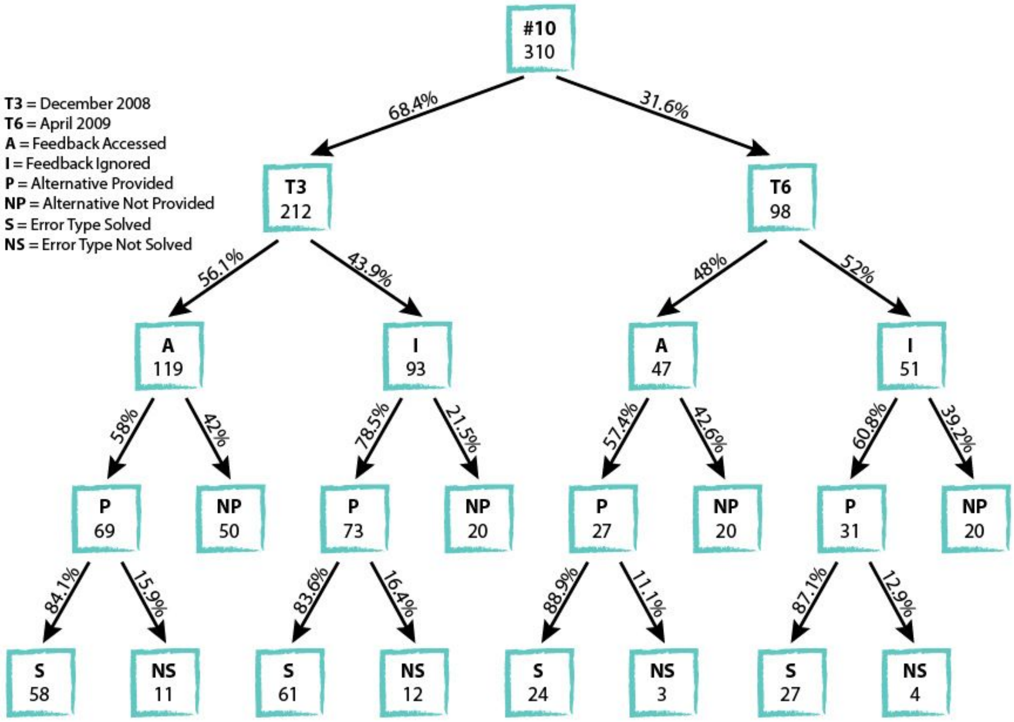
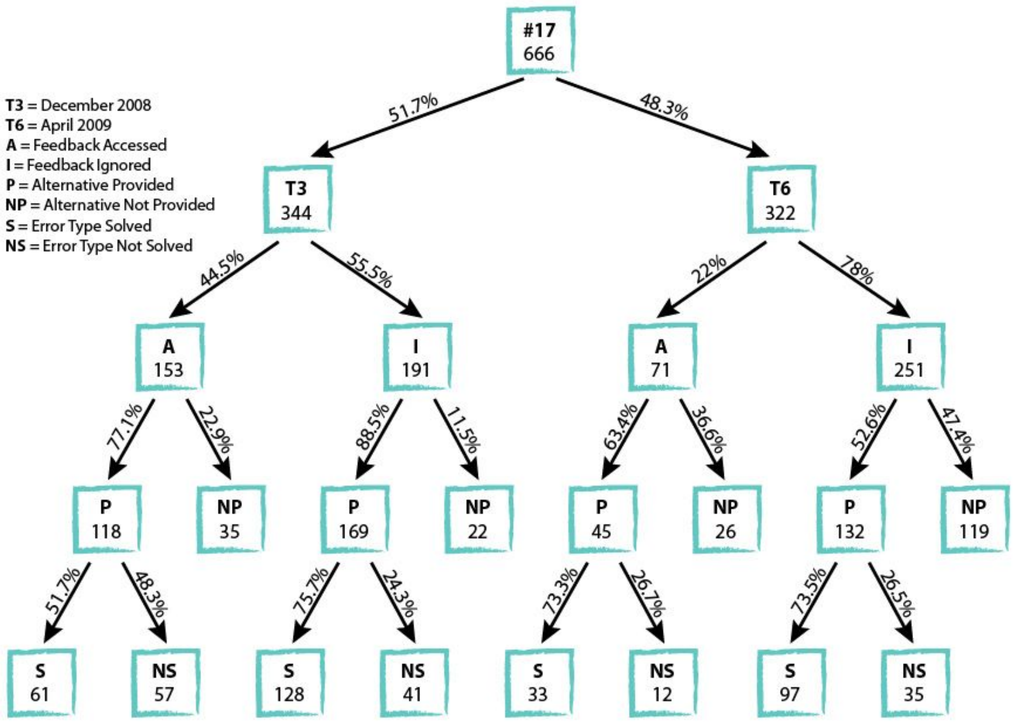


Figure 7.2. Overview of feedback access and alternatives provided: student #17



Although both learners #10 and #17 did not access much of their feedback at T3 and T6 (43.9% and 52% for student #10, and 55.5% and 78% for student #17), they nevertheless provided a relatively large amount of alternatives to all of their incorrect forms whether the assistance was accessed or not (66.98% and 59.18% for student #10, and 83.43% and 54.96% for student #17 at T3 and T6, respectively)[53]. In conclusion, the diagrams above reflect a relatively high participation from the learners at T3, where they both accessed more feedback and provided more alternatives, than at T6.

7.1.2. Variations in language sophistication and complexity

Table 7.3 below shows that the percentage of incorrect forms in student #10's texts decreased at T6 ranging from 15.2% at T3 to 11.3%, whereas student #17 produced more ill-formed words at T6 with 30.2% compared to 22.9% at T3.

Table 7.3. Average of incorrect form rate over time per student

Students	Time	Token count	Incorrect sequence	Incorrect token	Incorrect sequence rate	Incorrect token rate	Average of incorrect form rate
#10	T3	888	106	164	11.9%	18.5%	15.2%
	T6	613	49	90	8.0%	14.7%	11.3%
#17	T3	906	172	243	19.0%	26.8%	22.9%
	T6	643	161	228	25.0%	35.5%	30.2%

Although learner #17 produced a higher percentage of incorrect forms at T6, she increased her level of language sophistication and complexity during the same period, ranging from a coefficient of 24.76 to 27.78 (Table 7.4 below). In addition, the individual vectors reveal that learner #17 improved in all aspects of lexical and syntactical complexity, thus implying a greater lexical diversity, more inflectional and derivational affixes, and longer and more complicated sentences. In this respect, it may be considered

53. #10/T3: alternatives provided=(69+73)/212*100; #10/T6: alternatives provided=(27+31)/98*100; #17/T3: alternatives provided=(118+169)/344*100; #17/T6: alternatives provided=(45+132)/322*100.

that learner #17 did improve her language skills by attempting to reach a higher standard in written French. Student #10 who produced a lower percentage of incorrect forms also raised her levels of language sophistication and complexity, except for vector three (syntactic sophistication) whose coefficient was lowered to 12.65 at T6 instead of 13.49 at T3. Less incorrect forms and a higher level of balanced complexity measure may equally imply a better control of the French language in a written context.

Table 7.4. Students' progress in language complexity

Measure	**Time**	**Learner #10**		**Learner #17**	
		Mean	**Std. Deviation**	**Mean**	**Std. Deviation**
balanced	T3	23.38	0.27	24.76	0.43
	T6	25.02	.	27.78	.
vector 1	T3	5.36	0.12	6.10	0.32
	T6	5.40	.	6.13	.
vector 2	T3	4.30	0.14	4.21	0.27
	T6	4.52	.	4.38	.
vector 3	T3	13.49	2.73	18.42	2.81
	T6	12.65	.	20.21	.
vector 4	T3	11.78	0.52	12.31	0.21
	T6	13.05	.	14.94	.

Balanced: includes four vectors based on lexical and grammatical complexity and sophistication measures**;** **vector 1**: lexical complexity (text-length adjusted type token ratio); **vector 2**: lexical sophistication (mean length of a word); **vector 3**: syntactic sophistication (mean period unit length); **vector 4**: syntactic complexity (unique bigram ratio).

7.1.3. Variations in learners' actual and potential development

Table 7.5 lists the amounts and percentages of mistakes and errors made by both learners at the two points in time (T3 and T6). The percentages of errors gradually increased over the two time periods to reach 56.7% and 70.2% for student #10 and #17, respectively.

Table 7.5. Amount of correct and incorrect forms including errors and mistakes

Student	Time	Total tokens	Correct forms	Incorrect tokens		% of mistakes in incorrect tokens	% of errors in incorrect tokens
				Errors	Mistakes		
#10	T3	888	724	75	89	54.3%	45.7%
	T6	613	523	51	39	43.3%	56.7%
#17	T3	906	663	138	105	43.2%	56.8%
	T6	643	415	160	68	29.8%	70.2%

Given the fact that, as previously discussed (Section 7.1.1), both students provided less alternatives at T6, their incorrect forms had more chance to be considered as errors; indeed with no suggestions, the ill-formed word is automatically labeled as a competence issue since two correct alternatives are required to mark the incorrect form as a performance-related mistake. Yet, the possibility that there were mistakes cannot be completely ruled out. It might be that learners merely left the correction field blank, even if they knew correct replacements.

The aggregate data with respect to language accuracy, as illustrated in Table 7.6 below, shows no regular patterns over the two periods between learners. While learner #10 slightly improved her degree of language accuracy in actual development (going from 91.6% to 91.7%), student #17 obtained lower scores ranging from 84.8% to 75.1%, suggesting a regression in her capabilities. However, as mentioned above, learner #17 also demonstrated a change in her language standard, producing more complex and sophisticated sentences than at T3, which cannot be considered as regression. Rather, this attempt to produce a higher degree of language complexity in her written productions demonstrates that she constructed a new zone of proximal development with new objectives (potential development[54] ranging from 86.4% to 77% at T3 and T6,

54. Recall, the percentage of potential development is calculated by dividing all correct forms plus all errors for which the learners could provide correct alternatives with assistance by the total amount of correct and incorrect forms.

respectively). The latter (77%) denotes that she was not ready to accurately edit the changes she tried to make in her performance, even with assistance.

Table 7.6. Percentage of language accuracy in performance, actual and potential development

Student	Time	% of language accuracy in performance	% of language accuracy in actual development	% of language accuracy in potential development
#10	T3	81.5%	91.6%	93.1%
	T6	85.3%	91.7%	93.1%
#17	T3	73.2%	84.8%	86.4%
	T6	64.5%	75.1%	77%

Breaking down the scores per error category and error type, Tables 7.7, 7.8, and 7.9 demonstrate that the percentages obtained in language accuracy for both learners #10 and #17 at T3 and T6 do not follow any predictable patterns.

Table 7.7. Language accuracy in percentage per selection error type

Students			mu	lw	re	sn	cw	rv	ar	pr	mo	te	vo	cl
#10	T3	P	100	99.4	100	100	100	97.3	91.7	94	97.3	97.3	98.6	100
		AD	100	**99.7**	100	100	100	100	92.7	**95.5**	**97.3**	**97.3**	98.6	100
		PD		**99.7**					94.8	**95.5**	**97.3**	**97.3**	100	
	T6	P	100	98.8	100	100	100	93	100	95.3	100	97.6	97.6	100
		AD	100	**99.6**	100	100	100	**93**	100	**95.3**	100	100	**97.6**	100
		PD		**99.6**				**93**		**95.3**			**97.6**	
#17	T3	P	100	97.2	100	100	94.8	100	92.8	75.9	98.7	96.3	100	99.5
		AD	100	**97.6**	100	100	**94.8**	100	**95.9**	88.5	**98.7**	**98.8**	100	**99.5**
		PD		**97.6**			**94.8**		**96.9**	90.8	**98.7**	**98.8**		**99.5**
	T6	P	100	98	100	100	92.9	100	85.2	75.9	94.7	90	100	99.3
		AD	100	**98**	100	100	**92.9**	100	**88.9**	**77.6**	**94.7**	**92.5**	100	**99.3**
		PD		**98**			**92.9**		**88.9**	**77.6**	**94.7**	**92.5**		**99.3**

mu: nonstandard word; **lw**: inappropriate lexical choice; **re**: inappropriate register; **sn**: not understandable; **cw**: inappropriate connection word; **rv**: reflexive verb; **ar**: incorrect article type; **pr**: incorrect preposition; **mo**: incorrect mood; **te**: incorrect tense; **vo**: incorrect voice; **cl**: incorrect word class; **bold**: level of actual development equals level of potential development; **shaded areas**: difference between both levels; **P**: performance; **AD**: actual development; **PD**: potential development.

Table 7.8. Language accuracy in percentage per syntactic error type

Students			ad	om	wo	un
#10	T3	Performance	99.2	98.4	98.9	99.5
		Actual Dev.	**99.5**	**98.4**	99.3	**99.5**
		Potential Dev.	**99.5**	**98.4**	99.7	**99.5**
	T6	Performance	99.4	98.9	99.2	97.6
		Actual Dev.	100	**99.1**	99.2	**97.6**
		Potential Dev.		**99.1**	100	**97.6**
#17	T3	Performance	98.8	97.6	99.4	94.4
		Actual Dev.	**98.8**	**97.9**	**99.7**	**94.4**
		Potential Dev.	**98.8**	**97.9**	**99.7**	**94.4**
	T6	Performance	95.4	98.8	97.2	90.6
		Actual Dev.	95.9	**98.8**	**97.2**	**90.6**
		Potential Dev.	96.3	**98.8**	**97.2**	**90.6**

ad: word addition; **om**: word omission; **wo**: incorrect word order; **un**: not understandable; **bold**: level of actual development equals level of potential development; **shaded areas**: difference between both levels.

Table 7.9. Language accuracy in percentage per morphosyntactic error type

Students			dn	ge	na	pa	pp	sv	co	pl	wf
#10	T3	Performance	100	93.5	78.7	100	92.3	95.9	95.9	100	99.6
		Actual Dev.	100	**98.9**	**89.4**	100	**96.2**	**98.6**	98.6	100	**99.6**
		Potential Dev.		**98.9**	**89.4**		**96.2**	**98.6**	100		**99.6**
	T6	Performance	100	100	81.8	97.1	97.5	95.2	95.2	100	99.4
		Actual Dev.	100	100	84.8	**97.1**	100	100	95.2	100	**99.4**
		Potential Dev.			87.9	**97.1**			97.6		**99.4**
#17	T3	Performance	100	95.3	73.2	100	72.7	98.7	100	100	99.4
		Actual Dev.	100	100	**85.4**	100	95.5	**98.7**	100	100	99.4
		Potential Dev.			**85.4**		100	**98.7**			100
	T6	Performance	100	95.5	82.6	93.3	100	97.3	92.2	100	100
		Actual Dev.	100	97.7	**87.0**	93.3	100	100	**94.9**	100	100
		Potential Dev.		100	**87.0**	95.6			**94.9**		

dn: determiner noun agreement; **ge**: gender agreement; **na**: noun adjective agreement; **pa**: pronoun antecedent agreement; **pp**: past participle agreement; **sv**: subject verb agreement; **co**: incorrect conjugation form; **pl**: incorrect plural form; **wf**: incorrect word formation; **bold**: level of actual development equals level of potential development; **shaded areas**: difference between both levels.

The above results denote that the scores obtained in levels of actual and potential development do not follow a linear progression. Rather, they increase and/or decrease. With two points in time, whether the general progression is a straight line or a curve cannot be determined; this will be analysed more deeply in the diachronic intra-learner analysis in Section 7.2 below. Student #10, for instance, seemed to have regressed in noun adjective (na) agreement in both actual and potential development (actual development: from 89.4% to 84.8% at T3 and T6, respectively; potential development: from 89.4% to 87.9%). Opposed to learner #10's results, the scores obtained by learner #17 in noun adjective agreement (na) leads to a probable improvement in actual development (going from 85.4% to 87%). However, the lack of feedback access by student #17 failed to show her potential development at both T3 and T6, i.e., what she should be able to do with assistance. This observation is not an isolated case; the distance between actual and potential development is in most error types not tangible as learners did not use the assistance to correct themselves. As a direct consequence, the errors could not be labelled otherwise than low in the ZPD, as opposed to somewhere in-between high and low. The latter signals an error that could not be corrected even with feedback and the former indicates that the error could be self-edited with assistance.

7.1.4. Variations in interlanguage development

Measuring variations in interlanguage development between T3 and T6 is a difficult endeavour as both learners mostly wrote incorrect forms at T3 that they never correctly or incorrectly reproduced at T6. For example, student #10 produced at T3 the following error, which was never repeated at T6.

> [11]. **Aussi nous avons* dû créer une page individuelle (We also had to create an individual page) - correct form: nous avons dû aussi créer...*

The incorrect form was marked as an incorrect word order. While student #10 made two instances of this error type at T3 and one at T6, she never repeated an equivalent linguistic context, thus enabling a comparison between the incorrect forms. The incorrect linguistic feature listed in Line 11 above refers to the incorrect placement of the adverb[55]. Student #10 could not correct herself at level one; she did at levels two and three after consulting the error type once at L2. The fact that she could provide a correct alternative with little help denotes that the incorrect form holds a relatively high position in the ZPD, close to independent performance. However, since she never made a similar error nor used any adverbs in equivalent configurations at T6, the data does not enable to confirm or infirm whether this particular error could have developed over time to finally be fully internalised.

Another example of variations in interlanguage development is illustrated with learner #10's incorrect noun adjective agreements she produced at T3 and T6 (Table 7.10).

Table 7.10. Learner #10's incorrect noun adjective agreement produced at T3 and T6

Incorrect form status	**Incorrect forms at T3**	**Incorrect forms at T6**
1	• libre (free) • moderne (modern) • technologique (technological) • serviable (helpful) • quelque (some)	• différent (different)
0		• complémentaires (complementary)
-1	• différent (different) • première (first) • tout (all) • dernières (last) • inquiète (worried)	• complémentaires (complementary) • compliquer (complicated) • tout (all) • meilleur (better)

1: mistake confirmed**; 0:** error confirmed for which the learner could propose a replacement with assistance; **-1**: error confirmed for which the learner could not propose any replacement even with assistance.

55. Adverbs generally follow the tense-marked verb.

For example, both incorrect forms listed in Lines 12 and 13 below were marked as mistakes (labelled as 1 in Table 7.10); she could provide the correct forms without assistance.

> [12]. *Les amphis sont *moderne (The conference rooms are modern) - correct form: modernes*

> [13]. *Les profs sont très *serviable (The teachers are very helpful) - correct form: serviables*

Student #10 did not seem to experience any difficulties in correcting herself in the case of predicative adjectives. In such a case, the adjective (in French) must agree in number and gender with the subject. However, when the nominal group is composed of more than one noun, or when the verbal group includes auxiliary and past participle, her answers were uncertain. While learner #10 provided only one alternative, which was appropriate, to the incorrect form listed in Line 14 at level three after reading the feedback, she did not suggest any replacement for the ill-formed word listed in Line 15, despite the fact that she read the error type at level two.

> [14]. *Moi et l'autre membre sommes très *inquiète (Me and the other member are very worried) - correct form: inquiètes*

> [15]. *Les dernières semaines ont été *différent (The last weeks have been different) - correct form: différentes*

As a result, it could be advanced that learner #10 struggled with more complex situations than simple noun phrase agreement merely including a determinant, noun and adjective (Det N A), such as the mistake she made at T6 (Line 16).

> [16]. *Une ville *différent (a different city) - correct form: différente*

While she made no errors in simple constructions (Det N A) at T3, she could not correct herself without assistance at T6 in similar contexts (Lines 17 and 18).

> [17]. *J'ai parlé de l'information *complémentaires et de la conclusion (I talked about the additional information and the conclusion) - correct form: complémentaire*

> [18]. *J'ai donné l'information *complémentaires (I gave the additional information) - correct form: complémentaire*

The other errors she wrote at T6 confirmed her difficulties to make an agreement between adjectives and nouns when the former is not just next to the latter (Lines 19 and 20). While she read the assistance at level two (Line 19), she did not provide any alternatives at all. Additionally, she did not access any feedback for the incorrect form listed in Line 20, nor did she propose any replacement.

> [19]. *l'info n'est pas trop *compliquer (the information is not too complicated) - correct form: compliquée*

> [20]. *J'ai été un peu *meilleur (I did improve myself) - correct form: meilleure as the author of this statement is a female*

In summary, these examples listed above suggest that learner #10 did not improve over time the agreement error type between a noun and an adjective when the latter is not close to the former. The data also leads to the conclusion that learner #10 may have regressed in simple noun adjective agreement (Det N A), since she produced two incorrect occurrences marked as low in her ZPD at T6 whereas she never did at T3.

7.1.5. Discussion

In her study of the impact of metalinguistic explanations and clues on learner uptake, Heift (2010c) suggests that "as learners become more familiar with a CALL system and its metalinguistic feedback, the more likely they are to engage in the error correction process and thus benefit from the enhanced interaction" (p.211). The results obtained with regard to feedback access do not seem to follow the same line of observation, as it was demonstrated that both learners were more inclined to access their feedback during the first point in time (T3) rather than the second one (T6). As it was also established that both learners provided less alternatives at T6, it may be the case that, stressed by the submission deadline of their class assignment, they decided to shorten the self-editing task, thereby gaining time by not accessing the assistance nor providing alternatives to quite a few incorrect forms. Since learners suggested less replacements at T6 than at T3, and since most of their alternatives were not correct, it was expected to find more errors than mistakes in their texts in the second point in time. However, the fact that learners #10 and #17 did not provide any replacement for 41% and 45% of their incorrect forms at T6, respectively, does not signify that they did not know the correct answers. They may have known but did not take the time to correct themselves. Indeed, both learners #10 and #17 attempted the self-editing exercises on the afternoon they had to submit their assignment to their teacher for the end of semester grading purposes.

The diachronic analysis conducted within this section confirmed that errors could be random[56]. According to Corder (1967), errors are systematic as opposed to mistakes which are unsystematic. Yet, systematicity and non-systematicity were found in errors in the two time periods. With regard to systematicity in errors, learner #10 for instance

56. In contrast with errors, occurrences of mistakes being systematically wrong were equally observed and in particular in the misspelling category. For example, learner #10 consistently wrote the word **trés (very - correct form: très)* incorrectly at T3, which she could consistently correct without assistance (therefore a mistake). At T6, the word was written correctly in all occasions.

seemed to experience difficulties in noun adjective agreement in complex noun phrases. Such errors were consistently observed in both periods, thus inferring systematicity in her interlanguage competence. With regard to non-systematicity, the same learner made errors that appear only once over the two time periods. It was shown that learners made new ill-formed words at T6 that had never occurred before. Such errors were probably due to their attempts to use more sophisticated sentences and lexical choices.

The above analysis further suggests that merely comparing scores in language accuracy obtained at two points in time is not a sufficient indicator to measure learners' achievement and development. The case of learner #17 who scored less at T6 (75.1%) than at T3 (84.8%), as seen in Table 7.6 on page 204, illustrates the fact that even with a lower score in language accuracy at T6, a development may be observed. The analysis demonstrated that both learners improved their language in the sense that they both increased their coefficient of balanced complexity measures over time. Their texts showed evidence of their language enhancement with regard to lexical choices and syntactical constructions. However, the analysis of the errors written could not determine whether they were common between them or idiosyncratic, as learners generally did not repeat their errors, or at least similar contexts for their errors to be compared.

In summary, the results obtained in actual and potential development showed that the learners' language accuracy increased and decreased over time, and that no patterns could be established depending on learners themselves and/or error types. The learners' attempts to produce more complex and sophisticated language suggest that they constructed new zones of proximal development, where the potential development of T3 did not systematically lead to the actual development observed at T6. With two points in time, whether the general progression is a straight line or a curve could not be determined; this is further analysed and developed with the diachronic intra-learner analysis below.

7.2. Diachronic intra-learner analysis

Table 7.11 summarises the data produced by student #2 during the two academic semesters under investigation within this section. The time periods at which the texts were collected, labeled from T1 to T6, cover the months of October 2008 to April 2009.

Table 7.11. Data considered for the diachronic intra-learner analysis

Time	Text id	Text type	Text count	Token count	Incorrect sequence	Incorrect token	% of incorrect sequence-token		% of feedback not read
T1* Oct 2008	#5 #11 #12 #13 #14 #15 #16	forum	7	616	79	105	12.8	17.0	NA
	#6	bilan	1	168	21	28	12.5	16.7	NA
	#4 #10	email	2	141	25	37	17.7	26.2	NA
T2* Nov 2008	#19	bilan	1	534	210	117	39.3	21.9	NA
	#17 #20 #23 #25 #29 #30	email	6	154	30	46	19.5	29.9	NA
	#18 #24	écrire	2	457	60	82	13.1	17.9	NA
	#26 #27 #31 #32	wiki	4	1090	106	170	9.7	15.6	NA
T3 Dec 2008	#65	bilan	1	485	69	111	14.2	22.9	15.2
	#43 #59 #66	email	3	107	19	22	17.8	20.6	18.4
	#34 #48	wiki	2	481	40	57	8.3	11.9	33.8
	#54	parler	1	721	78	113	10.8	15.7	6.4
T4 Feb 2009	#70 #71 #73	email	3	219	45	51	20.5	23.3	20.0
	#72	écrire	1	53	4	6	7.5	11.3	25.0
T5 Mar 2009	#76 #77 #80	email	3	93	15	18	16.1	19.4	16.7
	#75 #79	écrire	2	809	171	207	21.1	25.6	26.3
T6 Apr 2009	#86	bilan	1	671	81	137	12.1	20.4	17.3
	#85 #83 #84 #87 #88	email	5	205	34	47	16.6	22.9	23.5
	#82	écrire	1	500	71	113	14.2	22.6	4.9
Total			47	8047	1021	1471	12.7	18.3	19.4

* Periods for which there is no information on feedback access.

It is important to mention that the data for which there is no information on the learner's behaviour with regard to feedback access, i.e., T1 and T2, is only considered in Section 7.2.1, which does not depend on feedback access to analyse the variations in language sophistication and complexity.

7.2.1. Variations in language sophistication and complexity

Each coefficient obtained with respect to the balanced complexity measure, along with their individual vectors, are listed in Table 7.12 below.

Table 7.12. Student #2's balanced complexity measures detailed per vector and text type

Time	Text type	Text count	Vector 1	Vector 2	Vector 3	Vector 4	Balanced
T1	forum	6	3.6929	4.3990	8.3868	5.0927	14.3896
	bilan	1	4.8407	4.8163	9.8000	8.2514	20.6081
	email	1	4.3333	4.4444	10.2857	5.7904	16.7348
T2	bilan	1	5.3565	4.4506	13.1351	13.0359	25.2295
	email	6	2.8402	4.6531	7.4361	3.1690	11.1607
	écrire	2	5.0489	4.2954	11.4488	9.0590	20.5155
	wiki	4	5.1961	4.5772	12.0184	9.7267	21.9835
T3	bilan	1	5.6697	4.2597	11.5526	12.6701	23.6742
	email	6	3.3364	4.3730	8.5611	3.6516	12.4253
	wiki	2	4.1737	4.3506	11.4545	7.6358	17.6941
	parler	1	6.1762	4.4009	14.3556	15.7031	27.2780
T4	email	3	4.1886	4.1632	11.5397	5.4581	15.6097
	écrire	1	3.4345	4.9184	9.8000	4.6949	14.2802
T5	email	3	3.2502	5.1349	6.3333	3.5351	12.8556
	écrire	2	6.3157	4.4690	14.7361	11.9805	25.1560
T6	bilan	1	6.3171	4.3847	14.7500	14.5096	27.5380
	email	5	3.0535	4.2440	7.5810	3.8349	11.9201
	écrire	1	6.2516	4.2604	14.7419	13.5433	26.2495

Balanced: includes four vectors based on lexical and grammatical complexity and sophistication measures**;** **vector 1**: lexical complexity (text-length adjusted type token ratio); **vector 2**: lexical sophistication (mean length of a word); **vector 3**: syntactic sophistication (mean period unit length); **vector 4**: syntactic complexity (unique bigram ratio).

Descriptive statistics (Explore) in SPSS, which include indicators such as means, medians, outliers, and percentiles, were used to visually determine whether the different text types were performed with various or similar levels of lexical and syntactical complexity. The means and medians of balanced complexity measures for each text type are listed in Table 7.13 below. The descriptive table as seen in SPSS is given in full in Appendix G.2 on page 325.

Table 7.13. Descriptive table of balanced complexity measures per text type

Text type	Mean	Median	Minimum	Maximum	Range	Std. Deviation
forum	14.58	12.20	10.00	21.31	11.30	4.53
email	12.70	12.53	8.40	17.86	9.46	2.28
bilan	24.26	24.45	20.61	27.54	6.93	2.91
écrire	21.98	22.98	14.28	26.46	12.18	4.70
wiki	20.55	22.49	11.75	23.64	11.90	4.54

A first reading of the values as listed in Table 7.13 shows that the means and medians for (a) forum and email, and (b) bilan, écrire, and wiki are virtually the same, ranging from 12.20 to 14.58 for the former and 20.55 to 24.45 for the latter. Table 7.14, displaying the scores for the M-Estimators, i.e., robust alternatives to means and medians, confirms the tendency of having two groups in terms of lexical and syntactical complexity. For instance, the results obtained from the Huber's M-Estimator range from 12.59 to 13.38 for forum and email, and from 22.32 to 24.45 for all the other text types.

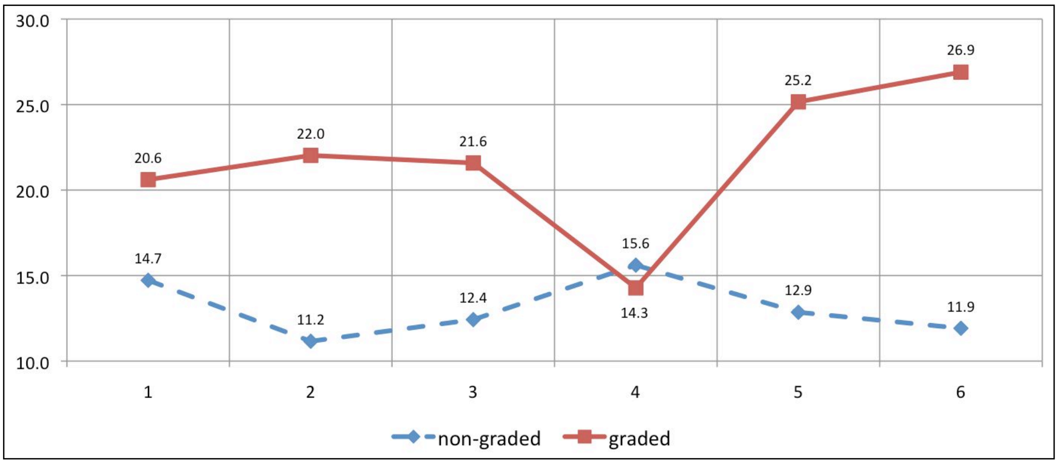
Table 7.14. M-Estimators

Text type	Huber's M-Estimator	Tukey's Biweight	Hampel's M-Estimator	Andrew's Wave
forum	13.38	12.53	13.60	12.45
email	12.59	12.48	12.55	12.48
bilan	24.45	24.40	24.40	24.40
écrire	22.62	22.58	22.38	22.57
wiki	22.32	22.70	22.55	22.70

One major characteristic that could be observed between these two groups relates to the fact that the one exhibiting the higher scores was formally assessed and graded by her teacher and the other was not. Accordingly, the two groups were called *graded* and *non-graded* groups. The graded group includes the *wiki*, *écrire* and *bilan* text types, all written with the objective of an academic assessment at the end of the course; the non-graded group contains the *email* and *forum* conversations which were written with less sophisticated and complicated sentences, and probably produced with less pressure as they were not assessed nor graded.

The learner's variations in language complexity for the two new constituted groups, i.e., graded and non-graded, during the two academic semesters (from T1 to T6) is illustrated in Figure 7.3 below. Focusing first on non-graded documents (dashed-line), the degree of complexity starts with a rather high coefficient of 14.7 and then decreases to 11.2 at T2. At T1, learner #2 used a more complex language in both the forum discussion and the email correspondence (balanced complexity measure: 14.39 and 16.73, respectively). Then, the forum discussion was not pursued, and the content of the email became less complex.

Figure 7.3. Student's overall progression with regard to language complexity in graded and non-graded texts



Line 21 and Line 22 are two examples of emails she sent at T1 and T2, respectively. They illustrate the evolution of style in terms of language complexity, where the learner adopted a more casual register with regard to sentence length and complexity after T1.

> [21]. [57] Bonjour Sylvia, Je me suis amusee le autre jour et je vous voudrais recontre pour une tasse de tee, peut etre l jeudi! La semaine prochain J'ai un exam en "introducing to Language". Donc, Je suis tres occupee avec mes devoir...

> [22]. Bonjour un autre exercice pour toi. Merci pour ton aide (*Hi, I have another exercise for you. Thanks for your help*)

The fact that learner #2 started to address her correspondent (this researcher) using the tu-form[58] (Line 22) indicated that she began reducing the distance between both of them, thus inferring a probable more spontaneous communication. As a result, the sentences were shortened to simple statements, the use of connection words was reduced and the punctuation was often missing. The slight increase in language complexity at T4 provides the second peak (15.6) on the non-graded documents. Learner #2 started writing long and more complicated sentences in her email, combining simple sentences by using connection words. However, the increased complexity in non-graded documents was observed only at that time; balanced complexity measures decreased afterwards.

The above figure (Figure 7.3) also demonstrated an increase in terms of lexical and syntactical complexity over time in graded documents ranging from 20.6 at T1 to 26.9 at T6. The reason for the decrease at T4 (14.3) could be justified by the fact that learner #2

---

57. Rough translation: Hi Sylvie, I had fun the other day and I would like to meet you for a cup of tea, maybe Thursday! Next week I have a test in "introducing to language". Therefore, I am very busy with my homework...

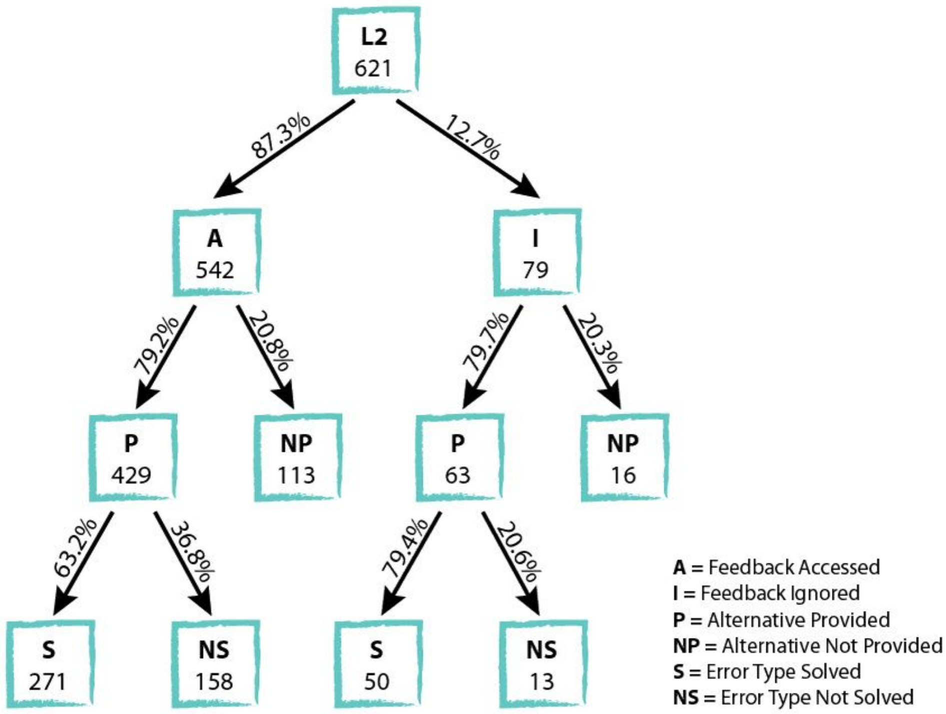
58. Douglass (2009) reports that "*tu* is to be used amongst students or young people [...], while *vous* [...] is to be used between strangers or people who are meeting for the first time or who do not know each other well and with people who are older or in a higher-level position" (p217).

submitted in all only 53 tokens for the graded text category at that period, whereas the other points in time record up to 2,081 tokens. It may be the case that the data collected at T4 was not sufficient to best represent the learner's capability to produce complex and sophisticated sentences. Consequently, it can be advanced that learner #2 gradually increased her level of balanced complexity measure in graded documents over time.

7.2.2. Learner's responses to interventions

Compared to the other participants, student #2 responded to a greater extent to the assistance provided at levels two and three (Figures 7.4 and 7.5), with an overall result of 81.6%[59] and 81.3%[60] of feedback accessed and alternatives provided, respectively.

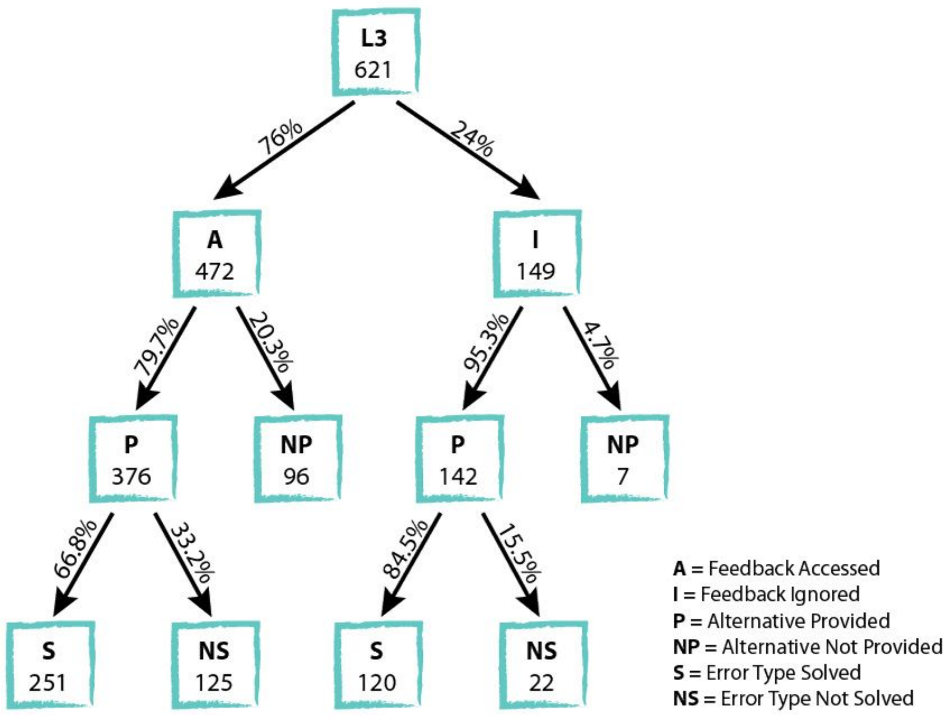
Figure 7.4. Overview of student #2's feedback access at level two (L2)



59. Feedback accessed at L2: N=542; at L3: N=472. (542+472)/1242*100=81.6% (Figures 7.4 and 7.5).

60. Alternatives provided at L2: N=429+63; at L3: N=376+142. (429+63+376+142)/1242*100=81.3% (Figures 7.4 and 7.5).

Figure 7.5. Overview of student #2's feedback access at level three (L3)



The figures given in Table 7.15 show the frequencies and percentages of feedback accessed per time period, i.e., T3 to T6 (for which the information on feedback access was available), and also displays how many times the assistance was retrieved. Most pop-up windows were opened at least once or twice. Some annotations were read six, seven and even ten times. For example, the assistance provided to the incorrect form **ágée (old)* at T3, where the correct orthography should be *âgée,* was accessed seven times at L2 and ten times at L3. The assistance provided was *incorrect or missing accent* and then *circumflex accent*. The first alternative provided at L1 (**ágé*) suggests that the learner had issues with the agreement between this adjective and the noun it qualifies – *étudiante (student, feminine form)* – since she removed the feminine mark from the adjective. After reading the feedback at L2, she entered the incorrect word *ágéé restoring thus the noun adjective agreement, and added an incorrect accent. Finally, the assistance provided at L3 enabled the student to provide the proper orthography (*âgée*).

Table 7.15. Student #2's responses to assistance at levels two and three per time period

Time	Level 2			Level 3		
	Feedback access	**Frequency**	**Percent**	**Feedback access**	**Frequency**	**Percent**
T3	0 time	23	11.2%	0 time	42	20.4%
	1 time	131	63.6%	1 time	56	27.2%
	2 times	42	20.4%	2 times	69	33.5%
	3 times	5	2.4%	3 times	26	12.6%
	4 times	4	1.9%	4 times	9	4.4%
	7 times	1	.5%	5 times	2	1.0%
	Total	206	100%	6 times	1	.5%
				10 times	1	.5%
				Total	206	100%
T4	0 time	14	29.2%	0 time	6	12.5%
	1 time	27	56.3%	1 time	16	33.3%
	2 times	6	12.5%	2 times	16	33.3%
	4 times	1	2.1%	3 times	7	14.6%
	Total	48	100%	4 times	3	6.3%
				Total	48	100%
T5	0 time	27	14.6%	0 time	67	36.2%
	1 time	111	60.0%	1 time	54	29.2%
	2 times	32	17.3%	2 times	34	18.4%
	3 times	10	5.4%	3 times	19	10.3%
	4 times	3	1.6%	4 times	7	3.8%
	5 times	2	1.1%	5 times	3	1.6%
	Total	185	100%	6 times	1	.5%
				Total	185	100%
T6	0 time	15	8.2%	0 time	34	18.7%
	1 time	131	72.0%	1 time	52	28.6%
	2 times	28	15.4%	2 times	65	35.7%
	3 times	7	3.8%	3 times	25	13.7%
	5 times	1	.5%	4 times	6	3.3%
	Total	182	100%	Total	182	100%

Breaking down the analysis per graded and non-graded categories, SPSS t-test for independent samples demonstrated no significant differences[61] in feedback access between the *non-graded* (M=1.26, SD=.959) and *graded* (M=1.35, SD=1.07) groups;

61. The group statistics and independent samples test results are given in full in Appendix G.1 on page 324.

(*t-test*(1240)=-1.223, *p*=.222, *p*>.05). These results suggest that learner #2 did not make any differences between text types when correcting herself. She accessed the assistance in a similar manner whether the texts were written for her final assessment or not.

7.2.3. Variations in learner's errors and mistakes over time

The total amount of incorrect and correct tokens, along with the percentage of mistakes and errors encountered in texts at each time period are displayed in Table 7.16 below. The amounts of confirmed and not confirmed errors and mistakes[62] are also given to provide an accurate representation of the learner's knowledge (see Section 5.5.3 on page 135 to distinguish between confirmed and unconfirmed errors/mistakes).

Table 7.16. Amount of correct and incorrect forms including errors and mistakes

Time	**Total tokens**	**Correct forms**	**Incorrect tokens**		**Errors not confirmed in error total**	**Mistakes not confirmed in error total**	**% of mistakes in incorrect tokens**	**% of errors in incorrect tokens**
			Error	**Mistake**				
1	925	755	111	59	46	0	34.71%	65.29%
2	2,235	1,820	269	146	55	2	35.18%	64.82%
3	1,794	1,491	200	103	0	0	33.99%	66.01%
4	272	215	37	20	0	0	35.09%	64.91%
5	902	677	136	89	0	0	39.56%	60.44%
6	1,376	1,079	205	92	0	0	30.98%	69.02%
Total	7,504	6,037	958	509	101	2	34.70%	65.30%

Table 7.16 indicates that there are a relatively large amount of errors that were not confirmed at T1 (N=46) and T2 (N=55). Recall, to determine whether an incorrect form is a confirmed error or a confirmed mistake, the amount of assistance required to provide the alternative must be declared. Given the fact that learner #2 highly accessed her feedback at T4 to T6, it may be assumed that she equally behaved at T1 and T2. If the

62. Appendix G.3 on page 327 details the amount of errors and mistakes confirmed and not confirmed per text type at each time period.

feedback is interpreted as being accessed, the unconfirmed errors and mistakes are considered as competence-dependent errors. Yet, it may be the case that these particular unconfirmed errors are in fact mistakes. Since the distinction could not be established with certainty at both T1 and T2, it was decided that not analysing the data for which there is incomplete information would be safer for the interpretation of the results.

Breaking down the analysis per text group, i.e., graded and non-graded as declared in Section 7.2.1, Table 7.17 below demonstrates that learner #2 generally made a higher percentage of mistakes in non-graded documents, as opposed to graded texts such as wiki or bilan. By contrast, learner #2 produced more errors in graded than non-graded texts.

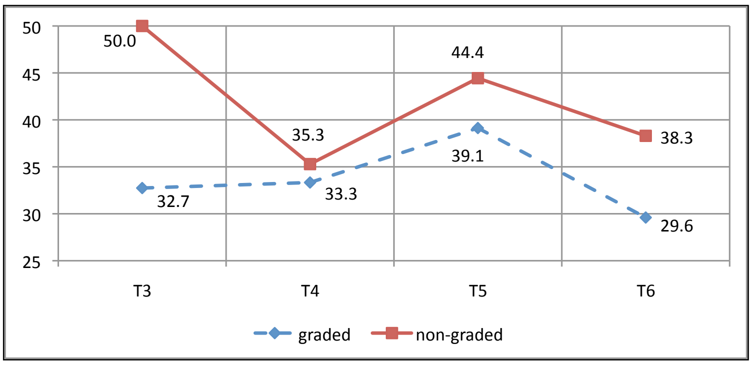
Table 7.17. Percentage of mistakes and errors in graded and non-graded documents

Time	Text group	Total tokens	Correct forms	Incorrect tokens		% of mistakes in incorrect tokens	% of errors in incorrect tokens
				Error	Mistake		
T3	graded	1687	1406	189	92	32.74%	67.26%
	non-graded	107	85	11	11	50.00%	50.00%
T4	graded	53	47	4	2	33.33%	66.67%
	non-graded	219	168	33	18	35.29%	64.71%
T5	graded	809	602	126	81	39.13%	60.87%
	non-graded	93	75	10	8	44.44%	55.56%
T6	graded	1171	921	176	74	29.60%	70.40%
	non-graded	205	158	29	18	38.30%	61.70%

The percentages of mistakes she made over time are illustrated in Figure 7.6 below. Excluding the peak at T5 (39.1%), the graphical representation demonstrates a relatively constant proportion of mistakes in incorrect forms produced in graded documents (dashed-line) and even a slight gradual decrease from T3 to T6, ranging from 32.7% to 29.6%. The results obtained for the non-graded texts (plain line), however, do not represent any gradual drop, fluctuating between 50% at the most and 35.3% at the least.

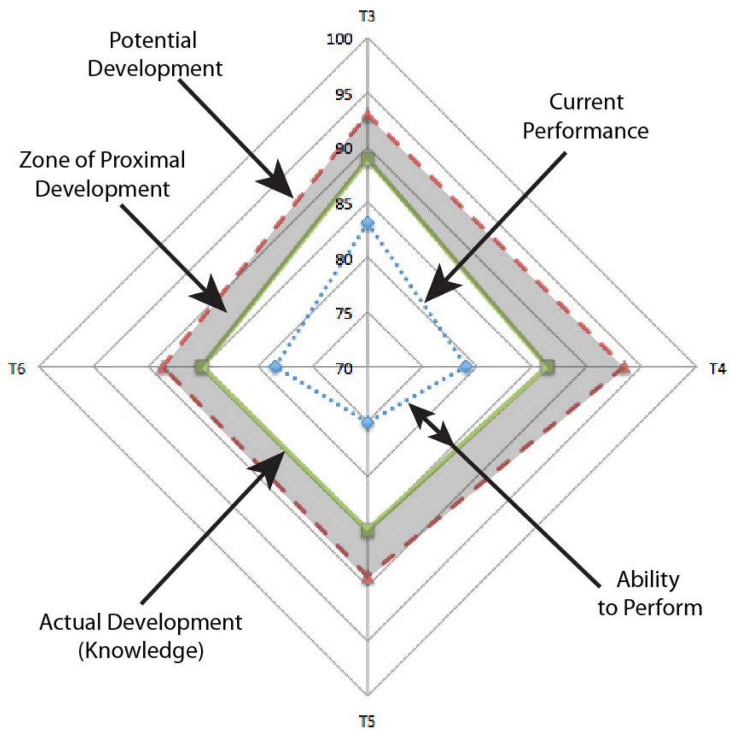
In all, the figure below demonstrates that learner #2 was more capable of performing at the level of her knowledge in a graded document than in a non-graded text.

Figure 7.6. Percentage of mistakes in graded and non-graded documents



Learner #2's overall performance, actual and potential development at each point in time, are graphically represented with Figure 7.7 below.

Figure 7.7. Learner #2's overall performance, actual and potential development



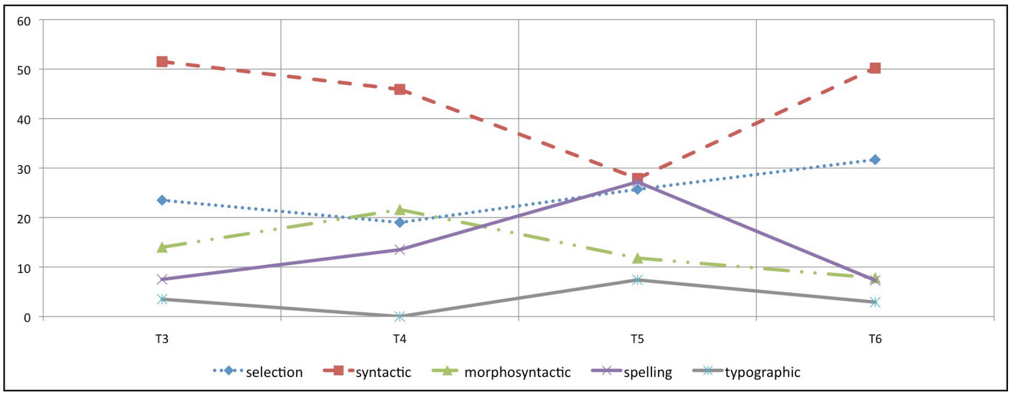
The spider chart illustrates the extent of (a) the learner's ability to perform her knowledge with respect to language accuracy, and (b) her capability to produce correct alternatives with assistance either at level two or three of the regulatory scale. Figure 7.7 shows that learner #2's percentage of success in performance was rather distant from her actual development, which denotes an incapability to perform at the level of her knowledge.

The chart further demonstrates that her potential development was observable, as opposed to the data analysed in the synchronic and diachronic inter-learner analysis (Section 6.3 on page 178 and Section 7.1.3 on page 202, respectively). In these sections, it was found that since learners did not access their feedback to correct themselves, their level of potential development (what they could have done with assistance) could not be determined. In other words, their zone of proximal development could not be created, and a fortiori observed. Learner #2, on the other hand, accessed the assistance to correct herself. By doing so, she gave an indication about what she will be able to do independently in a proximal time frame, thus revealing her potential development. As a result, her ZPD could be created and observed.

7.2.4. Variations in actual and potential development over time

Breaking down learner #2's errors per main error category gives an overview of the learner's domains of difficulties (Figure 7.8).

Figure 7.8. Percentage of errors made per error category over time



The categories considered in the above figure are selection (dotted-line), syntactic (dashed-line), morphosyntactic (dashed and dotted-line), spelling (plain line), and typographic (bottom plain line). From reading this figure, it may be interpreted that the categories for which the student had most difficulties were selection, syntactic, and morphosyntactic. The peak in spelling at T5 is mostly due to her overuse of one single word. She had to describe a Francophone country and chose Lebanon and always referred to it with an incorrect orthography (**Libanon* instead of *Liban*).

Table 7.18 below shows the proportion of errors per error category and illustrates the fact that learner #2 struggled in selection, syntax, and morphosyntax in all her documents whether they were graded or not. It was, however, expected to observe less syntactic errors in non-graded documents as learner #2 produced less complicated sentences in those types of texts. This might imply that her difficulties could be independent of text types. However, the fact that learners could avoid particular lexical or syntactical features should not be excluded. Laufer and Eliasson (1993) suggest that avoidance is “determined more by a systemic incongruence between the first language (L1) and the second language (L2) than by the inherent difficulty of L2 forms” (p.35). In other words, avoidance would be more related to the level of dissimilarity between a learner’s native language and his or her L2, than the degree of complexity occurring in the language to be learned. More recently however, Kormos (2011) found that “allowing students to generate their own content in narrative writing tasks does not lead to the avoidance of linguistic constructions that the students have not fully mastered yet." (p.159).

Table 7.18. Proportion of error per error category in graded and non-graded documents

		T3	T4	T5	T6
selection	graded	22.8%	0%	27.8%	20.7%
	non-graded	54.5%	21.2%	10%	33.5%
syntactic	graded	54.5%	75%	28.6%	41.4%
	non-graded	0%	42.4%	20%	51.7%
morphosyntactic	graded	13.8%	25%	10.3%	6.9%
	non-graded	0%	21.2%	20%	8%

Learner #2's scores in performance, actual and potential development with regard to language accuracy are displayed below. Table 7.19 enumerates the results obtained in the selection errors types, Table 7.20 gives the measures obtained in the syntactic error types, and Table 7.21 lists the scores obtained in the morphosyntactic category.

Table 7.19. Percentage of language accuracy in knowledge: selection

Student #2		mu	lw	re	sn	cw	rv	ar	pr	mo	te	vo	cl
T3	P	99.8	97.8	100	100	100	100	94.7	91.8	98.7	98.7	98.7	99.8
	AD	100	98.0	100	100	100	100	96.6	94.5	98.7	99.3	99.3	99.8
	PD		98.8					99.5	96.2	99.3	100	100	100
T4	P	98.4	95.5	100	100	100	100	100	95.7	95.5	95.5	100	100
	AD	**98.4**	95.5	100	100	100	100	100	95.7	**95.5**	100	100	100
	PD	**98.4**	96.6						100	**95.5**			
T5	P	100	96.8	100	98.4	93.2	100	89.4	83.0	100	100	100	98.9
	AD	100	**96.8**	100	**99.5**	100	100	95.9	88.3	100	100	100	**98.9**
	PD		**96.8**		**99.5**			99.2	89.4				**98.9**
T6	P	99.7	96.0	100	98.0	98.7	100	95.3	91.4	97.1	95.2	100	98.7
	AD	**99.7**	96.2	100	**98.0**	100	100	97.7	**93.1**	**97.1**	97.1	100	98.7
	PD	**99.7**	96.7		**98.0**			100	**93.1**	**97.1**	99.0		99.0

mu: nonstandard word; **lw**: inappropriate lexical choice; **re**: inappropriate register; **sn**: not understandable; **cw**: inappropriate connection word; **rv**: reflexive verb; **ar**: incorrect article type; **pr**: incorrect preposition; **mo**: incorrect mood; **te**: incorrect tense; **vo**: incorrect voice; **cl**: incorrect word class; **bold**: level of actual development equals level of potential development; **shaded areas**: difference between both levels; **P**: performance; **AD**: actual development; **PD**: potential development.

Table 7.20. Percentage of language accuracy in knowledge: syntax

Student #2		**ad**	**om**	**wo**	**un**
T3	Performance	99.2	97.1	99.6	96.6
	Actual Dev.	99.2	97.5	99.7	**96.8**
	Potential Dev.	99.4	98.8	100	**96.8**
T4	Performance	99.5	93.1	100	100
	Actual Dev.	**99.5**	93.1	100	100
	Potential Dev.	**99.5**	96.5		
T5	Performance	99.0	97.8	99.3	97.7
	Actual Dev.	99.3	98.0	**99.6**	97.7
	Potential Dev.	99.6	98.4	**99.6**	98.3
T6	Performance	98.4	97.4	99.2	94.4
	Actual Dev.	99.2	97.9	99.2	**94.6**
	Potential Dev.	100	98.8	99.8	**94.6**

ad: word addition; **om**: word omission; **wo**: incorrect word order; **un**: not understandable; **bold**: level of actual development equals level of potential development; **shaded areas**: difference between both levels.

Table 7.21. Percentage of language accuracy in knowledge: morphosyntax

Student #2		**dn**	**ge**	**na**	**pa**	**pp**	**sv**	**co**	**pl**	**wf**
T3	Performance	99.5	97.3	89.7	98	89.3	99.3	95.5	99.7	100
	Actual Dev.	100	98.9	93.1	**99.3**	89.3	100	96.1	99.7	100
	Potential Dev.		100	95.4	**99.3**	96.4		98.1	100	
T4	Performance	100	92.6	64.3	100	50.0	95.5	87.5	100	96.8
	Actual Dev.	100	100	71.4	100	50.0	100	95.8	100	**98.4**
	Potential Dev.			100		100		100		**98.4**
T5	Performance	96.9	95.9	78.6	93.3	55.6	98.0	92.3	100	100
	Actual Dev.	**97.9**	99.0	85.7	**93.3**	**88.9**	100	98.1	100	100
	Potential Dev.	**97.9**	100	88.1	**93.3**	**88.9**		100		
T6	Performance	98.7	99.4	83.1	98.9	94.7	99.0	94.3	99.5	98.0
	Actual Dev.	100	100	**94.9**	100	97.4	**99.0**	95.3	100	99.3
	Potential Dev.			**94.9**		100	**99.0**	96.2		99.7

dn: determiner noun agreement; **ge**: gender agreement; **na**: noun adjective agreement; **pa**: pronoun antecedent agreement; **pp**: past participle agreement; **sv**: subject verb agreement; **co**: incorrect conjugation form; **pl**: incorrect plural form; **wf**: incorrect word formation; **bold**: level of actual development equals level of potential development; **shaded areas**: difference between both levels.

Overall, none of the results in actual and potential development display stable, regular, or gradual changes over time. They all show periods of more or less success in language accuracy. For example, learner #2 seemed to have regressed and progressed in noun adjective agreement (na) scoring 93.1%, 71.4%, 85.7%, and 94.9% of success in language accuracy from T3 to T6 (Table 7.21). The three tables above further identify the numerous occasions the learner had in order to create ZPDs (shaded areas). There is a tangible zone of proximal development for a vast majority of all her incorrect types for which she needed assistance. By correcting herself at a specific point in time without assistance, she showed the extent of her actual development (e.g., T4/na: 71.4%). By accessing the assistance at levels two and/or three, she showed her potential development, i.e., the extent of her capabilities with assistance at the same point in time (e.g., T4/na: 100%).

7.2.5. Variations in interlanguage development

The following example illustrates the microgenetic development of a linguistic feature she wrote incorrectly, which was first produced at T3. Engaging in various text types, she had the possibility to repeat the exact same incorrect sequence on several occasions. The incorrect sequence, **avec autres (with others, correct form: avec d'autres)*, was tagged as a *word omission* error type. It was incorrectly written in three different text types, i.e., wiki, email, and bilan, at T3, T4 and T6, respectively. This strengthens the idea that a misconception is independent of the text types (graded or non-graded). The contexts in which the incorrect sequence occurred are listed below from Lines 23 to 26.

> [23]. *Parfois je vais au restaurant *avec autres* étudiants. (Sometimes I go to the restaurant with other students - correct form: avec d'autres)* - Time 3
>
> [24]. *Je travaille *avec autres* étudiants. (I work with other students - correct form: avec d'autres)* - Time 4
>
> [25]. *Travailler *avec autres* étudiants... (Working with other students... - correct form: avec d'autres*) - Time 6

[26]. *J'ai appris la géographie de la France et les frontières *avec autres* pays (I learned the geography of France and the borders with other countries - correct form: avec d'autres)* - Time 6

Student #2 produced altogether numerous sentences with the preposition *avec (with)*, all of them being correct apart from rare exceptions, such as **si c'est d'accord avec vous*, which is a direct translation of *if it is all right with you*. Mostly, she knew when to add a determinant after the preposition when necessary: e.g., *avec du sucre (with sugar)* or *avec les chutes de neige (with the snowfall)*. Yet, as soon as she employed the adjective *autre (other)* with the preposition *avec (with)*, she systematically omitted the determiner. Table 7.22 below summarises the alternatives she provided at each level along with her feedback access.

Table 7.22. Alternatives provided to the incorrect form *avec autres

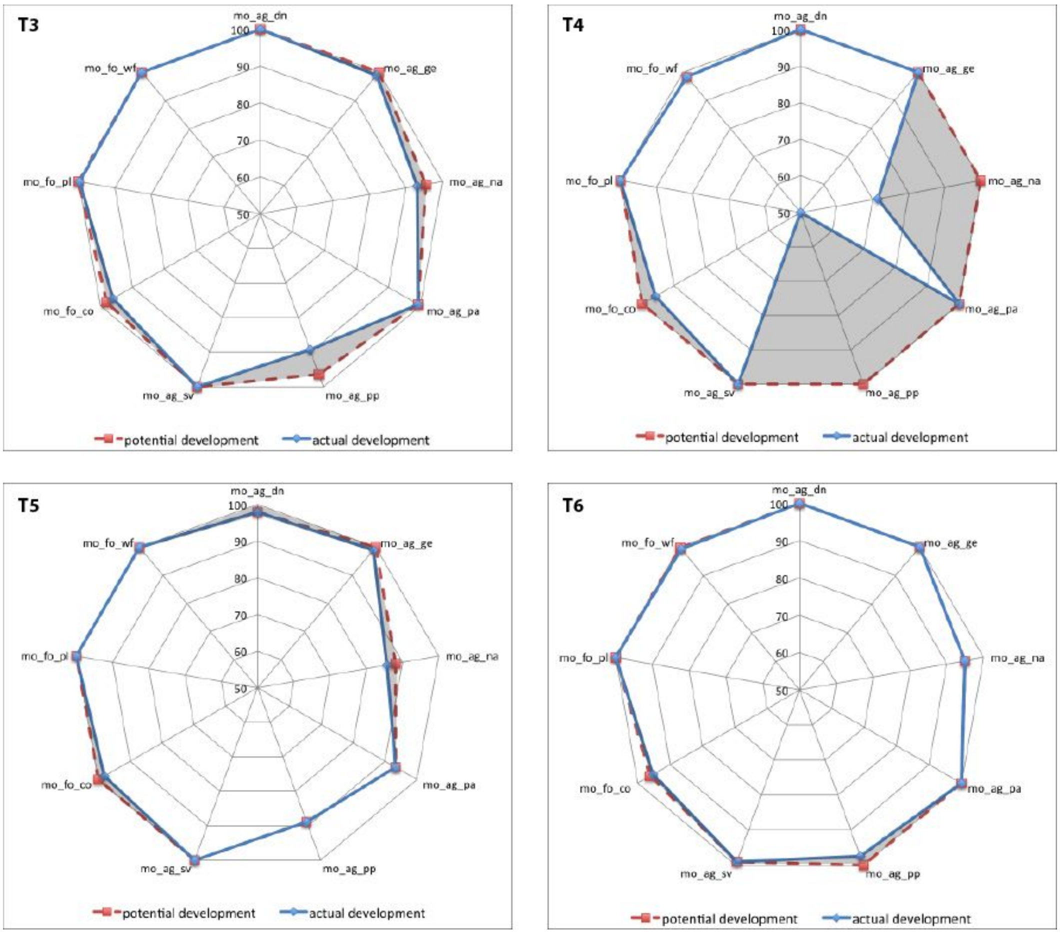
Time	Text type	Alternatives provided	Feedback accessed
T3 Line 23	wiki/graded	L1: blank L2: avec d'autres L3: *avec l'autres	NA 4 times 4 times
T4 Line 24	email/non-graded	L1: *aux autres L2: avec d'autres L3: *avec l'autre	NA 1 times 2 times
T6 Line 25	bilan/graded	L1: blank L2: avec d'autres L3: avec d'autres	NA 1 times 2 times
T6 Line 26	bilan/graded	L1: blank L2: avec d'autres L3: avec d'autres	NA 1 times 2 times

Learner #2 accessed the assistance and provided alternatives at each level. Up to T4, she entered correct alternatives at level two with the error type provided, and incorrect alternatives at level three with meta-linguistic annotations, which may suggest that the assistance provided at that level was probably not adapted to her needs. At T4 level one, she replaced the preposition *avec* by the preposition *à* combined to the article *les*, that is, *aux*, which is equally incorrect in this context. Then at T6, she provided correct replacements at both levels two and three, but left the field blank at level one. Consequently, and despite the fact that she still needed help at T6 to self-edit her incorrect form, a gradual movement in her potential development could be observed over time for this particular error. The incorrect form moved in the learner's ZPD from low to mid-high. As soon as she can produce the correct form at level one, the erroneous form will move up to a higher position in the ZPD to be finally considered as an internalised feature.

The example illustrated above was a rare occasion to observe a microgenetic development in the learner data. It was a difficult task to find incorrect features that were repeated over time. Most errors were never reproduced in the same linguistic context which prevented any comparison between them. As a direct consequence, it may be advanced that the potential development observed at a specific point in time is not systematically transformed into actual development at a later point.

Figure 7.9 below considers the learner #2's results obtained in morphosyntax from T3 to T6 with regard to levels of actual and potential development (see Table 7.21 on page 226 for the scores). The spider charts illustrate the fact that learner #2's potential development as observed at time *t* does not lead to her actual development at time *t+1*. For example, learner #2 attained 93.1% of success in noun adjective agreement (mo_ag_na_) at T3, and her potential development recorded at the same time indicates that she was able to edit 95.4% of the noun adjective agreement correctly with assistance. Her next levels of actual development did not demonstrate any evidence of her potential development as recorded at T3. Her scores in language accuracy decreased at T4 (71.4%) and increased again at T5 (85.7%) to finally reach 94.9% at T6 in noun adjective agreement error type. Consequently, it may be put forward that a learner's level of potential development does not help identify the learner's *future* actual development, i.e., his or her *future* knowledge. Rather, a learner's level of potential development provides, besides information on dependent current performance, the source necessary to unveil the learner's *present* actual development, that is, what the learner knows today, as opposed to what he or she is only able to perform.

Figure 7.9. Learner's actual and potential development in morphosyntax from T3 to T6



7.2.6. Discussion

In general, student #2 proposed a large amount of alternatives at each level of assistance during each self editing task, and from these alternatives entered, a relatively high quantity of them were correct replacements. Given the fact that learner #2 accessed most of her feedback (81.6% on average), the progression in proposing correct alternatives may indicate that the assistance provided was helping the student improve her self-corrections. Moreover, since learner #2 consistently read the assistance when needed at levels two and three (monitored from T3 to T6), it could be assumed that her behaviour with respect to feedback access was equivalent during the first two time periods when the monitoring system was not put into operation yet. However, as discussed in Section 7.2.3 on page 220, it was decided to put the data for which the information on feedback access was incomplete aside as it was estimated that this data could tamper with the findings.

Student #2 proposed fewer alternatives at level three at the end of the study than at the beginning. One possible reason for the change in behaviour might be that she experienced difficulties in correcting herself with and without assistance, given the fact that she improved her level of language complexity over time. The analysis also demonstrated that text types did not impact on the learner's difficulties and knowledge as learner #2 equally struggled in selection, syntax and morphosyntax in graded and non-graded documents. Additionally, it was found that language knowledge with regard to language accuracy may have linear as well as non-linear patterns over time. In her study of interlanguage variation from a computer assisted language learning angle, Heift (2008) outlines "the dynamic and non-linear process of language learning" (p.318). While the percentages of success measured in actual development varied over time going up or down, it was shown that the evolution of a specific error within its ZPD was relatively linear going from low to high. No examples of errors moving unpredictably within the ZPD could be observed in the corpus so far, and no incorrect forms considered as mistakes were found to require assistance afterwards to be corrected. This suggests that while the learner's actual development varies in an unsystematic way over time, the variations observed within the learner's zone of proximal development are linear, thus systematic. The time span, however, to achieve the higher position is at this stage unpredictable. Furthermore, the time frame for this student was definitely too short and the sample data too small to observe any instances of fossilisation, a phenomenon that affects language acquisition and development. In summary, it can be advanced that the leaner's knowledge construction at the microgenetic level was not linear, but rather cyclic, which implies dynamic and constant changes. The analysis of the data revealed that learner #2 increased her lexical and syntactical complexity which may have led to the production of new errors. In a cyclic progression, as defined by Rutherford (1987), "the learner is constantly engaged in reanalysing data, reformulating hypotheses, recasting generalisations, etc" (p.159).

Finally, it was demonstrated that the learner's *future* actual development was not necessarily an extension of her *present* actual and potential development. The diachronic intra-learner analysis, as opposed to the synchronic and diachronic inter-learner analysis, provided plenty of information on the learner's potential development, given the fact that she accessed her feedback in a systematic way throughout the course of this study. The analysis demonstrated that the learner's future knowledge could not be predicted from her actual and potential development. Predicting a *future* actual development is a challenging endeavour as it has been established that the future could not be estimated from past-to-present performances as suggested by past-to-present models of assessment (see discussion in Section 3.3.2 on page 64). The diachronic intra-learner analysis carried out within this section demonstrated that the learner's *future* actual development could not be derived from her potential development observed at time present, as suggested by present-to-future models of assessment. While Van der Veer and Valsiner (1991) claim that "the performance of [learners] cooperating with more knowledgeable others [is] characteristic of their future development: it reveal[s] the results of tomorrow" (p.338), the findings of this section have demonstrated that the *results of tomorrow* could not be systematically characterised by a projection of the learner's potential development.

7.3. Summary and conclusion

The aim of this chapter was to analyse learners' data diachronically. The first section, an inter-learner analysis, focused on texts written by two students at two points in time. One characteristic found was that both learners did not access their feedback systematically. As a direct consequence, their level of potential development could not be observed. Variations in their level of actual development were found to be unsystematic, going from low to high, and high to low scores. The analysis further demonstrated that competence-dependent errors generally occurred randomly as learners did not reproduce them; errors were mostly a one time event. Since errors have been found to be at times systematic and some other times random, it may be advanced that knowledge varies over time not only systematically, but also randomly.

The second section investigated various text types produced by one learner over a longer period of time. This learner accessed her feedback systematically, which enabled the creation and representation of her zone of proximal development for most of her errors. Text types were found to have no effect on her actual development. What she did not know in graded documents was equally found incorrect in non-graded texts. In addition to show unsystematic variations in the learner's actual and potential development over time with regard to language accuracy, it was also demonstrated that the learner's levels of actual and potential development did not lead to her future knowledge. While the findings point to the conclusion that actual and potential development follow a cyclic process, the variations occurring within the zone of proximal development tend to be linear going from a low to high position.

Chapter 8. Conclusion

The synchronic and diachronic analysis presented above in Chapter 6 and Chapter 7, respectively, provided the material to answer the three research questions listed in the introductory chapter. More specifically, they demonstrated that learners' responses to assistance was not consistent, given the fact that students were more inclined to correct themselves without assistance. As a direct consequence, the learners' level of potential development was not systematically observable. However, the fact that most learners did not access the feedback to correct themselves did not impact the error-mistake distinction, as the key information required was the identification of what they were able to do *without* assistance. In effect, the learners' competence could be differentiated from their performance.

The synchronic analysis revealed that the interlanguage competence was highly variable between students due to the learners' varying and unpredictable levels of ability to perform their knowledge. The diachronic analysis established that the interlanguage competence with regards to lexical and grammatical features did not develop gradually. Rather, it demonstrated periodicity in interlanguage development. While interlanguage is described by Corder (1981) as a constantly developing system, yet systematic (p.67), the findings presented in Section 7.1.5 on page 210 and 7.2.6 on page 230 established that errors could be random as well as systematic. Interlanguage competence is thus not as stable as Gregg (1990) suggested it, since it is also predisposed to randomness.

To conclude this thesis, the current chapter reviews the major findings in light of the research questions, and before discussing directions for future research and pedagogical implications, it outlines some limitations that were exposed during the course of this study.

8.1. Research questions revisited

Through the lens of a sociocultural perspective, this research represents an innovative way of investigating interlanguage variability by differentiating between interlanguage performance and competence. As discussed in Chapter 2, performance-related mistakes are frequently considered as competence-dependent errors by default. The present study follows Corder's (1967) reasoning that distinguishing between errors and mistakes can optimally represent the learners' interlanguage competence. However, the distinction is not operated through Corder's (1967) recommendations, i.e., systematicity and non-systematicity of the incorrect forms, but rather through the amount of assistance accessed by students to correct themselves, as called for by James (1998). In such a method, an incorrect form merely pointed out, if self-edited by the learner, is a mistake.

By devising a procedure and implementing the tools that helped separate competence from performance in L2 French learners' unrestricted written language, this thesis contributes to the learner interlanguage body of knowledge. The research questions that provided the structure for this study are listed below, so that the discussion is guided and the major findings based on the evidence established in Chapters 6 and 7 are reviewed.

8.1.1. Question 1

> *Can the learners' zone of proximal development, i.e., their actual and potential development, and the distance in-between, be represented and observed so that errors and mistakes can be distinguished?*

According to Vygotsky (1997a), "[t]he central fact of our psychology is the fact of mediation" (p.138). Mediation is indeed a key concept in the realisation of the zone of proximal development along with the double stimulation method (Sections 3.1.2 on page 47 and 3.1.5 on page 53, respectively). Both notions represent key elements of this thesis to differentiate between interlanguage performance and interlanguage competence. This doctoral research also elaborates on Ellis' (1985c, 2009) claim that a distinction between errors and mistakes cannot be achieved in practice, and provides empirical data – built on the work of Corder (1967), James (1998) and Poehner (2008) – to demonstrate that the distinction may indeed be realistic.

The methodology behind the error-mistake distinction is graphically illustrated with the error-mistake distinction model discussed in Section 3.3.2 on page 64. It indicates that the learners' potential development, i.e., their present dependent performance, is initially considered to help distinguish between errors and mistakes. The second step is determined by ruling out each incorrect form from the learners' independent performance for which they could correct themselves with minimum assistance, as defined in the regulatory scale of assistance, thus revealing their current knowledge. The proposed model does not reflect any new concept in itself. Rather, it is an expanded version of the present-to-future model (see Section 3.3.2. on page 64) to help represent the learners' present knowledge, as opposed to their future knowledge.

In regards to the question *Can the learners' actual development be represented and observed?,* the data explored in both synchronic and diachronic analysis demonstrated that the learners' actual development could be delineated on the condition that learners provided the system with an alternative each time they knew one. In the event of no replacements, the incorrect forms were marked as competence-dependent errors, regardless of the learners' reasons to leave a field blank. An empty box may result from a lack of knowledge, but it could equally be caused by an uninterested learner who decided to skip a few corrections. The learners' interactions during the dynamic assessment, whether assistance is accepted or ignored, is a necessary procedure to the error-mistake distinction. Without the learners' participation, the method to distinguish between performance and competence is ineffective in itself. Sections 6.1.4 on page 163, 7.1.1 on page 198, and 7.2.2 on page 217 demonstrated that while learners did not access much of their feedback on all occasions, they nevertheless provided alternatives to a large proportion of their incorrect forms, which denotes a high participation and therefore, a high level of accuracy in knowledge detection.

Concerning the question *Can the learners' potential development be represented and observed?,* the analysis carried out in both Chapters 6 and 7 established that the representation of the learners' potential development was conditioned by the learners' behaviour with respect to feedback access. Without knowing the amount of help learners required to attain the correct answer, the potential development could not be determined and observed. As it was shown that most students did not access the assistance during the self-editing task, especially during the synchronic and diachronic inter-learner analysis (Sections 6.1 on page 152 and 7.1 on page 197), what was principally known was whether learners could or could not edit their incorrect forms independently. This does not affect in any case the error-mistake distinction, as the objective was to determine what they were able to do without guidance. The diachronic intra-learner analysis (Section 7.2 on page 212) however, established that the potential development could be observable as the learner under investigation accessed the assistance when correcting herself. As a result, the distance between her actual and potential development, i.e., her zone of proximal development was realised and observed. This answers the final question listed within this section, which is *Can the learners' zone of proximal development be represented and observed?.*

In summary, the representation of the learners' zone of proximal development, i.e, their actual and potential development, along with the distance in-between, heavily relies on the learners' participation during the dynamic assessment stage. Including dynamic assessment during the process not only assists in unveiling the participants' zone of proximal development, but also brings to light the extent of their current strengths and weaknesses with regard to lexical and grammatical features. As a result, insights into learners' knowledge is gained and their ability to use it is revealed. In some cases, the error-mistake distinction demonstrated that the assessment of learners' performance would have exaggerated their difficulties as learners performed below their capabilities, thus exhibiting a large number of mistakes. On other occasions, learners were practically able to show the full extent of their knowledge, given the fact that most of their incorrect forms were competence-dependent errors.

To conclude this discussion, this thesis not only demonstrated that the error-mistake distinction was realisable in practice, but also empirically established that drawing conclusions on learners' knowledge by assessing performance only can be misleading. Indeed, the learners' level of knowledge may be far greater than their performance. The error-mistake distinction may have a direct repercussion in classroom practices not only in universities but also in other contexts such as primary and secondary schools, where assessment is often a summary of learners' performance, and whose outcomes are generally used afterwards to make major decisions about the learners' future (Poehner 2007 p.323). Assessing a student is indeed a critical matter that is taken by the National Council for Curriculum and Assessment (NCCA) in Ireland seriously. According to Donnelly (2011), the NCCA, whose current priority is on the revision of the junior and leaving certificates, "would see more continuous assessment of students by their own teachers, to take the focus off the single terminal exam" (section Criticism). Continuous assessment is certainly a big step forward, which does not exclude the fact that assessing learners' knowledge, as opposed to their performance, will portray their capabilities more accurately. Distinguishing between errors and mistakes is a significant distinction that teachers tend to overlook in favour of merely relying on the evaluation of their students' performance. As Corder (1981) claimed, competence-dependent errors (not incorrect forms) may tell teachers how far learners have progressed. Additionally, the level of assistance required to edit the errors (when known) provides a valuable indication on when the learners should be able to master the linguistic feature. As noted by Aljaafreh and Lantolf (1994), "the learner who is able to respond to such help must be considered to be at a more advanced developmental level than the one who fails to do so, because the learner who responds to help can be expected to show a more rapid rate of actual development" (p.468). This leads to the question of how to make learners access the assistance provided to them in a more active way. The discussion undertaken in the limitation section sheds light on a few points that could enhance the learners' participation. In particular, it considers the content of the meta-linguistic feedback and reflects on the computer-based application from a human-computer interactive perspective.

8.1.2. Question 2

> *Are interlanguage competence and performance variable across students, time, and text types?*

While a large body of research has investigated the probable causes of variability, mostly in learners' interlanguage performance, relatively little is known about variability in interlanguage competence, that is, what learners *know* about the grammar of the language they are learning, as opposed to what they *do* with it. Analysing variability with regard to learner knowledge is not without challenges as one "cannot be sure where 'use' ends and 'acquisition' begins" (Firth and Wagner 1998 p.91). The fact that inter-learner and intra-learner performance are variable is well documented in interlanguage literature, and is not questioned here with the current findings. However, as mentioned in the introductory chapter, the extent to which inter-learner and intra-learner competence are variable is not clearly addressed, as most studies investigate learner knowledge without distinguishing between performance-related mistakes and competence-dependent errors.

Pertaining to the question *Are interlanguage competence and performance variable across students?*, the synchronic inter-learner analysis (Chapter 6) not only confirmed the variable aspect of the interlanguage performance, but also exhibited unsystematic variations in interlanguage competence between students. It has been shown (Section 6.2 on page 165) that some students with comparable results with regard to language accuracy in performance had indeed quite different levels of actual development. The snapshot analysis demonstrated varying degrees of ability to perform knowledge. While it was expected that the removal of mistakes from the learners' performance would have smoothed the difference between learners in performance – as they all (a) performed at a similar level of language complexity and (b) had an equivalent percentage of incorrect forms per hundred words –, the process showed even more differences between learners' levels of language accuracy in actual development. The justification for this was that not all learners were able to show the extent of their knowledge. Therefore, the ability to perform what is known is a factor to be taken into consideration when investigating interlanguage variability.

To the question *Are interlanguage competence and performance variable across time?,* the diachronic analysis established that learners increased their level of language sophistication and complexity over time. This had an effect on their interlanguage development as their results in language accuracy showed ups and downs in terms of percentage. As a direct consequence of higher degrees of sophistication and complexity in language production, new errors that never occurred before were encountered at each new point of observation in time. It was found that errors could be systematic, but also random. This finding is in direct opposition of Corder's (1967) claim, which stipulates that an error is systematic and a mistake is random. One of the most significant key features of dynamic assessment is that it enables an insight into learners' potential development. In other words, it allows a focus on not only learners' current dependent performance, but also their next probable independent achievements. In line with

Vygotsky's (1978) quote, it was demonstrated that the learner's development was not a mere "gradual accumulation of separate changes", but rather "a complex dialectical process characterised by periodicity" (p.73). As mentioned by Robbins (2003), the learner's actual and potential development over time could not be represented with gradual straight lines. He further points out that if one would graphically represent the nature of development, the depiction would be made with the help of "wave-like curves" (Robbins 2003 p.26). While dynamic assessment was used to unveil the learners' level of actual development, no one could affirm that it contributed to their development. Although researchers, such as Poehner (2005), did establish the impact of interactions between teacher and learner on the latter's development (p.319), it may be the case that the students benefited from other sources of support.

It was also demonstrated that text types, either in the form of graded or non-graded documents, were found to have no impact on the learner's knowledge, which answers the following research question *Are interlanguage competence and performance variable across text types*?. However, this observation, derived from the data of one single participant, cannot be used to draw any decisive conclusions. More research would be necessary to determine whether text types have no definite influence on learners' interlanguage competence. Meanwhile, it might be advanced that when students do not have the skills to perform specific linguistic features in one text type, they do not have them in all circumstances. This is in line with Poehner's (2005) proposition: "if learners genuinely develop, they should be able to maintain their improved performance when the task changes" (p.324). In other words, if learners have internalised the lexical and grammatical features of their target language, they should be able to exhibit their linguistic knowledge independently of the task in which they are involved in. Further research would be necessary to confirm this hypothesis as variability across task types was not investigated within this research.

In summary, interlanguage competence may vary in an unpredictable way across students and over time, but does not across text types. Lantolf and Aljaafreh's (1995) observation that "development and performance in mental systems is not a smooth linear process, but simultaneously entails forward and *regression*, or what some L2 researchers refer to as *backsliding*" (p.619, emphasis in original) is in every respect applicable to the case of development and competence. These findings have clear implications for language teaching and instructional design. Given the fact that the error-mistake distinction was operated through the observation of the learners' zone of proximal development, this implies adopting dynamic assessment in classroom settings. Very recently, Lantolf and Poehner (2011) have stated that

> [t]o date, L2 DA research has not focused on implementation of the procedure during regular classroom instruction but has instead occurred in one-to-one sessions outside the classroom and has been implemented by a teacher/researcher with expertise in applied linguistics (p.13).

Providing feedback to incorrect forms depending on their placement within the learners' zone of proximal development implies reassessing the methods generally in use to evaluate students. As noted by McGarrell (2011), "the students are frequently unclear on how to use the comments to improve their writings" (p.140).

8.1.3. Question 3

> *Does the modeling of the learners' zone of proximal development provide further insight into their interlanguage development?*

As mentioned in Chapter 2, a learner model is a set of data that contains information about a student's strengths and weaknesses (Heift 2002 p.298). Learner models are usually integrated into intelligent tutoring systems. Researchers generally advocate systems based on educational principles. Heift (2010b), for instance, describes the design of the E-Tutor (discussed in Section 2.3.2 on page 38) as “strongly motivated by pedagogical considerations” (p.445). The limitation section below shows that the computer-based application used in this study was primarily designed for research purposes. For example, the comprehensive error annotation adopted in the methodology would not be suitable for students as they could easily be overwhelmed by the amount of corrections to be made (Van der Linden 1993). Ross, Morrison and Lowther (2010) claim that educational technology reflects a “broad variety of modalities, tools, and strategies for learning” and that its efficacy “depends on how well it helps teachers and students achieve the desired instructional goals” (p.19).

Through the means of the web-based dynamic assessment application developed for this thesis, all participants were prompted to correct their incorrect forms at all levels. As Heift (2010c) observed, prompting offers “interesting insights into the usage and effectiveness of learner feedback” (p.200), and so did the monitoring system implemented to check learners’ behaviours with regard to feedback access. The main comportment observed was that students mostly skipped the assistance. Heift (2002) identifies three distinct learner personas[63] in the error correction process: browsers, peekers, and adamants:

> *browsers* frequently browsed through the exercises without providing any input or making an attempt to respond to the task. [...] In contrast, *peekers* generally attempted to correct their input by making frequent use of system help options; however, they also peeked at the correct answer(s) provided by the system more often than they actually attempted to correct their errors. [...] Finally, *adamants* never requested a correct answer from the system but instead worked through the error correction process adamantly; they also

63. Learner personas may be defined as “archetypal users of a learning tool that represent the needs of larger groups of users in terms of their goals and personal characteristics” (Heift 2007 p.4).

made far less use of system help options (Heift 2007 p.8, emphasis in original).

The fact that a great majority of learners did not request assistance (adamants) was not considered as a limitation in itself, as learners had to correct themselves without assistance for an incorrect form to be considered as a mistake. However, acknowledging whether learners accessed the feedback was a primordial piece of information. Without knowing this, mistakes could not be differentiated from errors with certainty. Given the fact that a distinction between editing with and without assistance was made, an incorrect form corrected without assistance was characterised as an internalised feature; an error edited with assistance was labelled as an error in the process of internalisation, whereas an error not edited was considered as a linguistic feature without identified potential development. As discussed in Section 8.8.1 on page 235, observing the learners' actual development was conditioned on the fact that learners entered alternatives when they knew ones, whereas observing the learners' potential development was constrained by the learners' behaviours with regard to feedback access. As a result, information on learners' actual development was easily retrievable, whereas modeling their potential development was more challenging, as learners, in general, did not access their feedback.

In relation to the question *Does the modeling of the learners' zone of proximal development provide further insight into their interlanguage development?*, the diachronic intra-learner analysis (Sections 7.2.4 on page 223 and 7.2.5 on page 227) established that modeling the learners' current performances (including all incorrect forms) along with their levels of actual and potential development (what they can correct with and without assistance) gives valuable information about their actual and proximal strengths and weaknesses. Such information, however, did not assist much in predicting the learners' knowledge to come, i.e., their following levels of actual development. Results demonstrated that the learners knowledge recorded at a time T_{t+1} was not a mere continuation of the learners' levels of actual and potential development as observed at a time T_t. One specific error, however, has been found to have a linear progression within the ZPD going from a low to high position. This suggests that the time frame of this thesis was not sufficiently long enough to investigate variations within the learners' zone of proximal development, as no other examples could be outlined. Conversely, no examples of specific linguistic misconceptions moving unpredictably within the ZPD were hitherto observed in the learner corpus. Researchers, such as Long (2003), maintain that "true longitudinal studies are needed" in order to assess learners' progress and attainment, as well as the evaluation of their issues in achievement (p.497). A *true* longitudinal study could be illustrated with the work of Lardiere (2007), in which she recorded and documented her participant *Patty*'s conversation over a period of nearly nine years (pvii). However, interpreting longitudinal analysis in terms of duration is not always an easy task. While it is frequent to find longitudinal studies over a period of twelve months in research publications, Mackey and Gass (2005) use a study that is conducted over a period of two weeks as an example of a longitudinal investigation (p.112). The time frame of this doctoral research was two academic semesters at the most. This was not sufficient to investigate the errors' movements within the learners' zone of proximal development. Further research is needed to confirm whether their progression would be systematically linear as found within this thesis with one occurrence of incorrect forms.

8.2. Limitations of the thesis

As outlined above, the overall findings demonstrated that the idea of variable competence was justified as interlanguage competence displays variances that were not explainable and predictable. The strength of this study is that the level of actual development could be represented for each incorrect form encountered in L2 French learners' unrestricted texts, and this through the means of a computer-based application. As a result, the error-mistake distinction was indeed realisable in practice, but the process has nevertheless some limitations that are worth pointing out. First and foremost, the student sample with fourteen participants[64] was relatively small. While all fourteen students were considered for the cross-sectional analysis, only three of them were eligible for the longitudinal investigation. To provide a better insight into learner language development, results would have been strengthened with a larger collection of texts written by additional participants over an even longer period of time. While the fourteen participants expressed their interest in being a part of the study (during informal meetings), they also claimed that the reason for which they did not renew the experience during the second semester (for most of them) was because I was not their tutor anymore.

The fact that learners did not use the assistance at any time limited the access to their potential development. While learners provided an alternative for most of the incorrect forms for which they read the feedback, the analysis of the corpus showed that a probable explanation for not suggesting any replacement after accessing assistance could be due to a misunderstanding of the message read. Feedback was on some occasions probably too complicated to decipher in terms of linguistic jargon. Given that learners seemed to be familiar with the linguistic terms when addressing them orally in class, and given that most of the participants had an introductory course in linguistics, the study placed reliance on the fact that they would understand the terminology used. Yet, learners' ability to interpret the message might have been over-estimated, as it seems some of them may have struggled with terms describing specific aspects of the language, such as *auxiliary*, *direct object*, *indirect object*, or *pronoun*. In fact, students who come through the Irish educational system tend to have a deficit in terms of grammatical knowledge. A test prior to the start of the study would have detected participants with difficulties in terms of linguistic vocabulary. As found by Lee (1997), teachers often used "a wider

64. There were in fact 19 participants, but only 14 of them submitted at least one text for correction.

range of metalinguistic terms than students could understand" (p.471). Consequently, to remedy this lack of knowledge, a session dedicated to the explanation of the terminology in use when providing feedback should suitably prepare the students for the comprehension of the language descriptors. Additionally, these specific terms could be explained through the means of a glossary made available to students at all times. Moreover, future research should adjust more precisely the level of meta-linguistic annotation to each individual so as to ensure a proper understanding of the feedback itself.

From another perspective, and more particularly from a human-computer interactive approach, Bahr and Ford (2011) very recently have established that pop-up windows might be "annoying and frustrating" for participants (p.6). It may be thus relevant to investigate other means of dispensing meta-linguistic feedback to stimulate a more sustained engagement from learners, in other words, to avoid boredom and foster motivation. According to Dörnyei (2003),

> research on L2 motivation has considerable educational potential, particularly in two areas: (a) the systematic development of motivational strategies that can be applied to generate and maintain motivation in learners, and (b) the formulation of self-motivating strategies that enable L2 learners themselves to take personal control of the affective conditions and experiences that shape their subjective involvement in learning (Dörnyei 2003 p.23).

In addition, this study adopted a comprehensive error annotation based on the premises that marking all students' incorrect forms would provide a more complete picture of their knowledge than only marking a few specific error types. However, as observed by Ferris (2002), with a comprehensive annotation not only the corrector may end up exhausted in marking all incorrect forms, but the learners may also become overwhelmed (p.50).

Consequently, the learners' involvement may have been weakened in certain occasions. From an educational perspective, as opposed to a research point of view, feedback is more effective "when it focuses on patterns of error, allowing teachers and students to attend to, say, two or three major error types at a time, rather than dozens of disparate errors" (Ferris 2002 p.50). Lee (2003), also of the same opinion, claims that "selective marking" is more suitable for learners than "comprehensive marking" (p.228).

One further limitation to pinpoint was the assumption that learners would use the tools in the way they were designed to. More particularly, it was thought that they were going to access each feedback for each incorrect form at each level of assistance, so that the level of potential development could be represented with as much detail as possible. Further research must take account of the fact that learners do not behave as expected. In their study of students' performance in computer-assisted language learning (CALL) activities, Chapelle and Mizuno (1989) demonstrate that "students are often doing something different from what instructors believe they are doing" (p.42). More recently, Heift (2010b) states that "even the best team of CALL software designers cannot always anticipate the ways in which learners will use a CALL system" (p.445). The analysis of learners' behaviours with regard to feedback access (after the implementation of the monitoring system) demonstrated that students generally corrected themselves without assistance. From the learner knowledge point of view, the fact that the participants did not read all feedback did not compromise the error-mistake distinction, as what was needed was the alternatives they proposed without seeking help. However, the representation of the learners' potential development, and a fortiori the representation of their zone of proximal development, was limited as students, in most cases, did not access the assistance to propose alternatives. Indeed, the developmental level of learners can never be fully comprehended if one does not consider what they can do *with* the assistance of somebody else (Vygotsky 1978). This suggests that the research methodology was adequate to differentiate between errors and mistakes. However, it may be argued that the methodology in use was less suited to observe and analyse the learners' potential development. Future research should therefore consider a revision of the methods and tools to investigate this particular area. Students for instance, could be asked to systematically read the assistance in order for them to be allowed to move on in the program and to access the following level. In addition, learners could be invited to correct themselves twice at the first level of assistance in the regulatory scale in order to avoid lucky guesses.

As mentioned in Section 5.3 on page 117, students were allowed to spend as much time as needed to correct themselves during the self-editing exercises. This could be considered as a limitation as the time devoted at each level was different for each student[65]. For example, some of them could have looked for additional information, on the Internet or elsewhere to help them edit their texts, thus affecting the amount of assistance as well as the error-mistake distinction. Although the time parameter was not taken into account for the analysis of the data, the duration recorded for each correction phase would rather indicate that students did not try to find any additional source of information. Moreover, the fact that students were not monitored when composing their texts before sending them for correction might have had an effect on the results. Indeed, some students may have devoted more time than others to refine their independent performance, or others may have obtained different levels of assistance when composing their texts, thus affecting their amount of mistakes and errors. This shortcoming could be further addressed by giving all students the same treatment, that is, by imposing a time limit.

The final limitation addressed within this section relates to the tagging process. The primary aim of improving the part-of-speech tagging accuracy was to ensure that the

---

65. The time required by students to correct themselves at each level of the self-editing exercise was monitored and recorded.

learner interlanguage corpus was annotated with reliable part-of-speech tags so that learner knowledge can be measured, which was achieved. The secondary objective was to propose a means to improve the performance of the tagger. The method discussed in Section 5.1 on page 102 is not without drawbacks and has to be considered with caution. The main limitations refer to the use of (a) the word bank, which considerably slows down the system and may create time-out errors, and (b) the additional hand-written rules that cannot be generalised to other languages since the rules were specifically written for the French language. In addition, the method was tested on the training data, as opposed to be tested on an unseen corpus, which could explain the excellent results obtained with regard to part-of-speech tagging accuracy. Contrary to these drawbacks, the cross-reference approach, which compares the part-of-speech tags with the error type of the word, seems to be a worthwhile direction for further exploration, as it would be a method that could be generalised to all languages.

Considering the above limitations in further research and development will certainly refine the representation of the learners' performance, actual and potential development, as well as their zone of proximal development, thus giving deeper insights into L2 learner language development and acquisition.

8.3. Suggestions for further research

From a research perspective, and given the above findings regarding the creation and representation of the learners' zone of proximal development, comparing each ZPD in any error type not only provided significant information about each learner's actual and potential development (when observable), but also helped situate them in relation to the group. Hence, future research should look beyond the representation of one individual's ZPD and further extend the research on students as a group, as Vygotsky himself suggested it (1998 p.204). The aim behind grouping learners together is to establish a collective zone of proximal development, which broadens the search of weaknesses and strengths beyond individuals, thus enlarging the scope of the analysis to include the entire class. In the domain of L2 language learning, very few researchers have investigated dynamic assessment from the group perspective. For example, Poehner (2009) outlines the possibility of targeting the developmental level of groups as opposed to individuals to foster the interest in dynamic assessment in general education. As he explained, targeting the development of cognitive structures of individuals may be “an unrealistic model for classroom teachers” due to large numbers of learners (Poehner 2009 p.473). However, as he emphasised, groups should not be haphazardly constituted; learners in a group should be engaged in tasks that are challenging to them. In other words, they should not be able to complete the task without assistance of their peers (Poehner 2009 p.477). Thus, to know whether a task will be challenging for one specific student, each individual’s ZPD should be determined *before* the construction of the group to then be able to analyse the collective ZPD. Consequently, if the main objective is to reduce the teacher’s workload, group dynamic assessment is not the solution. Yet the approach of group dynamic assessment offers interesting perspectives with regard to future research. Once each learner’s ZPD is identified, the group could be formed by joining students with different ZPDs so that interactions within the group can create a new zone of proximal development that can be investigated. As a result, teachers should be able to direct specific questions to a particular student with the intention to engage the others in the reflection in a collaborative way. Collaboration may be defined as “the process of building and maintaining a shared understanding of a problem” (Laister and Koubek 2001 p.3). Lund (2008) adds that “[c]ollaboration involves participants jointly working together on a task and not by dividing it into individual subtasks that are later assembled” (p.36). He further suggests that

> [i]n a sociocultural view of SLA the learner is not just an individual who on encountering a foreign language processes and assimilates vocabulary and syntax. Rather, learners are seen to participate in different [...] communities

> where they draw on social resources (other participants, institutional affordances), material resources (PCs, networks, applications), and semiotic resources (signs, genres) (Lund 2008 p.39).

Research in collective ZPD should therefore consider the cultural, social, institutional, and historical contexts, which may imply to consider activity theory framework as an opportunity to broaden our knowledge in this domain. However, it is important to note that, according to Cole and Engeström (2007), the *context* in such a framework is the unit of analysis (p.485). They further clarify that "the activity is the context" (Cole and Engeström 2007 p.485). According to Blin (2004),

> [a]ctivity theory is not a method in itself, nor a theory in the usual sense of the term as it does not systematically allow us to predict phenomena, processes or outcomes. It does however provide a terminology and an analytical framework that help us to make sense of human and social practice in specific contexts, and more specifically where technological mediation plays a significant role (Blin 2004 p.393).

Along these lines, future research may also envisage a refinement of the tools in use to collect and process the data. The present study materialised the zone of proximal development of 14 learners of French; a revised edition of the tools should facilitate the representation and observation of their ZPDs. The method to differentiate between errors and mistakes is based on the amount of assistance learners require to correct themselves; it is language independent. While the method was used to determine the knowledge of students of French at university level, it could be equally applied to any language or education level. It might also be suitable for investigating L1 learners' knowledge. The system to represent this knowledge is however language dependent as it requires the use of a part-of-speech tagger when counting correct and incorrect forms. Not only must the tagger be adapted to the target language, it also necessitates various modifications as its tagging accuracy often correlates with the amount of incorrect forms encountered in written texts. As a result, the inclusion of additional rules is generally a necessity to minimise the decrease in accuracy. As mentioned earlier in the limitation section, further research is needed in this area given the fact that post-editing was a necessary step to increase the tagger accuracy. In particular, a further development of the cross-reference method, which compares part-of-speech tags with error annotations, should decrease the language dependency aspect, thus allowing for a better portability across languages. A supplemental area for future research should be directed towards the self-editing task, where it was mentioned that the mediation offered may have been unadapted to the learners' need or comprehension on some occasions. As Poehner (2005) noted, the advantage of dynamic assessment "lies in the timeliness of the mediation" (p.148), which is not without challenges when the assistance is provided through the means of a computer-based dynamic assessment application, also called computerised dynamic assessment (CDA) (Poehner 2008). As outlined in Chapter 3, the design of a CDA application is generally directed towards the reporting of learners' performance and the automatic provision of assistance going from implicit to explicit. A future and extremely ambitious direction may include research at the level of negotiated interactions, as opposed to interventions, between a computer and a learner.

8.4. Pedagogical implications

As discussed above, distinguishing between an error and a mistake using Vygotsky's concept of the zone of proximal development opens the door to a large body of research. It is an equally significant topic for methods and practices of teaching given the fact that the zone of proximal development "has been widely applied in both the general education and applied linguistics research literatures to illuminate processes of learner development" (Poehner 2011 pp.245-246). The findings of this doctoral research yield implications in terms of pedagogy.

For example, language teachers, when correcting learners' texts, should keep in mind that the learners' ability to build sentences and to express themselves in an accurate way did not develop simultaneously. The idea that both complexity and accuracy in written texts evolve concurrently is often an "underlying assumption" (Benevento and Storch 2011 p.98). While syntactical complexity has been found to increase gradually over time, language accuracy showed ups and downs. These results are consistent with other researchers, such as Benevento and Storch (2011), who find significant improvements in language complexity but not in language accuracy in texts written by learners' of French. Skehan (2009) proposes that there exists a "tension" between complexity and accuracy with respect to attention and information processing, and that this tension could be depicted as a "trade-off" (p.511). This suggests that learners, when performing a task, simplify certain aspects of it over others, and might decide (either consciously or not) to focus on accuracy rather than complexity, or vice versa. Therefore, teachers should alternate the learners' attention to either accuracy or complexity when assessing written language. This should enable teachers to better delineate what can, in effect, be expected from language learners.

Additionally, the findings of this research demonstrated that learners did not regard the assistance much. In particular, it was shown that most learners could correct themselves independently, which implies that they did not need help. However, other reasons for not accessing the assistance could be that learners were overwhelmed with the amount of corrections to be done or that they did not understand the feedback itself. Regarding feedback comprehension, the findings suggest that some learners might have experienced difficulties in understanding the terminology in use when accessing metalinguistic annotations. Teachers could enhance learners' linguistic awareness by testing their knowledge and by instructing the terms for which they have a blurred definition. Alternatively, instructors could provide students with book references or links on the Internet where meta-linguistic glossaries or dictionaries could be found. The pedagogical implications of acknowledging the fact that feedback is not systematically read nor understood may imply a revision of the way instructors correct texts and provide assistance to learners. It may be advanced that there is no need for teachers to provide feedback other than underlining the incorrect forms, if no-follow up is undertaken to ensure the reading and understanding of the metalinguistic annotations. But in any case, students should be given the opportunity to revise their incorrect forms with minimum help first – some of the incorrect forms might be very close to independent performance and even be considered as mistakes –, and then with more and more assistance until they propose correct suggestions to their errors. In practice, if teachers have no access to computer-based systems to correct their students' written language, they could first manually underline each incorrect form found on the learners' texts without providing any further information on the error type. Increasingly detailed explanations should be given to students only if they are unable to correct themselves with the previous level of assistance. As found by Nassaji and Swain (2000), help given in accord to learners' ZPD is more efficient for L2 learning than help provided without considering it.

Providing assistance to learners in line with their ZPD will enable teachers to distinguish between errors and mistakes. If learners require no or little assistance to correct themselves, the incorrect forms are said to be high in the ZPD and categorised as mistakes, otherwise, they are considered as errors. Distinguishing between both errors and mistakes have direct consequences for assessment practices. It not only implies that teachers should adopt dynamic assessment in the language classroom, it also presupposes their intention to assess learners' knowledge as opposed to performance. Poehner and Lantolf (2005) claim that the main difference between traditional psychometric approaches and dynamic assessment lies in "whether or not the administration of the assessment should have the expressed goal of modifying learner performance during the assessment itself" (p.235). While a teacher's intervention is considered as an issue in terms of validity in a summative assessment, the same intervention in dynamic assessment would help any instructor evaluate the extent of their students' knowledge as determined by independent and dependent performance. A major implication in distinguishing between lowly and highly placed incorrect forms in learners' ZPD would lead to the prevention of underrating or undervaluing students' abilities. In other words, differentiating between errors and mistakes would not only motivate students who always perform poorly, but also help instructors focus on learners' genuine difficulties rather than spending time and energy with linguistic explanations on something that is already internalised. As Wehlburg (2011) commented, "why spend time reviewing information that students already know?" (p.3). Teachers and students may thus easily avoid frustrations.

As a concluding remark, it may be advanced that being able to make a difference between errors and mistakes should ease any language teacher's workload, which is elegantly described by LaFontana (1996) as follows:

> The school day ends, and we [language] teachers trudge to our cars, laden with heavier book bags and larger piles of student papers than those of our colleagues in other disciplines. Most of us love the challenge of helping our students become better writers, but the task of providing meaningful, prompt, and frequent feedback to their efforts is enough to give us second thoughts about our career choice (LaFontana 1996 p.71).

References

Ableeva, R. 2008. Listening comprehension in foreign language instruction. (CALPER Professional Development Document 0810). University Park, PA: Center for Advanced Language Proficiency Education and Research Publications [Online]. Available from: <http://calper.la.psu.edu/publication.php?page=pdd6> [Accessed 25 September 2009].

Abney, S. 1997. Part-of-Speech Tagging and Partial Parsing IN: Young, S. and Bloothooft, G. (eds.) Corpus-Based Methods in Language and Speech Processing. Dordrecht, The Netherlands: Kluwer Academic Publishers.

Adair-Hauck, B., Willingham-McLain, L. and Earnest Youngs, B. 2000. Evaluating the Integration of Technology and Second Language Learning. CALICO Journal. 17 (2), pp269-306.

Adamson, H.D. 2009. Interlanguage Variation in Theoretical and Pedagogical Perspective. New York: Routledge.

Aljaafreh, A. and Lantolf, J.P. 1994. Negative feedback as regulation and second language learning in the zone of Proximal development. Modern Language Journal [Online]. 78 (4), pp465-483. Available from: <http://dx.doi.org/10.2307/328585> [Accessed 02 July 2009].

Amaral, L.A. and Meurers, D. 2006. Where does ICALL Fit into Foreign Language Teaching? Presentation at the 23rd Annual Conference of the Computer Assisted Language Instruction Consortium (CALICO), May 19, 2006. University of Hawaii. [Online]. Available from: <http://purl.org/net/icall/handouts/calico06-amaral-meurers.pdf> [Accessed 15 February 2010].

Amaral, L.A. and Meurers, D. 2007. Conceptualizing Student Models for ICALL IN: Conati, C., McCoy, K. and Paliouras, G. (eds.) User Modeling 2007: 11th international conference, UM 2007, Corfu, Greece, Proceedings. Berlin: Springer Verlag.

Amaral, L.A. and Meurers, D. 2008. From recording linguistic competence to supporting inferences about language acquisition in context. Computer Assisted Language Learning [Online]. 21 (4), pp323-338. Available from: <http://dx.doi.org/10.1080/09588220802343454> [Accessed 24 November 2009].

Amaral, L.A. and Meurers, D. 2009. Little Things With Big Effects: On the Identification and Interpretation of Tokens for Error Diagnosis in ICALL. CALICO Journal [Online]. 26 (3), pp580-591. Available from: <http://www.sfs.uni-tuebingen.de/~dm/papers/amaral-meurers-09.html> [Accessed 15 February 2010].

Anderson, J.R. 1983. The Architecture of Cognition. Cambridge, MA: Harvard University Press.

Anderson, J.R. 1996. Knowledge representation IN: Anderson, J.R. (ed.) Rules of the mind. Hillsdale: Lawrence Erlbaum Associates.

Anderson, J.R. 2000. Learning and memory : an integrated approach. 2nd ed. New York: John Wiley & Sons.

Antón, M. 2009. Dynamic assessment of advanced second language learners. Foreign Language Annals [Online]. 42 (3), pp576-598. Available from: <http://dx.doi.org/10.1111/j.1944-9720.2009.01030.x> [Accessed 15 September 2009].

Bachman, L.F. 1990. Fundamental considerations in language testing. Oxford: Oxford University Press.

Bachman, L.F. and Palmer, A.S. 1996. Language testing in practice. Oxford: Oxford University Press.

Bahr, G.S. and Ford, R.A. 2011. How and why pop-ups don’t work: pop-up prompted eye movements, user affect and decision making. Computers in Human Behavior [Online]. 27 (2), pp776-783. Available from: <http://dx.doi.org/10.1016/j.chb.2010.10.030> [Accessed 5 December 2010].

Bailin, A. 1990. Skills-in-Context and Student Modeling. CALICO Journal [Online]. 8 (1), pp7-22. Available from: <https://calico.org/html/article_474.pdf> [Accessed 29 March 2010].

Baker, R.S.J., Corbett, A.T. and Aleven, V. 2008. More Accurate Student Modeling through Contextual Estimation of Slip and Guess Probabilities in Bayesian Knowledge Tracing IN: Woolf, B.P., Aïmeur, E., Nkambou, R. and Lajoie, S. (eds.) Intelligent Tutoring Systems: 9th International Conference, ITS 2008 Montreal, Canada, June 2008, Proceedings. Berlin: Springer [Online]. Available from: <http://dx.doi.org/10.1007/978-3-540-69132-7> [Accessed 26 February 2010].

Beattie, G. 1983. Talk: an analysis of speech and non-verbal behaviour in conversation. Michigan: Open University Press.

Beck, J.E. 2007. Difficulties in inferring student knowledge from observations (and why you should care), 2007, July 9. IN: Proceedings of the AIED2007 Workshop on Educational Data Mining. Marina del Rey, CA [Online]. Available from: <http://www-2.cs.cmu.edu/~listen/pdfs/AIED2007_EDM_problems_with_knowledge_estimation.pdf> [Accessed 25 November 2009].

Beck, J.E. and Sison, J. 2006. Using knowledge tracing in a noisy environment to measure student reading proficiencies. International Journal of Artificial Intelligence in Education [Online]. 16 (2), pp129-143. Available from: <http://www.cs.cmu.edu/~listen/pubs.html> [Accessed 12 September 2008].

Beck, J.E., Stern, M. and Haugsjaa, E. 1996. Applications of AI in Education. Crossroads, the ACM Student Magazine [Online]. 3 (1), pp-. Available from: <http://www.acm.org/crossroads/xrds3-1/aied.html> [Accessed 20 February 2010].

Bell, A. 1984. Language style as audience design. Language in Society [Online]. 13 (2), pp145-204. Available from: <http://dx.doi.org/10.1017/S004740450001037X> [Accessed 27 November 2009].

Benevento, C. and Storch, N. 2011. Investigating writing development in secondary school learners of French. Assessing Writing [Online]. 16 (2), pp97-110. Available from: <http://dx.doi.org/10.1016/j.asw.2011.02.001> [Accessed 23 May 2011].

Bernsen, N.O. and Dybkjær, L. 2008. Modelling spoken multimodal instructional systems IN: Luppicini, R. (ed.) Handbook of conversation design for instructional applications. New York: Information Science Reference.

Bialystok, E. 1994. Representation and ways of knowing: three issues in second language acquisition IN: Ellis, N.C. (ed.) Implicit and explicit learning of languages. London: Academic Press.

Black, E. et al. 1992. Decision Tree Models Applied to the Labeling of Text with Parts-of-Speech. IN: Proceedings of the Workshop on Speech and Natural Language, Human Language Technology Conference, February 23 - 26, 1992. Morristown, NJ: Association for Computational Linguistics [Online]. Available from: <http://dx.doi.org/10.3115/1075527.1075554> [Accessed 27 March 2010].

Blin, F. 2004. CALL and the development of learner autonomy: Towards an activity-theoretical perspective. ReCALL [Online]. 16 (2), pp377-395. Available from: <http://dx.doi.org/10.1017/S0958344004000928> [Accessed 22 September 2009].

Bowker, L. and Pearson, J. 2002. Working with specialised language: a practical guide to use corpora. London: Routledge.

Brill, E. 1995. Transformation-based error-driven learning and natural language processing: a case study in part-of-speech tagging. Computational Linguistics. 21 (4), pp543-565.

Brown, A. and Ferrara, R.A. 1985. Diagnozing zones of proximal development IN: Wertsch, J.V. (ed.) Culture, Communication and Cognition. Vygotskian perspectives. Cambridge: Cambridge University Press.

Brown, R. 1973. A first language: the early stages. Cambridge, MA: Harward University Press.

Brusilovsky, P. and Millán, E. 2007. User Models for Adaptive Hypermedia and Adaptive Educational Systems IN: Bunt, A., Carenini, G. and Conati, C. (eds.) The Adaptive Web: Methods and Strategies of

Web Personalization. Berlin: Springer [Online]. Available from: <http://dx.doi.org/10.1007/978-3-540-72079-9_1> [Accessed 22 February 2010].

Budoff, M. 1987. Th e validity of learning potential assessment IN: Lidz, C.S. (ed.) Dynamic assessment: An interactional approach to evaluating learning potential. New York: Guilford.

Budoff, M. and Friedman, M. 1964. "Learning potential" as an assessment approach to the adolescent mentally retarded. Journal of Consulting Psychology. 28 (5), pp434-439.

Bull, S., Pain, H. and Brna, P. 1995. Mr. Collins: A collaboratively constructed, inspectable student model for intelligent computer assisted language learning. Instructional Science [Online]. 23 (3), pp65-87. Available from: <http://dx.doi.org/10.1007/BF00890446> [Accessed 28 March 2010].

Bull, S., Quigley, S. and Mabbott, A. 2006. Computer-based formative assessment to promote reflection and learner autonomy. Engineering education [Online]. 1 (1), pp8-18. Available from: <http://www.engsc.ac.uk/journal/index.php/ee/article/view/37/53> [Accessed 6 February 2010].

Burnard, L. 2005. Metadata for corpus work IN: Wynne, M. (ed.) Developing Linguistic Corpora: a Guide to Good Practice. Oxford: Oxford Books [Online]. Available from: <http://www.ahds.ac.uk/guides/linguistic-corpora/chapter3.htm> [Accessed 15 July 2010].

Bygate, M. 1999. Quality of language and purpose of task: patterns of learners' language on two oral communication tasks. Language Teaching Research [Online]. 3 (3), pp185-214. Available from: <http://dx.doi.org/10.1177/136216889900300302> [Accessed 27 November 2009].

Cameron, D. 2005. Language, gender, and sexuality: current issues and new directions. Applied Linguistics [Online]. 26 (4), pp482-502. Available from: <http://dx.doi.org/10.1093/applin/ami027> [Accessed 26 February 2010].

Carlson, J.S. and Wiedl, K.H. 1979. Toward a differential testing approach: Testing-the-limits employing the Raven matrices. Intelligence. 3 (4), pp323-344.

Carroll, J. 1964. Language and thoughts. Englewood Cliffs: Prentice-Hall.

Chaiklin, S. 2003. The Zone of Proximal Development in Vygotsky's Anankysis of learning and Instruction IN: Kozulin, A., Gindis, B., Ageyev, V.S. and Miller, S.M. (eds.) Vygotsky's Educational Theory in Culturel Context. Cambridge: Cambridge University Press [Online]. Available from: <http://lchc.ucsd.edu/MCA/ZOPEDS/Chailklin.pdf> [Accessed 29 July 2009].

Chandler, J. 2003. The efficacy of various kinds of error feedback for improvement in the accuracy and fluency of L2 student writing. Journal of Second Language Writing [Online]. 12 (3), pp267-296. Available from: <http://dx.doi.org/10.1016/S1060-3743(03)00038-9> [Accessed 02 February 2009].

Chandler, J. 2004. A response to Truscott. Journal of Second Language Writing [Online]. 13 (4), pp345-348. Available from: <http://dx.doi.org/10.1016/j.jslw.2004.05.001> [Accessed 02 July 2009].

Chapelle, C.A. 2009. The Relationship Between Second Language Acquisition Theory and Computer-Assisted Language Learning. The Modern Language Journal [Online]. 93 (s1), pp741-753. Available from: <http://dx.doi.org/10.1111/j.1540-4781.2009.00970.x> [Accessed 8 February 2010].

Chapelle, C.A. and Mizuno, S. 1989. Student's Strategies with Learner-Controlled CALL. CALICO Journal [Online]. 7 (2), pp25-47. Available from: <https://calico.org/html/article_458.pdf> [Accessed 20 July 2009].

Chomsky, N. 1965. Aspects of the theory of syntax. Cambridge: M.I.T Press.

Clarebout, G. and Elen, J. 2006. Tool use in computer-based learning environments: towards a research framework. Computers in Human Behavior [Online]. 22 (3), pp389-411. Available from: <http://dx.doi.org/10.1016/j.chb.2004.09.007> [Accessed 17 July 2009].

Codd, E.F. 1979. Extending the database relational model to capture more meaning. ACM Transactions on Database Systems [Online]. 4 (4), pp397-434. Available from: <http://dx.doi.org/10.1145/320107.320109> [Accessed 12 April 2010].

Codd, E.F. 1982. Relational database: a practical foundation for productivity. Communications of the ACM [Online]. 25 (2), pp109-117. Available from: <http://dx.doi.org/10.1145/358396.358400> [Accessed 12 April 2010].

Cohen, J. 1960. A Coefficient of Agreement for Nominal Scales. Educational and Psychological Measurement [Online]. 20 (1), pp37-46. Available from: <http://dx.doi.org/10.1177/001316446002000104> [Accessed 04 February 2009].

Cole, M. and Engeström, Y. 2007. Cultural-historical approaches to designing for development IN: Valsiner, J. and Rosa, A. (eds.) The Cambridge Handbook of Sociocultural Psychology. Cambridge: Cambridge University Press [Online]. Available from: <http://gse2.berkeley.edu/~uclinks/wp-content/uploads/2009/08/ch23-cole-and-engestrom.pdf> [Accessed 13 February 2011].

Conati, C., McCoy, K. and Paliouras, G. (eds.) 2007. User Modeling 2007: 11th international conference, UM 2007, Corfu, Greece, Proceedings Springer Verlag: Berlin.

Conati, C., Gertner, A. and VanLehn, K. 2002. Using bayesian networks to manage uncertainty in student modeling. User Modeling and User-Adapted Interaction. 12 (4), pp371-417.

Corbett, A.T. and Anderson, J.R. 1995. Knowledge tracing: modeling the acquisition of procedural knowledge. User Modeling and User-Adapted Interaction [Online]. 4 (4), pp253-278. Available from: <http://act-r.psy.cmu.edu/papers/893/CorbettAnderson1995.pdf> [Accessed 26 February 2010].

Corbett, A.T., McLaughlin, M. and Scarpinatto, K.C. 2000. Modeling student knowledge: cognitive tutors in high school and college. User Modeling and User-adapted Interaction. 10 (2-3), pp81-108.

Corder, S.P. 1967. The significance of learners errors. IRAL - International Review of Applied Linguistics in Language Teaching. 5 (4), pp161-170.

Corder, S.P. 1981. Error Analysis and Interlanguage. Oxford: Oxford University Press.

Corman, L. and Budoff, M. 1973. A Comparison of Group and Individual Training Procedures on the Raven Learning Potential Measure with Black and White Special Class Students. ERIC #: ED085969. Research Institute for Educational Problems [Online]. 3 (57), pp1-16. Available from: <http://www.eric.ed.gov/ERICWebPortal/contentdelivery/servlet/ERICServlet?accno=ED085969> [Accessed 17 Sep 2009].

Craggs, R. and Wood, M.M. 2005. Evaluating Discourse and Dialogue Coding Schemes. Computational Linguistics [Online]. 31 (3), pp289-296. Available from: <http://dx.doi.org/10.1162/089120105774321109> [Accessed 24 May 2010].

Cui, Y. and Bull, S. 2005. Context and learner modelling for the mobile foreign language learner. System [Online]. 33 (2), pp353-367. Available from: <http://dx.doi.org/10.1016/j.system.2004.12.008> [Accessed 9 April 2010].

Cutting, D. et al. 1992. A practical part-of-speech tagger. IN: Proceedings of the Third Conference on Applied Natural Language Processing, Trento, Italy, March 31 - April 03, 1992. Morristown, NJ: Association for Computational Linguistics [Online]. Available from: <http://dx.doi.org/10.3115/974499.974523> [Accessed 27 March 2010].

Daelemans, W. and Van der Bosch, A. 2005. Memory-based language processing. Cambridge: Cambridge University Press.

Dagneaux, E., Denness, S. and Granger, S. 1998. Computer-aided error analysis. System [Online]. 26 (2), pp163-174. Available from: <http://dx.doi.org/10.1016/S0346-251X(98)00001-3> [Accessed 12 April 2010].

Daniel, M.H. 1997. Intelligence testing: status and trends. American Psychologist [Online]. 52 (10), pp1038-1045. Available from: <http://dx.doi.org/10.1037/0003-066X.52.10.1038> [Accessed 14 July 2009].

Darus, S. and Subramaniam, K. 2009. Error Analysis of the Written English Essays of Secondary School Students in Malaysia: A Case Study. European Journal of Social Sciences [Online]. 8 (3), pp483-495. Available from: <http://www.eurojournals.com/ejss_8_3_12.pdf> [Accessed 17 January 2010].

De Beer, M. 2006. Dynamic testing: Practical solutions to some concerns. SA Journal of Industrial Psychology [Online]. 32 (4), pp8-14. Available from: <http://www.sajip.co.za/index.php/sajip/article/view/245/242> [Accessed 10 September 2009].

De Bot, K., Lowie, W. and Verspoor, M. 2007. A Dynamic Systems Theory approach to second language acquisition. Bilingualism: Language and Cognition [Online]. 10 (1), pp7-21. Available from: <http://dx.doi.org/10.1017/S1366728906002732> [Accessed 5 November 2009].

DeKeyser, R.M. 1998. Beyond focus on form: cognitive perspectives on learning and practicing second language grammar IN: Doughty, C. and Williams, J. (eds.) Focus on form in classroom second language acquisition. Cambridge: Cambridge University Press.

DeKeyser, R.M. 2003. Implicit and Explicit Learning IN: Doughty, C.J. and Long, M.H. (eds.) The Handbook of Second Language Acquisition. Malden, MA: Blackwell Publishing Ltd.

DeKeyser, R.M. 2009. Cognitive-psychological processes in second language learning IN: Long, M.H. and Doughty, C.J. (eds.) The Handbook of Language Teaching. West Sussex: Blackwell Publishing Ltd.

Di Eugenio, B. and Glass, M. 2004. The Kappa Statistic: A Second Look. Computational Linguistics [Online]. 30 (1), pp95-101. Available from: <http://dx.doi.org/10.1162/089120104773633402> [Accessed 24 May 2010].

Díaz-Negrillo, A. et al. 2010. Towards interlanguage POS annotation for effective learner corpora in SLA and FLT. Language Forum [Online]. 36 (1/2), pp1-15. Available from: <http://purl.org/dm/papers/diaz-negrillo-et-al-09.html> [Accessed 2 February 2011].

Dickerson, L.J. 1975. The Learner's Interlanguage as a System of Variable Rules. TESOL Quarterly. 9 (4), pp401-407.

Dickinson, M. and Herring, J. 2008. Developing Online ICALL Exercises for Russian. IN: The 3rd Workshop on Innovative Use of NLP for Building Educational Applications (ACL08-NLP-Education). Columbus, OH: ACL08-NLP-Education [Online]. Available from: <http://jones.ling.indiana.edu/~mdickinson/papers/dickinson-herring08.html> [Accessed 3 April 2010].

Donnelly, K. 2011. New junior cert to have maximum of eight subjects. July 09 2011. independent.ie [Online]. Available from: <http://www.independent.ie/national-news/new-junior-cert-to-have-maximum-of-eight-subjects-2817292.html> [Accessed 23 July 2011].

Dörnyei, Z. 2003. Attitudes, orientations, and motivations in language learning: advances in theory, research, and applications IN: Dörnyei, Z. (ed.) Attitudes, orientations, and motivations in language learning. Malden, MA: Blackwell Publishing.

Douglas, D. 2000. Assessing language for specific purposes. Cambridge: Cambridge University Press.

Douglas, D. and Hegelheimer, V. 2007. Assessing language using computer technology. Annual Review of Applied Linguistics [Online]. 27 (1), pp115-132. Available from: <http://dx.doi.org/10.1017/S0267190508070062> [Accessed 4 September 2009].

Douglass, K. 2009. Second-person pronoun use in French-language blogs: developing L2 sociopragmatic competence IN: Abraham, L.B. and Williams, L. (eds.) Electronic discourse in language learning and language teaching. Amsterdam: John Benjamins.

Dulay, H.C. and Burt, M.K. 1973. Should we teach children syntax? Language Learning [Online]. 23 (2), pp245-258. Available from: <http://dx.doi.org/10.1111/j.1467-1770.1973.tb00659.x> [Accessed 27 October 2009].

Ebadi, S. 2010. Exploring learners' microgenetic development in SCMC-based L2 dynamic assessment via Web 2.0. Unpublished PhD dissertation. Allameh Tabatabie University.

Elliott, J. 2003. Dynamic Assessment in Educational Settings: realising potential. Educational Review [Online]. pp15-32. Available from: <http://www.ingentaconnect.com/content/routledg/cedr/2003/00000055/00000001/art00002> [Accessed 17 September 2009].

Ellis, R. 1984. Classroom second language development: a study of classroom interaction and language acquisition. Pergamon: Oxford.

Ellis, R. 1985a. A variable competence model of second language acquisition. IRAL [Online]. 23 (1), pp47-59. Available from: <http://dx.doi.org/198503202653004> [Accessed 3 December 2010].

Ellis, R. 1985b. Sources of Variability in Interlanguage. Applied Linguistics [Online]. 6 (2), pp118-131. Available from: <http://dx.doi.org/10.1093/applin/6.2.118> [Accessed 22 January 2010].

Ellis, R. 1985c. Understanding second language acquisition. Oxford: Oxford University Press.

Ellis, R. 1989. Sources of intra-learner variability in language use and their relationship to second language acquisition IN: Gass, S.M., Madden, C., Preston, D. and Selinker, L. (eds.) Variation in second language acquisition Volume II: psycholinguistic issues. Philadelphia: Multilingual Matters.

Ellis, R. 1994a. A theory of instructed second language acquisition IN: Ellis, N.C. (ed.) Implicit and explicit learning of languages. San Diego: Academic Press.

Ellis, R. 1994b. The study of second language acquisition. Oxford: Oxford University Press.

Ellis, R. 1997. Second language acquisition. Oxford: Oxford University Press.

Ellis, R. 2003. Task-based language learning and teaching. Oxford: Oxford University Press.

Ellis, R. 2004. The Definition and Measurement of L2 Explicit Knowledge. Language Learning [Online]. 54 (2), pp227-275. Available from: <http://dx.doi.org/10.1111/j.1467-9922.2004.00255.x> [Accessed 19 February 2010].

Ellis, R. 2005. Planning and task-based performance: theory and research IN: Ellis, R. (ed.) Planning and task performance in second language. Amsterdam: John Benjamins Publishing Company.

Ellis, R. 2008a. Investigating grammatical difficulty in second language learning: implications for second language acquisition research and language testing. International Journal of Applied Linguistics [Online]. 18 (1), pp4-22. Available from: <http://dx.doi.org/10.1111/j.1473-4192.2008.00184.x> [Accessed 24 April 2009].

Ellis, R. 2008b. The study of second language acquisition. 2 ed. Oxford: Oxford University Press.

Ellis, R. 2009. Corrective Feedback and Teacher Development. L2 Journal [Online]. 1 (1), pp3-18. Available from: <http://www.escholarship.org/uc/item/2504d6w3> [Accessed 21 January 2010].

Ellis, R. and Barkhuizen, G. 2005. Analysing Learner Language. Oxford: Oxford University Press.

Ellis, R., Loewen, S. and Elder, C. 2009. Implicit and explicit knowledge in second language learning, testing and teaching. Bristol: Multilingual Matters.

Elsom-Cook, M. 1993. Student modelling in intelligent tutoring systems. Artificial Intelligence Review [Online]. 7 (3-4), pp227-240. Available from: <http://dx.doi.org/10.1007/BF00849556> [Accessed 20 February 2010].

Embretson, S.E. 2000. Multidimensional Measurement from Dynamic Tests: Abstract Reasoning Under Stress. Multivariate Behavioral Research [Online]. 35 (4), pp505-542. Available from: <http://dx.doi.org/10.1207/S15327906MBR3504_05> [Accessed 21 September 2009].

Engeström, Y. 2007. Putting Vygotsky to work: the change laboratory as an application of double stimulation IN: Daniels, H., Cole, M. and Wertsch, J.V. (eds.) The Cambridge Companion to Vygotsky. Cambridge: Cambridge University Press.

Erben, T., Ban, R. and Summers, R. 2008. Changing examination structures within a college of education: the application of dynamic assessment in pre-service ESOL endorsement courses in Florida IN: Lantolf, J.P. and Poehner, M.E. (eds.) Sociocultural Theory and the Teaching of Second Languages. London: Equinox Publishing Ltd.

Faaß, G. et al. 2009. Part-of-Speech tagging of Northern Sotho: Disambiguating polysemous function words IN: De Pauw, G., De Schryver, G.-M. and Levin, L. (eds.) Proceedings of the EACL 2009 Workshop on Language Technologies for African Languages, AfLaT 2009, 31 March 2009 Megaron Athens International Conference Centre Athens, Greece. Stroudsburg, PA: The Association for Computational Linguistics [Online]. Available from: <http://www.tshwanedje.com/publications/AfLaT2009.pdf #page=48> [Accessed 27 March 2010].

Ferris, D.R. 2002. Treatment of error in second language student writing. Ann Harbour: University of Michigan Press.

Ferris, D.R. 2004. The “Grammar Correction” Debate in L2 Writing: Where are we, and where do we go from here? (and what do we do in the meantime.?). Journal of Second Language Writing [Online]. 13 (1), pp49-62. Available from: <http://dx.doi.org/10.1016/j.jslw.2004.04.005> [Accessed 3 December 2010].

Ferris, D.R. 2010. Second language writing research and written corrective feedback in SLA. Studies in Second Language Acquisition [Online]. 32 (2), pp181-201. Available from: <http://dx.doi.org/10.1017/ S0272263109990490> [Accessed 16 May 2010].

Ferris, D.R. and Roberts, B. 2001. Error feedback in L2 writing classes: How explicit does it need to be? Journal of Second Language Writing [Online]. 10 (3), pp161-184. Available from: <http://dx.doi.org/ 10.1016/S1060-3743(01)00039-X> [Accessed 02 July 2009].

Feuerstein, R., Falik, L.H. and Feuerstein, R. 1998. The learning potential assessment device: an alternative approach to the assessment of learning potential IN: Samuda, R.J., Feuerstein, R., Kaufman, A.S., Lewis, J.E. and Sternberg, R.J. (eds.) Advances in cross-cultural assessment. Thousand Oaks, CA: Sage Publications.

Feuerstein, R. and Feuerstein, S. 1999. Mediated learning experience: a theoretical review IN: Feuerstein, R., Klein, P.S. and Tannenbaum, A.J. (eds.) Mediated learning experience (MLE): theoretical, psychosocial and learning implications. Tel Aviv: Freund Publishing House LTD.

Feuerstein, R., Rand, Y. and Hoffman, M.B. 1979. The dynamic assessment of retarded performers: the learning potential assessment device, theory, instruments, and techniques. Baltimore: University Park Press.

Firth, A. and Wagner, J. 1998. SLA Property: No Trespassing! The Modern Language Journal [Online]. 82 (1), pp91-94. Available from: <http://dx.doi.org/10.2307/328686> [Accessed 5 November 2009].

Fischer, R. 2007. How do we know what students are actually doing? Monitoring students’ behavior in CALL. Computer Assisted Language Learning [Online]. 20 (5), pp409 - 442. Available from: <http:/ /dx.doi.org/10.1080/09588220701746013> [Accessed 19 November 2009].

Flynn, E. and Siegler, R. 2007. Measuring change: current trends and future directions in microgenetic research. Infant and Child Development [Online]. 16 (1), pp135-149. Available from: <http://dx.doi.org/ 10.1002/icd.502> [Accessed 28 August 2009].

Fowley, C. 2011. Publishing the confidential: an ethnographic study of young Irish bloggers. Unpublished PhD dissertation. Dublin City University.

Freedman, R., Ali, S.S. and McRoy, S. 2000. Links: What is an Intelligent Tutoring System? Intelligence [Online]. 11 (3), pp15-16. Available from: <http://bit.ly/aqu4m3> [Accessed 20 February 2010].

Gao, S. 2009. Policies for Teachers towards Errors in College English Writing. International Education Studies [Online]. 2 (2), pp59-62. Available from: <http://journal.ccsenet.org/index.php/ies/article/view/1684/1629> [Accessed 17 January 2010].

Gass, S.M. and Selinker, L. 2001. Second language acquisition: an introductory course. 2nd ed. Mahwah, NJ: Lawrence Erlbaum Associates.

Gass, S.M. and Selinker, L. 2008. Second language acquisition: an introductory course. 3rd ed. New York: Routledge.

Gibbons, P. 2003. Mediating Language Learning: Teacher Interactions With ESL Students in a Content-Based Classroom. TESOL Quarterly [Online]. 37 (2), pp247-273. Available from: <http://www.ingentaconnect.com/content/tesol/tq/2003/00000037/00000002/art00003> [Accessed 25 Sep 2009].

Glaspey, A. 2008. Dynamic Assessment: Training and Reliability Across Clinicians. Paper presented at the International Child Phonology Conference. West Lafayette, IN [Online]. Available from: <http://www.units.muohio.edu/spa/documents/glaspey.international.child.phonology.conference.handout.2008.pdf> [Accessed 22 September 2009].

Glaspey, A. and Stoel-Gammon, C. 2007. A dynamic approach to phonological assessment. Advances in Speech Language Pathology [Online]. 9 (4), pp286-296. Available from: <http://dx.doi.org/10.1080/14417040701435418> [Accessed 20 September 2009].

Granger, S. 2002. A bird's-eye view of learner corpus research IN: Granger, S., Hung, J. and Petch-Tyson, S. (eds.) Computer learner corpora, second language acquisition and foreign language teaching. Amsterdam: John Benjamins Publishing Company.

Granger, S. 2003. Error-tagged learner corpora and CALL: a promising synergy. CALICO Journal. 20 (3), pp465-480.

Granger, S. 2009. The contribution of learner corpora to second language acquisition and foreign language teaching IN: Aijmer, K. (ed.) Corpora and language teaching. Amsterdam: John Benjmins Publishing Company.

Greene, B.B. and Rubin, G.M. 1971. Automated grammatical tagging of English. Department of Linguistics, Brown University.

Gregg, K.R. 1989. Second language acquisition theory: the case for a generative perspective IN: Gass, S.M. and Schachter, J. (eds.) Linguistic Perspectives on Second Language Acquisition. Cambridge: Cambridge University Press.

Gregg, K.R. 1990. The Variable Competence Model of Second Language Acquisition, and Why It Isn't. Applied Linguistics. 11 (4), pp364-383.

Grigorenko, E.L. 2009. Dynamic Assessment and Response to Intervention: Two Sides of One Coin. Journal of Learning Disabilities [Online]. 42 (2), pp111-132. Available from: <http://ldx.sagepub.com/cgi/content/abstract/42/2/111> [Accessed 16 September 2009].

Grigorenko, E.L. and Sternberg, R.J. 1998. Dynamic testing. Psychological Bulletin [Online]. 124 (1), pp75-111.

Guthke, J., Heinrich, A. and Caruso, M. 1986. The diagnostic program of 'syntactical rule and vocabulary acquisition' – a contribution to the psychodiagnosis of foreign language learning ability IN: Klix, F. and Hagendorf, H. (eds.) Human Memory and Cognitive Capabilities. Mechanisms and Performances. Amsterdam: Elsevier.

Hakuta, K. 1974. Preliminary Report on the Development of Grammatical Morphemes in a Japanese Girl Learning English as a Second Language (ED123871). Working Papers on Bilingualism [Online]. 3 (1), pp18-27. Available from: <http://www.eric.ed.gov/ERICWebPortal/contentdelivery/servlet/ERICServlet?accno=ED123871> [Accessed 14 July 2010].

Hamel, M.-J. 2008. Natural language processing tools in CALL IN: Ayoun, D. (ed.) Studies in French applied linguistics. Amsterdam: John Benjamins Publishing Company.

Han, Z. 2004. Fossilization in adult second language acquisition. Clevedon, England: Multilingual Matters.

Hawkins, R. and Towell, R. 2001. French grammar and usage. 2nd ed. London: Hodder Arnold.

Hayes, A.F. and Krippendorff, K. 2007. Answering the Call for a Standard Reliability Measure for Coding Data. Communication Methods and Measures [Online]. 1 (1), pp77-89. Available from: <http://www.informaworld.com/10.1080/19312450709336664> [Accessed 24 May 2010].

Haywood, H.C. and Lidz, C.S. 2007. Dynamic assessment in practice : clinical and educational applications. Cambridge: Cambridge University Press.

Heift, T. 2001. Intelligent Language Tutoring Systems for Grammar Practice. Zeitschrift für Interkulturellen Fremdsprachenunterricht [Online]. 6 (2), pp1-15. Available from: <http://www.spz.tu-darmstadt.de/projekt_ejournal/jg_06_2/beitrag/heift2.htm> [Accessed 21 February 2010].

Heift, T. 2002. Learner Control and Error Correction in ICALL: Browsers, Peekers, and Adamants. CALICO Journal [Online]. 19 (2), pp295-313. Available from: <https://calico.org/html/article_428.pdf> [Accessed 24 November 2010].

Heift, T. 2003. Multiple Learner Errors and Meaningful Feedback: A Challenge for ICALL Systems. IN: Heift, T. and Schulze, M. (eds.) Error Diagnosis and Error Correction in CALL. CALICO Journal. 20(3) (3), pp533-549.

Heift, T. 2004. Corrective feedback and learner uptake in CALL. ReCALL [Online]. 16 (02), pp416-431. Available from: <http://dx.doi.org/10.1017/S0958344004001120> [Accessed 20 November 2008].

Heift, T. 2007. Learner Personas in CALL. Calico Journal [Online]. 25 (1), pp1-10. Available from: <https://www.calico.org/html/article_666.pdf> [Accessed 2 September 2010].

Heift, T. 2008. Modeling learner variability in CALL. Computer Assisted Language Learning [Online]. 21 (4), pp305-321. Available from: <http://dx.doi.org/10.1080/09588220802343421> [Accessed 15 November 2008].

Heift, T. 2010a. Colloquium – Bridging computational and applied linguistics: Implementation challenges and benefits. Language Teaching [Online]. 43 (1), pp103-105. Available from: <http://dx.doi.org/10.1017/S0261444809990218> [Accessed 22 January 2010].

Heift, T. 2010b. Developing an Intelligent Language Tutor. CALICO Journal [Online]. 27 (3), pp443-459. Available from: <https://calico.org/page.php?id=5> [Accessed 21 April 2010].

Heift, T. 2010c. Prompting in CALL: A Longitudinal Study of Learner Uptake. The Modern Language Journal [Online]. 94 (2), pp198-216. Available from: <http://dx.doi.org/10.1111/j.1540-4781.2010.01017.x> [Accessed 24 November 2010].

Heift, T. and Nicholson, D. 2001. Web Delivery of Adaptive and Interactive Language Tutoring. International Journal of artificial Intelligence in Education [Online]. 12 (4), pp-. Available from: <http://iaied.org/pub/973/file/973_paper.pdf> [Accessed 21 February 2010].

Heift, T. and Rimrott, A. 2008. Learner responses to corrective feedback for spelling errors in CALL. System [Online]. 36 (2), pp196-213. Available from: <http://dx.doi.org/10.1016/j.system.2007.09.007> [Accessed 29 April 2011].

Heift, T. and Schulze, M. 2003. Student modeling and ab initio Language Learning. System [Online]. 31 (4), pp519-535. Available from: <http://dx.doi.org/10.1016/j.system.2003.02.003> [Accessed 29 March 2010].

Heift, T. and Schulze, M. 2007. Errors and intelligence in computer-assisted language learning; parsers and pedagogues. New York: Routledge.

Heritage, M. et al. 2009. From Evidence to Action: A Seamless Process in Formative Assessment? Educational Measurement: Issues and Practice [Online]. 28 (3), pp24-31. Available from: <http://dx.doi.org/10.1111/j.1745-3992.2009.00151.x> [Accessed 4 September 2009].

Herschensohn, J. and Arteaga, D. 2009. Tense and verb raising in advanced L2 French. Journal of French Language Studies [Online]. 19 (3), pp291-318. Available from: <http://dx.doi.org/10.1017/S0959269509990019> [Accessed 28 January 2010].

Holt, P. et al. 1994. The state of student modelling IN: Greer, J.E. and McCalla, G.I. (eds.) Student modeling: the key to individualised knowledge-based instruction. Berlin: Springer Verlag.

Holton, D. and Clarke, D. 2006. Scaffolding and metacognition. International Journal of Mathematical Education in Science and Technology [Online]. 37 (2), pp127-143. Available from: <http://dx.doi.org/10.1080/00207390500285818> [Accessed 8 September 2009].

Housen, A. and Kuiken, F. 2009. Complexity, Accuracy, and Fluency in Second Language Acquisition. Applied Linguistics [Online]. 30 (4), pp461-473. Available from: <http://dx.doi.org/10.1093/applin/amp048> [Accessed 14 July 2010].

Howard, M. 2006. Variation in Advanced French Interlanguage: A Comparison of Three (Socio)linguistic Variables. Canadian Modern Language Review [Online]. 62 (3), pp379-400. Available from: <http://tiny.cc/llmrG> [Accessed 1 February 2010].

Hull, R. and Gomez, F. 2002. Automatic acquisition of biographic knowledge from encyclopedic texts IN: Kent, A. and Williams, J.G. (eds.) Encyclopedia of microcomputers: volume 28. New York: Marcel Dekker, INC.

Hulstijn, J.H. 1989. A cognitive view on interlanguage variability IN: Eisenstein, M.R. (ed.) The Dynamic Interlanguage: Empirical Studies in Second Language Variation. New York: Plenum.

Hunt, K.W. 1965. Grammatical structures written at three grade levels, NCTE Research Report 3 (ED113735). Champaign, IL: National Council of Teachers of English [Online]. Available from: <http://www.eric.ed.gov/ERICWebPortal/contentdelivery/servlet/ERICServlet?accno=ED113735> [Accessed 14 July 2010].

Hwang, A., Ang, S. and Francesco, A.M. 2002. The Silent Chinese: The Influence of Face and Kiasuism on Student Feedback-Seeking Behaviors. Journal of Management Education. 26 (1), pp70-98.

Hyland, K. and Hyland, F. 2006. Feedback on second language students' writing. Language Teaching [Online]. 39 (2), pp83-101. Available from: <http://dx.doi.org/10.1017/S0261444806003399> [Accessed 15 February 2010].

James, C. 1998. Errors in language learning and use: exploring error analysis. Edinburgh: Pearson Education Limited.

James, M.A. 2007. Interlanguage variation and transfer of learning. IRAL [Online]. 45 (2), pp95-118. Available from: <http://dx.doi.org/10.1515/IRAL.2007.004> [Accessed 26 November 2009].

Jia, J. 2009. An intelligent web-based human-computer interaction system with natural language CSIEC and its integration into English instruction IN: Chang, M. (ed.) Learning Culture and Language Through ICTs: Methods for Enhanced Instruction. London: Information Science Reference.

Judson, E. 2006. How Teachers Integrate Technology and Their Beliefs About Learning: Is There a Connection? Journal of Technology and Teacher Education [Online]. 14 (3), pp581-597. Available from: <http://tinyurl.com/ledf7l> [Accessed 4 September 2009].

Jurafsky, D. and Martin, J.H. 2000. Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics and Speech Recognition. New Jersey: Prentice Hall.

Kalyuga, S. and Sweller, J. 2005. Rapid Dynamic Assessment of Expertise to Improve the Efficiency of Adaptive E-learning. Educational Technology Research & Development [Online]. 53 (3), pp83-93. Available from: <http://dx.doi.org/10.1007/BF02504800> [Accessed 21 September 2009].

Kang, Y.-S. and Maciejewski, A.A. 2000. A Student Model of Technical Japanese Reading Proficiency for an Intelligent Tutoring System. CALICO Journal [Online]. 18 (1), pp9-40. Available from: <http://calico.org/html/article_498.pdf> [Accessed 28 March 2010].

Kaniel, S. 2000. A metacognitive decision-making model for dynamic assessment and intervention IN: Lidz, C.S. and Elliott, J.G. (eds.) Advances in cognition and educational practice (volume 6); Dynamic assessment: prevailing models and applications; New York: Elsevier Science.

Karlsson, F. 1995. Designing a parser for unrestricted text IN: Karlsson, F., Voutilainen, A., Heikkilä, J. and Anttila, A. (eds.) Constraint Grammar: A Language-Independent System for Parsing Unrestricted Text. Berlin: Mouton de Gruyter.

Keane, K.J. 1987. Assessing deaf children IN: Lidz, C.S. (ed.) Dynamic assessment: An interactional approach to evaluating learning potential. New York: Guilford.

Kinginger, C. 2002. Defining the Zone of Proximal Development in US Foreign Language Education. Applied Linguistics [Online]. 23 (2), pp240-261. Available from: <http://dx.doi.org/10.1093/applin/23.2.240> [Accessed 29 July 2009].

Kodaganallur, V., Weitz, R.R. and Rosenthal, D. 2005. A Comparison of Model-Tracing and Constraint-Based Intelligent Tutoring Paradigms. International Journal of Artificial Intelligence in Education [Online]. 15 (2), pp117-144. Available from: <http://iaied.org/pub/1013/file/1013_Kodaganallur05.pdf> [Accessed 21 February 2010].

Kormos, J. 2011. Task complexity and linguistic and discourse features of narrative writing performance. Journal of Second Language Writing [Online]. 20 (2), pp148-161. Available from: <http://dx.doi.org/10.1016/j.jslw.2011.02.001> [Accessed 23 April 2011].

Kozulin, A. 2003. Psychological tools and mediated learning IN: Kozulin, A., Gindis, B., Ageyev, V.S. and Miller, S.M. (eds.) Vygotsky's Educational Theory in Cultural Context. Cambridge: Cambridge University Press.

Kozulin, A. and Garb, E. 2002. Dynamic Assessment of EFL Text Comprehension. School Psychology International [Online]. 23 (1), pp112-127. Available from: <http://dx.doi.org/10.1177/0143034302023001733> [Accessed 09 July 2009].

Krippendorff, K. 1970. Bivariate Agreement Coefficients for Reliability of Data. Sociological Methodology [Online]. 2 (1), pp139-150. Available from: <http://dx.doi.org/10.2307/270787> [Accessed 04 February 2009].

L'Haire, S. 2007. FipsOrtho: a spell checker for learners of French. ReCALL [Online]. 19 (2), pp137-161. Available from: <http://dx.doi.org/10.1017/S0958344007000420> [Accessed 3 April 2010].

Labov, W. 1971. The notion of "system" in Creoles languages IN: Hymes, D. (ed.) Pidginization and Creolization of languages. Cambridge: Cambridge University Press.

Labov, W. 1972. The internal evolution of linguistic rules IN: Stockwell, R.P. and Macaulay, R.K.S. (eds.) Linguistic change and generative theory : essays. London: Bloomington.

Lado, R. 1957. Linguistics across cultures: applied linguistics for language teachers.

LaFontana, V.R. 1996. Throw Away That Correcting Pen. The English Journal [Online]. 85 (6), pp71-73. Available from: <http://www.jstor.org/stable/819831> [Accessed 24 November 2010].

Laister, J. and Koubek, A. 2001. 3rd Generation Learning Platforms: Requirements and Motivation for Collaborative Learning. European Journal of Open, Distance and E-Learning [Online]. pp1-10. Available from: <http://www.eurodl.org/materials/contrib/2001/icl01/laister.pdf> [Accessed 13 March 2011].

Lakshmanan, U. and Selinker, L. 2001. Analysing interlanguage: how do we know what learners know? Second Language Research. 17 (4), pp393-420.

Lambert, E.B. and Clyde, M. 2003. Putting Vygotsky to the test IN: Lytle, D.E. (ed.) Play and educational theory and practice. Wesport: Praeger Publishers.

Lantolf, J.P. 2000. Introducing sociocultural theory IN: Lantolf, J.P. (ed.) Sociocultural theory and second language learning. Oxford: Oxford University Press.

Lantolf, J.P. 2004. Sociocultural theory and second and foreign language learning: an overview of sociocultural theory IN: Van Esch, K. and John, O.S. (eds.) New Insights Into Foreign Language Learning And Teaching. Frankfurt am Main, Germany: Peter Lang Verlag.

Lantolf, J.P. 2005. Sociocultural and second language learning research: an exegesis IN: Hinkel, E. (ed.) Handbook of research in second language teaching and learning. Mahwah, NJ: Lawrence Erlbaum.

Lantolf, J.P. 2007. Sociocultural source of thinking and its relevance for second language acquisition. Bilingualism: Language and Cognition [Online]. 10 (1), pp31-33. Available from: <http://dx.doi.org/10.1017/S1366728906002768> [Accessed 5 November 2009].

Lantolf, J.P. 2009. Dynamic assessment: the dialectic integration of instruction and assessment. Language Teaching [Online]. 42 (3), pp355-368. Available from: <http://dx.doi.org/10.1017/S0261444808005569> [Accessed 14 July 2009].

Lantolf, J.P. and Aljaafreh, A. 1995. Second language learning in the zone of proximal development: A revolutionary experience. International Journal of Educational Research [Online]. 23 (7), pp619-632. Available from: <http://dx.doi.org/10.1016/0883-0355(96)80441-1> [Accessed 21 January 2010].

Lantolf, J.P. and Poehner, M.E. 2004. Dynamic assessment of L2 development: Bringing the past into the future. Journal of Applied Linguistics [Online]. 1 (1), p49. Available from: <http://dx.doi.org/10.1558/japl.1.1.49.55872> [Accessed 10 July 2009].

Lantolf, J.P. and Poehner, M.E. 2009. The artificial development of second language ability: a sociocultural approach IN: Ritchie, W.C. and Bhatia, T.K. (eds.) The new handbook of second language acquisition. 2nd ed. Bingley, UK: Emerald Group Publishing Limited.

Lantolf, J.P. and Poehner, M.E. 2011. Dynamic assessment in the classroom: Vygotskian praxis for second language development. Language Teaching Research [Online]. 15 (1), pp11-33. Available from: <http://dx.doi.org/10.1177/1362168810383328> [Accessed 12 January 2011].

Lantolf, J.P. and Thorne, S.L. 2007. Sociocultural theory and second language learning IN: VanPatten, B. and Williams, J. (eds.) Theories in second language acquisition: an introduction. Mahwah, New Jersey: Lawrence Erlbaum Associates.

Lapkin, S., Swain, M. and Smith, M. 2002. Reformulation and the Learning of French Pronominal Verbs in a Canadian French Immersion Context. The Modern Language Journal [Online]. 86 (4), pp485-507. Available from: <http://dx.doi.org/10.1111/1540-4781.00157> [Accessed 12 September 2010].

Lardiere, D. 2007. Ultimate attainment in second language acquisition: a case study. Mahwah, New Jersey: Lawrence Erlbaum Associates.

Laufer, B. and Eliasson, S. 1993. What causes avoidance in L2 learning. Studies in Second Language Acquisition [Online]. 15 (1), pp35-48. Available from: <http://dx.doi.org/10.1017/S0272263100011657> [Accessed 15 July 2011].

Lee, I. 1997. ESL learners' performance in error correction in writing: Some implications for teaching. System [Online]. 25 (4), pp465-477. Available from: <http://dx.doi.org/10.1016/S0346-251X(97)00045-6> [Accessed 23 November 2010].

Lee, I. 2003. L2 writing teachers' perspectives, practices and problems regarding error feedback. Assessing Writing [Online]. 8 (3), pp216-237. Available from: <http://dx.doi.org/10.1016/j.asw.2003.08.002> [Accessed 23 November 2010].

Leech, G. 1997. Introducing corpus annotation IN: Garside, R., Leech, G. and McEnery, A. (eds.) Corpus annotation: linguistic information from computer text corpora. London: Addison Wesley Longman.

Leech, G. 2005. Adding Linguistic Annotation IN: Wynne, M. (ed.) Developing Linguistic Corpora: a Guide to Good Practice. Oxford: Oxford Books [Online]. Available from: <http://www.ahds.ac.uk/guides/linguistic-corpora/chapter2.htm> [Accessed 15 July 2010].

Leung, C. 2007. Dynamic Assessment: Assessment for and as Teaching? Language Assessment Quarterly [Online]. 4 (3), pp257-278.

Lidz, C.S. 1991. Practitioner's Guide to Dynamic Assessment. New York: Guilford.

Lightbown, P.M. and Spada, N. 1999. How languages are learned. Oxford: Oxford University Press.

Littlewood, W. 1984. Foreign and Second Language Learning: Language acquisition research and its implication for the classroom. Cambridge: Cambridge University Press.

Liu, C. 2006. Using Bayesian networks for student modeling IN: Alkhalifa, E.M. (ed.) Cognitively informed systems: utilizing practical approaches to enrich information presentation and transfer. London: Idea Group Publishing.

Long, M.H. 2003. Stabilization and Fossilization in Interlanguage Development IN: Doughty, C.J. and Long, M.H. (eds.) The Handbook of Second Language Acquisition. Malden, MA: Blackwell Publishing Ltd.

Lu, X. 2008. Automatic Measurement of Syntactic Complexity Using the Revised Developmental Level Scale. IN: Proceedings of the Twenty-First International FLAIRS Conference (2008). Coconut Grove: American Association for Artificial Intelligence [Online]. Available from: <http://www.aaai.org/Papers/FLAIRS/2008/FLAIRS08-043.pdf> [Accessed 15 July 2010].

Lu, X. 2009. Automating Measures of L2 Syntactic Complexity. IN: Workshop presentation. Automatic Analysis of Learner Language (AALL'09): From a Better Understanding of Annotation Needs to the Development and Standardization of Annotation Schemes, 10-MAR-09 - 11-MAR-09. Arizona State University, Tempe, AZ, US [Online]. Available from: <http://purl.org/dm/events/aall09/lu.pdf> [Accessed 15 July 2010].

Lu, X. 2010. Automatic analysis of syntactic complexity in second language writing. International Journal of Corpus Linguistics [Online]. 15 (4), pp474-496. Available from: <http://dx.doi.org/10.1075/ijcl.15.4.02lu> [Accessed 11 November 2010].

Lund, A. 2008. Wikis: a collective approach to language production. ReCALL [Online]. 20 (1), pp35-54. Available from: <http://dx.doi.org/10.1017/S0958344008000414> [Accessed 10 November 2010].

Luria, A. 1935. L.S. Vygotsky. Character and Personality, An International Psychological Quarterly [Online]. 3 (3), pp3-4. Available from: <http://luria.ucsd.edu/Character -Personality.pdf> [Accessed 15 July 2009].

Lyons, J. 1972. Human language IN: Hinde, R.A. (ed.) Non-verbal communication. New York: Cambridge University Press.

Lyster, R. 1998. Negotiation of form, recasts, and explicit correction in relation to error types and learner repair in immersion classrooms. Language Learning [Online]. 48 (2), pp183-218. Available from: <http://dx.doi.org/10.1111/1467-9922.00039> [Accessed 23 June 2011].

Lyster, R. 2007. Learning and teaching languages through content: A counterbalanced approach. Amsterdam: John Benjamins Publishing Company.

Lyster, R. and Ranta, L. 1997. Corrective feedback and learner uptake. Studies in Second Language Acquisition [Online]. 19 (1), pp37-66. Available from: <http://dx.doi.org/10.1017/S0272263197001034> [Accessed 23 June 2011].

Mackey, A. and Gass, S. 2005. Second language research: methodology and design. Mahwah, New Jersey: Lawrence Erlbaum Associates.

Mackey, A., Gass, S. and McDonough, K. 2000. How do learners percei

ve interactional feedback? Studies in Second Language Acquisition [Online]. 22 (4), pp471-497. Available from: <http://dx.doi.org/10.1017/S0272263100004022> [Accessed 02 July 2009].

Marcus, M.P., Marcinkiewicz, M.A. and Santorini, B. 1993. Building a large annotated corpus of English: the penn treebank. Computational Linguistics [Online]. 19 (2), pp313-330. Available from: <http://www.ldc.upenn.edu/acl/J/J93/J93-2004.pdf> [Accessed 20 January 2009].

Maritxalar, M., Diaz de Ilarraza, A. and Oronoz, M. 1997. From Psycholinguistic Modelling of Interlanguage in Second Language Acquisition to a Computational Model. IN: CoNLL97: Computational Natural Language Learning. [Online]. Available from: <http://www.ldc.upenn.edu/acl/W/W97/W97-1007.pdf> [Accessed 6 July 2010].

Màrquez, L., Padró, L. and Rodríguez, H. 2000. A Machine Learning Approach to POS Tagging. Machine Learning. 39 (1), pp59-91.

Marshall, I. 1983. Choice of Grammatical Word-Class without Global Syntactic Analysis: Tagging Words in the LOB Corpus. Computers and the Humanities [Online]. 17 (3), pp139-150. Available from: <http://www.jstor.org/stable/30204076> [Accessed 27 March 2010].

Martin, B. 1999. Constraint-Based Modeling: Representing Student Knowledge. New Zealand Journal of Computing [Online]. 7 (1), pp30-38. Available from: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.24.6266&rep=rep1&type=pdf> [Accessed 21 February 2010].

Martin, B. and Mitrovic, A. 2000. Tailoring feedback by correcting student answers IN: Gauthier, G., Frasson, C. and VanLehn, K. (eds.) Intelligent tutoring systems: 5th International Conference, ITS 2000, Montréal, Canada, June 2000, Proceedings. Berlin: Springer-Verlag.

Mathews, W.M. 1961. Study of the abnormal child: special session, 1960 (President W. Mason Mathews, Chairman, introducing Professor Alexander Romanovich Luria). American Journal of Orthopsychiatry [Online]. 31 (1), pp1–16. Available from: <http://dx.doi.org/10.1111/j.1939-0025.1961.tb02104.x> [Accessed 27 April 2010].

Mayo, M., Mitrovic, A. and McKenzie, J. 2000. CAPIT: An Intelligent Tutoring System for Capitalization and Punctuation. IN: Jesshope, C. and Okamoto, T. (eds.) International Workshop on Advanced Learning Technologies Iwalt2000: Advanced Learning Technology : Design and Development Issues : Proceedings 4-6 Dec 2000. Palmerston North, NZ: Institute of Electrical & Electronics Enginee.

McCarthy, C. 1996. The Architecture of Cognition. Mahwah, NJ: Lawrence Erlbaum Publishers.

McCoy, K.F. and Masterman, L.N. 1997. A Tutor for Teaching English as a Second language for Deaf Users of American Sign Language. IN: Proceedings of Natural Language Processing for Communication Aids, an ACL/EACL'97 Workshop, Madrid, Spain. [Online]. Available from: <http://www.eecis.udel.edu/research/icicle/pubs/McCoMast97.pdf> [Accessed 2 March 2010].

McCoy, K.F., Pennington, C.A. and Suri, L.Z. 1996. English error correction: A syntactic user model based on principled “mal-rule” scoring. IN: Proceedings of UM-96, the Fifth International Conference on User Modeling. Kailua-Kona, Hawaii, January 1996. [Online]. Available from: <http://www.eecis.udel.edu/research/icicle/pubs/McPeSu96.pdf> [Accessed 2 March 2010].

McEnery, T., Xiao, R. and Tono, Y. 2006. Corpus-based language studies: an advanced resource book. New York: Routledge.

McGarrell, H. 2011. AILA Matters: Writing Teacher Response (WTR). International Journal of Applied Linguistics [Online]. 21 (1), pp140-142. Available from: <http://dx.doi.org/10.1111/j.1473-4192.2011.00281.x> [Accessed 12 March 2011].

Meisel, J.M., Clahsen, H. and Pienemann, M. 1981. On determining developmental stages in natural second language acquisition. Studies in Second Language Acquisition [Online]. 3 (2), pp109-135. Available from: <http://dx.doi.org/10.1017/S0272263100004137> [Accessed 13 July 2010].

Meurers, D. and Wunsch, H. 2010. Linguistically Annotated Learner Corpora: Aspects of a Layered Linguistic Encoding and Standardized Representation. IN: Pre-Proceedings of Linguistic Evidence 2010. [Online]. Available from: <http://www.sfs.uni-tuebingen.de/~dm/papers/meurers-wunsch-10.pdf> [Accessed 16 July 2010].

Michaud, L.N. and McCoy, K.F. 1998. Planning Tutorial Text in a System for Teaching English as a Second Language to Deaf Learners. IN: Proceedings of Integrating Artificial Intelligence and Assistive Technology, an AAAI ‘98 Workshop, Madison, Wisconsin. [Online]. Available from: <http://www.eecis.udel.edu/research/icicle/pubs/MichMcCo98.pdf> [Accessed 2 March 2010].

Michaud, L.N. and McCoy, K.F. 2003. Evaluating a model to disambiguate natural language parses on the basis of user language proficiency. IN: Proceedings of the 9th international conference, User modeling, Johnstown, PA, USA, June 2003. Berlin: Springer-Verlag.

Michaud, L.N. and McCoy, K.F. 2006. Capturing the Evolution of Grammatical Knowledge in a CALL System for Deaf Learners of English. International Journal of Artificial Intelligence in Education [Online]. 16 (1), pp65-97. Available from: <http://www.eecis.udel.edu/research/icicle/pubs/Michaud-McCoy-2006.pdf> [Accessed 25 November 2009].

Michaud, L.N., McCoy, K.F. and Pennington, C.A. 2000. An intelligent tutoring system for deaf learners of written English IN: Tremaine, M., Cole, E. and Mynatt, E. (eds.) Proceedings of the fourth international ACM conference on Assistive technologies, 2000, Arlington, Virginia, United States, November 13-15. Michigan: Association for Computing Machinery.

Mihalcea, R. 2003. Unsupervised natural language disambiguation using non-ambiguous words IN: Nicolov, N., Bontcheva, K., Angelova, G. and Mitkov, R. (eds.) Recent advances in natural language processing III: selected papers from the RANLP conference 2003. Philadelphia: John Benjamins Publishing Company.

Mitchell, R. 1992. Testing for learning: how new approaches to evaluation can improve American Schools. New York: The Free Press.

Mitrovic, A. 1997. SQL-Tutor: a Preliminary Report. Technical Report TR-COSC 08/97. Computer Science Department, University of Canterbury [Online]. pp1-35. Available from: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.32.8265&rep=rep1&type=pdf> [Accessed 20 February 2010].

Mitrovic, A., Martin, B. and Mayo, M. 2002. Using Evaluation to Shape ITS Design: Results and Experiences with SQL-Tutor. User Modeling and User-Adapted Interaction [Online]. 12 (2-3), pp243-279. Available from: <http://dx.doi.org/10.1023/A:1015022619307> [Accessed 20 February 2010].

Mitrovic, A. and Ohlsson, S. 1999. Evaluation of a Constraint-Based Tutor for a Database Language. International Journal of Artificial Intelligence in Education [Online]. 10 (3), pp238-256. Available from: <http://www.ijaied.org/pub/987/file/987_paper.pdf> [Accessed 20 February 2010].

Moore-Brown, B. et al. 2006. Using Dynamic Assessment to Evaluate Children With Suspected Learning Disabilities. Intervention in School and Clinic [Online]. 41 (4), pp209-217. Available from: <http://dx.doi.org/10.1177/10534512060410040301> [Accessed 16 July 2009].

Nagata, N. 2002. BANZAI: An Application of Natural Language Processing to Web-based Language Learning. Calico Journal. 19 (3), pp583-599.

Nassaji, H. and Swain, M. 2000. A Vygotskian Perspective on Corrective Feedback in L2: The Effect of Random Versus Negotiated Help on the Learning of English Articles. Language Awareness. 9 (1), pp34-51.

Navarro, E. and Calero, M.D. 2009. Estimation of Cognitive Plasticity in Old Adults Using Dynamic Assessment Techniques. Journal of Cognitive Education and Psychology. 8 (1), pp38-51.

Ndiaye, M. and Vandeventer Faltin, A. 2003. A Spell Checker Tailored to Language Learners. Computer Assisted Language Learning. 16 (2/3), pp213-232.

Nemser, W. 1971. Approximative systems of foreign language learners. IRAL - International Review of Applied Linguistics in Language Teaching [Online]. 9 (2), pp115-124. Available from: <http://dx.doi.org/10.1515/iral.1971.9.2.115> [Accessed 6 January 2010].

Nemser, W. 1974. Approximative systems of foreign language learners (reprinted from IRAL, Vol 9/2, 1971) IN: Richards, J.C. (ed.) Error analysis: perspectives on second language acquisition. Essex: Longman Group Limited.

Nguyen, L. and Do, P. 2008. Learner Model in Adaptive Learning. World Academy of Science, Engineering and Technology [Online]. 45 (70), pp395-400. Available from: <www.waset.org/journals/waset/v45/v45-70.pdf> [Accessed 22 February 2010].

Norris, J.M. and Ortega, L. 2009. Towards an Organic Approach to Investigating CAF in Instructed SLA: The Case of Complexity. Applied Linguistics Applied Linguistics [Online]. 30 (4), pp555-578. Available from: <http://dx.doi.org/10.1093/applin/amp044> [Accessed 14 July 2010].

Ohlsson, S. 1992. Constraint-based Student Modeling. Journal of artificial intelligence in education. 3 (4), pp449-447.

Ohlsson, S. 1994. Constraint-based student modeling IN: Greer, J.E. and McCalla, G.I. (eds.) Student modeling: the key to individualised knowledge-based instruction. Berlin: Springer Verlag.

Ortega, L. 2003. Syntactic Complexity Measures and their Relationship to L2 Proficiency: A Research Synthesis of College-level L2 Writing. Applied Linguistics Applied Linguistics [Online]. 24 (4), pp492-518. Available from: <http://applij.oxfordjournals.org/cgi/content/abstract/24/4/492> [Accessed 14 July 2010].

Ortega, L. 2005. Methodology, Epistemology, and Ethics in Instructed SLA Research: An Introduction. The Modern Language Journal [Online]. 89 (3), pp317-327. Available from: <http://dx.doi.org/10.1111/j.1540-4781.2005.00307.x> [Accessed 17 July 2010].

Ortega, L. 2007. Second Language Learning Explained? SLA Across Nine Contemporary Theories IN: VanPatten, B. and Williams, J. (eds.) Theories in second language acquisition: an introduction. Mahwah, New Jersey: Lawrence Erlbaum Associates.

Oskoz, A. 2005. Students' Dynamic Assessment Via Online Chat. CALICO Journal. 22 (3), pp513-536.

Palak, D. and Walls, R.T. 2009. Teachers' Beliefs and Technology Practices: A Mixed-methods Approach. Journal of Research on Technology in Education. 41 (4), pp417-441.

Pallotti, G. 2009. CAF: Defining, Refining and Differentiating Constructs. Applied Linguistics Applied Linguistics [Online]. 30 (4), pp590-601. Available from: <http://dx.doi.org/10.1093/applin/amp045> [Accessed 14 July 2010].

Parr, J.M. and Timperley, H.S. 2010. Feedback to writing, assessment for teaching and learning and student progress. Assessing Writing [Online]. 15 (2), pp68-85. Available from: <http://dx.doi.org/10.1016/ j.asw.2010.05.004> [Accessed 23 November 2010].

Payne, S.J. and Squibb, H.R. 1990. Algebra Mal-Rules and Cognitive Accounts of Error. Cognitive Science Journal [Online]. 14 (3), pp445–461. Available from: <http://dx.doi.org/10.1207/s15516709cog1403_4> [Accessed 23 February 2010].

Peña, E.D. and Gillam, R.B. 2000. Dynamic assessment of children referred for speech and language evaluations IN: Lidz, C.S. and Elliott, J.G. (eds.) Dynamic Assessment: Prevailing Models and Applications, Volume 6 (Advances in Cognition and Educational Practice). New York: JAI Press.

Pérez, T.A. et al. 2000. Learning Basque in a Distance-Adaptive way IN: Ortega, M. and Bravo, J. (eds.) Computers and Education in the 21st Century. Dordrecht: Kluwer Academic Publishers [Online]. Available from: <http://dx.doi.org/10.1007/0-306-47532-4_23> [Accessed 28 March 2010].

Perie, M., Marion, S. and Gong, B. 2009. Moving Toward a Comprehensive Assessment System: A Framework for Considering Interim Assessments. Educational Measurement: Issues and Practice [Online]. 28 (3), pp5-13. Available from: <http://dx.doi.org/10.1111/j.1745-3992.2009.00149.x> [Accessed 4 September 2009].

Pica, T. 2005. Second language acquisition research and applied linguistics IN: Hinkel, E. (ed.) Handbook of research in second language teaching and learning. Mahwah, NJ: Lawrence Erlbaum.

Pienemann, M. 2007. Variation and dynamic systems in SLA. Bilingualism: Language and Cognition [Online]. 10 (1), pp43-45. Available from: <http://dx.doi.org/10.1017/S1366728906002793> [Accessed 5 November 2009].

Poehner, M.E. 2005. Dynamic assessment of oral proficiency among advanced L2 learners of French. Doctor of Philosophy. The Pennsylvania State University.

Poehner, M.E. 2007. Beyond the Test: L2 Dynamic Assessment and the Transcendence of Mediated Learning. The Modern Language Journal [Online]. 91 (3), pp323-340. Available from: <http://dx.doi.org/ 10.1111/j.1540-4781.2007.00583.x> [Accessed 02 March 2009].

Poehner, M.E. 2008. Dynamic Assessment: A Vygotskian Approach to Understanding and Promoting L2 Development. Milton Keynes: Springer.

Poehner, M.E. 2009. Group Dynamic Assessment: Mediation for the L2 Classroom. TESOL Quarterly. 43 (3), pp471-491.

Poehner, M.E. 2011. Validity and interaction in the ZPD: interpreting learner development through L2 dynamic assessment. International Journal of Applied Linguistics [Online]. 21 (2), pp244-263. Available from: <http://dx.doi.org/10.1111/j.1473-4192.2010.00277.x> [Accessed 17 July 2011].

Poehner, M.E. and Lantolf, J.P. 2005. Dynamic assessment in the language classroom. Language Teaching Research [Online]. 9 (3), pp233-265. Available from: <http://dx.doi.org/10.1191/1362168805lr166oa> [Accessed 08 April 2009].

Portes, P.R., Smith, T. and Cuentas, T.E. 1993. Parent child interaction patterns in dyadic problem solving: an application of Vygotsky's double stimulation method (ED361359). IN: Paper presentation at AERA 1993 annual meeting, Atlanta, GA. Atlanta [Online]. Available from: <http://www.eric.ed.gov/ ERICWebPortal/contentdelivery/servlet/ERICServlet?accno=ED361359> [Accessed 13 February 2011].

Pradhan, S.S., Ward, W. and Martin, J.H. 2008. Towards Robust Semantic Role Labeling. Computational Linguistics. 34 (2), pp556-563.

Pravec, N.A. 2002. Survey of learner corpora. International Computer Archive of Modern and Medieval English [Online]. 26 (2), pp81-114. Available from: <http://icame.uib.no/ij26/pravec.pdf> [Accessed 12 April 2010].

Preston, D.R. 1996. Variationist perspectives on second language acquisition IN: Bayley, R. and Preston, D.R. (eds.) Second language acquisition and linguistic variation. Amsterdam: John Benjamins.

Prior, M., Rogerson, S. and Fairweather, B. 2002. The ethical attitudes of information systems professionals: outcomes of an initial survey. Telematics and Informatics [Online]. 19 (1), pp21-36. Available from: <http://dx.doi.org/10.1016/S0736-5853(00)00014-9> [Accessed 8 April 2010].

Reif, F. and Allen, S. 1992. Cognition for Interpreting Scientific Concepts: A Study of Acceleration. Cognition & Instruction [Online]. 9 (1), pp1-44. Available from: <http://dx.doi.org/10.1207/s1532690xci0901_1> [Accessed 19 February 2010].

Richards, J.C. 1974. A Non-Contrastive Approach to Error Analysis (reprinted from English Language Teaching, Vol 25/3, 1971) IN: Richards, J.C. (ed.) Error analysis: perspectives on second language acquisition. Essex: Longman Group Limited.

Rimrott, A. and Heift, T. 2005. Language Learners and Generic Spell Checkers in CALL. CALICO Journal [Online]. 23 (1), pp17-48. Available from: <https://calico.org/html/article_130.pdf> [Accessed 16 July 2010].

Robbins, D. 2003. Vygotsky's and A.A. Leontiev's Semiotics and Psycholinguistics: Applications for Education, Second Language Acquisition, and Theories of Language. Westport: Praeger Publishers.

Robinson, P. 2001. Task complexity, task difficulty, and task production: exploring interactions in a componential framework. Applied Linguistics [Online]. 22 (1), pp27-57. Available from: <http://dx.doi.org/10.1093/applin/22.1.27> [Accessed 27 November 2009].

Rogers, C.B. 1995. The spider chart: a unique tool for performance appraisal. Annual Quality Congress [Online]. 49 (1), pp16-22. Available from: <http://www.asq.org/qic/display-item/index.html?item=10196> [Accessed 16 February 2009].

Romaine, S. 2003. Variation IN: Doughty, C.J. and Long, M.H. (eds.) The Handbook of Second Language Acquisition. Malden, MA: Blackwell Publishing Ltd.

Ross, S.M., Morrison, G.R. and Lowther, D.L. 2010. Educational Technology Research Past and Present: Balancing Rigor and Relevance to Impact School Learning. Contemporary Educational Technology [Online]. 1 (1), pp17-35. Available from: <http://cedtech.net/articles/112.pdf> [Accessed 12 March 2011].

Rothman, J. 2007. Sometimes They Use It, Sometimes They Don't: An Epistemological Discussion of L2 Morphological Production and Its Use as a Competence Measurement. Applied Linguistics [Online]. 28 (4), pp609-614. Available from: <http://dx.doi.org/10.1093/applin/amm035> [Accessed 14 February 2009].

Rothman, J. et al. 2010. On the (un)-ambiguity of adjectival interpretations in L2 Spanish: informing debates on the mental representations of L2 syntax. Studies in second language acquisition [Online]. 32 (1), pp47-77. Available from: <http://dx.doi.org/10.1017/S0272263109990258> [Accessed 25 July 2011].

Rutherford, W.E. 1987. Second language grammar: learning and teaching. Michigan: Longman.

Ryle, G. 1946. Knowing How and Knowing That: The Presidential Address. Proceedings of the Aristotelian Society [Online]. 46 (1945-1946), pp1-16. Available from: <http://www.jstor.org/stable/4544405> [Accessed 8 February 2010].

Samuelis, L. 2007. Notes on the Components for Intelligent Tutoring Systems. Acta Polytechnica Hungarica [Online]. 4 (2), pp77-85. Available from: <http://bmf.hu/journal/Samuelis_10.pdf> [Accessed 20 February 2010].

Samuels, M.T. 2000. Assessment of post-secondary students with learning difficulties: using dynamic assessment in a problem solving process IN: Lidz, C.S. and Elliott, J.G. (eds.) Dynamic Assessment: Prevailing Models and Applications, Volume 6 (Advances in Cognition and Educational Practice). New York: JAI, an imprint of Elsevier Science.

Schmid, H. 1994. Probabilistic part-of-speech tagging using decision trees. IN: Proceedings of International Conference on New Methods in Language Processing. September 1994. Manchester, UK [Online]. Available from: <http://www.ims.uni-stuttgart.de/ftp/pub/corpora/tree-tagger1.pdf> [Accessed 10 July 2009].

Schneider, E. and Ganschow, L. 2000. Dynamic assessment and instructional strategies for learners who struggle to learn a foreign language. Dyslexia [Online]. 6 (1), pp72-82. Available from: <http://dx.doi.org/10.1002/(SICI)1099-0909(200001/03)6:1<72::AID-DYS162>3.0.CO;2-B> [Accessed 15 September 2009].

Schulze, M., Wood, P. and Pokorny, B. forthcoming. Measuring balanced complexity. [Online]. ppxx-xx. Available from: <http://wcgs.ca/~mschulze/papers/complexity.pdf> [Accessed 29 July 2010].

Selinker, L. 1969. Language transfer. General Linguistics. 9 (2), pp67-92.

Selinker, L. 1972. Interlanguage. IRAL - International Review of Applied Linguistics in Language Teaching. 10 (3), pp209-232.

Selinker, L. 1974. Interlanguage (reprinted from IRAL, Vol 10/3, 1972) IN: Richards, J.C. (ed.) Error analysis: perspectives on second language acquisition. Essex: Longman Group Limited.

Selinker, L. 1988. Papers in Interlanguage. Occasional Papers No. 44 (ERIC Document Reproduction Service No. ED 321 549). Singapore: Seameo Regional Kanguage Centre [Online]. Available from: <http://bit.ly/5AapWk> [Accessed 8 January 2010].

Selinker, L. 1992. Rediscovering Interlanguage. New York: Addison Wesley Longman.

Selinker, L. and Lamendella, J.T. 1981. Updating the interlanguage hypothesis. Studies in Second Language Acquisition [Online]. 3 (2), pp201-220. Available from: <http://dx.doi.org/10.1017/S0272263100004186> [Accessed 7 January 2010].

Shaalan, K.F. 2005. An Intelligent Computer Assisted Language Learning System for Arabic Learners. Computer Assisted Language Learning [Online]. 18 (1-2), pp81-109. Available from: <http://dx.doi.org/10.1080/09588220500132399> [Accessed 28 March 2010].

Skehan, P. 2009. Modelling Second Language Performance: Integrating Complexity, Accuracy, Fluency, and Lexis. Applied Linguistics [Online]. 30 (4), pp510-532. Available from: <http://dx.doi.org/10.1093/applin/amp047> [Accessed 14 July 2010].

Skinner, B. and Madden, M.C. 2010. Help seeking in English language learning. ELT Journal [Online]. 64 (1), pp21-31. Available from: <http://dx.doi.org/10.1093/elt/ccp019> [Accessed 25 March 2010].

Sleeman, D. 1984. An Attempt to Understand Students' Understanding of Basic Algebra. [Online]. 8 (4), pp387–412. Available from: <http://csjarchive.cogsci.rpi.edu/1984v08/i04/p0387p0412/MAIN.PDF> [Accessed 23 February 2010].

Snow, R.E. 1990. Progress and Propaganda in Learning Assessment (A Review of Carol Schneider Lidz's (Ed.) "Dynamic Assessment: An Interactional Approach to Evaluating Learning Potential"). PsycCritiques formely Contemporary Psychology: APA Review of Books [Online]. 35 (12), pp1134-1136. Available from: <http://dx.doi.org/10.1037/029204> [Accessed 15 November 2010].

Sternberg, R.J. 2005. Intelligence, Competence, and Expertise IN: Elliott, A.J. and Dweck, C.S. (eds.) Handbook of competence and motivation. New York: The Guilford Press.

Sternberg, R.J. and Grigorenko, E.L. 2002. Dynamic Testing: The Nature and Measurement of Learning Potential. Cambridge: Cambridge University Press.

Stetsenko, A.P. 2010. Standing on the shoulders of giants a balancing act of dialectically theorizing conceptual understanding on the grounds of Vygotsky's project IN: Roth, W.-M. (ed.) Re/Structuring Science Education: Reuniting Sociological and Psychological Perspectives. London: Springer.

Stiggins, R. 2008. Assessment Manifesto: A Call for the Development of Balanced Assessment Systems. ETS Assessment Training Institute [Online]. pp2-12. Available from: <http://www.nmsa.org/portals/0/pdf/advocacy/other_resources/ AssessmentManifesto08.pdf> [Accessed 21 October 2009].

Subbotsky, E. 1994. Vygotsky's Distinction between Lower and Higher Mental Functions and Recent Studies on Infant Cognitive Development. IN: L.S.Vygotsky and the Contemporary Human Sciences International Conference, September, 5-8, 1994. Moscow [Online]. Available from: <http://psych.hanover.edu/vygotsky/subbot.html> [Accessed 25 July 2009].

Szerdahelyi, J. 2010. Multimedia Technologies for Responding to Students' Writing Assignments in an Online Environment. IN: Paper presentation at the World Conference on Educational Multimedia, Hypermedia and Telecommunications (EDMEDIA), 28-JUN-10 - 02-JUL-10, Toronto, Canada, 2010. Chesapeake, VA: AACE.

Tarone, E. 1979. Interlanguage as Chameleon. Language Learning [Online]. 29 (1), pp181-191. Available from: <http://dx.doi.org/10.1111/j.1467-1770.1979.tb01058.x> [Accessed 14 October 2009].

Tarone, E. 1985. Variability in interlanguage use: a study of style-shifting in morphology and syntax. Language Learning. 35 (3), pp373-403.

Tarone, E. 1988. Variation in interlanguage. London: Edward Arnold.

Tarone, E. 1994. Variation in interlanguage IN: Asher, R.E. and Simpson, J.M.Y. (eds.) The Encyclopedia of Language and Linguistics (vol 4). Oxford: Pergamon Press.

Tarone, E. 2000. Still wrestling with 'context' in interlanguage theory. Annual Review of Applied Linguistics [Online]. 20 (1), pp182-198. Available from: <http://dx.doi.org/10.1017/S0267190500200111> [Accessed 27 January 2010].

Tarone, E. 2007. Sociolinguistic Approaches to Second Language Acquisition Research —1997-2007. The Modern Language Journal [Online]. 91 (5), pp837-848. Available from: <http://dx.doi.org/10.1111/j.1540-4781.2007.00672.x> [Accessed 10 February 2009].

Tarone, E. and Parrish, B. 1988. Task-Related Variation in Interlanguage: The Case of Articles. Language Learning [Online]. 38 (1), pp21-44. Available from: <http://dx.doi.org/10.1111/j.1467-1770.1988.tb00400.x> [Accessed 26 November 2008].

Taylor, G. 1986. Errors and Explanations. Applied Linguistics [Online]. 7 (2), pp144-166. Available from: <http://dx.doi.org/10.1093/applin/7.2.144> [Accessed 21 January 2010].

TEI. 2009. TEI P5: Guidelines for Electronic Text Encoding and Interchange. [Online]. Available from: <http://www.tei-c.org/release/doc/tei-p5-doc/en/Guidelines.pdf> [Accessed 16 July 2010].

Thorne, S.L. 2005. Epistemology, Politics, and Ethics in Sociocultural Theory. The Modern Language Journal [Online]. 89 (3), pp393-409. Available from: <http://dx.doi.org/10.1111/j.1540-4781.2005.00313.x> [Accessed 24 July 2009].

Tlili-Guiassa, Y. 2006. Hybrid Method for Tagging Arabic Text. Journal of Computer Science [Online]. 2 (3), pp245-248. Available from: <http://www.scipub.org/fulltext/jcs/jcs23245-248.pdf> [Accessed 27 March 2010].

Toulmin, S. 1978. The Mozart of Psychology. New York review of books.

Truscott, J. 2004. Evidence and conjecture on the effects of correction: A response to Chandler. Journal of Second Language Writing [Online]. 13 (4), pp337-343. Available from: <http://dx.doi.org/10.1016/j.jslw.2004.05.002> [Accessed 02 July 2009].

Twidale, M. et al. 1992. Deep-knowledge acquisition for learner modelling in second language learning IN: Cerri, S.A. and Whiting, J. (eds.) Learning Technology in the European Communities. Dordrecht: Kluwer [Online]. Available from: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.49.895&rep=rep1&type=pdf> [Accessed 3 March 2010].

Valsiner, J. 2001. Process Structure of Semiotic Mediation in Human Development. Human Development [Online]. 44 (2/3), pp84-97. Available from: <http://dx.doi.org/10.1159/000057048> [Accessed 15 September 2009].

Van Compernolle, R.A. and Williams, L. 2009. Learner versus nonlearner patterns of stylistic variation in synchronous computer-mediated French. Studies in Second Language Acquisition [Online]. 31 (03), pp471-500. Available from: <http://dx.doi.org/10.1017/S0272263109090378> [Accessed 27 August 2009].

Van der Aalsvoort, G.M. and Lidz, C.S. 2002. Reciprocity in dynamic assessment in classrooms: taking contextual influences on individual learning into account IN: Van der Aalsvoort, G.M., Riesing, W.C.M. and Ruijssenaars, A.J.J.M. (eds.) Learning potential assessment and cognitive training. Amsterdam: Elsevier.

Van der Linden, E. 1993. Does feedback enhance computer-assisted language learning? Computers & Education [Online]. 21 (1-2), pp61-65. Available from: <http://dx.doi.org/10.1016/0360-1315(93)90048-N> [Accessed 1 April 2010].

Van der Veer, R. and Valsiner, J. 1991. Understanding Vygotsky: a quest for synthesis. Cambridge, MA: Blackwell Publishers.

VanLehn, K. 1988. Student Modeling IN: Polson, M.C. and Richardson, J.J. (eds.) Foundations of intelligent tutoring systems. Hillsdale, NJ: Lawrence Erlbaum Associates.

Vann, R.J., Meyer, D.E. and Lorenz, F.O. 1984. Error Gravity: A Study of Faculty Opinion of ESL Errors. TESOL Quarterly [Online]. 18 (3), pp427-440. Available from: <http://www.jstor.org/stable/3586713> [Accessed 21 January 2010].

VanPatten, B. 1992. Second-Language-Acquisition Research and Foreign Language Teaching, Part 1. ADFL Bulletin [Online]. 23 (2), pp52-55. Available from: <http://web2.adfl.org/adfl/bulletin/v23n2/232052.htm> [Accessed 22 January 2010].

Vanroose, P. 2001. Part-of-speech tagging from an information-theoretic point of view. IN: 22nd Symposium on Information Theory in the Benelux.

Veresov, N.N. 2005. Marxist and non-Marxist aspects of the cultural-historical psychology of L. S. Vygotsky. Critical Social Studies - Outlines [Online]. 7 (1), pp31-49. Available from: <http://www.outlines.dk/contents/Outlines051/Veresov05.pdf> [Accessed 21 July 2009].

Verspoor, M., Lowie, W. and Van Dijk, M. 2008. Variability in Second Language Development From a Dynamic Systems Perspective. The Modern Language Journal. 92 (2), pp214-231.

Virvou, M., Maras, D. and Tsiriga, V. 2000. Student Modelling in an Intelligent Tutoring System for the Passive Voice of English Language. Educational Technology & Society [Online]. 3 (4), pp139-150. Available from: <http://www.ifets.info/download_pdf.php?j_id=16&a_id=299> [Accessed 28 March 2010].

Voutilainen, A. 2003. Part-of-speech tagging IN: Mitkov, R. (ed.) The Oxford handbook of computational linguistics. Oxford: University Press.

Vygotsky, L.S. 1978. Mind in Society: The Development of Higher Psychological Processes. Cambridge, MA: Harvard University Press.

Vygotsky, L.S. 1987. The development of scientific concepts in childhood IN: Rieber, R.W., Carton, A.S. and Minick, N. (eds.) The Collected Works of L.S. Vygotsky: Volume 1: Problems of General Psychology, Including the Volume Thinking and Speech. New York: Plenum.

Vygotsky, L.S. 1997a. The problem of consciousness IN: Rieber, R.W. and Wollock, J. (eds.) The Collected Works of L.S. Vygotsky: Volume 3: Problems of the Theory and History of Psychology. New York: Plenum.

Vygotsky, L.S. 1997b. Genesis of higher mental functions IN: Rieber, R.W. (ed.) Collected works of L. S. Vygotsky: Volume 4: The history of the development of higher mental functions. New York: Plenum. (Original work published 1960).

Vygotsky, L.S. 1998. The problem of age IN: Rieber, R.W. (ed.) The Collected Works of L.S. Vygotsky: Volume 5: Child Psychology. New York: Plenum.

Wardhaugh, R. 2006. An introduction to sociolinguistics. 5th ed. Malden, MA: Blackwell Publishing.

Wehlburg, C.M. 2011. A scholarly approach to assessing learning (Invited Essays on SoTL). International Journal for the Scholarship of Teaching and Learning [Online]. 5 (2), pp1-4. Available from: <http://academics.georgiasouthern.edu/ijsotl/v5n2/invited_essays/PDFs/Wehlburg.pdf> [Accessed 28 July 2011].

Wei, F. et al. 2005. A Student Model for Object-Oriented Design and Programming. IN: CCSCNE, the tenth annual Consortium for Computing Sciences in Colleges Northeastern Conference, Providence College, April 22 - 23, 2005. Providence, RI [Online]. Available from: <http://www.cse.lehigh.edu/~cimel/papers/CCSCNE05.pdf> [Accessed 24 February 2010].

Wells, G. 1999. Dialogic inquiry: towards a sociocultural practice and theory of education. Cambridge: Cambridge University Press.

Wertsch, J.V. 1998. Mind as Action. New York: Oxford University Press.

Wertsch, J.V. 2007. Mediation IN: Daniels, H., Cole, M. and Wertsch, J.V. (eds.) The Cambridge Companion to Vygotsky. Cambridge: Cambridge University Press.

White, L. 2003. Second language acquisition and universal grammar. West Sussex, UK: Cambridge University Press.

Wilkins, D.A. 1972. Linguistics in language teaching. London: Edward Arnold.

Windschitl, M. and Sahl, K. 2002. Tracing teachers' use of technology in a laptop computer school: the interplay of teacher beliefs, social dynamics, and institutional culture. American Educational Research Journal. 39 (1), pp165-205.

Wolfe-Quintero, K., Inagaki, S. and Kim, H.-Y. 1998. Second Language Development in Writing: Measures of Fluency, Accuracy and Complexity. Honolulu: Second Language Teaching and Curriculum Center, University of Hawai'i Press.

Wood, D., Bruner, J.S. and Ross, G. 1976. The role of tutoring in problem solving. Journal of Child Psychology & Psychiatry & Allied Disciplines [Online]. 17 (2), pp89-100. Available from: <http://tinyurl.com/m5fmbs> [Accessed 7 September 2009].

Yan, Z. and Fischer, K. 2002. Always Under Construction. Human Development [Online]. 45 (3), pp141-160. Available from: <http://dx.doi.org/10.1159/000057070> [Accessed 27 August 2009].

Yuan, F. and Ellis, R. 2003. The Effects of Pre-Task Planning and On-Line Planning on Fluency, Complexity and Accuracy in L2 Monologic Oral Production. Applied Linguistics [Online]. 24 (1), pp1-27. Available from: <http://dx.doi.org/10.1093/applin/24.1.1> [Accessed 27 November 2009].

Ziefle, M. 1998. Effects of display resolution on visual performance. Human Factors [Online]. 40 (4), pp554-568. Available from: <http://dx.doi.org/10.1518/001872098779649355> [Accessed 29 July 2010].

Appendix A. Ethical considerations

A.1. Plain language statement

DCU

Helping language teachers provide better feedback to students

Sylvie Thouësny
School of Applied Language and Intercultural Studies
sylvie.thouesny2@mail.dcu.ie

Plain language statement

As part of my Ph.D. research, I am designing a program that will hopefully help language teachers correct assignments and provide useful feedback to students. For this program to be reliable and useful, I need to investigate whether it is possible to ask computers to differentiate between errors or mistakes when analyzing incorrect forms in a written document. While errors represent gaps in a language learner's knowledge, i.e. an aspect of the language that is not acquired, mistakes, on the other hand, denote occasional lapses in performance, i.e. an aspect of the language that is almost mastered, however, not systematically applied. With your help, I hope to collect some data that will help me test the program I am designing. More specifically, I need some texts in the foreign language by students before and after the teacher's correction throughout the year.

Your participation in this study will be greatly appreciated. Your decision to participate is strictly voluntary and will in no way affect your grade: none of the teachers involved in this course will have access to your errors/mistakes you may have made and they will not be communicated to them. And, if at any point you change your mind, you are free to withdraw from this study. All of the information collected will be confidential. This means that your identity will be kept anonymous.

If you agree to take part in this study, you will be given, via Moodle, a link to exercise pages. You will be invited to submit a copy of your written texts. Upon receipt of some feedback from me, you will be asked to correct them three times. For the first correction, you will be asked to correct all errors highlighted without any help. With the second correction exercise, you will be provided with comments that will indicate information about error types, and this, for each incorrect form. For the third exercise, you will be given detailed feedback on each incorrect form. As for the first exercise, you will be asked to correct all errors highlighted. Each correction will be performed in a restrained time period of roughly 10 minutes. A full statement of your errors and their correction, if any, will be sent to you after the completion of the exercises.

Your text and their correction (by myself and by yourself) will be processed by the computer, with a view to determine which language forms you have acquired and which ones are not yet fully mastered. It is thus expected that the findings will help differentiate between errors and mistakes automatically. In time, this will help your lecturers determine whether you need help in a particular area, thus enabling them to provide you with more useful feedback.

Anonymised extracts of your written productions may be presented at conferences or published in articles or reports. Should you wish to see the outcomes of this project, these will be made available to you.

Sylvie Thouësny

Sylvie Thouësny

November 2008

If you have concerns about this study and wish to contact an independent person, please contact:

The Secretary, Dublin City University Research Ethics Committee, c/o Office of the Vice-President for Research, Dublin City University, Dublin 9. Tel 01-7008000.

A.2. Informed consent form

DCU

Helping language teachers provide better feedback to students

Sylvie Thouësny
School of Applied Language and Intercultural Studies
sylvie.thouesny2@mail.dcu.ie

INFORMED CONSENT FORM

An in-depth evaluation of a program, designed to help language teachers provide useful feedback to students when submitting assignments electronically, is being conducted. The results will determine whether it is feasible for a computer to distinguish errors from mistakes in written assignments. If it is, your teacher will be able to target more precisely your needs in language learning. The evaluation process will consist of collecting your assignments written in foreign language throughout the year. You will be invited to correct your assignments during a lab session. Access to the data will be restricted to yourself and myself via Moodle.

Participation in this study is strictly voluntary and your decision will have no effect on your grade. If you decide to take part you are still free to withdraw at any time and without giving a reason. All of the information collected will be confidential. This means that your identity will be protected. Whenever sample data, drawn from your texts in this study, are published or presented via reports, journal articles, or conference presentations, your name will not be used.

Your participation is, therefore, being considered highly valuable and vital to the success of this study. I would greatly appreciate if you could answer the following questions (please circle Yes or No for each question) and add your signature at the bottom of this consent form along with the date and the name of your witness:

Have you read or had read to you the Plain Language Statement	*Yes*	*No*
Do you understand the information provided?	*Yes*	*No*
Have you had an opportunity to ask questions and discuss this study?	*Yes*	*No*
Have you received satisfactory answers to all your questions?	*Yes*	*No*
Do you give permission to use some of your language production?	*Yes*	*No*

I have read and understood the information in this form. My questions and concerns have been answered by the researcher, and I have a copy of this consent form. Therefore, I consent to take part in this research project.

Participants Signature: ______________________

Name in Block Capitals: ______________________

Witness: ______________________

Date: ______________________

1

Appendix B. Regulatory scale[66]

0. Tutor asks the learner to read, find the errors, and correct them independently, prior to the tutorial.

1. Construction of a “collaborative frame” prompted by the presence of the tutor as a potential dialogic partner.

2. Prompted or focused reading of the sentence that contains the error by the learner or the tutor.

3. Tutor indicates that something may be wrong in a segment (e.g., sentence, clause, line) – “Is there anything wrong in this sentence?”

4. Tutor rejects unsuccessful attempts at recognising the error.

5. Tutor narrows down the location of the error (e.g., tutor repeats or points to the specific segment which contains the error).

6. Tutor indicates the nature of the error, but does not identify the error (e.g., “There is something wrong with the tense mark here”).

7. Tutor identifies the error (“You can’t use an auxiliary here”).

8. Tutor rejects learner’s unsuccessful attempts at correcting the error.

9. Tutor provides clues to help the learner arrive at the correct form (e.g., “It is not really past but some thing that is still going on”).

10. Tutor provides the correct form.

11. Tutor provides some explanations for use of the correct form.

12. Tutor provides examples of the correct pattern when other forms of help fail to produce an appropriate responsive action.

66. Aljaafreh and Lantolf’s (1994 p.470) regulatory scale

Appendix C. Tagging process

C.1. Regular expression

The caret ^ and dollars $ symbols indicate the start and the end of a string. For example, the following regular expression ^abc$ matches any string that starts and ends with abc. The caret ^ symbol, when inside square brackets [ ] negates the whole expression. For example [^abc] means anything except a, b or c. The vertical bar | also called the alternation operator or pipe character is used to specify that several subexpressions can occur in a specific position in the expression. The regular expression ab|cd matches a string that has either ab or cd in it. The asterik * indicates zero or more occurrences of the preceding character. The plus + sign signals one or more occurrences of the preceding character and the question mark ? zero or 1 occurrence of the preceding character. Finally, the left and right parentheses ( ) are used for grouping parts of the sequences.

For example, a rule such as: ^(VER:conj)$->^(NOM)$/^(DET)__ signifies that any token tagged as a conjugated verb (VER:conj) placed after any type of determiner (e.g., demonstrative or possessive) (DET) will be now considered as a noun (NOM).

C.2. Rules to overcome consistent errors in tagging process

#rule	Description
#1	^(VER)->^(VER:conj)$/__ if ^(VER)!=^(VER:infi\|VER:pper\|VER:ppre)$
#2	^(NUM)$->^(ADJ)$/__ if token=[^(.*\d.*)] AND if token=(i(è|e|é)re*s*|i(è|e|é)mes*)$;
#3	ELSE ^(NUM)$->^(DET:ART)$/__ if token=[^(.*\d.*)] AND if token=^(une|un)$;
#4	ELSE ^(NUM)$->^(ADJ:NUM)$/__ if token=[^(.*\d.*)] AND if token!=^(c|m|x|i|l)$;
#5	^(VER:pper)$->^(ADJ)$/^(NOM)$__;
#6	ELSE ^(VER:pper)$->^(ADJ)$/^(ADJ) ^(PUN)__;
#7	ELSE ^(VER:pper)$->^(ADJ)$/^(ADJ) ^(KON)$__;
#8	ELSE ^(VER:pper)$->^(VER:conj)$/[^(VER:conj)] ^(PRO:PER)$__;

#11	^(PRO:IND)$->^(ADJ:IND)$/__ if token=^(chaque|maint([e|es])?)$;
#12	ELSE ^(PRO:IND)$->^(ADJ:IND)$/__^(NOM)$;
#13	ELSE ^(PRO:IND)$->^(ADJ:IND)$/__^(ADJ);
#14	ELSE ^(PRO:IND)$->^(ADJ:IND)$/__^(DET);
#19	^(ADJ)$->^(NOM)$/^(DET)__^(VER);
#43	ELSE ^(ADJ)$->^(ADV)$/^(DET)__^(peu)$;
#34	ELSE ^(ADJ)$->^(NOM)$/^(DET)__^(ADV)$;
#23	^(PRO)->^(ADJ)$/^(DET)__^(NOM)$;
#31	^(VER:conj)$->^(NOM)$/^(en)$__^(ADJ)$;
#33	ELSE ^(VER:conj)$->^(NOM)$/^(en)$__ if token=^(fait)$;
#24	ELSE ^(VER:conj)$->^(VER:ppre)$/[^(PRO:PER|NOM)] ^(en)$__;
#32	^(VER:conj)$->^(NOM)$/^(DET)__;
#25	^(PRO:REL)$->^(KON)$/^(KON)$__;
#26	ELSE ^(PRO:REL)$->^(KON)$/^(PRP)$__;
#46	ELSE ^(PRO:REL)$->^(KON)$/lemma ^(espérer)$__;
#9	^(PRO:DEM)$->^(DET:DEM)$/__^(NOM)$ if token=^(ce|ces|cet|cette)$;
#22	ELSE ^(PRO:DEM)$->^(DET:DEM)$/__^(PUN:cit)$ if token=^(ce|ces|cet|cette)$;
#10	ELSE ^(PRO:DEM)$->^(DET:DEM)$/__^(ADJ|NUM) if token=^(ce|ces|cet|cette)$;
#15	token ^(y)$->^(ADV)$/ ^(il|n)$\|^(SENT\|PUN)__^(a\|à\|á\|å\|avait\|aurait\|aura)$; ELSE token ^(y)$->^(ADV)$/^(ca|ça)$__^(est)$; ELSE token ^(y)$->^(ADV)$/__ lemma ^(connaître|entendre|prendre|tenir)$;
#16	tokens ^(ou|et|donc|ni)$->^(KON)$/__;
#17	token ^(pour)$->^(NOM)$/^(DET)__;
#20	ELSE token ^(pour)$->^(PRP)$/__;
#18	token ^(plein)$->^(ADV)$/^(tou(s|t))$__^(de\|des\|du\|d\')$;
#21	token ^(.*\d.*)$->^(NUM)$/__ if pos!=^(NUM|ABR)$;
#27	token ^(rien)$->^(PRO:IND)$/[^(DET)]__;
#28	token ^(apr(è|e|é)s)$->^(PRP)$/[^(DET)]__ if pos!=^(PRP|KON|ADV)$;
#38	ELSE token ^(apr(è|e|é)s)$->^(ADV)$/[^(DET)]__ if pos=^(KON)$;
#29	token ^(h(a|à|á|o|ô|é|è|e|i))$->^(INT)$/__;
#30	token ^(quel(|le|les|s))$->^(ADJ:IND)$/__ if pos!=^(ADJ);
#35	lemma ^(français|anglais|espagnol|irlandais)$->^(NOM)$/ ^(en|du|de|le|la|l\'|d\'|un)$__;
#42	ELSE lemma ^(français|anglais|espagnol|irlandais)$->^(NOM)$/lemma ^(parler|avoir|étudier)$__;

<table>
<tr><td>#36</td><td>token ^(des)$->^(DET:ART)$/
[^(NOM|PRO:DEM|PRO:POS|SYM|DET:ART)]__
if pos=^(PRP:det)$ AND
if previous lemma!=^(beaucoup|plein)$;</td></tr>
<tr><td>#41</td><td>ELSE token ^(des)$->^(PRP:det)$
^(NOM|PRO:DEM|PRO:POS|SYM|DET:ART)$__;</td></tr>
<tr><td>#44</td><td>ELSE token ^(des)$->^(PRP)$/^(beaucoup|plein)$__;</td></tr>
<tr><td>#37</td><td>token ^(neuf)$->^(ADJ:NUM)$/__^(heur(|e|s|es))$;</td></tr>
<tr><td>#39</td><td>lemma ^(intéresser)$->^(ADJ)$/[^(en)]__
if pos=^(VER:ppre)$;</td></tr>
<tr><td>#40</td><td>token ^(s\')$->^(KON)$/__^(il)$;</td></tr>
<tr><td>#45</td><td>token ^(importe)$->^(PRO:IND)$/__^(quoi)$;</td></tr>
</table>

C.3. TreeTagger French tag set

Tag code	Description	Tag code	Description
ABR	abbreviation	PRP:det	preposition plus article
ADJ	adjective	PUN	punctuation
ADV	adverb	PUN:cit	punctuation citation
DET:ART	article	SENT	sentence tag
DET:POS	possessive determinant	SYM	symbol
INT	interjection	VER:cond	verb conditional
KON	conjunction	VER:futu	verb futur
NAM	proper name	VER:impe	verb imperative
NOM	noun	VER:impf	verb imperfect
NUM	numeral	VER:infi	verb infinitive
PRO	pronoun	VER:pper	verb past participle
PRO:DEM	demonstrative pronoun	VER:ppre	verb present participle
PRO:IND	indefinite pronoun	VER:pres	verb present
PRO:PER	personal pronoun	VER:simp	verb simple past
PRO:POS	possessive pronoun	VER:subi	verb subjunctive imperfect
PRO:REL	relative pronoun	VER:subp	verb subjunctive present
PRP	preposition		

C.4. Updated French tag set

Tag code	Description	Tag code	Description
ABR	abbreviation	PRO:IND	indefinite pronoun
ADJ	adjective	PRO:PER	personal pronoun
ADJ:IND*	indefinite adjective	PRO:POS	possessive pronoun
ADJ:NUM*	numeral adjective	PRO:REL	relative pronoun
ADV	adverb	PRP	preposition
DET:ART	article	PRP:det	preposition plus article
DET:POS	possessive determinant	PUN	punctuation
DET:DEM*	demonstrative determinant	PUN:cit	punctuation citation
INT	interjection	SENT	sentence tag

KON	conjunction	SYM	symbol
NAM	proper name	VER:conj*	conjugated verb
NOM	noun	VER:infi	verb infinitive
NUM	numeral	VER:pper	verb past participle
PRO	pronoun	VER:ppre	verb present participle
PRO:DEM	demonstrative pronoun	*	new tags

C.5. Cross-referencing rules

<table>
<tr><th>#cr</th><th>Description</th></tr>
<tr><td></td><td>if error type=^(mo_fo_co_)$ - (incorrect conjugation form)</td></tr>
<tr><td>#9</td><td>[^(VER:pper)]->^(VER:pper)$/^(VER:conj|VER:conj ADV)$__;</td></tr>
<tr><td>#7</td><td>ELSE [^(VER:conj)]->^(VER:conj)$/[^(VER:conj)]__;</td></tr>
<tr><td>#27</td><td>ELSE [^(VER:pper)]->^(VER:pper)$/^(VER:inf)$__;</td></tr>
<tr><td></td><td>if error type=^(se_gr_mo_)$ - (incorrect mood)</td></tr>
<tr><td>#8</td><td>[^(VER:ppre)]->^(VER:ppre)$/lemma ^(en)$__;</td></tr>
<tr><td>#4</td><td>ELSE [^(VER:infi)]->^(VER:infi)$/^(PRP)$__;</td></tr>
<tr><td>#6</td><td>ELSE [^(VER:infi)]->^(VER:infi)$/lemma ^(vouloir|pouvoir|aller|faire|aimer)$__;</td></tr>
<tr><td>#19</td><td>ELSE [^(VER:infi)]->^(VER:infi)$/lemma ^(vouloir|pouvoir|aller|faire|aimer)$ ^(PRO:PER)$__;</td></tr>
<tr><td>#28</td><td>ELSE ^(VER:infi)->^(VER:conj)$/__;</td></tr>
<tr><td>#5</td><td>ELSE [^(VER|PRP)]->^(VER:conj)$/__;</td></tr>
<tr><td></td><td>if error type=^(se_gr_te_)$ - (incorrect tense)</td></tr>
<tr><td>#38</td><td>[^(ADV|VER:conj)]->^(VER:conj)$/[^(VER:conj)] [^(VER:conj)]__;</td></tr>
<tr><td>#39</td><td>ELSE [^(ADV|VER:pper|VER:inf)]->^(VER:pper)$/^(VER:conj ADV|VER:conj)$__;</td></tr>
<tr><td></td><td>if error type=^(se_gr_pr_)$ - (incorrect preposition)</td></tr>
<tr><td>#10</td><td>[^(PRP)]->^(PRP)$/[^(PRP)]__;</td></tr>
<tr><td></td><td>if error type=^(se_gr_rv_)$ - (reflexive verb)</td></tr>
<tr><td>#21</td><td>[^(VER:conj)]->^(VER:conj)$/^(PRO)__;</td></tr>
<tr><td>#22</td><td>ELSE [^(VER:inf)]->^(VER:inf)$/^(VER:conj)__;</td></tr>
<tr><td></td><td>if error type=^(se_gr_ar_)$ - (incorrect article type)</td></tr>
<tr><td>#11</td><td>^(PRO:POS|PRO:DEM)$->^(DET:POS|DET:DEM)$/__;</td></tr>
<tr><td>#14</td><td>ELSE [^(DET)]->^(DET:ART)$/^(PRP)$__;</td></tr>
<tr><td>#32</td><td>ELSE [^(PRP|PRP:det)]->^(PRP)$/__^(DET);</td></tr>
<tr><td>#33</td><td>ELSE ^(PRP)$->^(PRP:det)$/^(NOM)$__[^(DET)];</td></tr>
<tr><td>#12</td><td>ELSE [^(DET|PUN|PRP:det)]->^(DET:ART)$/ [^(NOM|ADJ)]__[^(DET)];</td></tr>
<tr><td>#44</td><td>ELSE [^(DET|PUN|PRP:det)]->^(PRP:det)$/__[^(DET)];</td></tr>
<tr><td></td><td>if error type=^(se_gr_gw_)$ - (incorrect connection word)</td></tr>
<tr><td>#13</td><td>^(PRO:REL)$->^(KON)$/^(VER:conj)$__;</td></tr>
<tr><td></td><td>if error type=^(mo_ag_dn_)$ - (determinant noun agreement)</td></tr>
<tr><td>#31</td><td>^(PRP)$->^(PRP:det)$/__;</td></tr>
<tr><td>#35</td><td>ELSE ^(VER)$->^(NOM)$/__;</td></tr>
</table>

<table>
<tr><td></td><td>if error type=^(mo_ag_na_)$ - (noun adjective agreement)</td></tr>
<tr><td>#23</td><td>[^(ADJ|DET|PRP)]->^(ADJ)$/__;</td></tr>
<tr><td></td><td>if error type=^(mo_ag_pa_)$ - (pronoun antecedent agreement)</td></tr>
<tr><td>#29</td><td>^(DET:POS|DET:DEM)$->^(PRO:POS|PRO:DEM)$/__;</td></tr>
<tr><td></td><td>if error type=^(mo_ag_sv_)$ - (subject verb agreement)</td></tr>
<tr><td>#37</td><td>^(NOM)$->^(VER:conj)$/^(PRO:REL)$__;</td></tr>
<tr><td></td><td>if error type=^(mo_ag_pp_)$ - (past participle agreement)</td></tr>
<tr><td>#20</td><td>[^(VER:pper)]->^(VER:pper)$/__;</td></tr>
<tr><td></td><td>if error type=^(mo_fo_wf_)$ - (incorrect word formation)</td></tr>
<tr><td>#34</td><td>token ^(sa|se)$->^(PRO:DEM)$/__;</td></tr>
<tr><td>#41</td><td>ELSE token ^(ca|ça)$->^(DET:POS)$/__;</td></tr>
<tr><td>#43</td><td>ELSE token ^(c\')$->^(PRO:PER)$/__;</td></tr>
<tr><td></td><td>if error type=^(sp_ac_)$ - (incorrect or missing accent)</td></tr>
<tr><td>#1</td><td>token ^(a|á)$->^(PRP)$/__;</td></tr>
<tr><td>#2</td><td>ELSE token ^(ou)$->^(PRO:REL)$/__;</td></tr>
<tr><td>#3</td><td>ELSE token ^(o(ù|ú))$->^(KON)$/__;</td></tr>
<tr><td>#18</td><td>ELSE token ^((Â|A|a|â|à|á)pr(e|é|è|ê)s)$->^(PRP)$/[^(DET)]__;</td></tr>
<tr><td>#24</td><td>ELSE token ^(la)$->^(ADV)$/__;</td></tr>
<tr><td>#30</td><td>ELSE token ^(s(u|ù|ú|û)r)$->^(ADV)$/^(bien)$__;</td></tr>
<tr><td>#42</td><td>ELSE token ^(sur)$->^(ADJ)$/^(bien)$__;</td></tr>
<tr><td></td><td>if error type=^(ty_ca_)$ - (capitalisation)</td></tr>
<tr><td>#26</td><td>token ^(NAM)$->^(NOM)$/__;</td></tr>
<tr><td></td><td>if error type=^(sp_hy_)$ - (hyphen)</td></tr>
<tr><td>#16</td><td>token ^(peut)$->^(ADV)$/__ token ^((ê|e)tre)$;</td></tr>
<tr><td></td><td>if error type=^(ty_ab_)$ - (abbreviation)</td></tr>
<tr><td>#15</td><td>token [^(ABR|NUM|PUN|SENT)]->^(ABR)$/__;</td></tr>
<tr><td></td><td>if error type=^(sp_ms_)$ - (spelling)</td></tr>
<tr><td>#36</td><td>token ^(peut)$->^(ADV)$/__;</td></tr>
</table>

C.6. Krippendorff's alpha coefficient

A strong argument in favour of Krippendorff's alpha coefficient is that the alpha algorithm is able to calculate the reliability coefficient regardless of missing data, sample size, number of categories or coders. To make the alpha coefficient practical and easily usable, Hayes and Krippendorff (2007) describe and illustrate how to use their macro developed for SPSS, also called KALPHA[67]

67. The KALPHA macro has been written by Andrew F. Hayes to promote the use of the Krippendorff' alpha coefficient since most statistical software packages do not include reliability coefficient. For an SPSS or SAS version of the script, and further explanations about how to execute the macro command, visit Andrew F. Hayes' website http://www.comm.ohio-state.edu/ahayes/.

macro. The agreement measure between two raters is generally presented under the following formulation:

$$\alpha = 1 - \frac{D_o}{D_e}$$

where D_o is the observed agreement and D_e is the agreement expected when the coding is attributable to chance.

C.7. Cohen's kappa coefficient

The calculation of the kappa coefficient differs from Krippendorff's alpha in the sense that the coefficient calculates the expected agreement based on the preferences of the raters rather than on the coding process. The formula is as follows:

$$\kappa = \frac{P_o - P_c}{1 - P_c}$$

where P_o is "the proportion of units in which the judges agreed", and P_c is "the proportion of units for which agreement is expected by chance" (Cohen 1960 p.39).

C.8. Recall, precision and F-measure

Whereas the recall measure computes the percentage of agreement by dividing the number of correct part-of-speech tags in TreeTagger by the number of correct part-of-speech tags in the Gold data, the precision measure calculates the percentage of agreement by dividing the number of correct part-of-speech tags in TreeTagger by the total number of part-of-speech tags in TreeTagger.

$$recall = \frac{number\ of\ correct\ pos\ tags\ in\ TreeTagger}{number\ of\ correct\ pos\ tags\ in\ Gold\ Data}$$

$$precision = \frac{number\ of\ correct\ pos\ tags\ in\ TreeTagger}{number\ of\ total\ pos\ tags\ in\ TreeTagger}$$

High scores in recall and precision may be considered as valuable only if the two measures are approximately the same high values. If the span between both results is too wide, for example 55% in recall and 98% in precision, both figures are combined together to normalize the compromise between recall and precision, which is called F-measure (Hull and Gomez 2002).

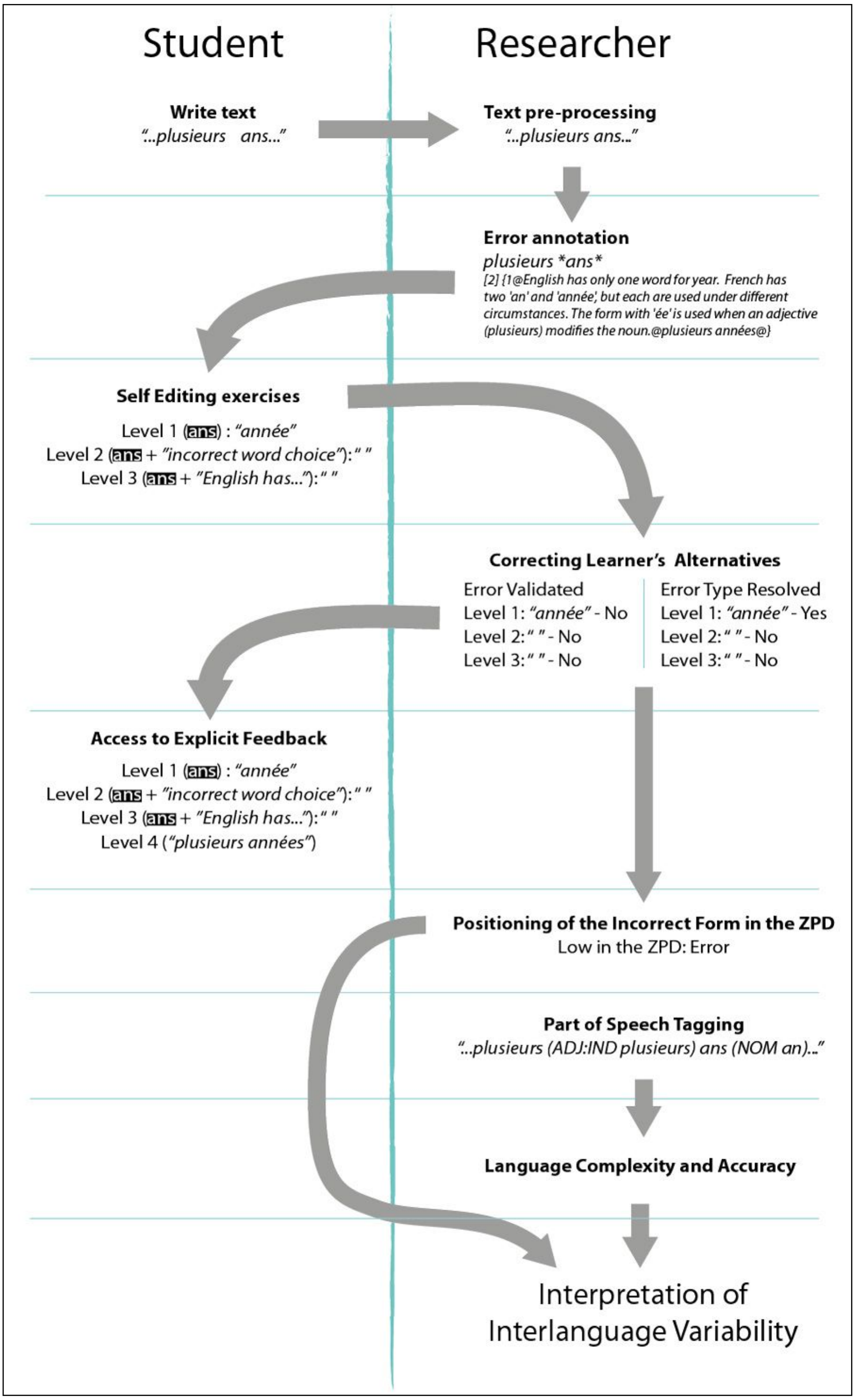
$$F - measure = \frac{(\beta + 1.0)(precision)(recall)}{\beta(precision) + (recall)}$$

where β gives varying weights to recall and precision measures.

Whereas $\beta = 1$ equally considers recall and precision, a β value of 0.5 put the accent on precision and a β value of 2 emphasises the recall perspectives.

Appendix D. Corpus

D.1. Sample text in full life-cycle



D.2. Text types and their database identification

	Bilan	Ecrire	Email	Forum	Parler	Survey	Wiki
id:	2	5	3	1	8	9	6

D.3. Error types and their database identification

Error types	Id.	Error types	Id.	Error types	Id.	Error types	Id.
se_vo_mu_	1	se_gr_vo_	10	mo_ag_dn	18	sp_hy_	28
se_vo_lw_	2	se_vo_re_	33	mo_ag_na_	19	sp_ms_	29
se_gr_cw_	3	se_vo_sn_	34	mo_ag_pa_	20	sp_ac_	26
se_gr_rv_	4	se_gr_cl_	35	mo_ag_pp_	21	ty_mp_	11
se_gr_ar_	5	sy_ad_	14	mo_ag_sv_	22	ty_wp_	12
se_gr_pr_	6	sy_om_	15	mo_fo_pl_	23	ty_es_	13
se_gr_ge_	7	sy_wo_	16	mo_fo_co_	24	ty_ca_	27
se_gr_mo_	8	sy_un_	17	mo_fo_wf_	25	ty_ab_	31
se_gr_te_	9						

D.4. Pre-determined meta-linguistic feedback

Error type	Description
se_vo_mu_ (nonstandard word or expression)	• What do you mean? • This word looks like a frenchified English word. • English word. • Not understandable.
se_vo_lw_ (incorrect word choice)	• English has only one word for year. French has two 'an' and 'année', but each are used under different circumstances. The form with 'ée' is usually used when an adjective modifies the noun. • English has only one word for day. French has two 'jour' and 'journée', but each are used under different circumstances. The form with 'ée' is usually used when an adjective modifies the noun. • The pronoun 'il' cannot usually be used to refer to events or actions. While 'ce' is normally used with 'être', 'cela' and 'ça' are used with other verbs. 'Cela' tends to be used in written French whereas 'ça' is widely used in the spoken language. • Word for word translation. • Inappropriate word in context. • Climatic conditions can be expressed by an impersonal use of 'faire' followed by an adjective or a noun. • This verb only exists in an impersonal form, which means that it only takes the pronoun 'il' as subject.

se_gr_cw_ (inappropriate connection word choice)	• When the non-specific head is understood to be non-human, 'ce qui' and 'ce que' are used. 'ce qui' is understood as the subject of the relative clause and 'ce que' as the object. • 'Qui' is the relative pronoun used when the noun phrase heading a relative clause is the implied subject of that relative clause. 'Que' is used when the noun phrase heading the relative clause is the implied object of the relative clause. • The head noun phrase is understood as the object in the relative clause, not the subject. • The head noun phrase is understood as the subject in the relative clause, not the object. • Comparative construction. • The construction 'ce que' is used with a non-specific head, which is not the case here. Your noun phrase is understood as the implied object of the relative clause. • Subordinating conjunctions introduce a clause which is dependent on another clause, not prepositions. • Coordinating conjunctions never introduce clauses in which the verb is in the subjunctive, whereas a number of subordinating conjunctions do.
se_gr_rv_ (reflexive verb)	• Reflexive pronouns agree in person with the subject of the verb. • This verb is being used reflexively, which means that you need a reflexive pronoun that refers to the subject of the verb.
se_gr_mo_ (incorrect mood)	• Verbs of wishing, ordering, expressing emotional states are normally followed by subjunctive subordinate clauses. • Infinitive clause introduced by a preposition. • This subordinating conjunction does not introduce hypothetical situations and, therefore, does not require the subjunctive in the subordinate clause. • Verbs of obligation, necessity and possibility (modal verbs) take infinitive complements without a preceding preposition. • Common idiom in which the preposition is 'à' followed by an infinitive complement. • Where coordinating conjunctions link two clauses, the verb in the second clause is always in the indicative. • Movement verbs such as 'aller' which do not have objects typically take infinitive complements (without a preceding preposition). • When the subject in the main clause expresses a belief in the relative probability of an event's occurring (whether in the past or the future), the indicative will be used. • The sentence does not include any conjugated verb. • When referring to events that would take place in the future if certain condition are met, we use the conditional tense. • This subordinating conjunction "jusqu'à ce que" is always followed by the subjunctive. • Verbs expressing personal attitude to something, such as 'aimer', typically take infinitive complements without a preceding preposition. • Verbs of obligation, necessity, and possibility take infinitive complements without a preceding preposition. • Imperatives in French are used very much as they are in English to give advice or orders, to express encouragement.

se_gr_ar_ (incorrect article type)	• Use the definite article to refer to a general class of phenomena. • Definite articles accompany nouns which already known from the context. • Definite articles fuse with preceding 'de' or 'à'. • A plural indefinite article is required since the sentence refers to an unspecified quantity of entities described by a plural count noun. • Do not confuse pronouns with possessive determiners. • Use the fused form of the preposition "de" and the definite article "les". • Definite articles fuse with preceding 'à'. Masculine : au, feminine : à la, masculine or feminine before a vowel : à l', plural : aux. • Partitive articles after the negative forms such as 'ne ... pas' or 'ne ... jamais' become 'de' or d' if the article immediately precedes a word beginning with a vowel or a silent 'h'. • The article 'des' is shortened since it immediately precedes a word beginning with a silent 'h'. • The normal way of describing events in which subjects do things to their own body is to use a pronominal verb, and the part of the body preceded by a definite or indefinite article, not a possessive determiner. • The masculin 'cet' appears only when the demonstrative determiner immediately precedes a noun or adjective beginning with a vowel or a silent 'h'. • Since the following word does not start with a vowel, the article does not need to be shortened with an apostrophe. • The articles le and la are shortened if they immediately precede an adjective or noun beginning with a vowel or a so-called 'silent h'. • A partitive article is required before mass nouns. Masculine: du, feminine: de la, masculine or feminine before a vowel: de l'. • One use of the indefinite article is when describing a general class of countable nouns.
se_gr_ge_ (incorrect gender)	• Possessive determiners agree in gender and number with the nouns they precede. • When there are a masculine and a feminine, the masculine wins over the feminine. • Incorrect gender.
se_gr_vo_ (incorrect voice)	• The subject is not the receiver of the action, it is the doer. Use an active form. • The subject is the receiver of the action, not the agent. Use a passive form here, that is être + past participle.
se_vo_re_ (inappropriate register)	• To address one person, 'tu' is used when there is no social distance between the speaker and the adressee, 'vous' is used when there is a certain social distance. Choose the form you want, but be consistent in the same text.
se_vo_sn_ (not understandable)	• What do you mean?
se_gr_cl_ (incorrect word class)	• Adjective inappropriate in context. • Incorrect word class. • Adjectives qualify nouns, adverbs qualify verbs.

se_gr_pr_ (incorrect preposition)	• Preposition indicationg a purpose or a beneficiary. • After a quantifier such as 'beaucoup' followed by a preposition, the partitive article is omitted. • Verbs of (self-)congratulation take an infinitive complement preceded by 'de'. • When the time expression refers to a period in the future in relation to the time of speaking, 'pour' is used. Otherwise 'pendant' is the usual form to describe stressed events that take place during the time period. • Within a certain period of time 'en' is used. This contrasts with 'dans' which occurs after a certain period of time has elapsed. • When the plural indefinite article is preceded by the preposition 'de', the article is omitted not the preposition. • For most countries of feminine gender, we use 'en' and for most countries of masculine gender, we use 'au'. • Incorrect use of preposition.
se_gr_te (incorrect tense)	• Whereas the simple past and compound past tenses refer to events completed in the past from the perspective of the writer, the plusperfect describes events completed at some point even before these past events. • To refer to a completed event in the past, you should use the compound past/passé composé. • The present tense is used to refer to an action which exists at the time of speaking/ writing. To refer to a completed event in the past, you should use the compound past/passé composé. • The tense of the result clause depends on the tense of the 'si' clause: si + present = result clause + present; si + imperfect = result clause + conditional. • The English present progressive will normally be translated into French by the simple present. • The future can be replaced by a present tense form of the verb aller + infinitive only where a greater certainty about the likelihood of an event taking place is implied • To refer to an habitual action in the past, use the imperfect tense. • To refer to an action which is habitual, use the present tense. • The imperfect is used to describe ongoing past events without reference to a time of starting or finishing. To refer to a completed event in the past, you should use the compound past/passé composé.
sy_ad_ (word addition)	• The preposition is not required. This structure looks like the conjunction 'at the time when', which is translated into one word in French. • Use the pronoun 'ce' when referring to another clause. What you are referring to here is the direct subject of your conjugated verb, which means a pronoun is not required. • When a noun alone follows the verb "to be", the article is omitted. • You use two articles, only one is required. • When referring to towns and cities, the preposition 'à' is directly followed by the town or city. • This verb is used in this context as a direct transitive verb with direct object, which means that the object is without preceding preposition.

sy_om_ (word omission)	• Your two clauses are linked by a coordinating conjunction. Their structure must be identical. The first clause starts with a subordinating conjunction, so should the second clause. • While it is common in spoken French for speakers to omit the 'ne' of negation in negated sentences, in written, however, the element 'ne' comes before the verb which is marked for tense in the sentence. • The verb is missing. • The auxiliary is missing. • The noun is missing. • The preposition is missing. • The determinant is missing. • This is a transitive verb that requires a direct object. • The subject is missing.
sy_wo_ (incorrect word order)	• When the verb is accompanied by the auxiliary verbs 'avoir' or 'être', direct and indirect object pronouns appear immediately before the auxiliary. • A direct object pronoun replaces a noun that is the object of a sentence. If the conjugated verb is followed by a verb in the infinitive, the pronoun comes before the infinitive. • The majority of the French adjectives follow the noun. • Manner, degree and time adverbs which consist of just one word usually immediately follow the tense-marked verb.
sy_un_ (syntax not appropriate)	• Word for word translation. • By using present participles preceded by 'en' (gerunds), you emphasize the fact that the event described in the main clause and the event described in the gerundive clause take place simultaneously. • Incorrectly formulated. • The word 'jamais' is mostly used with 'ne' to mean never. In this context, it means ever and is used without 'ne'.
mo_ag_dn (determinant noun agreement)	• When the plural indefinite article is preceded by the preposition 'de', the article is omitted. However the noun still agrees in number with its 'omitted' determinant. • Possessive determiners agree in gender and number with the nouns they precede. • A noun agrees in number with its determinant. • A determinant agrees in number and gender with the noun it qualifies.
mo_ag_na_ (noun adjective agreement)	• The adjective 'demi' agrees in gender with the noun. • Adjectives agree in number and gender with the nouns they qualify. • Numbers are invariable except for 'vingt' and 'cent' when no other numbers follow them. • Adjectives agree in number and gender with the pronouns they qualify.
mo_ag_pa_ (pronoun antecedent agreement)	• The gender of the object or person to which the pronoun refers to, determines the choice of the pronoun. • To refer to events or actions, 'ce', 'cela' or 'ça' is usually used. • A pronoun and its antecedent must agree in person, number and gender.
mo_ag_pp_ (past participle agreement)	• Past participles agree only with preceding direct objects in the compound tenses of verbs conjugated with 'avoir'" • 21 Past participles agree with the subject of intransitive verbs which select auxiliary 'être' in compound tenses.

mo_ag_sv_ (subject verb agreement)	• When a collective noun is the subject of a clause, the verb is usually singular. • Verbs agree with their subjects in person and number.
mo_fo_pl_ (incorrect plural form)	• Adjectives which end in -al generally change to -aux when the noun is masculine. • Most nouns ending in -ail have a regular plural -ails, but a number of -ail nouns also make their plural with -aux.
mo_fo_co_ (incorrect conjugation form)	• Incorrect ending for the present tense. • Incorrect conjugation of the verb 'avoir'. • Incorrect ending for the 'simple future' tense. • Incorrect formation for the imperfect. • Incorrect auxiliary. • Incorrect past participle formation.
mo_fo_wf_ (incorrect word formation)	• Word ill-formed and not really understandable. • Most adverbs ending in '-ment' are formed from the feminine form of the corresponding adjective. • When this word has the indefinite meaning 'everything, all', it is invariable. • Adjectives ending in -ant or -ent form the adverb with -amment or emment, respectively. • The pronoun 'que' is shortened when preceding a word beginning with a vowel. • Indefinite noun phrases like 'quelque chose de' can be followed by adjectives. The adjective is invariable in this construction. • Do not mistaken the conjunction with the verb "to be". • When referring to events, the demonstrative pronoun is used not the reflective pronoun. • When adverbs like souvent or bien are present, names of language become noun (rather than adverbial form), requiring the definite article and being masculine in gender. • Do not mistaken possessive with demonstrative pronouns.
sp_ac_ (incorrect or missing accent)	• Circumflex accent. • Inappropriate accent. • Grave accent. • Acute accent.
sp_hy_ (hyphen)	• The reflexive pronoun in the affirmative imperative conjugation follows the verb and is attached by a hyphen. • The parts of this compound word are linked by a hyphen. • A hyphen is used between the verb and the subject in interrogative forms by inversion.
sp_ms_ (misspelling)	• The parts of this compound word are not linked by a hyphen. • English word.

ty_ca_ (capitalisation)	• Languages do not start with upper case letters. • Months and days are written with a lower case initial letter in French. • When writing dates, months always begin with lower cases letters. • Capital letters are usually used at the start of a sentence, for titles and for proper nouns. • Adjectives that refer to nationalities are not capitalized. • Nouns indicating people of a given nation or civilization (the Germans, ...) require a capital letter.
ty_mp_ (missing punctuation)	• A punctuation mark is used on either side of an apposition. • When two clauses are joined by a coordinating conjunction (et/ou), the comma usually disappears. The comma is, however, used in front of et/ou if one of the two clauses already contains a conjunction. • The introductory word or clause has to be separated from the independent clause by a punctuation mark. • End all sentences with a punctuation mark. • Relative clauses are separated by a punctuation mark when the pronoun (qui/que) is too far from the noun it relates. • A punctuation mark is used when listing items.
ty_wp_ (inappropriate punctuation)	• Whereas in English the serial comma (the one before and in a list) is optional, it cannot be used in French. • The solidus (forward slash) is usually used to connect two words. • French quotes shoud be used for a title, an expression or a quotation. Spaces are required after the opening quote and before the closing quote. • Inapropriate use of a punctuation. • A period is usually an indicator for the end of a sentence.
ty_es_ (missing or inappropriate space)	• No spaces are required before a comma. • One space goes before every exclamation point, semicolon, ellipsis, question mark, and colon (but not period). • One space goes after every exclamation point, semicolon, ellipsis, question mark, colon and period.
ty_ab_ (abbreviation)	• Incorrect conventions. • In French, the period is not used as a decimal point. • When writing numbers, either a period or a space may be used to separate every three digits. • The convention is to use the currency symbol after the number, and leave a space between them. • Incorrect time format. The convention is to write numbers with digits and 'heure' either in full or abbreviated.

D.5. Annotated corpus: text #88 (42 words)

<?xml version="1.0" encoding ="UTF-8" standalone="yes"?>
<text>
<text_id>88</text_id>
<student_id>2</student_id>
<text_type>email</text_type>
<original_text>Bonjour Sylvia. J'ai fini mes examens en francais , les grammaires était bien,

mais l'écrit était difficile. Merci pour ton aide. Si, le vendredi, tu es à l'université, je
te renconterai à onze heures ou apres treize heures! A bientot </original_text>
<errorEncoded_text>Bonjour Sylvia. J'ai fini mes examens en
francais[29]{1@Misspelling.@français@} , *les grammaires*[2]{2@Inappropriate word in
context.@les exercices de grammaire@} *était*[22]{3@Verbs agree with their subjects in
person and number.@étaient@} bien, mais l'écrit était difficile. Merci pour ton aide. Si, le
vendredi, tu es à l'université, je te *renconterai*[29]{4@Misspelling.@rencontrerai@} à onze
heures ou *apres*[26]{5@Grave accent.@après@} treize *heures!*[13]{6@One space goes before
every exclamation point, semicolon, ellipsis, question mark, and colon (but not
period).@heures !@} A *bientot*[26]{7@Circumflex accent.@bientôt !@} </errorEncoded_text>
<time_span_1>00:01:57</time_span_1>
<time_span_2>00:01:05</time_span_2>
<time_span_3>00:00:34</time_span_3>
<chunk>
<token_id>18930</token_id>
<token>Bonjour</token>
<error_tagged_as>correct</error_tagged_as>
<pos_5>NOM</pos_5>
<position>0</position>
</chunk>
<chunk>
<token_id>18931</token_id>
<token>Sylvia</token>
<error_tagged_as>correct</error_tagged_as>
<pos_5>NAM</pos_5>
<position>8</position>
</chunk>
<chunk>
<token_id>18932</token_id>
<token>.</token>
<error_tagged_as>correct</error_tagged_as>
<pos_5>SENT</pos_5>
<position>14</position>
</chunk>
<chunk>
<token_id>18933</token_id>
<token>J'</token>
<error_tagged_as>correct</error_tagged_as>
<pos_5>PRO:PER</pos_5>
<position>17</position>
</chunk>
<chunk>
<token_id>18934</token_id>
<token>ai</token>
<error_tagged_as>correct</error_tagged_as>
<pos_5>VER:conj</pos_5>
<position>19</position>
</chunk>
<chunk>
<token_id>18935</token_id>
<token>fini</token>
<error_tagged_as>correct</error_tagged_as>
<pos_5>VER:pper</pos_5>
<position>22</position>
</chunk>
<chunk>
<token_id>18936</token_id>
<token>mes</token>
<error_tagged_as>correct</error_tagged_as>
<pos_5>DET:POS</pos_5>
<position>27</position>
</chunk>
<chunk>
<token_id>18937</token_id>
<token>examens</token>
<error_tagged_as>correct</error_tagged_as>
<pos_5>NOM</pos_5>
<position>31</position>
</chunk>
<chunk>
<token_id>18938</token_id>
<token>en</token>
<error_tagged_as>correct</error_tagged_as>
<pos_5>PRO:PER</pos_5>

<position>35</position>
</chunk>
<chunk>
<token_id>18939</token_id>
<token>francais</token>
<error_tagged_as>sp_ms_</error_tagged_as>
<incorrectForm>
<incorrectForm_id>2818 </incorrectForm_id>
<incorrect_seq>francais</incorrect_seq>
<stu_corr_level_1>français</stu_corr_level_1>
<stu_corr_level_2>français</stu_corr_level_2>
<stu_corr_level_3>français</stu_corr_level_3>
<val_student_1>1</val_student_1>
<val_student_2>1</val_student_2>
<val_student_3>1</val_student_3>
<val_errorType_1>1</val_errorType_1>
<val_errorType_2>1</val_errorType_2>
<val_errorType_3>1</val_errorType_3>
<fb_level_2>misspelling</fb_level_2>
<fb_level_3>Misspelling.</fb_level_3>
<fb_level_4>français</fb_level_4>
<feedback_read_count_level_2>1</feedback_read_count_level_2>
<feedback_read_average_time_level_2>1.46</feedback_read_average_time_level_2>
<feedback_read_count_level_3>2</feedback_read_count_level_3>
<feedback_read_average_time_level_3>1.15</feedback_read_average_time_level_3>
<submissionDate>2009-05-01</submissionDate>
</incorrectForm>
<pos_5>NOM</pos_5>
<position>42</position>
</chunk>
<chunk>
<token_id>18940</token_id>
<token>,</token>
<error_tagged_as>correct</error_tagged_as>
<pos_5>PUN</pos_5>
<position>51</position>
</chunk>
<chunk>
<token_id>18941</token_id>
<token>les</token>
<error_tagged_as>se_vo_lw_</error_tagged_as>
<incorrectForm>
<incorrectForm_id>2819 </incorrectForm_id>
<incorrect_seq>les grammaires</incorrect_seq>
<stu_corr_level_1>la grammaire</stu_corr_level_1>
<stu_corr_level_2>la grammaire</stu_corr_level_2>
<stu_corr_level_3>_blank</stu_corr_level_3>
<val_student_1>0</val_student_1>
<val_student_2>0</val_student_2>
<val_student_3>0</val_student_3>
<val_errorType_1>0</val_errorType_1>
<val_errorType_2>0</val_errorType_2>
<val_errorType_3>0</val_errorType_3>
<fb_level_2>inappropriate word choice</fb_level_2>
<fb_level_3>Inappropriate word in context.</fb_level_3>
<fb_level_4>les exercices de grammaire</fb_level_4>
<feedback_read_count_level_2>1</feedback_read_count_level_2>
<feedback_read_average_time_level_2>1.53</feedback_read_average_time_level_2>
<feedback_read_count_level_3>2</feedback_read_count_level_3>
<feedback_read_average_time_level_3>1.6</feedback_read_average_time_level_3>
<submissionDate>2009-05-01</submissionDate>
</incorrectForm>
<pos_5>DET:ART</pos_5>
<position>53</position>
</chunk>
<chunk>
<token_id>18942</token_id>
<token>grammaires</token>
<error_tagged_as>se_vo_lw_</error_tagged_as>
<incorrectForm>
<incorrectForm_id>2819 </incorrectForm_id>
<incorrect_seq>les grammaires</incorrect_seq>
<stu_corr_level_1>la grammaire</stu_corr_level_1>
<stu_corr_level_2>la grammaire</stu_corr_level_2>

```xml
    <stu_corr_level_3>_blank</stu_corr_level_3>
    <val_student_1>0</val_student_1>
    <val_student_2>0</val_student_2>
    <val_student_3>0</val_student_3>
    <val_errorType_1>0</val_errorType_1>
    <val_errorType_2>0</val_errorType_2>
    <val_errorType_3>0</val_errorType_3>
    <fb_level_2>inappropriate word choice</fb_level_2>
    <fb_level_3>Inappropriate word in context.</fb_level_3>
    <fb_level_4>les exercices de grammaire</fb_level_4>
    <feedback_read_count_level_2>1</feedback_read_count_level_2>
    <feedback_read_average_time_level_2>1.53</feedback_read_average_time_level_2>
    <feedback_read_count_level_3>2</feedback_read_count_level_3>
    <feedback_read_average_time_level_3>1.6</feedback_read_average_time_level_3>
    <submissionDate>2009-05-01</submissionDate>
  </incorrectForm>
  <pos_5>NOM</pos_5>
  <position>57</position>
</chunk>
<chunk>
  <token_id>18943</token_id>
  <token>était</token>
  <error_tagged_as>mo_ag_sv_</error_tagged_as>
  <incorrectForm>
    <incorrectForm_id>2820 </incorrectForm_id>
    <incorrect_seq>était</incorrect_seq>
    <stu_corr_level_1>_blank</stu_corr_level_1>
    <stu_corr_level_2>_blank</stu_corr_level_2>
    <stu_corr_level_3>_blank</stu_corr_level_3>
    <val_student_1>0</val_student_1>
    <val_student_2>0</val_student_2>
    <val_student_3>0</val_student_3>
    <val_errorType_1>0</val_errorType_1>
    <val_errorType_2>0</val_errorType_2>
    <val_errorType_3>0</val_errorType_3>
    <fb_level_2>subject verb agreement</fb_level_2>
    <fb_level_3>Verbs agree with their subjects in person and number.</fb_level_3>
    <fb_level_4>étaient</fb_level_4>
    <feedback_read_count_level_2>1</feedback_read_count_level_2>
    <feedback_read_average_time_level_2>3.86</feedback_read_average_time_level_2>
    <feedback_read_count_level_3>1</feedback_read_count_level_3>
    <feedback_read_average_time_level_3>3.86</feedback_read_average_time_level_3>
    <submissionDate>2009-05-01</submissionDate>
  </incorrectForm>
  <pos_5>VER:conj</pos_5>
  <position>68</position>
</chunk>
<chunk>
  <token_id>18944</token_id>
  <token>bien</token>
  <error_tagged_as>correct</error_tagged_as>
  <pos_5>ADV</pos_5>
  <position>74</position>
</chunk>
<chunk>
  <token_id>18945</token_id>
  <token>,</token>
  <error_tagged_as>correct</error_tagged_as>
  <pos_5>PUN</pos_5>
  <position>78</position>
</chunk>
<chunk>
  <token_id>18946</token_id>
  <token>mais</token>
  <error_tagged_as>correct</error_tagged_as>
  <pos_5>KON</pos_5>
  <position>80</position>
</chunk>
<chunk>
  <token_id>18947</token_id>
  <token>l'</token>
  <error_tagged_as>correct</error_tagged_as>
  <pos_5>DET:ART</pos_5>
  <position>85</position>
```

</chunk>
<chunk>
<token_id>18948</token_id>
<token>écrit</token>
<error_tagged_as>correct</error_tagged_as>
<pos_5>NOM</pos_5>
<position>87</position>
</chunk>
<chunk>
<token_id>18949</token_id>
<token>était</token>
<error_tagged_as>correct</error_tagged_as>
<pos_5>VER:conj</pos_5>
<position>93</position>
</chunk>
<chunk>
<token_id>18950</token_id>
<token>difficile</token>
<error_tagged_as>correct</error_tagged_as>
<pos_5>ADJ</pos_5>
<position>99</position>
</chunk>
<chunk>
<token_id>18951</token_id>
<token>.</token>
<error_tagged_as>correct</error_tagged_as>
<pos_5>SENT</pos_5>
<position>108</position>
</chunk>
<chunk>
<token_id>18952</token_id>
<token>Merci</token>
<error_tagged_as>correct</error_tagged_as>
<pos_5>NOM</pos_5>
<position>110</position>
</chunk>
<chunk>
<token_id>18953</token_id>
<token>pour</token>
<error_tagged_as>correct</error_tagged_as>
<pos_5>PRP</pos_5>
<position>116</position>
</chunk>
<chunk>
<token_id>18954</token_id>
<token>ton</token>
<error_tagged_as>correct</error_tagged_as>
<pos_5>DET:POS</pos_5>
<position>121</position>
</chunk>
<chunk>
<token_id>18955</token_id>
<token>aide</token>
<error_tagged_as>correct</error_tagged_as>
<pos_5>NOM</pos_5>
<position>125</position>
</chunk>
<chunk>
<token_id>18956</token_id>
<token>.</token>
<error_tagged_as>correct</error_tagged_as>
<pos_5>SENT</pos_5>
<position>129</position>
</chunk>
<chunk>
<token_id>18957</token_id>
<token>Si</token>
<error_tagged_as>correct</error_tagged_as>
<pos_5>ADV</pos_5>
<position>132</position>
</chunk>
<chunk>
<token_id>18958</token_id>
<token>,</token>

```xml
  <error_tagged_as>correct</error_tagged_as>
  <pos_5>PUN</pos_5>
  <position>134</position>
</chunk>
<chunk>
  <token_id>18959</token_id>
  <token>le</token>
  <error_tagged_as>correct</error_tagged_as>
  <pos_5>DET:ART</pos_5>
  <position>136</position>
</chunk>
<chunk>
  <token_id>18960</token_id>
  <token>vendredi</token>
  <error_tagged_as>correct</error_tagged_as>
  <pos_5>NOM</pos_5>
  <position>139</position>
</chunk>
<chunk>
  <token_id>18961</token_id>
  <token>,</token>
  <error_tagged_as>correct</error_tagged_as>
  <pos_5>PUN</pos_5>
  <position>147</position>
</chunk>
<chunk>
  <token_id>18962</token_id>
  <token>tu</token>
  <error_tagged_as>correct</error_tagged_as>
  <pos_5>PRO:PER</pos_5>
  <position>149</position>
</chunk>
<chunk>
  <token_id>18963</token_id>
  <token>es</token>
  <error_tagged_as>correct</error_tagged_as>
  <pos_5>VER:conj</pos_5>
  <position>152</position>
</chunk>
<chunk>
  <token_id>18964</token_id>
  <token>à</token>
  <error_tagged_as>correct</error_tagged_as>
  <pos_5>PRP</pos_5>
  <position>155</position>
</chunk>
<chunk>
  <token_id>18965</token_id>
  <token>l'</token>
  <error_tagged_as>correct</error_tagged_as>
  <pos_5>DET:ART</pos_5>
  <position>157</position>
</chunk>
<chunk>
  <token_id>18966</token_id>
  <token>université</token>
  <error_tagged_as>correct</error_tagged_as>
  <pos_5>NOM</pos_5>
  <position>159</position>
</chunk>
<chunk>
  <token_id>18967</token_id>
  <token>,</token>
  <error_tagged_as>correct</error_tagged_as>
  <pos_5>PUN</pos_5>
  <position>169</position>
</chunk>
<chunk>
  <token_id>18968</token_id>
  <token>je</token>
  <error_tagged_as>correct</error_tagged_as>
  <pos_5>PRO:PER</pos_5>
  <position>171</position>
</chunk>
```

```xml
<chunk>
  <token_id>18969</token_id>
  <token>te</token>
  <error_tagged_as>correct</error_tagged_as>
  <pos_5>PRO:PER</pos_5>
  <position>174</position>
</chunk>
<chunk>
  <token_id>18970</token_id>
  <token>renconterai</token>
  <error_tagged_as>sp_ms_</error_tagged_as>
  <incorrectForm>
    <incorrectForm_id>2821 </incorrectForm_id>
    <incorrect_seq>renconterai</incorrect_seq>
    <stu_corr_level_1>_blank</stu_corr_level_1>
    <stu_corr_level_2>recontrerai</stu_corr_level_2>
    <stu_corr_level_3>recontrerai</stu_corr_level_3>
    <val_student_1>0</val_student_1>
    <val_student_2>0</val_student_2>
    <val_student_3>0</val_student_3>
    <val_errorType_1>0</val_errorType_1>
    <val_errorType_2>0</val_errorType_2>
    <val_errorType_3>0</val_errorType_3>
    <fb_level_2>misspelling</fb_level_2>
    <fb_level_3>Misspelling.</fb_level_3>
    <fb_level_4>rencontrerai</fb_level_4>
    <feedback_read_count_level_2>2</feedback_read_count_level_2>
    <feedback_read_average_time_level_2>1.88</feedback_read_average_time_level_2>
    <feedback_read_count_level_3>3</feedback_read_count_level_3>
    <feedback_read_average_time_level_3>1.44</feedback_read_average_time_level_3>
    <submissionDate>2009-05-01</submissionDate>
  </incorrectForm>
  <pos_5>VER:conj</pos_5>
  <position>177</position>
</chunk>
<chunk>
  <token_id>18971</token_id>
  <token>à</token>
  <error_tagged_as>correct</error_tagged_as>
  <pos_5>PRP</pos_5>
  <position>189</position>
</chunk>
<chunk>
  <token_id>18972</token_id>
  <token>onze</token>
  <error_tagged_as>correct</error_tagged_as>
  <pos_5>ADJ:NUM</pos_5>
  <position>191</position>
</chunk>
<chunk>
  <token_id>18973</token_id>
  <token>heures</token>
  <error_tagged_as>correct</error_tagged_as>
  <pos_5>NOM</pos_5>
  <position>196</position>
</chunk>
<chunk>
  <token_id>18974</token_id>
  <token>ou</token>
  <error_tagged_as>correct</error_tagged_as>
  <pos_5>KON</pos_5>
  <position>203</position>
</chunk>
<chunk>
  <token_id>18975</token_id>
  <token>apres</token>
  <error_tagged_as>sp_ac_</error_tagged_as>
  <incorrectForm>
    <incorrectForm_id>2822 </incorrectForm_id>
    <incorrect_seq>apres</incorrect_seq>
    <stu_corr_level_1>après</stu_corr_level_1>
    <stu_corr_level_2>après</stu_corr_level_2>
    <stu_corr_level_3>après</stu_corr_level_3>
    <val_student_1>1</val_student_1>
```

```
    <val_student_2>1</val_student_2>
    <val_student_3>1</val_student_3>
    <val_errorType_1>1</val_errorType_1>
    <val_errorType_2>1</val_errorType_2>
    <val_errorType_3>1</val_errorType_3>
    <fb_level_2>incorrect or missing accent</fb_level_2>
    <fb_level_3>Grave accent.</fb_level_3>
    <fb_level_4>après</fb_level_4>
    <feedback_read_count_level_2>0</feedback_read_count_level_2>
    <feedback_read_average_time_level_2>0</feedback_read_average_time_level_2>
    <feedback_read_count_level_3>1</feedback_read_count_level_3>
    <feedback_read_average_time_level_3>1.25</feedback_read_average_time_level_3>
    <submissionDate>2009-05-01</submissionDate>
  </incorrectForm>
  <pos_5>PRP</pos_5>
  <position>206</position>
</chunk>
<chunk>
  <token_id>18976</token_id>
  <token>treize</token>
  <error_tagged_as>correct</error_tagged_as>
  <pos_5>ADJ:NUM</pos_5>
  <position>212</position>
</chunk>
<chunk>
  <token_id>18977</token_id>
  <token>heures</token>
  <error_tagged_as>ty_es_</error_tagged_as>
  <incorrectForm>
    <incorrectForm_id>2823 </incorrectForm_id>
    <incorrect_seq>heures!</incorrect_seq>
    <stu_corr_level_1>heure</stu_corr_level_1>
    <stu_corr_level_2>heures !</stu_corr_level_2>
    <stu_corr_level_3>heures !</stu_corr_level_3>
    <val_student_1>0</val_student_1>
    <val_student_2>1</val_student_2>
    <val_student_3>1</val_student_3>
    <val_errorType_1>0</val_errorType_1>
    <val_errorType_2>1</val_errorType_2>
    <val_errorType_3>1</val_errorType_3>
    <fb_level_2>missing or inappropriate space</fb_level_2>
    <fb_level_3>One space goes before every exclamation point, semicolon, ellipsis, question
      mark, and colon (but not period).</fb_level_3>
    <fb_level_4>heures !</fb_level_4>
    <feedback_read_count_level_2>1</feedback_read_count_level_2>
    <feedback_read_average_time_level_2>2.64</feedback_read_average_time_level_2>
    <feedback_read_count_level_3>0</feedback_read_count_level_3>
    <feedback_read_average_time_level_3>0</feedback_read_average_time_level_3>
    <submissionDate>2009-05-01</submissionDate>
  </incorrectForm>
  <pos_5>NOM</pos_5>
  <position>219</position>
</chunk>
<chunk>
  <token_id>18978</token_id>
  <token>!</token>
  <error_tagged_as>ty_es_</error_tagged_as>
  <incorrectForm>
    <incorrectForm_id>2823 </incorrectForm_id>
    <incorrect_seq>heures!</incorrect_seq>
    <stu_corr_level_1>heure</stu_corr_level_1>
    <stu_corr_level_2>heures !</stu_corr_level_2>
    <stu_corr_level_3>heures !</stu_corr_level_3>
    <val_student_1>0</val_student_1>
    <val_student_2>1</val_student_2>
    <val_student_3>1</val_student_3>
    <val_errorType_1>0</val_errorType_1>
    <val_errorType_2>1</val_errorType_2>
    <val_errorType_3>1</val_errorType_3>
    <fb_level_2>missing or inappropriate space</fb_level_2>
    <fb_level_3>One space goes before every exclamation point, semicolon, ellipsis, question
      mark, and colon (but not period).</fb_level_3>
    <fb_level_4>heures !</fb_level_4>
    <feedback_read_count_level_2>1</feedback_read_count_level_2>
```

<feedback_read_average_time_level_2>2.64</feedback_read_average_time_level_2>
<feedback_read_count_level_3>0</feedback_read_count_level_3>
<feedback_read_average_time_level_3>0</feedback_read_average_time_level_3>
<submissionDate>2009-05-01</submissionDate>
</incorrectForm>
<pos_5>SENT</pos_5>
<position>225</position>
</chunk>
<chunk>
<token_id>18979</token_id>
<token>A</token>
<error_tagged_as>correct</error_tagged_as>
<pos_5>PRP</pos_5>
<position>229</position>
</chunk>
<chunk>
<token_id>18980</token_id>
<token>bientot</token>
<error_tagged_as>sp_ac_</error_tagged_as>
<incorrectForm>
<incorrectForm_id>2824 </incorrectForm_id>
<incorrect_seq>bientot</incorrect_seq>
<stu_corr_level_1>bientôt</stu_corr_level_1>
<stu_corr_level_2>bientôt</stu_corr_level_2>
<stu_corr_level_3>bientôt</stu_corr_level_3>
<val_student_1>1</val_student_1>
<val_student_2>1</val_student_2>
<val_student_3>1</val_student_3>
<val_errorType_1>1</val_errorType_1>
<val_errorType_2>1</val_errorType_2>
<val_errorType_3>1</val_errorType_3>
<fb_level_2>incorrect or missing accent</fb_level_2>
<fb_level_3>Circumflex accent.</fb_level_3>
<fb_level_4>bientôt !</fb_level_4>
<feedback_read_count_level_2>0</feedback_read_count_level_2>
<feedback_read_average_time_level_2>0</feedback_read_average_time_level_2>
<feedback_read_count_level_3>1</feedback_read_count_level_3>
<feedback_read_average_time_level_3>0.9</feedback_read_average_time_level_3>
<submissionDate>2009-05-01</submissionDate>
</incorrectForm>
<pos_5>ADV</pos_5>
<position>231</position>
</chunk>
</text>

Appendix E. Analyses and measures

E.1. Distinguishing between errors and mistakes


```
<?php
mb_internal_encoding('UTF-8');
include 'database_var.php';
include 'subroutines.php';
include 'variables.php';
?>
<!DOCTYPE HTML PUBLIC "-//W3C//DTD HTML 4.01 Transitional//EN"
"http://www.w3.org/TR/html4/loose.dtd">
<html>
    <head>
    <meta http-equiv="Content-Type" content="text/html; charset=UTF-8">
    <META HTTP-EQUIV="Pragma" CONTENT="no-cache">
    <META Http-Equiv="Expires" Content="0">
    <META name="ROBOTS" content="NOIMAGEINDEX,NOIMAGECLICK,NOFOLLOW">
    <META name="Author" content="Sylvie THOUESNY">
    <link rel="stylesheet" type="text/css" href="css/styles.css">
    <title>error versus mistake</title>

    </head>
<body bgcolor="#FFFFFF">

<?php
//this subroutine retrieves the data from the database
//and determines whether the incorrect form is an error or a mistake
//depending on alternatives provided and level of feedback required to provide this alternative

// IncorrectForm    level2 -               level3                 Error or mistake
// 0/-              0/-                    0/-                    error
// 0/-              0/-                    1 (feedback read)      error (potential development)
// 0/-              0/-                    1 (feedback not read)  error
// 0/-              1 (feedback read)      0/-                    error (potential development)
// 0/-              1 (feedback not read)  0/-                    error
// 0/-              1                      1(feedback read)       error (potential development)
// 0                1 (feedback read)      1(feedback read)       error (potential development)
// -                1 (feedback read)      1(feedback not read)   mistake
// 0/-              1 (feedback not read)  1(feedback not read)   mistake
// 1                0/-                    0/-                    error
// 1                0/-                    1 (feedback read)      error  (potential development)
// 1                0/-                    1 (feedback not read)  mistake
// 1                1 (feedback read)      0                      error  (potential development)
// 1                1 (feedback read)      -                      mistake
// 1                1 (feedback not read)  0/-                    mistake
// 1                1                      1                      mistake

// 0 means incorrect alternative provided
// - means no alternative
//1 means correct alternative


//connect to mysql
mysql_select_db($var_database,$db) or die("Couldn't open $db: ".mysql_error());

$ara_full=array();
$ara_incorrectForm=array();
```

DB_error_mistake.php 2010-09-27

```
//retrieve incorrect forms from db and store them in an array by incorrectForm identification
$result=mysql_query("SELECT incorrectForm.incorrectForm_id, incorrectForm.text_id,
incorrectForm.errorType_id, incorrectForm.student_id, incorrectForm.incorrect_seq,
incorrectForm.stu_corr_level_1, incorrectForm.stu_corr_level_2, incorrectForm.stu_corr_level_3,
incorrectForm.val_4, incorrectForm.val_5, incorrectForm.val_6, incorrectForm.foot_number,
incorrectForm.fb_level_3, errorType.description, errorType.error_id
    FROM incorrectForm, errorType where incorrectForm.errorType_id=errorType.error_id") OR
exit(mysql_error());
while($row=mysql_fetch_array($result)){
    $ara_incorrectForm[$row['incorrectForm_id']]="".$row['text_id']."@".$row['errorType_id']."@".$row[
student_id']."@".$row['incorrect_seq']."@".$row['stu_corr_level_1']."@".$row['stu_corr_level_2']."@".$r
w['stu_corr_level_3']."@".$row['val_4']."@".$row['val_5']."@".$row['val_6']."@".$row['foot_number']."@"
$row['fb_level_3']."@".$row['description']."@";
}


// for each incorrect form, determine whether the feedback has been accessed yes or no
foreach($ara_incorrectForm as $key=>$value){
    //echo "1: $key=>$value<br>";
    $ara_feedbackRead=isFeedbackRead($key,$value);
    //split returned value as it contained the count for levels 2 and 3 and add them to the array at
incorrect form index
    foreach($ara_feedbackRead as $key2=>$value2){
        //echo "2: $key2=>$value2<br><br>";
        list($feedback_read_level_1, $feedback_read_level_2,$_)=split("@",$value2,3);

$ara_full[$key2]=$ara_incorrectForm[$key2].$feedback_read_level_1."@".$feedback_read_level_2."@";
    }
}
//$text_id, $errorType_id, $student_id, $incorrect_seq, $stu_corr_level_1, $stu_corr_level_2,
$stu_corr_level_3, $val_4, $val_5, $val_6, $foot_number, $fb_level_3, $fb_level_2,
$feedback_read_level_2, $feedback_read_level_3)
//determining whether the incorrect form is an error or a mistake
//error = 0 with potential development
//mistake = 1
//error not confirmed = 2
//mistake not confirmed =3
//error without potential development =-1
$ara_incorrect_form_status=array();
foreach($ara_full as $key=>$value){
    list($text_id, $errorType_id, $student_id, $incorrect_seq, $stu_corr_level_1, $stu_corr_level_2,
$stu_corr_level_3, $val_4, $val_5, $val_6, $foot_number, $fb_level_3,
$fb_level_2,$feedback_read_level_2, $feedback_read_level_3,$_)=split("@",$value,16);

    //[0/-] [0/-] [0/-]=>error
    if($val_4==0 && $val_5==0 && $val_6==0){
        $ara_incorrect_form_status[$key]="-1";
        //echo "$incorrect_seq: [0/-] [0/-] [0/-]<br>";
    }

    //[0/-] [0/-] [1]=>error and potential development if feedback read
    else if($val_4==0 && $val_5==0 && $val_6==1){
        if($feedback_read_level_3>0){
            $ara_incorrect_form_status[$key]="0";
            //echo "$incorrect_seq: [0/-] [0/-] [1R]<br>";
```

- 2/7 -

DB_error_mistake.php 2010-09-27

```
        }
        else{
            $ara_incorrect_form_status[$key]="-1";
            //echo "$incorrect_seq: [0/-] [0/-] [1]<br>";
        }
    }

    //[0/-] [1] [0]=>error and potential development if feedback read
    else if($val_4==0 && $val_5==1 && $val_6==0){
        if($feedback_read_level_2>0){
            $ara_incorrect_form_status[$key]="0";
            //echo "$incorrect_seq: [0/-] [1R] [0]<br>";
        }
        else{
            $ara_incorrect_form_status[$key]="-1";
            //echo "$incorrect_seq: [0/-] [1] [0]<br>";
        }
    }

    //[0/-] [1] [1R]=>error and potential development if feedback read
    else if($val_4==0 && $val_5==1 && $val_6==1 && ($feedback_read_level_3>0 ||
$feedback_read_level_3=="NA")){
        if($feedback_read_level_3=="NA"){
            $ara_incorrect_form_status[$key]="<span class=\"mainTextColor\">2</span>";
            //echo "$incorrect_seq: [0/-] [1] [1?]<br>";
        }
        else{
            $ara_incorrect_form_status[$key]="0";
            //echo "$incorrect_seq: [0/-] [1] [1R]<br>";
        }
    }

    // [0] [1R] [1NR]=>error and potential development if feedback read
    else if($val_4==0 && $stu_corr_level_1!="" && $val_5==1 && $val_6==1 && ($feedback_read_level_2>0
|| $feedback_read_level_2=="NA") && ($feedback_read_level_3==0 || $feedback_read_level_3=="NA")){
        if($feedback_read_level_2=="NA" || $feedback_read_level_3=="NA"){
            $ara_incorrect_form_status[$key]="<span class=\"mainTextColor\">2</span>";
            //echo "$incorrect_seq: [0] [1?] [1?]<br>";
        }
        else{
            $ara_incorrect_form_status[$key]="0";
            //echo "$incorrect_seq: [0] [1R] [1NR]<br>";
        }
    }

    // [-] [1R] [1NR]=>mistake
    else if($stu_corr_level_1=="" && $val_5==1 && $val_6==1 && ($feedback_read_level_2>0 ||
$feedback_read_level_2=="NA") && ($feedback_read_level_3==0 || $feedback_read_level_3=="NA")){
        if($feedback_read_level_2=="NA" || $feedback_read_level_3=="NA"){
            $ara_incorrect_form_status[$key]="<span class=\"mainTextColor\">3</span>";
            //echo "$incorrect_seq: [-] [1?] [1?]<br>";
        }
        else{
            $ara_incorrect_form_status[$key]="1";
            //echo "$incorrect_seq: [-] [1R] [1NR]<br>";
```

- 3/7 -

```
        }
    }

    // [0/-] [1NR] [1NR]=>mistake
    else if($val_4==0 && $val_5==1 && $val_6==1 && ($feedback_read_level_2==0 ||
$feedback_read_level_2=="NA") && ($feedback_read_level_3==0 || $feedback_read_level_3=="NA")){
        if($feedback_read_level_2=="NA" || $feedback_read_level_3=="NA"){
            $ara_incorrect_form_status[$key]="<span class=\"mainTextColor\">3</span>";
            //echo "$incorrect_seq: [0/-] [1?] [1?]<br>";
        }
        else{
            $ara_incorrect_form_status[$key]="1";
            //echo "$incorrect_seq: [0/-] [1NR] [1NR]<br>";
        }
    }

    // [1] [0/-] [0/-]=>error
    else if($val_4==1 && $val_5==0 && $val_6==0){
        $ara_incorrect_form_status[$key]="-1";
        //echo "$incorrect_seq: [1] [0/-] [0/-]<br>";
    }

    // [1] [0/-] [1R]=>error and potential development if feedback read
    else if($val_4==1 && $val_5==0 && $val_6==1 && ($feedback_read_level_3>0 ||
$feedback_read_level_3=="NA")){
        if($feedback_read_level_3=="NA"){
            $ara_incorrect_form_status[$key]="<span class=\"mainTextColor\">2</span>";
            echo "$incorrect_seq: [1] [0/-] [1?]<br>";
        }
        else{
            $ara_incorrect_form_status[$key]="0";
            //echo "$incorrect_seq: [1] [0/-] [1R]<br>";
        }
    }

    // [1] [0/-] [1NR]=>mistake
    else if($val_4==1 && $val_5==0 && $val_6==1 && ($feedback_read_level_3==0 ||
$feedback_read_level_3=="NA")){
        if($feedback_read_level_3=="NA"){
            $ara_incorrect_form_status[$key]="<span class=\"mainTextColor\">3</span>";
            //echo "$incorrect_seq: [1] [0/-] [1?]<br>";
        }
        else{
            $ara_incorrect_form_status[$key]="1";
            //echo "$incorrect_seq: [1] [0/-] [1NR]<br>";
        }
    }

    // [1] [1R] [0]=>error and potential development if feedback read
    else if($val_4==1 && $val_5==1 && $stu_corr_level_3!="" && $val_6==0 && ($feedback_read_level_2>0
|| $feedback_read_level_2=="NA")){
        if($feedback_read_level_2=="NA"){
            $ara_incorrect_form_status[$key]="<span class=\"mainTextColor\">2</span>";
            //echo "$incorrect_seq: [1] [1?] [0]<br>";
        }
```

```
        else{
            $ara_incorrect_form_status[$key]="0";
            //echo "$incorrect_seq: [1] [1R] [0]<br>";
        }
    }

    // [1] [1R] [-]=>mistake
    else if($val_4==1 && $val_5==1 && $stu_corr_level_3=="" && ($feedback_read_level_2>0 ||
$feedback_read_level_2=="NA")){
        if($feedback_read_level_2=="NA"){
            $ara_incorrect_form_status[$key]="<span class=\"mainTextColor\">3</span>";
            //echo "$incorrect_seq: [1] [1?] [-]<br>";
        }
        else{
            $ara_incorrect_form_status[$key]="1";
            //echo "$incorrect_seq: [1] [1R] [-]<br>";
        }
    }

    // [1] [1NR] [0/-]=>mistake
    else if($val_4==1 && $val_5==1 && $val_6==0 && ($feedback_read_level_2==0 ||
$feedback_read_level_2=="NA")){
        if($feedback_read_level_2=="NA"){
            $ara_incorrect_form_status[$key]="<span class=\"mainTextColor\">3</span>";
            //echo "$incorrect_seq: [1] [1?] [0/-]<br>";
        }
        else{
            $ara_incorrect_form_status[$key]="1";
            //echo "$incorrect_seq: [1] [1NR] [0/-]<br>";
        }
    }

    // [1] [1] [1]=>mistake
    else if($val_4==1 && $val_5==1 && $val_6==1){
        $ara_incorrect_form_status[$key]="1";
        echo "$incorrect_seq: [1] [1] [1]<br>";
    }
}
//echo "incorrectForm_id=>error/mistake<br>";
//foreach($ara_incorrect_form_status as $key=>$value){
//  echo "4: $key=>$value<br>";
//}


//retrieve tokens from db and add incorrect form status
$result=mysql_query("SELECT token.token_id, token.text_id, token.incorrectForm_id, token.pos_5,
token.error_tagged_as, text.text_id, text.student_id, text.artefact_id, text.submissionDate,
pos.pos_id, pos.description
    FROM token, text, pos where token.text_id=text.text_id and pos.description=token.pos_5") OR
exit(mysql_error());
while($row=mysql_fetch_array($result)){
    if($row['error_tagged_as']!="correct"){$status=$ara_incorrect_form_status[$row['incorrectForm_id']
;}
    else{$row['error_tagged_as']="0";}
```

DB_error_mistake.php 2010-09-27

```
    $row['submissionDate']=mb_ereg_replace('-','/',$row['submissionDate']);

    $ara_forAnalysis[$row['token_id']]="".$row['token_id']."@".$row['text_id']."@".$row['student_id'].
@".$row['artefact_id']."@".$row['submissionDate']."@".$row['pos_id']."@".$row['error_tagged_as']."@".$s
atus."@";
    $status="";
}
//print output for SSPSS
echo "token_id@text_id@student_id@artefact_id@submissionDate@pos_5@error_tagged_as@status@<br>";
foreach($ara_forAnalysis as $key=>$value){
    echo "$value<br>";
}

//echo sizeof($arr_text)."<br>";
//echo "1: $key=>$value<br>";


//subroutines
//=========================================

function isFeedbackRead($key, $value){

    $wpm=180;
    $word_per_milli_second=(60*1000)/$wpm;
    $ara_incorrect_form_roll_count=array();
    $ara_incorrect_form_roll_average=array();
    $words_3=array();
    $words_2=array();
    $ara_result=array();
    $incorrectForm_id=$key;

    //counting the amount of words in feedback
    list($text_id, $errorType_id, $student_id, $incorrect_seq, $stu_corr_level_1, $stu_corr_level_2,
$stu_corr_level_3, $val_4, $val_5, $val_6, $foot_number, $fb_level_3,
$fb_level_2,$_)=split("@",$value,14);
    $words_3=split(' ',$fb_level_3);
    $words_2=split(' ',$fb_level_2);

    //amount of words for feedback level 2 and 3
    $var_word_feedback_3=sizeof($words_3);
    $var_word_feedback_2=sizeof($words_2);

    //retrieve information from roll_over table to determine whether student read feedback during the
process
    //make results available to students
    for($i=2;$i<4;++$i){
        $check=0;
        $ara_roll_over=array();
        $result=mysql_query("SELECT millisecond FROM roll_over where text_id='$text_id' and
footnote='$foot_number' and level='$i'") OR exit(mysql_error());
        while($row=mysql_fetch_array($result)){
            ++$check;

$ara_roll_over[$incorrectForm_id]=$ara_roll_over[$incorrectForm_id].$row['millisecond']."@";
        }
```

```
        //if data retrieval is equal to zero => 0
        //or if text not monitored => NA
        //or count
        if($text_id<42 &&
$text_id!=81){$ara_result[$incorrectForm_id]=$ara_result[$incorrectForm_id]."NA@";}
        else if($check==0){$ara_result[$incorrectForm_id]=$ara_result[$incorrectForm_id]."0@";}
        else{
            //incorrect form id = millisecond at each level
            ksort($ara_roll_over);

            foreach($ara_roll_over as $key_ara_roll_over=>$value_ara_roll_over){
                $list_time=array();
                $count=0;
                $sum=0;
                $average=0;

                //echo "each rollover:  $key_ara_roll_over=>$value_ara_roll_over<br>";
                //$key_ara_roll_over= incorrect form id

                //echo "word: ".${"ara_word_feedback_".$i}[$key_ara_roll_over]."<br>";
                $time_required=round(${"var_word_feedback_".$i}*$word_per_milli_second,0);
                //echo "time required: $time_required<br>";
                $list_time=split("@",$value_ara_roll_over);
                for($j=0;$j<sizeof($list_time);$j++){
                    if($list_time[$j]>=$time_required){
                        ++$count;
                        $sum=$sum+$list_time[$j];
                    }
                }
                //average in seconds
                if($count>0){$average=round((($sum/$count)/1000),2);}
                else{$average=0;}
            }
            $ara_result[$incorrectForm_id]=$ara_result[$incorrectForm_id].$count."@";
        }
    }
    return $ara_result;
}

//===================================================================

//close database
mysql_close();
?>
</body>
</html>
```

E.2. Balanced complexity measure

DB_vector_balanced.php 2010-09-27

```
<?php
mb_internal_encoding('UTF-8');
include 'database_var.php';
?>
<!DOCTYPE HTML PUBLIC "-//W3C//DTD HTML 4.01 Transitional//EN"
"http://www.w3.org/TR/html4/loose.dtd">
<html>
    <head>
    <meta http-equiv="Content-Type" content="text/html; charset=UTF-8">
    <META HTTP-EQUIV="Pragma" CONTENT="no-cache">
    <META Http-Equiv="Expires" Content="0">
    <META name="ROBOTS" content="NOIMAGEINDEX,NOIMAGECLICK,NOFOLLOW">
    <META name="Author" content="Sylvie THOUESNY">
    <META name="Copyright" content="Copyright 2008 onwards Sylvie THOUESNY">
    <META name="Description" content="research on errors">
    <META name="keywords" content="language learner model, learner model, sylvie thouesny, feedback">
    <link rel="stylesheet" type="text/css" href="css/styles.css">
    <title>pos-tagger</title>

    </head>
<body bgcolor="#FFFFFF">

<?php

//connect to mysql
mysql_select_db($var_database,$db) or die("Couldn't open $db: ".mysql_error());

//vector 1
$ara_type_count_in_text=array();
$ara_word_count_in_text=array();
$ara_adjusted_type_token_ratio=array();
//vector 2
$ara_letter_count_in_text=array();
$ara_word_count_in_text=array();
$ara_mean_length_word_ratio=array();
//vector 3
$ara_period_count_in_text=array();
$ara_period_unit_length=array();
//vector 4
$ara_unique_bigram=array();
$ara_total_bigram_in_text=array();

$ara_on_text=array();


//compute balanced complexity


//retrieve text id
$result=mysql_query("SELECT text_id, student_id, artefact_id, submissionDate from text") OR
exit(mysql_error());
while($row=mysql_fetch_array($result)){
    $row['submissionDate']=mb_ereg_replace('-','/',$row['submissionDate']);
    $ara_on_text[$row['text_id']]="".$row['text_id']."@".$row['student_id']."@".$row['artefact_id']."@
.$row['submissionDate']."@";
```

- 1/5 -

```
}
//foreach($ara_on_text as $key=>$value){
//  echo "$key=>$value<br>";
//}

//one text at the time
//$ara_on_text=array("4"=>"1");


foreach($ara_on_text as $key=>$value){
    $ara_lemma=array();
    $arr_temp=array();
    $arr_each_word_in_text=array();

    $result=mysql_query("SELECT token_id, token, lemma_5, text_id FROM token where text_id='$key' AND
pos_5!='PUN' AND pos_5!='PUN:cit' AND pos_5!='SENT' AND pos_5!='NUM'") OR exit(mysql_error());
    while($row=mysql_fetch_array($result)){
        ++$ara_word_count_in_text[$key];

        //vector 2
        //replace each accentuated letter by a non accentuated letter
        //strlen counts a é as 2 chars
        $var_word_without_accent=normalize($row['token']);
        $var_letter_count_word=strlen($var_word_without_accent);
        $ara_letter_count_in_text[$key]=$ara_letter_count_in_text[$key]+$var_letter_count_word;

        //vector 1
        if($row['lemma_5']!="unknown"){
            if(array_key_exists($row['lemma_5'],$ara_lemma)){
                //do nothing
            }
            else{
                $ara_lemma[$row['lemma_5']]="1";
                ++$ara_type_count_in_text[$key];
            }
        }
        else{
            if(array_key_exists($row['token'],$ara_lemma)){
                //do nothing
            }
            else{
                $ara_lemma[$row['token']]="1";
                ++$ara_type_count_in_text[$key];
            }
        }
        //vector 3
        //to chech whether the last token is a period, and if not, virtually add one
        array_push($arr_temp,$row['token_id'].$row['token']);
        //vector 4
        array_push($arr_each_word_in_text,$row['token']);
    }

    //echo "words: ".$ara_word_count_in_text[$key]."<br>";

    //vector 3
```

```
    //to chech whether the last token is a period
    sort($arr_temp);
    //echo  $arr_temp[sizeof($arr_temp)-1]."<br>";
    if(mb_ereg_match("(\.|\?|!)",$arr_temp[sizeof($arr_temp)-1])){$var_add_pun=0;}
    else{$var_add_pun=1;}
    $ara_period_count_in_text[$key]=$ara_period_count_in_text[$key]+$var_add_pun;
    //count periods in the text
    $result=mysql_query("SELECT token, lemma_5, text_id FROM token where text_id='$key' AND
(pos_5='PUN' or pos_5='PUN:cit' or pos_5='SENT')") OR exit(mysql_error());
    while($row=mysql_fetch_array($result)){
        //echo $row['token']."<br>";
        if(mb_ereg_match("(\.|\?|!)",$row['token'])){
             ++$ara_period_count_in_text[$key];
        }
    }
    //vector 4
    //substract one from the count as what was counted was the amount of words
    $ara_total_bigram_in_text[$key]=$ara_word_count_in_text[$key]-1;
    //echo "total bigram: ".$ara_total_bigram_in_text[$key]."<br>";
    //count the amount of bigram
    $ara_bigram=array();
    for($i=0;$i<sizeof($arr_each_word_in_text)-1;++$i){
        $var_bigram=$arr_each_word_in_text[$i]."@".$arr_each_word_in_text[$i+1];
        $var_bigram=mb_strtolower($var_bigram);
        $var_bigram_without_accent=normalize($var_bigram);
        //echo "bigram: $var_bigram_without_accent<br>";
        if(array_key_exists($var_bigram_without_accent,$ara_bigram)){
            //do nothing
            //echo "exists: $var_bigram_without_accent<br>";
        }
        else{
            $ara_bigram[$var_bigram_without_accent]=1;
            //echo "do not exist: $var_bigram_without_accent<br>";
             ++$ara_unique_bigram[$key];
            //echo "count: ".$ara_unique_bigram[$key]."<br>";
        }
    }
}

//adjusted type token ratio
//type/sqr 2*total words
foreach($ara_on_text as $key=>$value){

    $var_word=$ara_word_count_in_text[$key];

    //vector 1
    $var_type=$ara_type_count_in_text[$key];
    $sqr_2_word=sqrt(2*$ara_word_count_in_text[$key]);
    $ara_adjusted_type_token_ratio[$key]=$var_type/$sqr_2_word;
    $ara_on_text[$key]=$ara_on_text[$key].$ara_adjusted_type_token_ratio[$key]."@";
    $max= $ara_adjusted_type_token_ratio[$key];
    $min= $ara_adjusted_type_token_ratio[$key];
    //echo "vector 1: ".$ara_adjusted_type_token_ratio[$key]."<br>";
```

```
    //vector 2
    $var_letter=$ara_letter_count_in_text[$key];
    $ara_mean_length_word_ratio[$key]=$var_letter/$var_word;
    $ara_on_text[$key]=$ara_on_text[$key].$ara_mean_length_word_ratio[$key]."@";
    if($max<$ara_mean_length_word_ratio[$key]){$max=$ara_mean_length_word_ratio[$key];}
    if($min>$ara_mean_length_word_ratio[$key]){$min=$ara_mean_length_word_ratio[$key];}
    //echo "vector 2: ".$ara_mean_length_word_ratio[$key]."<br>";

    //vector 3
    $var_period=$ara_period_count_in_text[$key];
    $ara_period_unit_length[$key]=$var_word/$var_period;
    $ara_on_text[$key]=$ara_on_text[$key].$ara_period_unit_length[$key]."@";
    if($max<$ara_period_unit_length[$key]){$max=$ara_period_unit_length[$key];}
    if($min>$ara_period_unit_length[$key]){$min=$ara_period_unit_length[$key];}
    //echo "vector 3: ".$ara_period_unit_length[$key]."<br>";

    //vector 4
    $var_unique_bigram=$ara_unique_bigram[$key];
    $sqr_2_bigram=sqrt(2*$ara_total_bigram_in_text[$key]);
    $ara_unique_bigram_ratio[$key]=$var_unique_bigram/$sqr_2_bigram;
    $ara_on_text[$key]=$ara_on_text[$key].$ara_unique_bigram_ratio[$key]."@";
    if($max<$ara_unique_bigram_ratio[$key]){$max=$ara_unique_bigram_ratio[$key];}
    if($min>$ara_unique_bigram_ratio[$key]){$min=$ara_unique_bigram_ratio[$key];}
    //echo "vector 4: ".$ara_unique_bigram_ratio[$key]."<br>";

    //balance measure
    $var_v1=$ara_adjusted_type_token_ratio[$key];
    $var_v2=$ara_mean_length_word_ratio[$key];
    $var_v3=$ara_period_unit_length[$key];
    $var_v4=$ara_unique_bigram_ratio[$key];
    //echo "max:$max - min:$min<br>";
    $ara_balanced_measure[$key]=(($var_v1-1)+($var_v2-(1/$sqr_2_word))+($var_v3-(1/$sqr_2_word))+($var
v4-1))-($max-$min);
    $ara_on_text[$key]=$ara_on_text[$key].$ara_balanced_measure[$key]."@";
    //echo "vector balanced: ".$ara_balanced_measure[$key]."<br>";
}

echo "text_id@student_id@artefact_id@submissionDate@vector_1@vector_2@vector_3@vector_4@balanced@<br>";
foreach($ara_on_text as $key=>$value){
    echo $ara_on_text[$key]."<br>";
}

function normalize ($string) {
    $table = array(
        'Š'=>'S', 'š'=>'s', 'Đ'=>'Dj', 'đ'=>'dj', 'Ž'=>'Z', 'ž'=>'z', 'Č'=>'C', 'č'=>'c', 'Ć'=>'C',
'ć'=>'c',
        'À'=>'A', 'Á'=>'A', 'Â'=>'A', 'Ã'=>'A', 'Ä'=>'A', 'Å'=>'A', 'Æ'=>'A', 'Ç'=>'C', 'È'=>'E',
'É'=>'E',
        'Ê'=>'E', 'Ë'=>'E', 'Ì'=>'I', 'Í'=>'I', 'Î'=>'I', 'Ï'=>'I', 'Ñ'=>'N', 'Ò'=>'O', 'Ó'=>'O',
'Ô'=>'O',
        'Õ'=>'O', 'Ö'=>'O', 'Ø'=>'O', 'Ù'=>'U', 'Ú'=>'U', 'Û'=>'U', 'Ü'=>'U', 'Ý'=>'Y', 'Þ'=>'B',
'ß'=>'Ss',
        'à'=>'a', 'á'=>'a', 'â'=>'a', 'ã'=>'a', 'ä'=>'a', 'å'=>'a', 'æ'=>'a', 'ç'=>'c', 'è'=>'e',
'é'=>'e',
        'ê'=>'e', 'ë'=>'e', 'ì'=>'i', 'í'=>'i', 'î'=>'i', 'ï'=>'i', 'ð'=>'o', 'ñ'=>'n', 'ò'=>'o',
```

```
'ó'=>'o',
        'ô'=>'o', 'õ'=>'o', 'ö'=>'o', 'ø'=>'o', 'ù'=>'u', 'ú'=>'u', 'û'=>'u', 'ý'=>'y', 'ý'=>'y',
'þ'=>'b',
        'ÿ'=>'y', 'Ŕ'=>'R', 'ŕ'=>'r', '\''=>'', '"'=>'',
    );

    return strtr($string, $table);
}


//close database
mysql_close();
?>
</body>
</html>
```

E.3. Repeated measures: wiki and bilan

Within-Subjects Factors

Measure:MEASURE_1

type	Dependent Variable
1	bilan
2	wiki

Descriptive Statistics

	Mean	Std. Deviation	N
bilan	24.2900	1.35481	6
wiki	21.8550	2.72737	6

Multivariate Tests[b]

Effect		Value	F	Hypothesis df	Error df	Sig.
type	Pillai's Trace	.436	3.860[a]	1.000	5.000	.107
	Wilks' Lambda	.564	3.860[a]	1.000	5.000	.107
	Hotelling's Trace	.772	3.860[a]	1.000	5.000	.107
	Roy's Largest Root	.772	3.860[a]	1.000	5.000	.107

a. Exact statistic
b. Design: Intercept
Within Subjects Design: type

Tests of Within-Subjects Effects

Measure:MEASURE_1

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
type	Sphericity Assumed	17.788	1	17.788	3.860	.107
	Greenhouse-Geisser	17.788	1.000	17.788	3.860	.107
	Huynh-Feldt	17.788	1.000	17.788	3.860	.107
	Lower-bound	17.788	1.000	17.788	3.860	.107
Error(type)	Sphericity Assumed	23.039	5	4.608		
	Greenhouse-Geisser	23.039	5.000	4.608		
	Huynh-Feldt	23.039	5.000	4.608		
	Lower-bound	23.039	5.000	4.608		

Tests of Within-Subjects Contrasts

Measure:MEASURE_1

Source	type	Type III Sum of Squares	df	Mean Square	F	Sig.
type	Level 1 vs. Level 2	35.575	1	35.575	3.860	.107
Error(type)	Level 1 vs. Level 2	46.078	5	9.216		

Appendix F. Synchronic analyses

F.1. Method to estimate learners’ amount of mistakes and errors

Following the discussion on learners’ responses to interventions, we know that even if learners do not access assistance for 53% of all incorrect forms, they nevertheless propose an alternative for many of them (71.5%). From these 71.5% attempted, 75.49% will be correct. For example, learner #2 wrote 29 incorrect forms, the feedback of 15 of these will hypothetically never be opened. From these 15 incorrect forms, 10 (71.5%) will be attempted. From these 10 incorrect forms attempted without reading the feedback, 7 (75.49%) will be correct, that is, there will be 7 mistakes within this learner’s incorrect forms, which corresponds to the learner’s real figure. Extending the experiment further, 7 mistakes over 15 incorrect forms, from which the feedback will not be accessed means that the 8 remaining incorrect forms will be errors without potential development. Knowing that the learner will read the feedback of 14 incorrect forms, the percentage of attempts when the feedback is read equals 74.9%. Consequently, learner #2 will provide alternatives to 10 incorrect forms. From these 10 alternatives proposed with assistance, 65.1% of them (6 in total for selection error types) will be correct. These 6 correct alternatives will refer to errors with potential development, as they will be corrected with assistance. As a result, the 8 remaining incorrect forms attempted or not with assistance, and incorrect if tried, will be considered as errors without potential development. The count thus far is 7 mistakes, 6 errors *with* potential development, and 16 errors *without* potential development; in total 29 incorrect forms in selection error types, which reflects the real situation of the example given here. However, this calculation is not accurate enough for most learners, which stresses the idea of individual learner analysis.

F.2. Learner #16’s incorrect pronoun antecedent agreement

token_id	text_id	student_id	genre_id	submissionD...	pos_5	error_tagged_as	status	P	K1	K2	K3	K4	PD
12739	53	16	bilan	2008/12/13	PRO:PER	mo_ag_pa_	1	1	0	0	0	0	0
12499	52	16	wiki	2008/12/13	PRO:PER	mo_ag_pa_	0	1	1	1	1	1	0
12878	53	16	bilan	2008/12/13	PRO:PER	mo_ag_pa_	-1	1	1	1	1	1	1

incorrectForm_id	level	student_id	text_id	incorrect_seq	errorType...	stu_corr	alternative_provided	fb_final	error_type_solved	average	count	token_id
1272	1	16	52	il	mo_ag_pa_	ils	1	elle	0	NA	.	12499
1294	1	16	53	il	mo_ag_pa_	qu'ils	1	qu'ils aident	1	NA	.	12739
1304	1	16	53	il	mo_ag_pa_	que ca	1	qu'elle	0	NA	.	12878
1272	2	16	52	il	mo_ag_pa_	ils	1	elle	0	2.54	4	12499
1294	2	16	53	il	mo_ag_pa_	ils	1	qu'ils aident	1	2.65	1	12739
1304	2	16	53	il	mo_ag_pa_		0	qu'elle	0	2.09	1	12878
1272	3	16	52	il	mo_ag_pa_	elle	1	elle	1	7.9	1	12499
1294	3	16	53	il	mo_ag_pa_		0	qu'ils aident	0	4.35	1	12739
1304	3	16	53	il	mo_ag_pa_	ca	1	qu'elle	0	0	0	12878

F.3. Language accuracy in performance, actual and potential development: selection error types

Student		Performance		Knowledge		Potential development	
		Frequency	Percent	Frequency	Percent	Frequency	Percent
#2	correct	798	96.7	804	97.5	814	98.7
	incorrect	27	3.3	21	2.5	11	1.3
	total	825	100.0	825	100.0	825	100.0
#5	correct	354	94.4	361	96.3	361	96.3
	incorrect	21	5.6	14	3.7	14	3.7
	total	375	100.0	375	100.0	375	100.0
#6	correct	1008	92.2	1013	92.7	1017	93.0
	incorrect	85	7.8	80	7.3	76	7.0
	total	1093	100.0	1093	100.0	1093	100.0
#7	correct	838	92.4	849	93.6	861	94.9
	incorrect	69	7.6	58	6.4	46	5.1
	total	907	100.0	907	100.0	907	100.0
#8	correct	690	92.0	692	92.3	692	92.3
	incorrect	60	8.0	58	7.7	58	7.7
	total	750	100.0	750	100.0	750	100.0
#9	correct	357	93.0	370	96.4	370	96.4
	incorrect	27	7.0	14	3.6	14	3.6
	total	384	100.0	384	100.0	384	100.0
#10	correct	724	96.0	731	96.9	736	97.6
	incorrect	30	4.0	23	3.1	18	2.4
	total	754	100.0	754	100.0	754	100.0
#11	correct	597	95.7	614	98.4	616	98.7
	incorrect	27	4.3	10	1.6	8	1.3
	total	624	100.0	624	100.0	624	100.0
#12	correct	293	91.8	300	94.0	309	96.9
	incorrect	26	8.2	19	6.0	10	3.1
	total	319	100.0	319	100.0	319	100.0
#14	correct	284	91.0	289	92.6	297	95.2
	incorrect	28	9.0	23	7.4	15	4.8
	total	312	100.0	312	100.0	312	100.0

#15	correct	316	93.2	326	96.2	329	97.1
	incorrect	23	6.8	13	3.8	10	2.9
	total	339	100.0	339	100.0	339	100.0
#16	correct	824	93.4	846	95.9	848	96.1
	incorrect	58	6.6	36	4.1	34	3.9
	total	882	100.0	882	100.0	882	100.0
#17	correct	663	90.9	688	94.4	691	94.8
	incorrect	66	9.1	41	5.6	38	5.2
	total	729	100.0	729	100.0	729	100.0
#19	correct	370	97.4	374	98.4	374	98.4
	incorrect	10	2.6	6	1.6	6	1.6
	total	380	100.0	380	100.0	380	100.0

F.4. Language accuracy in performance, actual and potential development: syntactic error types

Student		Performance		Knowledge		Potential development	
		Frequency	Percent	Frequency	Percent	Frequency	Percent
#2	correct	798	92.5	810	93.9	828	95.9
	incorrect	65	7.5	53	6.1	35	4.1
	total	863	100.0	863	100.0	863	100.0
#5	correct	354	97.3	354	97.3	354	97.3
	incorrect	10	2.7	10	2.7	10	2.7
	total	364	100.0	364	100.0	364	100.0
#6	correct	1008	91.3	1012	91.7	1019	92.3
	incorrect	96	8.7	92	8.3	85	7.7
	total	1104	100.0	1104	100.0	1104	100.0
#7	correct	838	91.9	852	93.4	861	94.4
	incorrect	74	8.1	60	6.6	51	5.6
	total	912	100.0	912	100.0	912	100.0
#8	correct	690	95.7	692	96.0	692	96.0
	incorrect	31	4.3	29	4.0	29	4.0
	total	721	100.0	721	100.0	721	100.0
#9	correct	357	96.5	360	97.3	360	97.3
	incorrect	13	3.5	10	2.7	10	2.7
	total	370	100.0	370	100.0	370	100.0
#10	correct	724	96.0	729	96.7	732	97.1
	incorrect	30	4.0	25	3.3	22	2.9
	total	754	100.0	754	100.0	754	100.0
#11	correct	597	98.4	601	99.0	601	99.0
	incorrect	10	1.6	6	1.0	6	1.0
	total	607	100.0	607	100.0	607	100.0
#12	correct	293	90.4	293	90.4	294	90.7
	incorrect	31	9.6	31	9.6	30	9.3
	total	324	100.0	324	100.0	324	100.0

#14	correct	284	95.6	290	97.6	291	98.0
	incorrect	13	4.4	7	2.4	6	2.0
	total	297	100.0	297	100.0	297	100.0
#15	correct	316	96.6	320	97.9	320	97.9
	incorrect	11	3.4	7	2.1	7	2.1
	total	327	100.0	327	100.0	327	100.0
#16	correct	824	91.6	826	91.8	837	93.0
	incorrect	76	8.4	74	8.2	63	7.0
	total	900	100.0	900	100.0	900	100.0
#17	correct	663	90.8	667	91.4	667	91.4
	incorrect	67	9.2	63	8.6	63	8.6
	total	730	100.0	730	100.0	730	100.0
#19	correct	370	96.1	385	100.0	385	100.0
	incorrect	15	3.9	-	-	-	-
	total	385	100.0	385	100.0	385	100.0

F.5. Language accuracy in performance, actual and potential development: morphosyntactic error types

Student		Performance		Knowledge		Potential development	
		Frequency	Percent	Frequency	Percent	Frequency	Percent
#2	correct	798	97.4	808	98.7	811	99.0
	incorrect	21	2.6	11	1.3	8	1.0
	total	819	100.0	819	100.0	819	100.0
#5	correct	354	97.0	358	98.1	358	98.1
	incorrect	11	3.0	7	1.9	7	1.9
	total	365	100.0	365	100.0	365	100.0
#6	correct	1008	97.5	1011	97.8	1014	98.1
	incorrect	26	2.5	23	2.2	20	1.9
	total	1034	100.0	1034	100.0	1034	100.0
#7	correct	838	96.9	843	97.5	847	97.9
	incorrect	27	3.1	22	2.5	18	2.1
	total	865	100.0	865	100.0	865	100.0
#8	correct	690	97.7	692	98.0	692	98.0
	incorrect	16	2.3	14	2.0	14	2.0
	total	706	100.0	706	100.0	706	100.0
#9	correct	357	93.9	369	97.1	369	97.1
	incorrect	23	6.1	11	2.9	11	2.9
	total	380	100.0	380	100.0	380	100.0
#10	correct	724	96.4	739	98.4	741	98.7
	incorrect	27	3.6	12	1.6	10	1.3
	total	751	100.0	751	100.0	751	100.0
#11	correct	597	98.5	601	99.2	602	99.3
	incorrect	9	1.5	5	.8	4	.7
	total	606	100.0	606	100.0	606	100.0

#12	correct	293	98.0	296	99.0	298	99.7
	incorrect	6	2.0	3	1.0	1	.3
	total	299	100.0	299	100.0	299	100.0
#14	correct	284	96.3	288	97.6	293	99.3
	incorrect	11	3.7	7	2.4	2	.7
	total	295	100.0	295	100.0	295	100.0
#15	correct	316	98.1	319	99.1	319	99.1
	incorrect	6	1.9	3	.9	3	.9
	total	322	100.0	322	100.0	322	100.0
#16	correct	824	98.4	832	99.4	833	99.5
	incorrect	13	1.6	5	.6	4	.5
	total	837	100.0	837	100.0	837	100.0
#17	correct	663	95.7	679	98.0	682	98.4
	incorrect	30	4.3	14	2.0	11	1.6
	total	693	100.0	693	100.0	693	100.0
#19	correct	370	98.4	375	99.7	376	100.0
	incorrect	6	1.6	1	.3	-	-
	total	376	100.0	376	100.0	376	100.0

Appendix G. Diachronic analyses

G.1. T-test for independent samples: feedback access between graded and non-graded texts

SPSS t-test for independent samples demonstrated no significant differences between the *non-graded* (M=1.26, SD=0.959) and *graded* (M=1.35, SD=1.07) groups; (*t-test*(1240)=-1.223, *p*=.222, *p*>.05).

Group statistics for the balanced complexity measure

Text type group	N	Mean	Std. Deviation	Std. Error Mean
non-graded	224	1.26	.959	.064
graded	1018	1.35	1.068	.033

A Sig. value of 0.214 > 0.05 designates that the variability in-between the two groups is almost identical. Since the variability is not significant, the result of the t-test is obtained with the reading of the row that stipulates *equal variances assumed*.

Independent samples test

	Levene's Test for Equality of Variances		T-test for Equality of Means				
balanced measure	**F**	**Sig.**	**t**	**df**	**Sig. (2-tailed)**	**Mean Difference**	**Std. Error Difference**
Equal variances assumed	1.547	.214	-1.223	1240	.222	-.095	.077
Equal variances not assumed			-1.310	355.672	.191	-.095	.072

A Sig value (2-tailed) of 0.222 > 0.05 indicates that the means for the two text groups are not statistically different.

G.2. Descriptive statistics: exploring the effect of text types on balanced complexity measures

Descriptives[a]

	genre_id			Statistic	Std. Error
balanced	forum	Mean		14.5750	1.71407
		95% Confidence Interval for Mean	Lower Bound	10.3808	
			Upper Bound	18.7692	
		5% Trimmed Mean		14.4552	
		Median		12.1987	
		Variance		20.566	
		Std. Deviation		4.53499	
		Minimum		10.00	
		Maximum		21.31	
		Range		11.30	
		Interquartile Range		8.92	
		Skewness		.738	.794
		Kurtosis		-1.394	1.587
	bilan	Mean		24.2625	1.45388
		95% Confidence Interval for Mean	Lower Bound	19.6356	
			Upper Bound	28.8893	
		5% Trimmed Mean		24.2835	
		Median		24.4519	
		Variance		8.455	
		Std. Deviation		2.90775	
		Minimum		20.61	
		Maximum		27.54	
		Range		6.93	
		Interquartile Range		5.59	
		Skewness		-.351	1.014
		Kurtosis		.230	2.619
	email	Mean		12.7000	.45638
		95% Confidence Interval for Mean	Lower Bound	11.7581	
			Upper Bound	13.6420	
		5% Trimmed Mean		12.6543	
		Median		12.5271	
		Variance		5.207	
		Std. Deviation		2.28192	
		Minimum		8.40	
		Maximum		17.86	
		Range		9.46	
		Interquartile Range		2.75	
		Skewness		.392	.464
		Kurtosis		.266	.902
	écrire	Mean		21.9788	1.91921
		95% Confidence Interval for Mean	Lower Bound	17.0453	
			Upper Bound	26.9123	
		5% Trimmed Mean		22.1574	
		Median		22.9820	
		Variance		22.100	
		Std. Deviation		4.70110	
		Minimum		14.28	
		Maximum		26.46	
		Range		12.18	
		Interquartile Range		8.54	
		Skewness		-.896	.845
		Kurtosis		-.016	1.741
	wiki	Mean		20.5537	1.85378
		95% Confidence Interval for Mean	Lower Bound	15.7884	
			Upper Bound	25.3190	
		5% Trimmed Mean		20.8714	
		Median		22.4877	
		Variance		20.619	
		Std. Deviation		4.54082	
		Minimum		11.75	
		Maximum		23.64	
		Range		11.90	
		Interquartile Range		5.61	
		Skewness		-1.978	.845
		Kurtosis		3.965	1.741

a. balanced is constant when genre_id = parler. It has been omitted.

Case Processing Summary

	genre_id	Valid N	Valid Percent	Missing N	Missing Percent	Total N	Total Percent
balanced	forum	7	100.0%	0	.0%	7	100.0%
	bilan	4	100.0%	0	.0%	4	100.0%
	email	25	100.0%	0	.0%	25	100.0%
	écrire	6	100.0%	0	.0%	6	100.0%
	wiki	6	100.0%	0	.0%	6	100.0%
	parler	1	100.0%	0	.0%	1	100.0%

M-Estimators[e]

	genre_id	Huber's M-Estimator[a]	Tukey's Biweight[b]	Hampel's M-Estimator[c]	Andrews' Wave[d]
balanced	forum	13.3775	12.5336	13.5999	12.4470
	bilan	24.4519	24.4016	24.4005	24.4010
	email	12.5854	12.4829	12.5469	12.4779
	écrire	22.6183	22.5789	22.3802	22.5749
	wiki	22.3244	22.7030	22.5500	22.7002

a. The weighting constant is 1.339.
b. The weighting constant is 4.685.
c. The weighting constants are 1.700, 3.400, and 8.500
d. The weighting constant is 1.340*pi.
e. balanced is constant when genre_id = parler. It has been omitted.

Percentiles[a]

		genre_id	5	10	25	50	75	90	95
Weighted Average (Definition 1)	balanced	forum	10.0009	10.0009	11.1210	12.1987	20.0443	.	.
		bilan	20.6081	20.6081	21.3746	24.4519	26.9609	.	.
		email	8.6835	9.6279	11.1274	12.5271	13.8804	16.7052	17.5225
		écrire	14.2802	14.2802	17.7578	22.9820	26.3027	.	.
		wiki	11.7455	11.7455	17.7273	22.4877	23.3386	.	.
Tukey's Hinges	balanced	forum			11.3943	12.1987	17.8659		
		bilan			22.1411	24.4519	26.3838		
		email			11.3243	12.5271	13.8678		
		écrire			18.9170	22.9820	26.2495		
		wiki			19.7212	22.4877	23.2373		

a. balanced is constant when genre_id = parler. It has been omitted.

G.3. Amount of errors and mistakes confirmed and not in text types

Time	Text types	Correct tokens	Incorrect tokens	Error with no potential development	Error	Mistake	Error not confirmed	Mistake not confirmed
1	forum	511	105	39	-	37	29	-
	bilan	140	28	15	-	11	2	-
	email	104	37	11	-	11	15	-
2	bilan	417	117	58	-	43	14	2
	email	108	46	23	-	17	6	-
	écrire	375	82	38	-	35	9	-
	wiki	920	170	93	-	51	26	-
3	bilan	374	111	50	24	37	-	-
	email	149	39	10	9	20	-	-
	wiki	424	57	10	14	33	-	-
	parler	608	113	63	28	22	-	-
4	email	168	51	15	18	18	-	-
	écrire	47	6	3	1	2	-	-
5	email	75	18	3	7	8	-	-
	écrire	602	207	95	31	81	-	-
6	bilan	534	137	79	15	43	-	-
	email	158	47	19	10	18	-	-
	écrire	387	113	57	25	31	-	-