The Development of Oral Fluency and Rhythm during a Study Abroad Period

Margalida Valls Ferrer

TESI DOCTORAL UPF / 2011

DIRECTORS DE LA TESI

Dra. Carmen Pérez Vidal (Departament de Traducció i Ciències del Llenguatge, Universitat Pompeu Fabra)

Dr. Joan C. Mora (Departament de Filologia Anglesa i Alemanya, Universitat de Barcelona)

DEPARTAMENT DE TRADUCCIÓ I CIÈNCIES DEL LLENGUATGE

UNIVERSITAT POMPEU FABRA

A mon pare i ma mare

A en Marc

Acknowledgments

First and foremost I want to thank my two PhD advisors, Carmen Pérez Vidal and Joan Carles Mora. The faith and hope they had in me from the very beginning and until the very last stages of writing this dissertation is most appreciated. I am most grateful to Carmen for giving me the opportunity to become an assistant researcher and for her unconditional support and guidance through the completion of this PhD. Her continued encouragement, suggestions and advice during these years have been invaluable. Likewise, my most sincere thank you to Joan Carles for his generous dedication and the time spent reading and re-reading drafts of this dissertation. His genuine enthusiasm, inspiration and great efforts to explain things clearly and simply, have been crucial for the development of my academic work.

This dissertation has highly benefited from the two research visits granted during the last three years. Thank you to the School of Languages, Literatures and Cultures at the University of Maryland, and in special to Prof. Robert DeKeyser for giving me the opportunity to attend the highly interesting and challenging PhD courses he and Prof. M. Long were teaching when I was there. They were most helpful to lay the foundations for this work. Regarding the second research visit, a very special thanks to all the people in the Department of Speech, Hearing and Phonetic Sciences at the University College London. I am particularly indebted to Dr. Volker Dellwo for his invaluable assistance with the analysis of rhythm and his overall generous help.

I gratefully acknowledge the funding sources that made these visits and all my PhD work possible. Thank you to the Department of Translation and Language Sciences at the Universitat Pompeu Fabra and the ALLENCAM research group for funding this research and for enabling me to attend a great number of international conferences. I am most grateful to the Ministerio de Educación for funding my research visit to the UCL, which made my PhD eligible for European mention.

All these years I have had the privilege to share my time, experiences and learning with many people at the UPF, from the supportive head of the department, Enric Vallduví, to my wonderful colleagues Helena, Teodora, Gilmar, Germán, Cristina, Diana, Pilar, Rebecca and many others. It has been a special pleasure to share most interesting and fruitful discussions with Elisa Barquin in the many hours we spent in the office. A big thanks to all of them for making my time at the UPF memorable.

I am extremely grateful to all the members of the SALA project and the ALLENCAM research group, especially Maria Juan and Aurora Bel, for

their encouragement and support. Furthermore, I would also like to thank the students who participated in the three data collection times that made this study possible.

On a more personal basis, I would like to express my most sincere gratitude to Dr. Ma. Antònia Pujol, whose stories shared during summer breaks in Ses Covetes first motivated my interest on academia.

For their love, trust and encouragement I heartily thank my family in Mallorca and Barcelona. I especially wish to thank my husband Marc for bringing into my life his ever contagious optimism. Thank you for your encouragement, patience and enthusiastic support during the final stages of this PhD.

Finalment, agraïr als meus pares, Miquel i Francisca, la confiança i el suport incondicional que sempre m'han demostrat. Si he arribat aquí és gràcies a ells. Gràcies per ensenyar-me que amb dedicació, esforç i constància es pot aconseguir qualsevol cosa. Però sobretot gràcies per animar-me i donar-me l'oportunitat de "fer lo que m'agrada".

Margalida Valls Ferrer
Barcelona, July 2011

Abstract

The present study examines the development of L2 fluency and rhythm in the extemporaneous speech of advanced EFL learners during a study abroad (SA) period, and the relationship between utterance fluency and rhythm and listeners' perceived fluency. Perceived fluency is also assessed in relation to the listeners' backgrounds, as well as the effect of learners' *initial fluency levels* (pre-SA fluency level) on SA outcomes in both fluency and rhythm. Speech samples were collected at different points in time before and after a 6-month formal instruction (FI) and a 3-month SA period. Subsequently, these speech samples were analyzed for changes in utterance fluency, perceived fluency and rhythm between data collection times. Further analyses of the relationships between these perceptual and productive dimensions, as well as the impact of participants' *initial fluency levels* were undertaken. Both fluency and rhythm, operationalized as temporal fluency measures and rhythm metrics, were found to improve significantly during the SA period. Similarly, significant gains in fluency during the SA were perceived by 69 listeners of different L1 backgrounds and degrees of experience with Catalan/Spanish-accented speech. Furthermore, scores in the temporal fluency measures, rhythm metrics, and listeners' fluency ratings were found to be related to one another. Mean length of run, a temporal fluency measure which encompasses both speed and breakdown fluency, as well as two rhythm measures (%V and varcoC) accounted for a substantial amount of variation in perceived fluency. Finally, the *initial fluency level* was found to significantly predict fluency gains during SA. Overall, the findings in this study suggest that a 3-month SA period is highly beneficial for advanced EFL learners' development of L2 fluency and rhythm.

Resum

Aquest estudi investiga el desenvolupament de la fluïdesa oral i del ritme en la parla extemporània d'estudiants d'anglès de nivell avançat durant una estada a l'estranger (EE), així com la relació entre fluïdesa oral i ritme, i la percepció de la fluïdesa. Així mateix, també s'avalua la percepció de la fluïdesa segons les característiques de l'oient, i el nivell inicial de fluïdesa oral (pre-estada a l'estranger). Les mostres de parla es van recollir en diferents temps: abans i després d'un període d'instrucció formal (IF) de 6 mesos i d'una estada a l'estranger de 3 mesos. L'anàlisi longitudinal d'aquestes mostres de parla permet observar canvis en la fluïdesa oral, el ritme i la percepció de la fluïdesa durant el període de temps analitzat. Igualment, s'examina la relació entre les dimensions productives i la perceptiva, així com l'impacte del nivell inicial de fluïdesa oral. Els

resultats revelen una millora significativa en la fluïdesa i el ritme durant el període de 3 mesos a l'estranger. Aquesta millora és percebuda per un grup d'oients independentment de la primera llengua i de la seva experiència amb parla amb accent català/castellà. A més, s'observa una relació entre els resultats de les mesures temporals de fluïdesa, les mesures de ritme, i les valoracions dels oients. Finalment, el nivell inicial de fluïdesa explica en gran mesura els guanys en fluïdesa obtinguts durant l'EE. Els resultats d'aquest estudi indiquen que una EE de tres mesos té un efecte de millora en el desenvolupament de la fluïdesa oral i del ritme d'una L2 en estudiants d'anglès de nivell avançat.

Table of Contents

List of Abbreviations

ACT: Adaptive Control of Thought (theory)
AH: At home
AOA: Age of Arrival
AOL: Age of Learning
AR: Articulation Rate
CLIL: Content Language Integrated Learning
DysRat: Dysfluency Ratio
EFL: English as a Foreign Language
EU: European Union
FA: Foreign Accent
FI: Formal Instruction
FL: Fluency
IL: Interlanguage
IM: Immersion Programs
L1: first language
L2: second language (also third language and so forth)
LCP: Language Contact Profile
LOR: length of Residence
MLoR: Mean Length of Runs
NNS: Non-native Speaker
nPVI-V: normalized Pairwise Variability Index
NS: Native Speaker
NT: Naturalistic Contexts
Pause_i_Dur: Pause Internal Duration
PauseDur: Pause Duration
PauseFreq: Pause Frequency
PhonRat: Phonation Time Ratio
rPVI-C: raw Pairwise Variability Index
RT: Reaction Time
SA: Study Abroad
SALA: Stay Abroad and Language Acquisition (project)
SLA: Second Language Acquisition
SD: Standard Deviation
SFL: Spanish as a Foreign Language

SR: Speech Rate
TL: Target Language
UPF: Universitat Pompeu Fabra
varcoC: rate-normalize standard deviation of consonantal interval
varcoV: rate-normalize standard deviation of vocalic interval
VOT: Voice Onset Time
%V: percentage vowel
ΔC: standard deviation of duration of consonantal interval
ΔV: standard deviation of duration of vocalic interval

Introduction

In a globalized world, universities world-wide are increasingly encouraging their students to follow Study Abroad programs to complete their education, especially those majoring in languages or language related studies. In the European context, such programs are geared towards implementing the Schengen Treaty which establishes mobility of students and workers. That is, national and international linguistic policies have been developed to promote the circulation of citizens across national borders.

The EU has set a series of guidelines for the promotion of multilingualism among its citizens. Within the measures adopted to fulfill the objectives on multilingualism, study abroad (SA) programs such as the Erasmus scheme have flourished with an ever-growing number of students going abroad each year.

This current social and political situation has prompted an increased interest in what happens to students' linguistic skills when going abroad. Within this framework, studies on second language acquisition (SLA) have become paramount in order to throw light on this issue. In a reciprocal way, SLA research currently benefits from the number of students who embark on SA by gaining opportunities to test central factors in SLA in a fairly under-researched learning context, SA, in comparison to the larger number of research studies being undertaken within formal instruction (FI) contexts.

The SA learning context has certain characteristics which make it unique. It promotes the development of specific language abilities and skills (L2[1] fluency and sociopragmatic competence, among others) which are generally difficult to teach in FI. Language teaching and learning in FI tends to focus on certain linguistic aspects such as grammar and metalinguistic awareness. This does not imply that the learning of certain features is context specific, but rather that the linguistic experience that learners undergo in each context is different, to the extent that different linguistic outcomes can be expected for each learning context.

In the past 20 years, SLA research has been exploring learning contexts other than FI. Regarding the SA context, however, conclusive evidence is by and large scarce. The main finding so far regarding the role of SA on SLA is that it facilitates the development of oral production skills, especially fluency. However, fine-grained studies on what these gains in fluency mean are yet to be undertaken. Other findings from previous research identify listening skills and vocabulary acquisition as areas of notable gains, whereas grammatical and phonological accuracy seem to benefit more from FI.

Notwithstanding the aforementioned findings, SA research arguably still has a long way to go towards tackling the main issues in the SLA research field. However, as a different learning context, SA is a promising area of research that might provide a broader understanding of these central aspects in SLA, as some authors have already pointed out (Collentine & Freed, 2004; DeKeyser, 2007; Collentine, 2009), such as the role of age,

[1] For the purpose of simplicity, L2 is used throughout the dissertation to refer to the non-native language –in some cases it might be a second language, a third, fourth etc. For the participants in this study, L2 refers to the English language they have been learning after their native Catalan and Spanish.

length of residence, input, cognitive factors and developmental threshold level, which remain largely under-researched within this learning context.

As briefly mentioned above, fluency has often been claimed to be the main winner when SA gains in linguistic skills and knowledge are assessed. Indeed, this claim is put forward by a fairly large number of studies documenting these gains (Lennon, 1990; Freed, 1995b; Towell, Hawkins & Bazergui, 1996; Segalowitz & Freed, 2004). Although in many cases ground-breaking, these early L2 fluency studies, especially those undertaken before the year 2000, often suffer from methodological shortcomings, mainly because they are often based on a small sample of participants, or the technical aids to identify pauses were almost inexistent, or they made limited use of statistical analyses. Moreover, most of these studies have relied on measures of utterance fluency to assess gains in the SA context, leaving perceived fluency under-researched. This is an area which one would expect to be extremely important as the listeners' impressions would seem to be highly relevant when evaluating the success of SA programs.

The study of fluency has been important in SA research when seeking a better understanding of the fluency construct (in a context where changes in fluency abound). Fluency is equally important from the practical perspective of the learner trying to cope with the communicative demands of language use in conversational interactions because "it facilitates natural and successful interaction between L2 users and the NS and NNS of their communities." (Rossiter, 2009:408). Thus, fluency is important for the integration of L2 users into the target language (TL) community.

Factors contributing to L2 fluency development include speakers' proficiency level, degree of exposure to the target language, and cognitive factors, among others (Segalowitz et al., 1998). Kormos & Dénes (2004)

suggest that fluency research suffers from a lack of studies investigating the combination of temporal variables with other type of variables such as linguistic, interactional, and phonological ones. Little research has been conducted in the area of phonological aspects of fluency up to this moment (exceptions are Hieke, 1984; Wennestrom, 2000; Vanderplank, 1993). And yet, it seems plausible to think that the changes in the quality of learners' productions in relation to fluency are also likely to occur at the suprasegmental level.

In this same vein, Trofimovich & Baker (2006) call for more studies on suprasegmentals in different learning contexts, such as naturalistic and SA, in order to shed light on the relationship between suprasegmental learning and L2 experience. They should, above all, contribute to a fine-grained definition of L2 experience and an accurate account of the relationship between L2 experience and specific aspects of L2 phonology.

Further research has also been called for on the effects of SA on advanced learners in relation to the developmental threshold level at which SA seems to become more effective (Collentine, 2009) as well as well-designed longitudinal studies aimed at the investigation of L2 development (Ortega & Iberri-Shea, 2005).

The present study aims to contribute to the understanding of the development of fluency and rhythm during a period of 15 months including two different learning contexts: a 6-month period of FI at home followed by a 3-month SA period. The study is conducted on a group of bilingual Catalan/Spanish university advanced EFL learners.

More specifically, we investigate whether a significant change in the fluency of non-native speakers (NNSs) during a 3-month SA period, as captured by utterance fluency measures (i.e. speed and breakdown fluency measures), is also perceived by 5 distinct groups of listeners differing in L1

(English vs. Catalan/Spanish, and experience with Catalan/Spanish accented speech). We also undertake the analysis of temporal variables underlying L2 fluency with a special focus on one prosodic feature.

That is, in addition to utterance fluency, we want to contribute to the understanding of L2 suprasegmental learning, by examining whether the 3-month SA period can also instigate a change in NNSs' rhythmic performance towards target language norms. This is something which would otherwise be very unlikely to happen in the regular classroom without specific instruction on suprasegmentals.

Moreover, this study explores the relationship between utterance fluency and rhythm and perceived fluency. Rhythm is thus regarded as a phonological aspect of fluency that deserves to be investigated in detail, just as Hieke's (1984) analysis of linking phenomena and Wennerstrom's (2000) intonational analysis. Finally, the impact of *initial fluency level* on SA is assessed.

For the purposes of this study, we consider L2 fluency to be an automatic procedural skill (Schmidt, 1992) on the part of the speaker and a perceptual phenomenon in the listener (Derwing et al., 2004). English being the language under investigation, we deemed important that the raters were not only NSs but also NNSs of the TL, since learners most probably will be involved in conversations with both NSs and NNSs of the language (Derwing & Rossiter, 2003:14).

In short, the present study aims to shed new light on the SLA field with a longitudinal study examining the development of fluency and rhythm in advanced learners in two different learning contexts. This study differs from most previous studies on fluency in SA, and purports to make a step forward in the general field of SLA. As regards SLA research, it seeks to make two contributions. First, the listeners' perceptions are used to assess

fluency gains during the period abroad. Moreover, not only do NSs take part as listeners, but also NNSs of different backgrounds. Second, the development of rhythm during a SA is assessed and suggested as a phonological fluency variable which would affect listeners' fluency ratings. As regards SA research focusing on fluency, our study intends to contribute to the ongoing debate on the existence of a threshold level to (maximally) benefit from the SA experience.

The dissertation is divided into two parts. Following this introduction, the first part provides an overview of the most relevant literature the study builds on. Subsequently, the second part presents the empirical study undertaken in order to answer the research questions.

Once the arguments which have motivated the study have been presented and the main objectives outlined, Part I follows with the theoretical background overview in Chapter 1, devoted to the study of learning contexts within SLA research. The first part of the chapter provides a thorough conceptualization of the different learning contexts, including the most relevant aspects encountered in each context for the study of SLA. Then, the chapter continues with a review of the studies undertaken within SA and a description of the main findings. The last part deals with specific factors in SLA for which the SA context can provide a better understanding.

Chapter 2 deals with L2 speech production. The chapter starts with a brief overview of the speech production models and approaches in SLA, going from L1 to L2 research. The theories of automatization are then presented (rule-based vs. item-based approaches) to finish with the review of four phenomena central to the study of L2 speech production (automaticity, output, practice and memory).

Chapter 3 is dedicated to fluency. The first part includes a brief overview of L1 fluency development, follow by a more thorough description of L2 fluency development, which is subdivided into three sections. The first section presents various definitions of fluency as used in SLA research. The second reviews the studies on L2 fluency within SLA, and the third section presents the studies on L2 fluency undertaken within SA research which have generally shown the SA context to contribute to progress in L2 fluency.

Chapter 4 presents an overview of research on rhythm. The chapter starts by presenting what is understood by rhythm in the L1 and its main characteristics, with a special section on vowel reduction. It is followed by a section which introduces the most widely used rhythm metrics in the study of rhythm. The chapter continues with a review of the studies on L2 rhythm.

Following a recapitulation section, chapter 5 presents the objectives and the research questions which have motivated the study. Subsequently, chapter 6 deals with the methods used to carry out the study. First a brief contextualization of the educational background in which the study takes place is presented, followed by a description of the participants and the design of the study. Then the chapter is subdivided into the two tasks presented, the first one dealing with production and the second with perception. Within each task a brief description of the participants, instruments and procedures is presented together with a detailed description of the measures and analyses undertaken.

Chapter 7 contains the results of the analyses undertaken to respond to the research questions formulated in chapter 5. And chapter 8 discusses these results in the light of previous findings, organized following the research questions.

The study closes with Chapter 9, in which the conclusions are presented together with the limitations of the study and a section on further research.

PART I. BACKGROUND TO THE STUDY

1

Learning Contexts: Study Abroad

Since second language acquisition (SLA) first started as a separate discipline from applied linguistics[2], research has been undertaken on both naturalistic and instructed learning. Early studies in SLA focused on either one context or the other. It was during the decade of the 80s, that new different types of settings, characterized by a mixture of elements from naturalistic and instructed contexts emerged worthy of investigation: *immersion programs* and *study abroad*. Notwithstanding a series of groundbreaking studies undertaken in Canada by Swain and colleagues on immersion, study abroad (SA) is the context that has received most attention of the two, especially from the 1990s onwards. SA studies have evolved from being purely linguistic in nature, to taking a more socially oriented perspective. More recently in Europe, the attention of research has been directed towards a new context, often called Content and Language Integrated Learning (CLIL). In spite of its recency, much research is currently conducted within this context, especially in Europe (see Pérez-Vidal, 2011 and Pérez-Vidal & Juan-Garau, 2010).

On a social level, the main boost of SA research has occurred in great measure due to the changes taking place world-wide which have shaped a new reality, to a great extent with regards to globalization and

[2] Not all scholars agree considering SLA as a discipline on its own. (for further reading on this matter refer to Towell & Hawkins, 1994; Ellis, 2000; Sharwood Smith, 1994)

internationalization. Indeed, this has been happening hand in hand with the emergence of low-cost companies in the last decade which have put travelling within reach of almost everyone. Consequently, the increase in number of people travelling generated an ever growing interest in other languages and cultures. This phenomenon has been reinforced by the need of mobility for professional purposes, mainly due to people's interest in promoting their professional careers. More often than not, a certain proficiency level in the language spoken in the target language (TL) culture is required when travelling for work or academic purposes. In order to meet this requirement, people strive to find language courses that satisfy their needs, sometimes opting for spending a period of time studying abroad.

When looking into study abroad programs with an institutional perspective, we realize that they are promoted worldwide by higher education institutions trying to give their students the opportunity of an international education. This experience is believed to benefit students' future careers and provide them with more competitive professional profiles, in addition to a better command of the language spoken in the target culture which is what most of them envisage, as already mentioned. Nevertheless, universities do not stand alone in the promotion of students' mobility among countries. Governments have taken action and are enhancing new policies by awarding scholarships for university students who want to spend a period of their studies abroad.

The European Union (EU) promotes a highly successful scheme to stimulate students' mobility around Europe, the Erasmus scheme. This scheme was created with the goal of enabling students to study part of their degree in another country with the recognition of academic achievements abroad. It was first established in 1987 by the EU and during its first year, a total of 3.244 students participated in the experience.

The increase in number of students from its first year to 2010[3] has been exponential, reaching a maximum of 213.266 students who went abroad last year[4]. With this impressive number of students going abroad, studies on the possible benefits of spending a period abroad and how to enhance them are urgently needed.

North-american exchange programs have been the source of data for most of the early research conducted on SA, even though it has not been until this last decade that SA programs have experienced a rise in the United States (see Kinginger, 2009 for a thorough review of this issue). In recent years, new studies have started to focus on Asian learners going abroad, since they represent a high percentage of the population studying abroad (Collentine, 2009).

There are basic differences between these geographically distinct contexts which make their direct comparison sometimes challenging, at the same time that they also offer the opportunity to broaden the context. One of the most commonly cited differences is the L2 procifiency level of the population under study. Studies undertaken with American students going abroad have been mainly conducted with learners at the initial stages of acquisition, whereas European studies have concentrated on the Erasmus population with a more advanced level of the language (Coleman, 1997; Kinginger, 2007; Collentine, 2009). A further difference concerns the type of programs (Churchill, 2006; Collentine, 2009; Llanes, 2011). While American university programs are largely fostered programs where students go abroad with a group of classmates and participate in both

[3] This is the last year from which information is provided by the European Union.
[4] Spain sent the largest number of students (*N*=31 158) abroad the year 2009/2010 and was also the most popular host country (35 389 Ss.), followed by France (26 141 Ss.) and the UK (22 650 Ss.). Source: European Comission Education and Training (6 June 2011)

academic and leisures activities together on-site supervised by instructors and program directors, the European experience is much more individualistic, with students enrolled in regular host university classes, each being responsible for their out-of-class activities (Murphy-Lejeune, 2002; Kinginger, 2009). Murphy-Lejeune (2002) provides a full ethnographic description of what the experience abroad in the European Erasmus context means. She does not focus on linguistic gains but on the different social and psychological aspects of the experience *per se.* as viewed by the students themselves.

For SLA researchers, SA provides a rich context of investigation on how the acquisition of an L2 develops. According to Collentine (2009), the context of learning "affects the nature and the extent to which learners acquire a second language (L2)" (p. 218). Students who participate in these programs are immersed in a situation very different from the one in formal instruction (FI) at home. Phenomena such as quantity and quality of input, output, practice, learning opportunities, interaction, etc. make the two contexts differ. These differences, together with learners' cognitive processes taking place in such circumstances are what determine the rate and route of acquisition.

This chapter provides an account of research on learning contexts, with a special focus on SA. The first part of the chapter is devoted to the description of the different learning contexts (section 1.1.), followed by a section on previous studies on learning contexts (section 1.2.).

1.1. Learning Contexts in SLA

The relevance of learning contexts and their impact on language development has not gone unnoticed for SLA researchers. In an attempt to describe Second Language Acquisition, Ellis writes: "'L2 acquisition'

can be defined as the way in which people learn a language other than their mother tongue, inside or outside the classroom, and 'Second Language Acquisition' (SLA) as the study of this." (Ellis, 2000). This definition already mentions the existence of different learning contexts in the study of SLA. Some years later, Collentine & Freed (2004) claim the importance of learning contexts for a more comprehensive view of SLA, saying that "The study of SLA within and across various contexts of learning forces a broadening of our perspective of the most important variables that affect and impede acquisition in general" (p. 158).

Four learning contexts are most usually cited: the 'at home' (AH), which usually takes the form of formal instruction (FI), the domestic immersion programs (IM), in which the L2 is studied intensively throughout the curriculum, the study abroad (SA), most often combining FI and in-country residence, and the naturalistic learning environments (NT), usually with no FI provided. As mentioned at the beginning of the chapter, during the last decade a new context has emerged, the Content and Language Integrated Learning (CLIL), which in a way can be understood as a combination of FI at home and IM, and has sometimes been referred to as immersion programs. In this context, regular content subjects are taught AH, but in a foreign language, aiming at the learning of both content and the foreign language (for a comprehensive review refer to Pérez-Vidal, 2010; Roquet i Pugès, 2011).

The next subsection provides a thorough description of each learning context (section 1.1.1.). In order to do so, key factors in SLA research and also in describing the SA context are reviewed (sections 1.1.1.a. to d.).

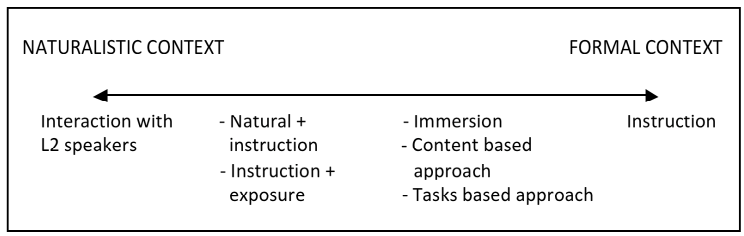
1.1.1. Description of the contexts.

The contexts above mentioned have been the focus of study for many

SLA researchers during different periods of time; from early studies based on one of the dichotomies, naturalistic versus instructed learning, to later research comparing several other in-between contexts. This section is devoted to describing the characteristics of each of these contexts in an integrated fashion.

According to Cenoz & Perales (2000), the traditional distinction between learning contexts, naturalistic and instructed, is not as clear-cut as has been postulated by some researchers. As they put it, "the big diversity of instruction's practices, the diversity of natural contexts of acquisition and the individual differences present in this process do not allow considering the difference between formal and natural contexts in terms of an absolute dichotomy." (Cenoz & Perales, 2000). They view these terms as a continuum where the ends are formal and naturalistic contexts, and intermediate positions would be accepted (see Figure 1.1). In the intermediate positions we find 'mixed' contexts, representing these situations where formal instruction is provided while being abroad or those where formal instruction is complemented by natural exposure of some sort.

Figure 1.1. Learning Contexts and Second Language Acquisition.

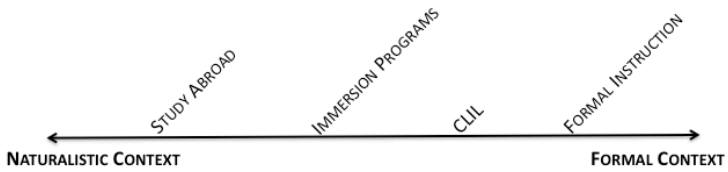


Source: translated and adapted from Cenoz & Perales, 2000:110.

Different types of formal instruction can also be found in these intermediate positions, such as domestic immersion programs and the content based or task based approaches. Figure 1.2. shows the five

learning contexts distributed over a context continuum going from naturalistic to formal settings.

Figure 1.2. Learning Contexts Continuum.



Within the SA context, Collentine (2009), based on Batstone (2002) differentiation, further distinguished two sub-contexts: *communicative contexts* and *learning contexts*. In the first one the focus is on meaning, with the exchange of information as the main objective, whereas in the second form is the focus, together with the intentional steps that leaners take to improve in the target language (TL). He claims that learners can indulge in both during SA.

Several central factors in SLA research are key in the characterization of the different learning contexts: input, interaction, output and type of practice.

1.1.1.a. Input

The role of input in SLA remains a central question even though the field is becoming more and more hetereogeneous and prolific. Recently, several books have been dedicated to this environmental factor. This is how Piske and Young-Scholten (2008) call 'input' in a book which provides a comprehensive review of what is known about input in current SLA research from different perspectives. That same year Rast (2008) published a book dedicated to foreign language input in relation to initial processing. The emphasis here is given to the importance of the first contact with input from the foreign language in order to find out how learners convert

this input into intake and what strategies learners use in relation to both the L1 and the TL to process the input.

First introduced by Krashen (1985), the role of input has been expanded since and, apart from the *comprehensible input*, it currently is understood as a whole together with cognitive processes such as attention, awareness, noticing and intake. The main claim of Krashen's *Input Hypothesis* was that we move along a developmental continuum by receiving comprehensible input. Comprehensible input is the kind of input received in the form of *i + 1*. This means that the L2 input is just beyond the learner's L2 current competence, in terms of syntactic complexity. Only input of the type *i + 1* will successfully contribute to acquisition. If learners are provided with input which is not comprehensible or it is too easy, no acquisition will take place.

However, research in the last decade has focused more on the way learners process this input to convert it into intake. There was a first wave of reasearch on the importance of noticing and attention (see for instance Robinson, 1995, Schmidt, 1990, 2001). They found that comprehensible input itself was not enough, but that attention and noticing were necessary for learning to take place (we refer back to this issue in section 2.3.2.).

A second wave has focused on different contexts and how they afford learners different exposure to the target language. Krashen's (1985) claim that comprehensible input is *all* that learners need to acquire a language, is no longer accepted as the absolute main motor of acquisition in SLA (see Long, 1996; Izumi, 2002). In this set of studies, input is considered to be the 'entrance-gate' to other processes leading to L2 acquisition. This, in relation to contexts of acquisition and different amounts and types of input accessed leads to different cognitive processes and different learning opportunities for L2 learners. Therefore, we could argue that the amount

and type of input a learner receives will condition the cognitive processes required for L2 acquisition to take place, such as awareness, attention, noticing, etc. Taking the idea a bit further, it could also be posited that the quality of the input a learner receives varies depending on the context where the exposure occurs (Moyer, 2008), the subtlety and experience on the part of the interlocutor to cope with non-proficient L2 learners (teacher-talk, foreigner-talk, non-modified talk), the formality of the situation (Moyer, 2004), time-limits (DeKeyser, 2010), the interest of the interlocutor in what is being said (Lennon, 1990), etc… In this respect, not only the quantity (usually measured as time-on-task), but also the quality of the input received should be taken into account, since depending on the type of input, different cognitive processes can be activated. In Moyer's (2008) words, "multiple and complex opportunities for L2 use ensure rich input and meaningful communication" (p.166).

1.1.1.b. Interaction

Interaction, closely linked to both input and output, has to do with the modifications that both interlocutors communicating introduce in their speech (output) building on the other's linguistic and non-linguistic signals (input through reformulations, recasts and repairs). Through interaction, input, as an environmental factor, becomes mediated by cognitive processes such as those mentioned above (awareness, attention, noticing).

First proposed by Long (1983) and further developed in Long (1996), the *Interaction Hypothesis* builds on Krashen's *Input Hypothesis* and takes it a step futher by including the notion of 'negotiation of meaning' through interaction. According to Long (1996), it is through this negotiation of meaning that learners can make sense of the message they receive and get the type of comprehensible input they need to advance their learning of the foreign language. When engaging in interaction, learners not only

receive input from the interlocutors, but they often receive input which has been accommodated to their needs, as is the case with motherese or caretakerese.

However, the importance of the interaction is not only on account of the adequacy of the input that learners receive, but also of the negative evidence that these obtain from communication breakdowns and misunderstandings due to faulty output. Once these problems occur, the learner gains awareness that something has happened that has hindered communication, and by signalling the interlocutors' mismatch (either by providing linguistic or gestural information), he/she receives modified input from which to extract information both for form and meaning. So, interaction is highly important for learners to 'notice the gap' between what they know and what they don't know (Schmidt & Frota, 1986).

1.1.1.c. Output

Since Merrill Swain (1985) first advanced the *Ouput Hypothesis*, the crucial role of output in SLA has been advanced by others (Swain, 1985; Doughty, 2001; Izumi, 2002; DeJong, 2008). Output has mainly been claimed to be important for hypothesis testing, getting feedback or negative evidence, and to enhance automaticity. However, its main contribution is probably in pushing syntactic processing. Section 2.3.2. is dedicated to this factor in relation to L2 speech learning.

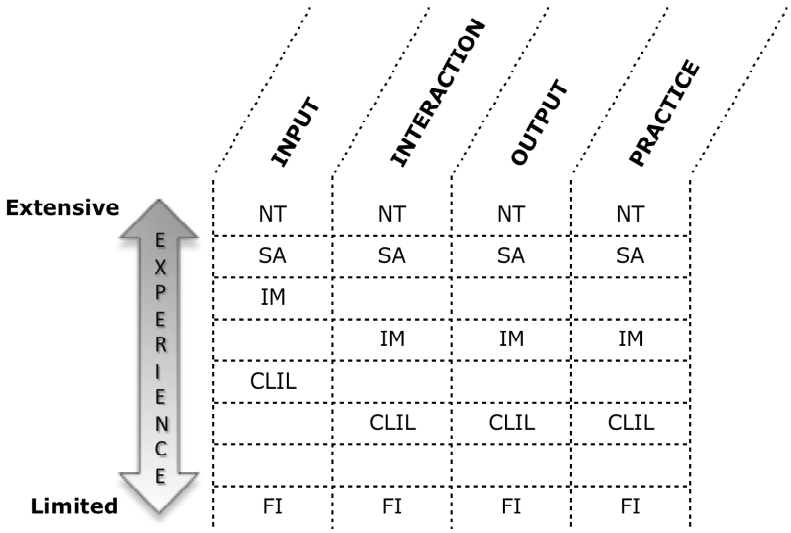
1.1.1.d. Practice

The role of practice in SLA research has gone a long way from the early behaviourist drills, to the current cognitive approaches which regard practice as crucial for language processing. However, in order to have an effect on the learners L2, both the quantity and quality of practice are important. A more thorough review of practice is given below (section

2.3.3.) in relation to L2 speech learning.

These briefly described four factors (input, interaction, output and practice) have been taken together in an attempt to characterize each learning context and understand their core features. A visual description of the ensuing two-dimensional construct is displayed in Figure 1.3. In the vertical axis, the contexts are included in a *L2 Experience Continuum* with extensive experience with the language at one end, and limited experience at the other, as already shown in Figure 1.2. above. In the horizontal axis, the four factors above mentioned categorize each learning context.

Figure 1.3. Exposure to input-interaction-output-practice by learning context.



Note: NT: naturalistic; SA: study abroad; IM: immersion; CLIL: content-language integrated learning; FI: formal instruction.

Nevertheless, in addition to these four key factors in SLA research, there are several other variables which are, in turn, central factors in the description of the learning contexts: type of language (authentic-elaborated-simplified), social domain (open to all social domains-restricted to classroom), formality of the situation (formal-semiformal-informal), L1 use and linguistic focus (focus on meaning-focus on form). These have all

been taken into account in the general description of the learning contexts (FI, CLIL, IM, SA) that follows:

Formal Instruction: The social domain in FI is restricted to the classroom, a formal or semiformal situation. The type of language used in this context is non-authentic, sometimes simplified and at other times elaborated. It responds to what in the literature has been called teacher-talk or classroom-talk, in which language is adapted to learners' level and needs. The type of instruction is mainly focused on form, leaving meaning to a second position. The use of the L1 in this context is quite common, especially among peers.

> ***Input***: Limited amount and type of input; mainly restricted to metalinguistic talk.
>
> ***Interaction***: Limited to short interactions with the teacher and peers (depending on the methodology used).
>
> ***Output***: Limited to classroom-talk and sporadic.
>
> ***Practice***: Limited amount and type of practice, especially for oral productive skills.

CLIL (Content Language Integrated Learning): The social domain is restricted to the classroom, with a quite formal or semiformal situation. A specific or several content subjects in the curriculum are taught through the target language, while all the others remain taught in the learners' L1. The language is semi-authentic (teacher-talk), being most of the time elaborated or simplified, as well as adapted to learners' level and needs. On the other hand, written tests do tend to be authentic. The focus is mainly on meaning and very little reference is made to form. Content + language are integrated. The use of L1 among peers is the norm (see Pérez-Vidal, 2007, 2008; Muñoz, 2007 for descriptions of the core features in CLIL programs).

Input: Limited amount and type of input; mainly restricted to a formal or semi-formal register and focused on a limited range of topics.

Interaction: Limited to short interactions with teachers and rarely with peers. Often limited to academic discourse (Pérez-Vidal, 2011).

Output: Limited to classroom-talk and mainly in the written form.

Practice: Limited amount and type of practice, especially for oral productive skills.

Immersion Programs: Social domain restricted to the school interaction and premises, a formal or semiformal situation. All subjects are taught through the target language, and this is the language of communication within the school context. The type of language used is authentic, but basically limited to classroom-talk or academic talk, and hardly adapted to the learners' level and needs. The focus is usually on meaning, but at the beginning of the experience, it can also be on form. The use of the L1 is highly limited, but may happen among peers when not supervised.

Input: Considerable amount of input but of a limited type.

Interaction: Limited interactions with teachers and peers.

Output: Considerable amount of output but limited in type.

Practice: Considerable amount and type of practice, especially for written skills.

Naturalistic context: Wide diversity of social domains, with both formal and informal situations. The type of language used is authentic, real-life talk. Language may sometimes be simplified or elaborated, foreigner talk, but this is not always the norm. The focus is on communication, so the importance is on meaning, not so much on form. The use of L1 tends to be limited, or at least reserved to specific situations, such as at home or at

certain meetings with L1 peers.

Input: Large amount and varied input.

Interaction: Considerable amount of interactions with a variety of interlocutors.

Output: Extensive output, but sometimes limited to certain types of language.

Practice: Extensive amount and varied practice, especially for oral skills. Writing skills are often very limited.

Study Abroad: Large diversity of social domains, with both formal and informal situations. The language used is authentic, real-life talk, and it is hardly adapted to the speakers' level (some foreigner talk may occur). Language instruction is usually provided, but the amount and type vary from experience to experience. The focus is usually on meaning, but during instruction it may also turn to form. The L1 use is usually limited, but it might vary highly from person to person. A certain homogeneity in educational level is common in this context.

Input: Large amount and varied input.

Interaction: Considerable amount of interaction with a variety of interlocutors and using different registers (interaction and social roles).

Output: Extensive output in a variety of registers, but often formulaic (DeKeyser, 2007c).

Practice: Extensive amount and varied practice, especially for oral skills, perceptive and productive, and reading skills. Often formulaic in nature (DeKeyser, 2007c).

From the descriptions, it can be inferred that the differences between one context and the other are not always clearcut, and that they often share

many of the core features (see Valls-Ferrer, Roquet-Pugès, & Pérez-Vidal, 2010 for a comparison between three learning context: SA, FI and CLIL). DeKeyser (1991) suggests that "a semester in the native-speaking environment, following or combined with a (high) intermediate course with some focus on the explicit teaching of grammar, provides a prolonged opportunity for an ideal mix of focus on form and focus on meaning" (p.116). In addition, the context itself does not provide any benefits, it is the learner who has the last word and decides to what extent he/she will take advantage of the opportunities that the context provides. As Howard (2005) claims, "the instructed learner assumes the status of the naturalistic learner during a period of residence in the TL community" (p.496).

Considering more specifically the SA context, in early research, the study abroad context was defined by Brecht, Davidson & Ginsberg (1993) as a "blend of structured and unstructured learning […] which necessarily relies on a combination of academic programs, individual study, and intensive interaction with native speakers of the language." (p. 2). The authors placed emphasis on the blend or combination of factors which were considered to be part of the study abroad: the formal instruction, the interaction with native speakers (NSs) in the TL community, and the importance of the learner. More than a decade later, Regan, Howard & Lemee (2009) provide a more comprehensive description of what SA programs are:

> "Whilst the participants often follow a course of instruction whilst abroad, study abroad programmes are principally a means of allowing the instructed learner to acquire 'pseudo-naturalistic' status, by engaging in more informal acquisition in the TL community, through naturalistic contact with the L2 in everyday social situations." (p. 20)

When learners go abroad, they take all the knowledge they have acquired during formal instruction at home, in most of the cases through explicit

instruction, and put it into practice. By put it into practice we mean that they have to meet the requirements of social communication and deal with the online processing that spontaneous speech requires. In DeKeyser's (2010) words: "Because of the ensuing time pressure and the complexity of the knowledge involved, this is a highly demanding cognitive process." (p.82). Therefore, the SA provides students with the best opportunity to practice the knowledge they bring from home. However, a word of caution has to be inserted, since in order to benefit from this opportunity, the knowledge must have reached a threshold level which permits learners to build on that knowledge (DeKeyser, 2007c, 2010); a level at which much of the proceduralization process has been completed and so practice abroad can help automatize it.

We now turn to the presentation of research undertaken within the SA learning context.

1.2. Study Abroad and SLA

As stated in the previous section, studies in SLA have been mainly undertaken by scholars basing their research on data collected in formal instructional settings, especially at tertiary level. Hence, much of what we know about most of the studied phenomena in SLA is based on studies carried out with either language elicited in a formal situation, such as FI, immersion and CLIL, or in highly controlled laboratory settings. Without underestimating the valuable information that this research provides, in the last two decades SLA research has widened its scope to include an ever growing number of studies undertaken in SA settings.

As described in the previous section, language learned in a study abroad setting presents certain characteristics that are different from language acquired in a formal setting (Regan, 1995; Regan et al., 2009). As

summarized by Collentine (2009) referring to the Collentine & Freed (2004) review article, "research in a SA context provides an important contextualization for understanding the interaction between cognitive, sociolinguistic, and socio-cultural factors in the construction of a comprehensive theory of SLA".

With the intention to find out how learning differed from one context to another, Freed (1995a) compiled a collection of studies exploring the effects of learning abroad compared to formal classroom at home. These studies analyzed different phenomena, such as fluency (Freed, 1995), communicative strategies (Lafford, 1995), predictors of foreign language gains (Brecht, Davidson & Ginsberg, 1995), etc. This collection is often cited in academic articles as the formal beginning of SA research as we understand it today.

The decade of the 90s marked the beginning of SA as a very prolific field. A number of studies were presented during this period and several others were being developed at the time. Multi-dimensional projects and collections of individual studies emerged, focusing not only on the linguistic advantages of SA, but also on the SA experience *per se*.

One of the best known multi-dimensional studies is that of the landmark American Council of Teachers of Russian/National Foreign Language Center (ACTR/NFLC) project (Brecht & Robinson, 1993; Brecht et al., 1995; Miller & Ginsberg, 1995). Brecht and his colleagues worked on a study of American students who spent a semester or more in Russia to learn the language. The first phase consisted in measuring linguistic gains in three language skills (reading, listening and speaking). The second phase of the study focused on "an ethnographic study of the in-country language-learning process, documented by self-report diaries, observations, interviews, and recordings" (Brecht & Robinson, 1993). These variables

were used to measure the effect that they had on linguistic gains. Overall they found that a semester abroad was beneficial for all three skills under analysis, but improvement was conditioned by other variables such as proficiency level; the higher the initial level, the lower the gains in reading and listening.

DeKeyser (1991) followed a group of American students during a six months SA period in Spain and compared them to a control group of students who remained in the US. He looked at the development of grammar, vocabulary and oral proficiency (including fluency), and found that gains between the two groups were very similar for grammar, but a greater improvement in vocabulary and fluency was found for the group abroad. Although no changes were found in oral communication strategies for any of the groups, he pointed out the high variability in their use among students.

In a preliminary study focusing on the acquisition of fluency during a period abroad, Lennon (1990) analyzed the speech of 4 German learners of English using a large array of utterance fluency measures and native speaker (NS) listeners' judgements. Improvement was found after a 6-month SA period for some of the measures, as well as the listeners' ratings.

As briefly mentioned above, Freed's (1995a) book is often considered a turning point for SA research. In her own study, Freed (1995b), two groups were compared in two different learning context, at home and stay abroad in order to identify linguistic gains in each learning context. Few differences were found between them in overall gains, but higher fluency gains were detected for the group abroad as judged by NS listeners and temporal fluency measures, especially for those who started with a lower fluency level.

Milton & Meara (1995), in one of the few studies which have looked at vocabulary growth, analysed the speech of learners of English from different European countries spending a semester in the UK. They found that learners' vocabularies were expanded as a result of the period spent abroad. Students with lower levels on the entrance test were the ones that improved the most, and students with higher levels, improved less (some even went backwards). In addition, they also referred to the differences among individuals.

Towell, Hawkins & Bazergui (1996) developed a very detailed study on the fluency of British learners of French who spent a year abroad. They found that gains in speech rate and especially the increase in mean length of runs were responsible for the good results in students' fluency after the period abroad. In a later study, Towell (2002) tried to explain why some learners attained higher scores on temporal variables than others and why the ones with a low initial level increased the most but never reached the levels of the former.

The first years of the new century started with several publications on SA, culminating in a monographic presented in 2004 by the journal *Studies on Second Language Acquisition*, dedicated to SA. Collentine and Freed were responsible for assembling the special issue, and provided some new insights into the field. Up to that moment research had been mainly conducted taking a purely linguistic perspective. With the 2004 issue, a broader perspective was introduced in the field. The authors in the monograph were not only interested in measuring linguistic gains, but in observing which variables conditioned these gains, and both individual differences and contextual variables were found to exert an influence.

In the last years, several books have been published on SA taking different perspectives (ethnographic and self-construction: Pellegrino-Aveni, 2005;

social and pragmatic: DuFon & Churchill, 2006; socio-cultural: Byram, 2006; socio-cultural and identity: Kinginger, 2009; sociolinguistic: Regan et al., 2009). From these studies the social turn in SLA becomes evident, at least in SA research. On the more cognitive line of research, book chapters have also been devoted to the SA context (practice: DeKeyser, 2007c; handbook of language teaching: Collentine, 2009;).

From all these years of research, robust findings have been found for some areas (i.e. fluency), but many more questions have arisen in others (i.e. individual cognitive factors). In what follows, the main findings are presented.

The area where most improvement has been found during SA periods is that of oral proficiency and, especially, fluency. Early studies reported large gains for oral proficiency during a semester abroad (Brecht et al. 1993, 1995; Lapkin, Hart & Swain, 1995). These studies were then followed by studies focusing on fluency, and in general, most of the studies have reported large gains in fluency for students who spent a period abroad (Lennon, 1990; Freed, 1995b; Towell et al., 2006; Freed et al., 2004; Segalowitz & Freed, 2004; Juan-Garau & Pérez-Vidal, 2007), even for stays as short as 3-4 weeks (Llanes & Muñoz, 2009). See chapter 3, section 3.1.2d for a thorough review of fluency studies in SA research.

Research in the domain of phonology is scarce, and has often produced mixed results (Díaz-Campos, 2004; Mora, 2008). Díaz-Campos, 2004 did not find differences in improvement for segmental phonemes in Spanish between a group who went abroad and the group at home. In Mora (2008), significant improvement in perceptual phonology was only found after the formal instruction period learners went through, but not after the SA. Nontheless, in the same study participants' productions were assessed in terms of Voice Onset Times (VOTs), and a consistent (although non-

significant) increase was found during the SA period.

Listening skills and lexical acquisition have also been reported as increasing the most during SA periods (Listening: Beattie, 2008; Kinginger, 2008; Llanes, 2009; Lexical: Milton & Meara, 1995; Ife, Vives Boix & Meara, 2000, Foster & Tavakoli, 2009). In addition, sociopragmatic aspects of language use appear to develop substantially through SA (Huebner, 1995; Regan, 1998; Lafford, 2004; Pellegrino Aveni, 2005; Schauer, 2007; Regan et al., 2009).

However, as far as grammatical accuracy goes, most studies have reported no differential effects between both learning contexts (Collentine, 2004; DeKeyser, 1991). Relatively few studies have examined lexical and grammatical complexity, reporting limited or no gains in both these dimensions during SA (Pérez-Vidal & Juan-Garau, 2009) and greater development in the at home context (Howard, 2005). However, the SA context seems to include greater variability in the use of some grammatical aspects such as verb morphology (Howard, 2001, 2005).

As far as written skills are concerned results are extremely scarce, with some studies reporting small differential gains in written fluency between AH and SA groups (Freed, So, & Lazar, 2003; Sasaki, 2007) and some finding greater improvement after SA (Pérez-Vidal & Juan-Garau, 2009) particularly in the domain of fluency, lexical complexity, formulaic language and accuracy.

Studies on individual differences affecting the SA outcomes have been mainly undertaken in the last years tapping on cognitive aspects such as phonological memory and lexical access (Segalowitz & Freed, 2004; O'Brien, Segalowitz, Freed & Collentine, 2007; O'Brien, Segalowitz, Collentine & Freed, 2006; Sunerman & Kroll, 2009). Most of these studies have found that gains in oral production are related to L2 cognitive

abilities, but the relationship is not a straightforward one (cf. Lord, 2006).

Having this overall picture of SA research findings in the background, we now turn to two important factors which have been claimed to affect the language gains, or lack of, obtained during a SA period: a) contact with the target language and b) participants' profiles and threshold levels.

1.2.1. Contact with the target language.

The degree of contact with the target language is a key variable in SLA research[5]. As mentioned above, studies on American learners going to Russia (Brecht & Robinson, 1993; Brecht et al., 1995) focused on contact variables as predictors of language gains. Freed (1990, 1995) also related several contact variables to learners' linguistic gains during periods abroad. However, these contact variables are often difficult to operationalize. In an attempt to do so, Freed, Dewey, Segalowitz & Halter (2004) developed the Language Contact Profile (LCP). This questionnaire was designed to gather information and assess second language contact for students in any language learning context (FI, SA, IM).

Although the LCP questionnaire provided useful data on language contact for several studies (Freed, Segalowitz & Dewey, 2004; Segalowitz & Freed, 2004; Dewey, 2004; Magnan & Back, 2007; Hernandez, 2010), results show that the contact measures obtained do not always correlate with language gains (Segalowitz & Freed, 2004; Magnan & Back, 2007). Collentine (2009) warns about the self-evaluation nature of the questionnaire and suggests that data provided by the LCP should be

[5] In SA research often operationalized as 'time-on-task', and in Psycholinguistics and L2 Speech Learning research, language contact has often been referred to as L2 experience.

triangulated with other instruments, or validated by third-party informants such as host families or program organizers abroad.

A modified version of the LCP was used in Juan-Garau & Pérez-Vidal (2007), Pérez-Vidal & Juan-Garau (2009), and Valls-Ferrer (2008). In these studies, the strength of the correlation between language contact measures and language gains for high achievers was found to be dependent on the linguistic domain under study. For written skills, Pérez-Vidal & Juan-Garau (2009) found that living in families or flats, carrying out academic work, being eager to learn and being emotionally aware were associated with written proficiency. Regarding oral proficiency, Juan-Garau & Pérez-Vidal (2007) found that working in an international setting, involving oneself in independent study and listening to the media were good indicators of gains in oral skills. Similarly, in a study focusing on the development of fluency, Valls-Ferrer (2008) found that certain contact variables such as living with NSs and NNSs from different L1 backgrounds, amount of contact with media and degree of self-confidence discriminated high achievers from the rest of participants.

Although the above mentioned studies have found some kind of relationship between certain contact variables and language gains, the evidences are far from conclusive, and direct relationships cannot be established. Contact with the target language may be too broad a term to be operationalized as a variable. Indeed, contact with the target language actually means, to a great extent, amount and quality of input and interaction. Although this is what most questionnaires developed to capture language contact while abroad have intended to assess, researchers should probably aim at designing more fine-grained questionnaires where information about different types of input and interaction can be obtained.

Most early SA research focused on quantitative rather than qualitative aspects of language contact. However, recent studies within the socio-cultural framework have used qualitative research methods such as ethnographic observation (Ginsburg & Miller, 2000; DuFon & Churchill, 2006; Kinginger, 2007, 2008, 2009). These studies usually examine a reduced number of participants, but the analyses provide in-depth accounts of the participants' social moves while abroad in relation to their language use. They provide useful information to understand how the contact with the target language occurs, or does not occur, and why.

SA research can greatly benefit from both quantitative and qualitative methods. Qualitative methods might be used to describe the language context where learning takes place, as well as the type and amount of input and interaction that learners get involved in. Simultaneously, quantitative methods can provide the instruments to objectively measure the linguistic gains and the amount of contact which take place during the periods abroad. Their combined use (as in mix-methods research) will provide a better understanding of the relationship between contact variables and language gains in SA.

1.2.2. Participants' profiles and threshold levels.

Regarding learner profile, age of learning (AOL) or age of arrival (AOA), meaning the age of first extensive exposure to the L2, has received some attention lately. Trofimovich & Baker (2006) found that AOA influenced some aspects of temporal fluency (speech rate, pause frequency and duration), and amount of experience (or length of residence) influenced one suprasegmental (stress timing). Llanes (2010) also found an effect of age when comparing a group of adult and child learners abroad. However, these results might be confounded since they could be due to the differences in proficiency level as well as age.

Furthermore, in relation to learners' profiles, in recent years, literature reviews on SA have underscored the importance of a proficiency threshold level on developmental gains during SA periods (Collentine, 2009; Llanes, 2011; Pérez-Vidal & Juan-Garau, 2011). Collentine (2009) suggests that the use of threshold levels, highly benefitial for administrative purposes, is probably too broad in scope for research on second language acquisition. He claims that "there are most likely specific domains that require a particular developmental threshold for overall gains to occur." (p. 221). This research area has been informed by several studies.

Brecht et al. (1993) found that grammar and reading achievement scores before the SA were significant predictors of gains in speaking proficiency, listening and reading after the SA. They suggest that grammar during the first years of learning a language results in advances in speaking and listening skills at upper-intermediate and advanced levels.

Towell (2002) observed a group of 12 English students who went to France and found that an initially low-scoring group had a much higher percentage of improvement than the higher performers' group in fluency. However, even though both groups increased their scores, the initially low-scoring group never caught up in absolute terms with the high group in terms of mean length of run, phonation time ratio and pause length.

Segalowitz & Freed (2004) analysed a wide array of cognitive and linguistic factors which would impinge on gains during SA. They found that a certain level of ability in word recognition and lexical access processing had to be reached for significant gains in oral proficiency and fluency to take place while abroad. In a related study, O'Brien, Segalowitz, Freed & Collentine (2007) provided further evidence for the role of cognitive and linguistic abilities, in this case phonological short term memory, in the development of the L2 abroad (cf. Towell & Dewaele, 2005).

Golonka (2006) building on Brecht et al. (1993, 1995) identified a series of predictor variables, both linguistic (grammar, vocabulary and accuracy) and metalinguistic (self-corrected errors and sentence repair), which were indicators of advanced-level threshold gains.

From the few studies briefly described here we can conclude that learners' initial level is paramount for the developmental gains during the period abroad. These studies inform the current SLA agenda on the role of a threshold level as a factor conditioning the linguistic development while abroad. In an inverse relationship, the lower the level, within a functional range of language competence (DeKeyser, 2007c), the higher the gain. However, further research is needed to provide a complete picture of the issue.

We now turn to the second chapter of this study which deals with speech production in SLA.

2

L2 Speech Production

Research in the last decades has shown that acquiring the sound system of one's native language is a very different issue from acquiring an L2 sound system. Studies on L2 speech production have been grounded in previous findings in L1 speech production research (see Kormos, 2006 for an overview). Most of this research has been undertaken by either cognitive psychologists, or by SLA researchers with a cognitive orientation. The major difference between L1 and L2 speech acquisition seems to lie in the processing capacity of the human mind. When speaking the L1, many of the processes necessary to produce speech are automatic, hence, producing a rapid and accurate performance. However, when trying to produce L2 speech, the learners' cognitive capacity can be overwhelmed by the many processes going on in their mind, which have not yet been fully automatized, due to incompleteness of the L2 learner's knowledge of the target language. This circumstance also results in a somewhat slower production by the L2 speakers, due to shortage of storage. Consequently, not only the processing capacity of the human mind seems to play an important role, but also the storage capacity of its memory (especially short-tem memory). The present chapter presents an overview of research dealing with these phenomena.

This chapter starts with a description of the theories and models of speech production and automatization present in SLA research. Two models and theories are first presented in section 2.1., spreading activation and modular theories. They are discussed in relation to L1 and L2 speech

production. In addition, theories of automatization are presented in section 2.2 in an attempt to link them to speech production models. Finally, four central phenomena to speech production, which contribute to the better understanding of how speech is produced, are presented in section 2.3.: a) automaticity, in connection with the theories of automatization, b) output, as a facilitator of acquisition through production, c) practice, as an indispensable phenomenon for L2 speech to develop, and d) memory, in relation to language processing.

2.1. Speech Production Models and Theories: from L1 to L2

Two main theories dominate the current speech production research: *spreading activation* (Dell, 1986) and *modular theories* (Levelt, 1989). These theories mainly differ in that spreading activation allows the backward flow of activation from a subordinate level to a superordinate, while modular theories do not. Within the spreading activation theories, it is assumed that frames for sentences and phonetic representations are constructed, and only then, speakers choose the appropriate words or phonetic features to fill in these frames. On the other hand, modular theories are lexically driven, meaning that words activate the syntactic building procedures. Lexical encoding has to take place before syntactic encoding, and phonological encoding can only occur when lexico-syntactic processes are completed.

These theories were first developed to explain L1 speech production, but have been accommodated to explain L2 speech production as well (for modular theories: de Bot, 1992; Towell et al., 1996; Kormos 2006; for spreading activation theories: Costa et al., 1999; Doughty, 2001; Poulisse & Bongaerts, 2004;). The most widely applied model in SLA studies, on

account of its robustness and empirical basis, is Levelt's (1989, 1993) modular model, initially designed to account for speech production in monolingual speakers, and now used for bi-, tri-, multilingual speakers (see de Bot, 1992 for an adaptation to a bilingual model below). The model is characterized by having two principal components (the rhetorical/semantic/syntactic system and the phonological/phonetic system), and three knowledge stores (the mental lexicon, the syllabary and the knowledge of the external and internal world). A distinction is also made between declarative knowledge (knowledge 'that'), and procedural knowledge (knowledge 'how').

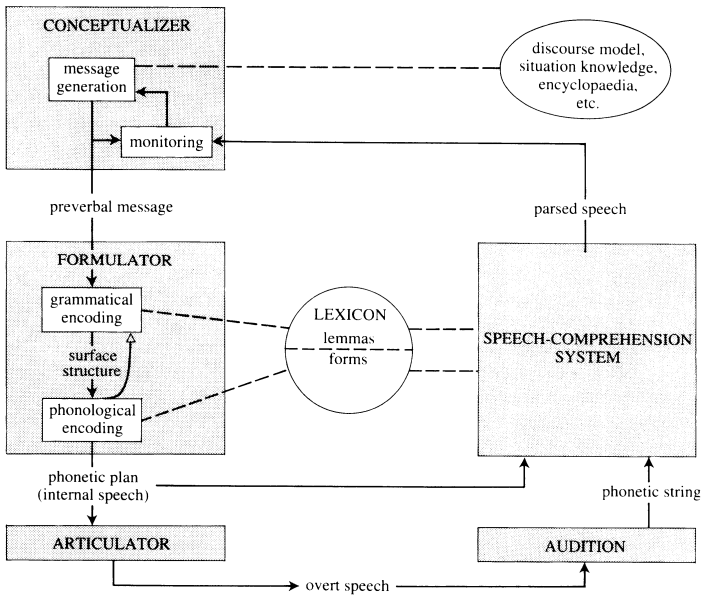


Figure 2.1. Levelt's speech production model (taken from DeBot, 1992:3).

As figure 2.1. shows, according to Levelt, in order to produce speech, several steps are needed. First, the pre-verbal plan is prepared in the conceptualizer, containing all the information which is necessary for the conversion into language. This pre-verbal plan is then passed on as input to the formulator, where the grammatical encoding, then the morpho-

phonological encoding, and last, the phonetic encoding take place. It is in the phonological encoding where the metrical and segmental features are set. The last step occurs in the articulator, where the articulatory score is generated and becomes actual speech. The model assumes that processing takes place in these three different components, and that they can be carried out at the same time, due to the specificity of each process. As stated by Kormos (2006): "the incremental, parallel, and automatized nature of processing needs to be assumed in order to account for the great speed of language production." (p. 8).

Monitoring, as one of the components of language production, together with the conceptualization, formulation and articulation of speech, is also part of the model. The monitor allows for the checking of the output produced in the different components of the model. Three monitoring loops have been defined by Levelt; the first is located right before the formulator, the second, before the articulator, and the third, right after the articulator, once speech has been produced. Monitoring in the L2 is highly different from monitoring in the L1, mainly due to attentional resources' limitations.

Even though the model has been much followed among researchers, attention has been called to some limitations. Poulisse & Bongaerts (1994) point out a problem with Levelt's 1989 model regarding the morphological encoding, arguing that it is very difficult to account for all inflected word forms to have their own separate representations in the lexicon when looking at unintentional code-switching. They suggest that the verb stems and inflectional morphemes are accessed separately, and are later on combined together. So, their main proposal is that "lemmas for the base forms and inflectional morphemes are stored and accessed separately" (Poulisse & Bongaerts, 1994:50).

De Bot (1992) represents a first attempt to fully adapt Levelt's model to bilingual speakers. He takes the steps described by Levelt to produce speech and tries to accommodate them into a bilingual system. He argues that Levelt's model proves to be appropriate to describe bilingual speech production, when only a few changes are made. The most relevant of these changes is that the choice of language to be produced is made in the conceptualizer. When the language-specific preverbal plan meets the formulator, this one works very much in the same fashion as in Levelt's monolingual model. However, for the bilingual model, two speech plans are produced, even though only one is selected. The other stays in the background. As for the lexicon, he suggests that there is only one lexicon where lexical items are stored together. And in contrast to the formulator, he claims the articulator to be non-language specific, "the bilingual speaker must have models for all sounds/syllables in the different languages." (de Bot, 1992:16).

Poulisse & Bongaerts (1994) point out two main problems with de Bot's account of bilingual speech production. The first one being the assumption that language choice is made in the conceptualizer, but then, two languages specific speech plans are formulated. "It is not clear how two alternative speech plans can be formulated in parallel if the information in the pre-verbal message raises the activation level of one of the languages only." (Poulisse & Bongaerts, 1994:40). The other drawback is economy. Having several alternative plans seem to be quite uneconomical for the production of real-time speech.

Poulisse & Bongaerts (1994) suggest that the pre-verbal message contains a further language component, so more than one language is activated, in line with the postulated of spreading activation models. As in these models, Poulisse and Bongaerts also suggest that high-frequency words and words that have been just produced need less activation than low-

frequency words. They go a bit further by claiming that L2 lexical items "receive more activation than the corresponding L1 lexical items when the pre-verbal message contains the specification [+L2]." (p. 42). They also suggest that "inflected word forms are stored in the lexicon both fully and in decomposed form and there is a checking device used to intercept forms that are not represented in the lexicon." (p. 52), accounting for Levelt's limitation on this issue.

The models and theories presented here were designed to account for how language is processed and produced. However, they were not conceived to provide an explanation for how knowledge is developed in the process of producing L2 speech, a central issue in SLA research, and for this study in particular. In order to account for how this development occurs, researchers in SLA have drawn on theories of automatization, by trying to relate them to speech production models (Towell et al., 1996; DeKeyser, 2001, 2007b; Kormos, 2006). If we want to investigate the development of L2 learners' fluency and prosody, we need to consider these theories and understand how they work. The following section is devoted to the description of these theories of automatization and how they can be used to describe the development of knowledge in the process of producing speech.

2.2. Theories of Automatization and SLA

Theories of automatization were first developed within the field of psychology, and as just mentioned, they were adapted by SLA researchers to provide new insights into the process of L2 development in speech production. Kormos (2006) propounds a classification of cognitive theories into 'rule-based' approaches, and 'item-based' approaches. The main distinction between the two being whether learning draws on rule-based (conceptually-driven) or item-based (data-driven) processes. In this

subsection, the two approaches are described, together with a brief presentation of the main theories comprised in each approach.

2.2.1. Rule-based approaches

The main focus of rule-based approaches lies in the conversion from declarative to procedural knowledge. In order to convert this declarative knowledge to procedural rules, practice is required. At the beginning of practice, the application of rules is a conscious, controlled process, which relies heavily on several cognitive resources, especially attention. With practice there is a gradual withdrawal of attention, hence the process becomes proceduralized and then, automatized. Fluency is affected at each stage of this conversion process, ranging from slow and hesitant speech to fully automatized, rapid, non-hesitant speech.

Anderson's (1983) ACT (Adaptive Control of Thought) theory, proposed by Crookes (1991) as potentially applicable to SLA, was the first adapted to attempt to describe how the conversion from declarative to procedural knowledge occurs. Anderson contends that for an automatic process to develop, both a quantitative and a qualitative change are needed. These changes can only be instigated by practice. And it is through practice that the automaticity necessary to accomplish the real-time requirement of fluent speech can be achieved.

Apart from speed of performance, Cheng (1985) adds to the approach the creation of new mechanisms, such as the restructuring of tasks. Moreover, Segalowitz & Segalowitz (1993) further supports this idea of practice that motivates both quantitative and qualitative changes. They investigated this qualitative change by analyzing data taken from different tasks, and looked at the variability of performance under different practice conditions. At early stages, performance was reported as much more variable and

errorfull than later on in the processes of proceduralitzation, when speech was less variable and contained fewer errors.

In their 1996 study, Towell et al. made an attempt to explain how fluent speech is developed, using Anderson's (1993) idea of 'proceduralization of knowledge' within Levelt's (1989) model of speech production. Using data from English learners studying French abroad, they performed both a quantitative and a qualitative analysis of participants' speech and suggest that proceduralization and automatization occurs in the formulator, being the place where declarative knowledge is turned into procedural rules. Several measures to track fluency development were used. Among them, mean length of runs (MLoR) proved to be the most appropriate to identify the conversion of declarative to procedural knowledge taking place in the formulator.

Another key issue in rule-base approaches is how automaticity develops from the learning of rules. Do learners simply add up rules and practice them until they become automatic? This question has been answered based on the creation of macroproductions from smaller units. Anderson (1983) accounts for how composition is conceptualized by using macroproductions, or chunking, in which single items are combined in larger units to produce lexical items easier to retrieve.

Anderson's ACT theory belongs to the traditionally called information processing approaches, and especially to the Skill Acquisition Theory. Anderson's (1983) ACT theory was adopted by SLA researchers to explain how skill acquisition works. However, there are two researchers who have most contributed to the expansion of the Skill Acquisition Theory in SLA research: Segalowitz & Segalowitz (1993), Segalowitz (2000, 2003) and DeKeyser (1997, 2001, 2007a). Both DeKeyser (2001) and Segalowitz (2003) suggest that an integration of rule-based and item-based processes

would enhance the production of fluent L2 speech.

Segalowitz & Segalowitz (1993) investigated how reaction times (RTs) could be used as a measure of qualitative change, instead of equating it only to fast speech. They introduced a coefficient of variability (CV) that included RTs and standard deviation to track the qualitative change. They conclude by saying that: "if the RT is reduced with training and changes in CV correlate significantly with changes in the RT, then we may conclude that some kind of restructuring or reorganization has occurred to reduce the SD over and above what would be expected by virtue of the RT becoming faster." (Segalowitz & Segalowitz, 1993:374). So, both the increase and restructuring of speech are results of the automatization process having taken place. A decade later, Segalowitz (2003) further investigated the issue by focusing on the relationship between automaticity, attention and skill in L2 speech production. He made the observation that "as one's skill level in a domain increases, the amount of attention and effort required to perform generally appears to decrease." (p. 382). However, he suggests that automaticity not only benefits by relieving attentional resources, and fast information processing, but also by promoting fluency. This fluency can be promoted by: "automatic execution of certain aspects of L2 performance such as pronunciation, grammatical processing, and word recognition" (Segalowitz, 2003:401). Nevertheless, Segalowitz (2003) also claims that automatic processing is not enough to promote fluency, controlled processes are also necessary for its development.

DeKeyser's research has been very much linked to the application of skill acquisition theories to SLA. Within this framework, and in order to find out how principles work in L2 acquisition, he developed several studies, covering aspects having to do with automaticity, such as practice, or skill-specificity, among others. DeKeyser's (1997) main aim was to find out

whether L2 learning could be equated to the learning of other skills. He found positive evidence for that with participants following the power law of practice. He then went on to investigate whether the type of practice undergone by each group would be transferred from one skill to the other. In this case, findings revealed that practice is skill-specific, its effects not transferring from comprehension to production and vice versa. DeKeyser (2001) focused on the transferability of practice from one skill to another. Once knowledge has been proceduralized, it cannot be generalized to other aspects. However, declarative knowledge can be easily generalized from one use to another. In order to explore one of the aspects closer to automaticity, DeKeyser (2007a) edited a collection of studies on practice, giving a very thorough account of how practice can benefit SLA. (this issue is further described in section 2.3.3)

Apart from rule-based theories and models, which rely on the proceduralization of knowledge from declarative to procedural rules, a different group of researchers has explored the issue of automatization taking a different perspective (Logan, 1988; MacKay, 1983; Ellis, 2003). For these researchers, automatization of knowledge does not involve the conversion from declarative to procedural knowledge, but rather memory retrieval of previously encountered items. These are explained in the following section.

2.2.2. Item-based approaches

Item-based approaches are data-driven, bottom-up approaches. They are not concerned with the conversion from declarative to procedural learning, but with memory retrieval. For these approaches, memory plays a central role in the development of automaticity. Moreover, memory capacity conditions the amount of learning that takes place. Here again practice has a central role but as a strengthener of associations, and

resulting memory traces.[6]

Logan's (1988) instance theory pioneered the item-based approach for the development of automaticity. As Logan himself puts it, "automatization reflects a transition from algorithm-based performance to memory-based performance." (p. 493). Memory process is seen as a race between these two types of performance. Only with practice can the memory-based performance prevail over the algorithm-based. Automaticity will be reached when a problem is solved time and time again, and the solution is stored as one unit that can be further retrieved when the problem is encountered again. This is how the approach accounts for the acquisition of formulaic language as well. First, L2 speakers apply rules to the production process because "linguistic units are not yet sufficiently encoded in memory" (Kormos, 2006:46). However, with practice, the formulas become strongly rooted in memory helping to speed up the retrieval process, hence benefiting fluent speech production.

MacKay (1982) strength theory also belongs to the item-based approaches, since it relies completely on memory. It is a node-structure approach which views practice as the condition to strengthen links between the nodes. Put in MacKay's own words, "repeated activation increases linkage strength, which results in a higher asymptotic level of priming and a faster accrual of priming per unit time across one particular connection." (MacKay, 1982:490). It is when strong links have been created in memory between concepts and lexical items that retrieval is faster.

Chunking theories have also been used to explain how speech production is automatized. They first started by analyzing how children acquire their

---

[6] By memory traces we refer to the information concerning word meaning.

L1, but have later been applied to L2 learning. Chunking can be regarded as the creation of large units of meaning from already existing chunks, or the break down of long phrases into shorter meaningful units. N. Ellis (2003) claims the importance of chunking and memory in language acquisition:

> "The notion of chunking has been at the core of short-term memory research since Miller (1956) first proposed the term: while the chunk capacity of short-term memory (STM) is fairly constant at 7±2 chunks, its information capacity can be increased by chunking, a useful representational process in that low-level features that co-occur can be organized together and thence referred to as an individual entity." (Ellis; 2003:23)

It is through the repetition of these chunks, which are then retrieved as one single unit, that the process of language production can be speeded up, hence fluency improved. "Native-like competence and fluency demands such idiomaticity." (Ellis 2003:12) At later stages, learners start to analyze these chunks, and deduce rules from them, using the language in a creative manner.

A more recent theory in SLA, which can also account for automaticity in language learning, but does not belong to the information processing theories, is the emergentist approach. Language acquisition is regarded as an associative, probabilistic, and usage-based learning (Ellis, 2003). As this author puts it: "emergentists believe that many of the rule-like regularities that we see in language emerge from the mutual interactions of the billions of associations that are acquired during language usage." (p. 4). Frequency of occurrences in the input to which L2 learners are exposed is central for the understanding of this approach. Through this input learners automatically extract information about the frequency and the co-occurrence of items, which is then stored in memory. Using this information, and taking into account contextual factors, learners can make predictions of what the output is going to look like. "The interactions that constitute language are associations, billions of connections which co-exist

within a neural system like organisms co-exist within an ecosystem. And systematicities emerge as a result of their interactions and mutual constraints." (Ellis 2003:32) The development of fluency is accounted by the large number of hours performing a task, creating associations between and within linguistic elements that will later help fast retrieval.

In spite of the many differences between rule-based and item-based approaches, there is important common ground shared by both views: the support for the power-law of practice. The power law of practice states that reaction time and error rate decrease linearly with practice, until optimal performance is reached. Once optimal performance is reached, improvement stops. (see DeKeyser, 2001, 2007 for a complete description).

The goal of both types of approaches in SLA is to describe how L2 speech becomes automatic. However, as we have seen, there is no agreement about the means to do so. Certain aspects of fluency are strongly related to automaticity. Segalowitz (2003) reads: "To the extent that fluency represents the ability to speak or read quickly, accurately, and without undue hesitation, then automatic execution of certain aspects of L2 performance such as pronunciation, grammatical processing, and word recognition would, by definition, promote fluency." (p. 401). Controlled speech tends to be non-fluent, and it is only when new knowledge has been practiced and automatized, that speech can become fluent. The following section is devoted to the factors required when producing fluent speech.

2.3. Central Phenomena in Speech

Up to this point, the main theories and models of speech production related to SLA research have been described, in conjunction with the

automatization theories which have been adopted to account for the development and processing of L2 speech production. These theories and models provide the framework to understand where the construct of fluency comes from, and the role it plays. However, apart from the theories and models presented, specific factors are required to create the optimal conditions for fluency to develop. The phenomena contributing to the development of fluency in L2 speakers are: automaticity, output, practice, and memory.

2.3.1. Automaticity

Even though the term automaticity is claimed to be key in the study of speech production, to our knowledge, a sole attempt has been made to interrelate theories of automaticity and speech production models. Kormos (2006) dedicates a section of her comprehensive book to assemble these two domains. In order to bring them together, she suggests some considerations which need to be made: "whether language learning is a rule- or item-based process and whether language production is mainly creative or memory based." (p. 44). Then, she goes on to illustrate how theories of automaticity can account for learning formulas (see table 2.1) and for the development of language production processes.

The use of formulas, or formulaic language in L2 speech production is regarded as beneficial for the development of fluency. Table 2.1. illustrates how different theories account for the use of formulaic language in speech production. Modular theories differ from the other two, in that formulaic language seems to rely on controlled processing by undergoing the process of syntactic encoding. On the other hand, chunking theories and instance theories are based on memory retrieval, with formulaic language accessed as one unit.

Table 2.1. Theories of Automaticity's approach to learning of formulaic language.

	Modular Models	**Chunking Theories**	**L. Instance Theory**
Formulaic Language	- Taking place at the conceptual level and retrieved as any other lexical item.	- Lexical items that often occur together tend to form chunks.	-Competition between rule-based processing and memory retrieval.
	- Produced by the same process of syntactic encoding.	-Chunks are retrieved as one unit.	- Formulas accessed in memory as one unit.

Adapted from Kormos 2006.

Kormos (2006) refers to memory strength theories as the most appropriate ones to explain the automatization of lexical encoding. On the other hand, for an explanation of how syntactic and phonological encoding takes place, both Anderson's ACT and chunking theories are regarded as the most appropriate.

Within SLA studies, DeKeyser (2007b) observes that the term automatization has been used with different senses. In the broadest sense, automatization refers to the whole process of converting the rules present in declarative knowledge to "the final stage of fully spontaneous, effortless, fast, and errorless use of that rule, often without being aware of it anymore." (p.3). As for the narrow sense, "it refers to the slow process of reducing error rate, reaction time, and interference with other tasks that take place after proceduralization." (p.3). He distinguishes a still more specific sense, which is merely a "quantitative change in the subcomponents of procedural knowledge to the exclusion of any qualitative change or restructuring." (p. 3).

Automaticity is perhaps the phenomenon most closely related to fluency. As we have seen, once the encoding and retrieval processes have been automatized, fluency is enhanced. However, for automatization to take place, it is necessary certain other phenomena occur. The production of output is the first step towards automatization, and improvement of fluency. Second, this output needs to be repeated, hence practiced, if the language is to be automatized. It is through practice that learner's L2 speech production will become more fluent. And finally, each individual's memory capacity will also play a role in the language that can be processed at a time. To all these phenomena the following sections are devoted, beginning with output.

2.3.2. Output

This subsection dwells on the notion of output, briefly presented in Chapter 1 above. The notion of output in SLA research emerged naturally from discontent with the (at the time) dominant Krashen's Input hypothesis. First, Long (1981) argued that unidirectional input was not enough to account for the acquisition of an L2. He presented his *Interaction hypothesis*, which claimed that interaction was important between interlocutors in order to negotiate meaning. It is through this negotiated meaning that the input the learner receives becomes more directed to his/her developmental needs. In his 1996 revised version of the *interaction hypothesis*, Long added the notion of *noticing* and *attention* (initially proposed by Schmidt, 1995) as crucial for acquisition to be facilitated.

After a first short incursion made by Bialystok (1978) on the role of output in her model of L2 learning, Swain (1985, 1995) has been the researcher who has most contributed to expand the relevance of the role of output in SLA. The origin of the *output hypothesis* came from evidence from the analysis of results obtained in the Swain's Canadian immersion

programs study that led her to reject the idea that comprehensible input was enough for succesful interlanguage development. She introduced the notion of output as an essential element to promote acquisition (see Gass & Mackey, 2006; Izumi, 2002; Shehadeh, 1999, for an overview). According to Swain's *Output Hypothesis*, production causes learners to engage in syntactic processing and in doing so, acquisition is promoted.

Even though the role of output in second language acquisition was not as widely accepted as that of input for many years, in the last decades it has gained more adherents. Even Ellis, who was very skeptical on this idea in most of his work, in his 2008 book writes: "It is becoming clear that output contributes to language acquisition." In Swain's revised form of 'pushed output', learners have to produce messages that are concise and socially appropriate. Ellis also supports this idea by writing that "production requires learners to process syntactically; they have to pay attention to form." (2003:113). This production will then assist interlanguage development. Some researchers as de Bot (1996) and VanPatten (2004) still view production output as merely enhancing fluency and automaticity of processing. In contrast, when trying to account for the way L2 output affects cognitive processes involved in SLA, Swain identified the following specific functions of output in L2 learning: a) noticing, b) hypothesis formulation and testing, c) metalinguistic function, and d) syntactic processing. Building on this categorization, Skehan (1998) adds three more roles: e) help to automatize already existing L2 knowledge, f) promote the development of discourse skills, and g) the creation of a 'personal voice'.

Noticing is central to output. Muranoi (2007) states that noticing "trigger[s] important cognitive processes such as selective attention and cognitive comparison" (p.57). When speaking in an L2, learners are likely to notice that they lack the knowledge to produce part of the message they

want to convey. It is through this that they become aware of their limitations, and focus their attention on relevant input to overcome these limitations. Hence, the production of output promotes noticing a 'gap' in the interlanguage (IL) system and it also helps noticing the 'gap' between what they are able to say and the actual target language.

Hypothesis formulation and testing refers to learners using output in order to try new forms (hypothesis) to transmit meaning (Swain, 1998). This produced output serves to create feedback for the same learners, since their hypotheses can be tested against the feedback they receive from either external resources or internal knowledge (Schachter, 1986). Support to the positive effect of external feedback on learners' output has been given by interaction studies (Long, Inagaki & Ortega 1998; Muranoi, 2000; among others).

The metalinguistic function described by Swain means that learners use language to reflect on language (Swain, 1995, 1998). The metatalk helps learners to reflect on their language, making them aware of forms and linguistic rules (Swain, 1998). Related to the metalinguistic function is syntactic processing. Swain (1985) claims that learners need to turn to syntactic processing when producing language in order to convey their intended meaning.

It is by producing speech that learners can practice the knowledge they have already acquired (or semi-acquired), hence pushing the automatization of linguistic knowledge and promoting the development of discourse skills. This is in turn important for the creation of a 'personal voice', since this voice is created through personal engagement in topics the learner finds interesting and motivating.

2.3.3. Practice

As has been mentioned in the last paragraph of the previous section, practice seems to be key to promote automatization of both linguistic knowledge and discourse skills. Hence, output and practice go hand by hand in the acquisition of language.

The concept of practice has gained renewed attention in the SLA research in the last years, especially with DeKeyser's (2007a) book. The term is not new for SLA researchers. Behaviorists believed that practicing highly mechanistic, form-focused exercises (commonly known as controlled drills) led to linguistic achievement. Chomsky (1986) and Ellis (1993) considered practice as relevant only for performance or procedural knowledge, having nothing to do with grammar rules. In contrast, cognitive theories of acquisition, and especially the skill acquisition theory, consider practice as essential (although not sufficient) for the development of second language skills. It is within this frame that DeKeyser (2007b) defines practice: "specific activities in the second language, engaged in systematically, deliberately, with the goal of developing knowledge of and skills in the second language" (p. 1).

DeKeyser's (2007a) approach to practice is derived from an amalgam of cognitive psychology, educational psychology and applied linguistics. He expresses his dissatisfaction with the few attempts made in applied linguistics to explain the role of practice as useful only to enhance fluency (Ellis, 1993; VanPatten & Cardierno, 1993). Nonetheless, a shift seems to have taken place in the last years towards an increased importance given to output practice, especially to stimulate noticing (VanPatten, 2004), reflect on structure (Ellis 2003), and formulate and test hypotheses (Swain, 1985; Muranoi, 2007) as has just been explained.

In SLA research the issues of skill-specificity and transfer of practice are the ones which have drawn more attention. One of the most oft-cited articles on this issue is DeKeyser's (1997) study on the cognitive processes needed to learn an artificial language. His main findings include the confirmation that his data followed the power law of practice, and that practice is skill-specific. Lack of transfer between receptive and productive skills has also been reported by DeKeyser & Sokalski (1996) and Izumi (2002). The problem of transfer can be of two types: transfer from declarative knowledge to procedural skill, as tackled by the studies just mentioned, and transfer from knowledge and skill from one context to another. Studies on the second aspect are scarce and have centered around research on SA. So far, results are not conclusive but seem to suggest that transfer may not always take place (Brecht et al., 1995; DeKeyser, 2007b).

For Instance Theories, practice is the means to reach automaticity. Their basic premise is that if an item is presented many times, it will then be easy to retrieve. Memory retrieval becomes faster with practice. Practice helps to lessen information-load effects and dual-task interference, hence speeding up processing and reducing variability. Logan (1988) refers to the last phenomenon claiming that "the instance theory predicts that the standard deviation will decrease as a power function of practice." (p. 495). He also claims that automatic processing is specific, hence practice helps to improve only on those items that were practiced, but not others, even though they are similar. Logan's (1988) instance theory has been criticized for this assumption. However, Kormos (2006) refers to a new version of the theory, the exemplar-based random-walk model (Nosofsky & Palmeri, 1997 cited in Kormos 2006:46) which allows for memory retrieval of similar stimuli.

For skill acquisition theory, practice has a crucial role in transforming declarative/explicit knowledge into procedural/implicit knowledge. This

procedural knowledge can be later transformed into automatized knowledge. The only way for learners to achieve the automatization stage is through large amounts of practice. As DeKeyser (2007c) describes it, "automaticity can be gradually increased under pressure; the initial stages of proceduralization, on the contrary, require careful, deliberate use of the relevant declarative knowledge in the execution of the target task." (p. 216).

For learners to fully acquire an L2 two different types of practice are required: practice aiming at a *fluent rule use*, or a *fluent use of chunks*. "In the case of rule use, what is needed is repeated rule retrieval under increasingly demanding task conditions after initial proceduralitzation. In the case of chunk use, ample short-term recycling of the same sentences with minimal change is required." (DeKeyser, 2007d:293). If learners are still in the declarative-procedural stage when going on a SA, they are likely to base their practice almost exclusively on the *use of chunks*. Students with a functional level of the language would probably insert the use of both types of practice (DeKeyser, 2007c). We now turn to the factor memory.

2.3.4. Memory

The construct of memory, just as automaticity, output and practice, plays a central role in models of speech production, and adaptations made in SLA (within a cognitivist perspective). Two types of memory are involved in the processing of language: long-term memory and working memory.

Long-term memory is characterized by the possibility of storing information for long periods of time, and for its unlimited capacity. Within long-term memory, a further division is made between explicit-declarative memory and implicit-procedural memory. The main difference between the two being that knowledge stored in the explicit memory is

knowledge that can be recalled consciously, whereas knowledge stored in the implicit memory is acquired unconsciously. In contrast, working memory has limited capacity, and controls processing.

Anderson's ACT model, adopted by many SLA researchers, provides a good account of how memory is integrated into a production model. The interaction among the three different kinds of memory (declarative, production, and working memory) is exemplified. Long-term memory can only be accessed through working memory. Implicit and explicit knowledge are combined in production in order to produce fluent speech at a high speed.

As we have previously seen, for the ACT model, all knowledge is assumed to be initially declarative, and conversion from declarative to procedural knowledge is essential in order to speak fluently. This conversion is required due to the special nature of the working memory. Working memory has limited capacity and declarative knowledge requires a lot of attention, hence a large amount of space which is not available. Besides, knowledge stored in declarative form is retrieved by interpretive mechanisms which are controlled by the speaker. On the contrary, procedural knowledge does not need the attention of the speaker, since it consists of productions in the form IF/THEN, which are accessed by match and execution. Hereby, procedural knowledge can be processed by the working memory in larger units without exhausting its capacity.

A different approach to the study of memory in SLA has been taken by those researchers interested in implicit learning as a means by which an L2 might be acquired. Their main concern is to find out whether it is possible to extract rules from L2 data without being given any explicit instruction on these rules. Robinson (1997) designed a study directed to answer this question among others. He used a design of four levels of learning

conditions. He concluded that it is possible to learn implicitly, but he himself recognizes that his conclusions are only provisional due to lack of comparable studies.

The role of memory, specially working memory, has also received much attention in the area of individual differences, especially in relation to aptitude. SLA researchers are investigating whether working memory capacity has an effect on L2 achievement: grammar, vocabulary, listening, overall proficiency, speaking, etc. Up to this moment, results are still preliminary, and conclusions seem to point to a more complex panorama than may appear at first sight. Studies by O'Brien and colleagues have looked at the relationship between phonological memory and L2 speech production. In their 2006 study, results showed that memory was related to success in different areas depending on the L2 proficiency level. In O'Brien et al.'s (2007) study, the relationship between phonological memory and fluency is investigated. By observing the performance of English speaking adult learners of Spanish in a phonological memory task and temporal/hesitation phenomena in two different learning contexts, the researchers conclude that there is a contribution of phonological memory in L2 oral fluency development. However, they also suggest that "the role of phonological memory in language acquisition is more closely related to the learner's level of language competence rather than to the learner's age as such." (O'Brien et al., 2007:578).

In a doctoral dissertation entirely devoted to the study of working memory, Mizera (2006) sought to extend on the study of the relationship between working memory and second language speech production. He hypothesized that individual differences in working memory capacity would correlate significantly with L2 oral fluency. Results did not confirm his hypothesis. As we have mentioned above, the relationship between working memory and language achievement may not be so

straightforward as has been often claimed, and other variables may be at play. The researcher himself points at some possible factors that may have interfered: the complex nature of speaking a foreign language fluently, which may need more than a good working memory, personal and affective factors, and the nature of the working memory itself.

As we have seen in this section on speech production models and approaches in SLA, there are several conditions that need to be met for L2 speech production to develop towards a fluent native-like performance. Speech production and automaticity theories provide the basis to explain how L2 speech is produced and develops along time. However, in order to precisely understand how this occurs, several phenomena need to be taken into account: the role of automaticity in the development of fluency, the importance of output production to push this development, the central role of practice as an instigator of automaticity and fluency, and the relationship of memory and processing capacity. As Segalowitz (2003) mentioned, "in language learning, increased performance efficiency can be seen as contributing to fluency, that is, the ability to use language rapidly, smoothly, and accurately." (Segalowitz, 2003:383). To this construct of fluency the following section is devoted.

3

Fluency in SLA

As specified in the previous chapter, fluency is the product of the operation of the speech production system of both native and non-native speakers. When speaking a first language, fluency in speech seems to be taken for granted; speech is delivered in a fast, effortless, natural manner. On the other hand, when speaking an L2, the development of fluency requires initially much effort, and then large amounts of practice for automatized language to emerge (Segalowitz & Segalowitz, 1993; DeKeyser, 2007b). But, how does fluency develop in an L2? Which conditions have to be met for fluency to develop towards native-like standards?

Apart from the theoretical and applied linguistics' interest in fluency, pedagogical research has also shown an interest in the study of this issue. An illustration of this is Chambers (1997) claim for the need to understand how fluency develops in order to "create the conditions in which foreign language learners increase their fluency" (p. 535). Brown (2003) pictures fluency as a learnable skill, which is basically a continuum that learners have to cover: "Fluency is probably not an absolute characteristic that students either have or do not have. If, in fact, fluency is a matter of degrees, students at any level of proficiency can probably achieve some degree of fluency" (p.7).

The following sections attempt to provide a comprehensible picture of how the study of fluency has evolved from early L1 studies to current

research on L2 fluency. Firstly, a review of the most significant studies on L1 and L2 fluency, together with a summary of the ongoing debate on a definition of L2 fluency is provided. A special mention is made of the role of fluency in study abroad (SA) research (section 3.1.). Secondly, the issue of measures in fluency research is considered; especially with respect to perceived, temporal and phonological fluency (section 3.2.).

3.1. The development of fluency: from L1 to L2

This section presents a summary of the most significant findings in L1 fluency in relation to L2 fluency (section 3.1.1.). Next, subsection 3.1.2. is entirely devoted to L2 fluency research. First, an attempt to compile the different definitions of fluency used in SLA is made (section 3.1.2.a) followed by a section on models of fluency used in SLA (section 3.1.2.b). Then, section 3.1.2.c provides an overview of the studies on L2 fluency undertaken within SLA research, followed by section 3.1.2.d devoted to the studies of L2 fluency in SA contexts.

3.1.1. L1 fluency

Native speakers of any language are expected to speak their L1 fluently, as well as accurately, and socially and pragmatically appropriately. Fluency in native speech is taken for granted, in spite of evidence arguing against this extended belief (Jefferson, 1979; Schegloff, 1979; Grosjean, 1980, cited in Riggenbach, 1991).

Lennon (1990) claims that not all native speakers perform in exactly the same way regarding fluency and that "any individual native speaker may be more or less fluent according to topic, interlocutor, situation, "noise," stress, and other factors." (p. 392) Furthermore, Chambers (1997) argues that albeit non-fluent phases in L1 speakers exist, they respond to a need

to accurately articulate thought.

> "Non-fluent phases in L1 speech production are more likely to be meaning-oriented, in an attempt to bring language and thought as close as possible, whereas foreign language learners may struggle both with expressing their meaning and focusing on basic morphological production. (Chambers, 1997:538)

Apart from Chambers' argument, other researchers have looked at the differences between L1 and L2 speech production in relation to fluency, and they all agree that speed makes up for the largest distinction between one and the other. Native speakers speak faster than non-native speakers, basically due to the automatic nature of the process by the former. Non-native speakers, on the contrary, have to cope with attentional effects underlying their production, apart from planning and monitoring, common to all (native and non-native).

A further distinction between L1 and L2 speech is the use of pausing phenomena. L1 speakers tend to pause at utterance or clause boundaries, while L2 speakers pause within those boundaries as well (Riggenbach, 1991; Mora & Valls-Ferrer, *submitted*). Hilton (2009), in a methodological article analyzing temporal aspects of spoken fluency, points to temporal dysfluencies present in native speakers speech "but in relatively stable or predictable quantities" (p. 645). Hilton reports that NSs in her study hesitate 70% of the time at syntactic or discursive boundaries, and only 30% within these boundaries. For NNSs, the percentages are rather different, with approximately half of the hesitations occurring clause-internally[7]. In addition, Raupach (1980) argues that L1 pausing profile tends to be transferred to the L2, so that when speaking an L2, speakers

[7] Hilton (2009) distinguishes between a group of non-native fluent learners and non-native dysfluent learners. For the fluent learners, the percentages are 54% of hesitations at boundaries, and 46% within clauses. The percentages are reversed for dysfluent learners, with 56% within clauses and 44% at clause boundaries.

will never exceed L1 fluency.

In a study that investigated the effects of task complexity on native and non-native speakers' fluency, De Jong et al. (2007) found that NSs' fluency was enhanced when performing a complex task, if measured as breakdown fluency (phonation time ratio). However, NSs' performance was not affected by the cognitive complexity of the task performed when measured as speed fluency (syllables per second). This no-effect, or positive effect of task complexity on NSs' fluency contrasted with NNSs' results, which were negatively affected by increases in task complexity.

Tauroza and Allison (1990) conducted a study which established an estimate of British English speech rates for different types of situations. The study is a replication (and expansion) of Pimsleur et al., (1977). Findings differred from the original study by Pimsleur, and they established a new estimate of standard rates (given in syllables per minute); the English NS speech ranged from fast, above 320 syll./min., to slow, below 190 syll./min., depending on the degree of formality of the situation.

3.1.2. L2 fluency

As already mentioned in the previous subsection, L2 fluency differs from L1 fluency in some of its main traits, especially the speed at which it is delivered, the attentional processes involved and the distribution of pauses in speech. These differences are basically grounded in the automatic nature of the oral production process for NSs and the non-automatic nature of NNSs' speech production, which requires devoting varying amounts of attentional resources to the speech production task as a function of speaking ability and proficiency level. Kormos (2006) puts it in the following manner: "Due to the lack of automaticity, processes of L2

production cannot run in parallel as in L1, which slows speech down to a considerable extent." (p. 154).

3.1.2.a Definitions of fluency in SLA

Many are the researchers who turn to the division first made by Lennon (1990) between broadly and narrowly defined fluency. Fluency, understood in the broad term, equals in many cases global oral proficiency. Fillmore (1979) gives a complete four-step definition of fluency in this sense, including aspects such as pragmatic ability, creativity and imagination. When narrowly delimited, fluency is mostly understood in terms of temporality. Lennon (2000) provides a precise definition in its narrow sense: "a working definition of fluency might be that rapid, smooth, accurate, lucid, and efficient translation of thought or communicative intention into language under the temporal constraints of on-line processing" (p.26). Chambers (1997) writes: "For fluency to be a useful concept for research into foreign language oral production, it needs to be clearly differentiated from overall language proficiency and from communicative competence." (p. 543).

In addition, Tavaloki & Skehan (2005) claim fluency is multifaceted, consisting of: speed fluency, breakdown fluency and repair fluency. Segalowitz (2010) would refer to these three aspects of fluency that Tavaloki & Skehan (2005) propose as part of utterance fluency, which are then complemented with cognitive fluency and perceived fluency. For Segalowitz (2010), a cognitive science approach to fluency would define L2 fluency as a property of the system linking cognitive fluency and utterance fluency. Chambers (1997) focuses on the role of fluency in communicative language teaching. In this sense, fluency is understood as a construct that is opposed to accuracy, and the use of language, independently of the proficiency level, and the ultimate outcome, is the

main purpose. Richards, Platt, & Weber (1985) defined fluency as: "the features which give speech the qualities of being natural and normal, including native-like use of pausing, rhythm, intonation, stress, rate of speaking, and use of interjections and interruptions." (p. 108). This definition of fluency takes into account both temporal and phonological dimensions of the speech production process.

The definition that Derwing, Rossiter, Munro & Thomson (2004) elaborate from Schmidt: "an automatic procedural skill on the part of the speaker and a perceptual phenomenon in the listener" (p.656) is probably one of the most complete definitions of fluency (understood in the narrow sense), since it focuses on both the speaker and the listener. This is the working definition of fluency that we will use in the present study. Therefore, we will relate utterance fluency to perceived fluency to gain a better understanding of the fluency construct.

Many studies examining fluency in second language acquisition have focused either on speakers' productions (Raupach, 1980; Towell, Hawkins & Bazergui, 1996; Riggenbach, 2000; Freed, Segalowitz & Dewey, 2004; among others), or to a lesser extent, on listeners' perceptions, using judgments from raters (Ejzenberg, 1992; Trofimovich & Baker, 2006). Additionally, a number of studies have compared both, L2 speakers' productions and raters' judgments (Lennon, 1990; Riggenbach, 1991; Freed, 1995; Derwing et al., 2004; Kormos & Dénes, 2004). These studies aim at reaching a working definition of L2 fluency and at providing an account of the processes underlying fluent speech production. Results do not totally coincide. However, a common pattern seems to emerge with respect to several fine-grained measures, such as speech rate and pause duration (and to some extent fluent speech runs), as best correlating with raters' judgments. This is an issue which we fully explore in section 3.2.1 below.

3.1.2.b Models of Fluency in SLA

The development of L2 fluency has been investigated within several speech production models and theories of automatization. Within information-processing models, two different approaches have been proposed; the *Universalist approach* and the *Individual Differences approach*.

The *individual differences approach*, developed by Skehan (2002) proposes that differences in fluency among learners are due to the differences in the memory component of aptitude. In his previous work (Shekan, 1989, 1998), he proposed that analysis-oriented learners prioritize accuracy whereas memory-oriented learners prioritize fluency. Some support for this approach has been given by Kormos (1999), with learning styles also being a key part of it.

The *Universalist approach* regards fluency as the automatization of encoding processes, previous proceduralization from declarative knowledge. Development in fluency evolves as the result of practice, enabling the automatization of encoding processes involved in speech production. Studies in SLA supporting this model have been undertaken by DeKeyser (1997) and Towell et al. (1996).

However, these studies do not specify the type of practice that may be necessary to promote fluency. More recently, Segalowitz (2000) attempts to do that by suggesting that practice has to be transfer-appropriate, in the sense that the cognitive operations that have to be activated are the ones that the learner will shortly put into practice. In the same direction, DeKeyser (2007a) provides different approaches to what kind of practice may be more appropriate for acquiring a second language. He claims that good practice needs to involve real operating conditions as soon as possible, which means comprehending and expressing real thoughts, and

this necessarily involves a variety of structures, some of which will be much further along the declarative-procedural- automatic path than others. (DeKeyser 2007d:292)

On the opposite side, we find the memory-based theories, like instance theory, strength theories of automatization, and theories of chunking, that view fluency mainly as a process consisting of retrieval from memory, and to a lesser extent the application of rules. The deduction of rules from previously acquired chunks, or formulaic language, is claimed to be possible by these theories. They all agree with the previously mentioned theories in that practice plays a vital role in the development of fluency, either by helping with the automatization process or with retrieval from memory.

In a recent book devoted to the study of L2 fluency from a cognitive science perspective, Segalowitz (2010) provides a thorough account of how the study of L2 fluency has been approached and which directions should it take to advance in its understanding. He proposes that the following distinction should be made between three senses of fluency: cognitive fluency, utterance fluency and perceived fluency. Cognitive fluency is defined as "the speaker's ability to efficiently mobilize and integrate the underlying cognitive processes responsible for producing utterances with the characteristics that they have." (p. 48). Utterance fluency is the fluency which has to do with the features of the utterance, the characteristics and properties that an utterance has, often operationalized as temporal fluency. And perceived fluency refers to the inferences that speakers make from the degree of utterance fluency perceived in speech.

3.1.2.c. Studies on L2 fluency in SLA

The research undertaken by Dechert and Raupach at the University of Kassel is among the first attempts to fully account for the component of fluency in speech production and its reflection on temporal variables. Dechert (1980) reports on a case study of a student who went abroad for three months. He focused on the production of pauses and intonation in the second language of this student, reported to have become "less hesitant and more fluent" (p.273) after the SA. Raupach (1980) studies the performance of 5 French and 5 German undergraduate students by measuring several temporal variables both in their L1 and L2. Raupach suggests that the variables used (speech rate, silent pauses, articulation rate, length of runs and length of silent pauses) allow different degrees of fluency to be determined.

As suggested by Lennon (1990), Möhle was the first researcher to adapt temporal variables used in psycholinguistics to oral production research in EFL studies on fluency. Building on Möhle's (1984) suggestions, Lennon (1990) designed a study to find out to what extent certain variables were good indicators of perceived fluency. In order to do so, he used both instrumental analyses and raters' judgments. Whereas generalizations cannot be made, due to the exploratory nature of the study, with merely 4 subjects, findings are quite informative and they have been very suggestive for subsequent research, which has extensively built on Lennon's conclusions.

Riggenbach (1991) approached the study of fluency by performing a microanalysis of the oral productions of 6 Chinese learners of English. She compared ratings from English instructors with utterance and discourse level analysis, including a pragmatic qualitative analysis of the data. This was done by dividing the participants into two groups: a fluent group and

a non-fluent group. Then, Riggenbach compared the performance of both groups at the level of each individual. The author suggests that speech rate and unfilled pauses are the variables that best correlate with judgements of non-fluency, and that in the case of her study, discriminated better between the two groups of speakers. However, the qualitative analysis allows the author to conclude that fluency is a complex phenomenon, not so straightforward as it may seem at first, with non-fluent speakers at the heart of this complexity.

Ejzenberg (2000) followed Riggenbach's steps by arguing that a qualitative analysis of fluency is necessary. She analyzed the speech of Brazilian learners of English in Sao Paulo performing four different tasks (cued and uncued monologue and dialogue). Holistic ratings on fluency were given by four trained raters. Results showed a context-dependency (according to the four types of task) effect on fluency, as measured by the raters. As in Riggenbach (1991), two groups of speakers were established (the 3 most fluent speakers and the 3 least fluent speakers) and a detailed analysis of the productions of these speakers was conducted. Coinciding with Riggenbach, speech rate accounts for differences between the two groups. Ejzenberg goes a step further by claiming the necessity to add other features to fluency such as ease of expression and continuity.

Derwing, Rossiter, Munro & Thomson (2004) also observed the development of fluency by comparing different tasks (picture description, monologue and dialogue), in this case performed by mandarin-speakers. Their results confirm Ejzember's (2000) finding that perception of fluency varies across tasks. Apart from that, they also measured the relationship between temporal measures and raters' assessments of fluency. Their main conclusion is that temporal measures seems to account relatively well for listener judgments, at least for low-proficiency speakers.

3.1.2.d. SA context as fluency booster

Several of the SA studies mentioned in the first chapter have focused on the development of learners' L2 oral production. When considering SA research within SLA, the most widely reported finding is the improvement of learners' fluency during a SA period, picturing the context as ideal to improve this skill. Nevertheless, a close examination of these SA studies reveals that results are far more complex than the generalizations often made.

The common belief that going on a SA automatically enhances learners' fluency in the L2 was swiftly spreading during the 90s, a decade when SA research was experiencing a great boost. However, up to that moment, no study had scientifically tested that widely held impression. Studies focusing on the effect of the SA context on L2 fluency basically started in the 1990s (eg. Lennon, 1990; DeKeyser, 1991; Freed, 1995).

Before the 1990s, Dechert & Raupach (1980), and Moehle (1984), conducted several studies on German and French students going abroad. However, their main concern was not so much on the effect of the learning context on fluency, but the development of an understanding of L2 fluency, and the measures that could better define it (see the previous section 'Studies on L2 fluency in SLA' for a more complete description of their work). Lennon (1990) elaborated on these studies and observed a group of 4 German EFL students before and after a 6-month SA period in Britain. Again, his main aim was to quantify the characteristics of fluency, and no special attention was given to the context. However, the findings show that the 6-month SA period resulted in an increased speech rate and fewer filled pauses per T-Unit. Even though very inspirational for many researchers, most of these early studies suffer from several methodological shortcomings, either by having too few participants, by the limitations in

the type of data collected, or the instruments used.

Although quantitative in nature, DeKeyser (1991) includes a somehow more qualitative approach to SA for the first time in the field. Whereas his main aim was on grammar development, he also gave some insights on fluency. After observing a group of American students in Spain, he claims that the students became more fluent after the semester abroad, but he does not report the type of instruments used for the analysis. He finishes the chapter by acknowledging some of the limitations of the study, mainly the small number of subjects and the criterion measure for assessment. Nevertheless, the importance of this study falls on being the first study that truly deals with (and discusses) the effect of SA contexts on learners' performance, using a control group that stays at home.

In her 1995 book, Freed devotes a chapter to report on a study she conducted on the effects of learning contexts on students' fluency. The study was designed with the intention to find out whether NS raters would be able to differentiate between two groups of students, one that went on a SA and one that remained AH. Next, the qualities of these groups' speech, which contributed to the impression of fluency, were isolated. Results show that raters were able to distinguish between two groups, but only when considering students who had started with low fluency levels, and not so when highly fluent learners were included. When asked to rank a list of fluency components, according to the importance they assigned to them, 83% of the judges agreed that 'rate of speech and smoother speech with fewer false starts' were among the most important factors. The last part of the study identified particular qualities of the speech that affected the perception of fluency. Only one feature revealed significant differences between the two groups: speech rate. A clear tendency was in the direction of SA students producing fewer dysfluent pauses and longer speech runs. Nevertheless, Freed finishes her study by pointing out the importance of

individual differences within subjects in the same group for the interpretation of the results. The small number of participants in the study makes the results somehow limited. However, findings have been very suggestive for many researchers, and numerous studies have been undertaken after hers.

Towell, et al. (1996) also developed a very complete study on fluency for learners of French who spent a year abroad. They found that gains in speech rate and especially the increase in mean length of runs were responsible for the good results in students' fluency after the period abroad. In a later study, Towell (2002) investigated why some learners attained higher scores on temporal variables than others and why those with a lower initial level increased the most and never caught up with higher-level starters. He suggests that modification of pausing behavior is highly responsible for increases in fluency between groups of learners. In the Towell & Dewaele's (2005) study, they tried to uncover the role of psycholinguistic factors on the development of fluency by looking into automatization, working memory, neurobiological factors and implicit versus explicit learning. Against their initial predictions they found that working memory was not a predictor of increased fluency (as opposed to O'Brien, Segalowitz, Freed & Collentine, 2007), although when compared to results in the L1, they suggest that some individual differences on capacity limitations may have applied for the lower level group. A clear finding was the increase in speaking rate for all groups after the SA period, with lower learners increasing the most, in consonance with Freed's (1995) findings.

Segalowitz & Freed (2004) examined both fluency and oral proficiency in a study that contrasted two learning conditions, SA in Spain and formal instruction (FI) at home (AH) in an English speaking country. Apart from looking at temporal and hesitation measures of fluency and ratings in the

OPI test, they also examined the relation between these measures and L2 specific cognitive factors (speed and efficiency of lexical access and attention control). Results reveal differential greater gains in fluency for SA learners compared to those at home. However, when checked against amount of out-of-class contact, no significant differences were found. The researchers attribute this finding to two possible causes: the amount of contact being too little, or the nature of interactions being too repetitive and banal. Regarding the cognitive abilities, although they demonstrated a significant correlation with oral fluency, no differential effects were attributable to context. A final remark was made by the researchers on the link between initial oral performance levels and learner's predisposition to communicate outside class. So, both the context and the individuals' predisposition, together with the cognitive factors seem to be crucial to the understanding of the complex process that it is involved in acquiring a second language.

In the same issue on SA in *Studies in Second Language Acqusition* (SSLA), Freed, Segalowitz & Dewey (2004) presented a study examining the acquisition of fluency in three different learning contexts, SA, AH and IM (immersion programs). Significant gains in fluency were reported for both SA and IM programs, but not for students who remained AH. Interestingly, more gains were obtained by the IM group than the SA group. This finding agrees with reported amount of time using the L2 (French), much higher for the IM group than the others. Contrary to Segalowitz and Freed's (2004) results on the relationship between context and time-on-task, in this study, the researchers report that "hours per week spent writing French significantly predicted speaking rate gains." (p. 295), suggesting a benefit from output production.

Within the multi-dimensional SALA project, which provided the context for this thesis, several studies have looked at the development of fluency

over a 3-month SA from different perspectives. Juan-Garau & Pérez-Vidal (2007) examined overall oral proficiency, assessed as fluency, accuracy, formulaic speech and complexity in a role-play task. They reported substantial gains in oral fluency for the period spent abroad compared to no gains during the formal instruction period at home. The researchers depict learner's speech after the period abroad as: "speak[ing] at a faster rate and with greater confidence, putting together more error-free clauses and utilizing more appropriate linguistic formulas while attempting to express more complex ideas." (p. 15). More within a discourse analysis perspective, Trenchs (2009) analyzed the development of the native-like oral behavior of 19 learners, comparing them to NS productions. A detailed analysis of learners' dysfluencies (number of filled and unfilled pauses, self-repetitions, and lexical and non-lexical pause fillers) was performed. The researcher reports learners' behavior developed towards NSs norms, especially after the SA, when they tend to rely less on unfilled pauses and self-repetitions, and more on lexical pause fillers.

Still, two more studies within the SALA project have focused on fluency development of bilingual university EFL learners. In a research project leading to the obtention of the DEA (Diploma d'Estudis Avançats; equivalent to a master thesis), the author examined the development of fluency for 30 students during a 15-month period comprising two learning contexts, SA and AH (Valls-Ferrer, 2008). In this study a temporal perspective to language fluency was adopted. First, a fine-grained analysis of 4 fluency measures and 4 hesitation phenomena was performed, followed by the identification of a group of 'high fluency speakers'. Second, a detailed analysis of this group's performance and behavior was undertaken, and a final analysis established the correlations between fluency gains and individual and contextual variables. Findings on the temporal dimension of fluency showed a significant development of learners' fluency. Different developmental patterns emerged, the most

common being the slight decrease in fluency during the FI period, followed by a significant increase in fluency during the period abroad. The measures that turned out to be more critical in the development of fluency were speech rate and mean length of run, as already suggested by other researchers. However, in this study, internal pause duration was also found to critically contribute to the fluency composite. For the group of high fluency speakers, speech rate and articulation rate were found to reach a plateau. Nonetheless, learners in this group went on improving in mean length of run, suggesting an ongoing proceduralization still present. In relation to contact variables, three of them discriminated between types of speakers: accommodation, contact with the media, and confidence.

The last study dealing with fluency within the SALA project, Mora & Valls-Ferrer (*submitted*) examined the oral productions of 30 students by presenting a detailed analysis of learners' fluency (8 measures), accuracy and complexity (4 measures each) in a semi-structured interview performed in pairs. Results show a significant increase in fluency for all variables measured during SA, but no gains in accuracy and complexity. Gains in fluency did not occur at the expense of the other two domains. The researchers suggest that the nature of the task may have influenced the findings obtained.

To sum up, fluency has gained attention among SLA researchers over the last two decades, providing a better understanding of how it develops, especially in certain learning contexts such as the SA. There is agreement on the crucial role of practice to enhance fluency, and for some researchers to promote proceduralization and automatization. Most generally, several temporal measures (speech rate, mean length of runs, and pause frequency and duration) are reported as good predictors of fluency by most researchers, but many also recognise the limitations of their studies due to small number of participants, and the necessity to

replicate early studies with the help of new technologies. A further line of research is directed to the study of fluency from a more qualitative perspective, and the use of other fine-grained measures attempting to describe the language of NNS. The next section deals with measures in fluency studies, from an utterance fluency perspective (including temporal and phonological phenomena) to a perceived fluency perspective.

3.2. Measures in Fluency Studies

The use of appropriate fluency measures has been an issue of discussion among researchers. Kormos (2006) establishes four different approaches existing in the investigation of L2 learner's speech: 1) temporal aspects of speech production (e.g. Lennon, 1990), 2) the combination of the temporal variables with the study of interactive features such as turn-taking mechanisms (e.g. Riggenbach, 1991), 3) phonological aspects of fluency (e.g. Hieke, 1984), and 4) the analysis of formulaic speech (Wood, 2006). To this categorization, a fifth approach could be added; the use of listeners' judgments (usually NS or near-natives) to assess the level of fluency of NNSs.

There are several studies devoted to find out which measures are more appropriate to represent the construct of fluency, especially in the last decade (Raupach, 1980; Cucchiarini et al., 2002; Kormos & Dénes, 2004;. Derwing et al., 2004; Hilton, 2009, among others). Up to this moment, findings seem to rely mainly on temporal measures and hesitation phenomena as the best objective descriptors of fluency. However, there are several researchers who have already claimed the need for a more qualitative analysis of the language produced by both high-fluency and low-fluency speakers (Riggengach, 1991; Towell et al., 1996; Munro & Derwing, 2001; Wood, 2006). Nevertheless, in spite of the relatively increased attention lately given to this issue, a common understanding

seems not to have been reached yet.

Interesting findings have been reported in studies looking at formulaic language (Towell et al., 1996; Ejzenberg, 2000; Wood, 2006), and interactive features (Lennon, 1990; Riggenbach, 1991; among others) in relation to fluency. Nevertheless, here the main emphasis is put on perceived fluency (section 3.2.1.), and utterance fluency (3.2.2.) including both temporal measures (section 3.2.2.a) and phonological measures (section 3.2.2.b), to which the following sections are devoted.

3.2.1. Perceived Fluency

Within SLA research, studies on perceived fluency are not many and they have mainly focused on which are the specific aspects of utterance fluency that listeners attend to when rating L2 speech samples. That is, which utterance fluency aspects contribute to perceived fluency (Lennon, 1990; Riggenbach, 1991; Freed, 1995; Cucchiarini et al., 2000, 2002; Freed, So & Lazar, 2003; Derwing et al., 2004; Kormos & Dénes, 2004). In order to do so, researchers have examined the relationship between utterance fluency and perceived fluency scores. Apart from this most common focus, researchers have also undertaken studies which tried to contribute to the understanding of perceived fluency from different perspectives; the influence of L1 and experience in listeners' ratings (Cucchiarini et al., 2000; Derwing & Munro, 2001; Rossiter 2009), the influence of task type (Ejzemberg, 2000; Derwing et al., 2004) and nature of the data as read vs. extemporaneous speech (Cucchiarini et al., 2002), the influence of context of learning (Freed, 1995; Lafford, 2004; Segalowitz & Freed, 2004) and changes in fluency over time (Derwing et al., 2006, 2008; Rossiter, 2009).

In most of the aforementioned studies, ratings were provided by either 'expert' judges such as trained examiners, phoneticians, speech therapists

(Cucchiarini et al., 2000, 2002; Wennerstrom, 2000), language teachers (Lennon, 1990; Riggenbach, 1991; Cucchiarini et al., 2002), or raters trained by the researchers (Ejzenber, 1992). To our knowledge few studies have been undertaken using non-expert listeners (Derwing et al., 2004; Rossiter, 2009).

Apart from Cucchiarini et al. (2002) and Kormos & Dénes (2004), all the studies above have been conducted using read speech. Moreover, the number of participants in these studies (both the learners producing the speech samples and the listeners providing the ratings) tend to be limited (cf. Rossiter, 2009).

A compilation of the existing studies undertaken within SLA which have related utterance fluency to perceived fluency are presented in Table 3.1. This table contains information about the number of participants, their L1, L2 and language proficiency level, the number and L1 of listeners, the learning context, type of task used to retrieve speech, the fluency scale used to provide the ratings and the main findings. Some of the studies are referred to in greater detail in the following paragraphs.

Table 3.1. Studies on perceived fluency in SLA.

Study	Partic-ipants	Profi-ciency	L1	L2	Lis-tener	Context	Task	FL Scale	Findings
Lennon (1990)	4 EFL learners	Adv.	Ger.	Eng.	9 NSs	SA 6-month	narr.	no report.	- Perceived gains after the SA (no statist. analysis provided) - Utt. FL: Improvement found in speech rate, filled pauses/T-unit and %T-Units followed by pause.
Riggenbach (1991)	6 EFL learners	Adv. Low-inter.	Chin.	Eng.	12 NSs	no reported	dialg.	7 point	- Speech rate and unfilled pauses contributed to fluency judgements to differentiate btw. low and high FL groups.
Freed (1995)	30 univ. students (15 abroad – 15 AH)	mix	Eng.	Fr.	6 NSs	AH-SA	OPI	7 point	- No gains perceived after the SA semester. - Utterance FL: Students abroad speak faster and with fewer clusters of disfluencies.
Ejzenberg (2000)	46 EFL learners	Low Inter. High	Port.	Eng.	4 NSs traine raters	FI in the home country	dialg. – monolg	6 point	- Monologue obtained lower scores in perceived FL than dialogue.
Cucchiarini et al. (2000)	60 NNs 20 NSs	Begin. Inter. Adv.	Mix	Dut.	9 NSs expert	FI in TL country	read	10 point	- Automatically calculated speech rate as the best predictor of expert ratings. - Utt. FL: NSs obtained sig. better results than NNSs.

Wennerstrom (2000)	10 NNS	Inter. Adv.	Mix	Eng.	2 NSs exam.	Intensiv. program	dialog.	3 point	- Fluent speakers better than non-fluent in signalling relationships among words and phrases and segment speech turns.
Cucchiarini et al. (2002)	60 NNSs 57 NNSs	Begin. Inter. (Adv. 1 group)	Mix	Dut.	9 NSs expert 10NSs teach.	FI in TL country	read - spont.	10 point	- Predictive power of utt. FL measures higher for read than spontaneous speech. - MLoRs best predictor of FL ratings for the spontaneous speech.
Derwing & Rossiter (2003)	48 NNSs	Interm.	Mix	Eng.	6 NSs teach.	FI in TL country	narr.	9 point	- Improvement in perceived FL for a group given prosody instruction. - Impressions on prosody: improved over time.
Kormos & Dénes (2004)	16 EFL learners	Adv. Low-inter.	Hung	Eng.	3 NSs 3NNs teach.	FI in the home country	narr.	5 point	- No difference btw. ratings from NSs and NNSs - SR, MLoR and pace as best predictors of perceived FL
Derwing et al. (2004)	20 Mandarin immigr.	Low level	Mand arin	Eng.	28NSs	FI in TL country	monolg dialg. narr.	9 point	- Narrative judged to be less fluent than monologue and dialogue. - Utt. FL: pauses/sec. and pruned syll./sec. best predictors of perceived FL.
Derwing et al. (2006)	20 Slavic 20 Mandarin immigr.	Low level	Slavic & Mand arin	Eng.	40NSs	FI in TL country	narr.	7 point	- Perceived improvement found for the Slavic speakers only. - Slavic speakers significantly more contact with Eng. than Mandarin speakers.

Derwing et al. (2008)	16 Slavic 16 Mandarin immigr.	Low level	Slavic & Mandarin	Eng.	33NSs	FI in TL country	narr.	7 point	- Perceived improvement in FL and comprehensibility after 2 years found for the Slavic speakers only.
Rossiter (2009)	24 EFL learners	inter.	mix	Eng.	6 NSs exprt. 15Nss non-exprt. 15NNS	FI in TL country	narr.	9 point	- No sig. effect for time (10 months btw. T1-T2). - Novice NS higher FL ratings than NS experts, and these higher than NNSs.

Derwing & Munro (2001) and Munro and Derwing (2001) provide evidence that slowing down the speech of L2 speakers does not always lead to better ratings from judges. They found that for advanced learners, unmodified-rate productions were preferred to slowed down ones, and that for very slow productions, speeded up versions were preferred. But learners usually preferred to listen to speech that was a little bit slower than naturally occurring NS speech, at least in the case of Mandarin-accented English.

Derwing et al. (2004) investigated the influence of task type on the fluency ratings of untrained NS judges (*N*=28). A brief definition of what they were expected to attend to when rating the samples for fluency was provided. They used a 9-point scale. They found that task type influenced judges' ratings, specifially speech samples from a picture description task were judged to be significantly poorer in fluency than speech elicited through a monologue and a conversation, which were found not to differ in fluency. Moreover, they found that pausing and pruned syllables accounted for 69% of the variance in the fluency ratings for the picture description task and 65% for the monologue. Pauses per second did not make any significant contribution to explaining variability.

Kormos & Dénes (2004) analysed the speech of 16 Hungarian speakers of English performing a short narrative task for temporal fluency phenomena, as well as NS and NNS ratings. Temporal measures were correlated with perceived fluency measures and they found that speech rate (SR), mean length of run (MLoR), phonation time ratio, and number stressed words (pace) correlated strongly with perceived fluency. Among the previous, SR, MLoR and pace were found to be the best predictors of perceived fluency. They accounted for between 60 and 80 % of the variance in fluency scores. No differences were found for fluency ratings between native and non-native listeners.

Derwing, Munro, Thomson & Rossiter (2009) investigated the relationship between L1 fluency and L2 fluency development by analyzing speech from the participants in their previous studies (Derwing et al., 2004, 2006). In this case they added the analysis of L1 productions for the two groups of speakers (Slavic and Mandarin). The findings revealed a complex relationship between L1 and L2 fluency; while they seem to be related at early stages of language development, this relationship seems to vanish at later stages. No conclusive remarks are made regarding this issue. Regarding the similarity in temporal measures for L1 and L2 found only at T2, the researchers attribute it to an initial reliance on the L1 which vanishes with time. They also touch upon the issue of transfer from the L1 to the L2 when discussing the findings on vowel reduction. It seems that better results for the Slavic group could be attributed to the high temporal similarity between this language and English vowel length, contrary to what happens with Mandarin. Regarding the relationship between temporal measures and L2 fluency ratings, they corroborate previous findings on the high correlations usually found. They even emphasize that the best predictor of L1 fluency, pruned SR, seems to be one of the best predictors of L2 fluency ratings, suggesting that raters can indeed focus on aspects of fluency independently of other dimensions such as foreign accent or grammatical accuracy.

Derwing, Munro & Thomson (2008) studied the development of fluency and comprehensibility over time and its relation to willingness to communicate for two groups of immigrants in Canada(Mandarin and Slavic speakers). Speech samples from different data collection times were rated for fluency and comprehensibility by a group of 33 English NS listeners through a 7-point Likert scale. Some information was provided regarding what they meant by fluency. An improvement on fluency was found for the Slavic speaker group but not the Mandarin speaker group after an 8-month period living in the TL country. This difference in

fluency improvement between the two groups was attributed to differences in willingness to communicate. A more qualitative analysis of the data showed that speakers in the Slavic speaker group interacted and listened to the radio significantly more often than the Mandarin speaker group, hence, highly benefiting from the input from and interaction with the TL.

Rossiter (2009) conducted a study in which the raters did not perceive an improvement in fluency over the 10 weeks of the study. The author offers several tentative explanations for no improvement having taken place, one of them being that 10 weeks is too short a period of time. The other two explanations are either that participants had already attained a threshold in speaking fluency, or that they did not get enough exposure outside classroom to make a difference in perceived fluency. Rossiter also points to the fact that participants did not receive enough instruction on fluency in class as one of the reasons why they might have not improved. She also found that ratings of novice NSs were higher than ratings of NNSs. However, ratings from expert NS did not differ from those of the other two groups (novice NSs and NNSs).

Cucchiarini et al. examined the relationship between utterance fluency in read (Cucchiarini et al., 2000) and spontaneous speech (Cucchiarini et al., 2002) using an automatic measurement system and assessed perceived fluency using expert raters. They found that pause frequency was more relevant than pause duration for perceived fluency in read speech. For spontaneous speech, and the intermediate level group, the variable which explained the greatest amount of variance in perceived fluency was MLoR (R^2= 0.42). They then added pause frequency, but the extra variance explained by this measure is quite marginal. AR was not included in the model since the correlation the authors find between perceived fluency and this measure were relatively low. When discussing the importance of

MLoR to perceived fluency ratings the authors argued that "the importance of this variable seems to suggest that pauses are tolerated, provided that sufficiently long uninterrupted stretches of speech are produced." (p. 2871).

3.2.2. Utterance Fluency

After reviewing the studies conducted on perceived fluency, this section is devoted to the utterance fluency measures which have been used by researchers to assess L2 fluency. First, a subsection on temporal phenomena is presented in relation to the studies where they have been used (section 3.2.2.a). Second, research on phonological phenomena related to L2 fluency is dealt with in section 3.2.2.b.

3.2.2.a Temporal phenomena

Fluency as a temporal phenomena is closely related to the processing of speech. Temporal measures are regarded as highly objective and have been largely used to assess fluency gains of second language learners in the field of second language acquisition. Findings from these studies contribute to the understanding of how fluency can be measured in a valid and reliable way.

Raupach, in one of his early studies (1980), reported on temporal measures in speech production and concluded that they "allow us to determine different degrees of fluency" (pp. 269-270). Likewise, Towell (2002) claims that temporal variables provide "objective measurements of the output of the productions which must lie behind language processing" (p.119) and, in second language acquisition, they show "what developments in fluency have taken place" (pp.119-120).

Many researchers have tried to establish which variables are the best

predictors of fluency (Raupach, 1980; Lennon, 1990; Freed, 1995; Segalowitz & Freed, 2004; among others). Speech rate (SR), defined as the number of syllables spoken per minute, has been regarded as one of the most reliable measures to track fluency gains (Freed, 1995; Lennon, 1990; Riggenbach, 1991; Towell et al., 1996; Kormos & Denés, 2004). However, the same researchers have acknowledged that SR alone cannot account for the development of a learner's L2 fluency. Several other variables have been claimed to stand in a central position for the study of fluency. Towell et al. (1996) was the first study to demonstrate the relationship of mean length of runs (MLoR) with speech production in general, and fluency in particular. MLoR, number of syllables produced in utterances between pauses[8], is regarded as the optimal measure to detect proceduralitzation of knowledge. In this same study, the researchers also found that SR and phonation time ratio (PhonRat), percentage of time spent speaking as a proportion of total time of speech, were good predictors of fluency.

In their 2004 study, Kormos and Dénes investigated which linguistic and temporal variables predicted the perceptions of fluency by both NS and NNS. They used a wide range of variables: SR, AR, PhonRat, MLoR (pauses of 0.25 sec.), silent pauses per minute, mean length of pauses, filled pauses per minute, dysfluencies per minute, pace and space). They attest SR, MLoR and PhonRat, together with pace, number of stressed words per minute, as the best predictor of fluency scores. The first three variables had already been reported as good predictors of fluency by other researchers (Lennon, 1990; Riggenbach, 1991; Freed, 1995; Towell et al., 1996). However, the novelty resides in the last variable, called pace. This

[8] Different researchers have used different cut-off points to account for what counts as a pause. Early studies counted pauses of 0.1, 0.2, 0.25 and 0.30 milliseconds. Lately, in SLA, researchers tend to count pauses of 0.4 ms and above as dysfluent pauses.

variable was first introduced by Vanderplank (1993), but for measuring difficulty in listening, not production. Kormos & Dénes (2004) were the first to use this variable in L2 oral production research and found it to be a good predictor of fluency, together with the three other variables just mentioned. We should bear in mind, however, that this variable is not a pure temporal measure, since it relies mainly on intonational phenomena, or stress, and could thus be considered a phonological measure. (our study has also considered the importance of such a factor: this is dealt with further in the next chapter).

Pausing behavior and hesitation phenomena research has produced somehow mixed results. Most studies have not been able to demonstrate a significant relation between hesitation phenomena, such as repetitions and dysfluencies, and fluency scores (Kormos, 2004). However, several studies have pointed at the difference between fluent and non-fluent learners falling on the placement of pauses. As already mentioned fluent speakers tend to pause at clause boundaries, whereas non-fluent speakers pause within clauses as well (Lennon, 1990; Towell et al., 1996; Mora & Valls-Ferrer, *submitted).* In Hilton's (2009) study mentioned in section 3.1.1., she claims the importance of looking at hesitation phenomena when analyzing L2 speech, since they are more representative of encoding difficulties highly present in L2 productions. By using the PAROLE (PARallèle, Oral en Langue Etrangère 'parallel oral foreign language') corpus, a fine-grained analysis of hesitation phenomena was performed for English as an L2. Of special interest are her findings comparing NS and learner pausing behavior, supporting previous research on the placement of pauses. Apart from hesitation frequency and placement, Hilton also presents results on the duration of hesitations, which differ greatly between NS and NNS. "NS productions contain proportionally more hesitations lasting from 200-600 ms, whereas our learners (overall) produced more hesitations lasting from 900 ms to over 3 seconds." (p. 654).

In a recent study, Fullana & MacKay (2010) examined the fluency of Catalan/Spanish non-native speakers with different degrees of experience within a formal instruction context. They found that apart from pausing phenomena, the use of segment lengthening contributed to low speech rates. Participants tended to overgeneralize the use of lengthening (both for consonants and vowels) in word-final position.

3.2.2.b Phonological phenomena

Although the study of L2 fluency has been very prolific in the last two decades, the area of phonological fluency is still particularly under-researched. Its exploration would not only be of high interest for the understanding of how L2 phonological fluency develops, but it would also contribute to the ongoing debate of which variables can best describe L2 fluency. Hieke (1985) was already advocating the importance of connected speed processes to define fluency. He writes: "fluent speech is the cumulative result of dozens of different kinds of processes. These can be classified according to the severity with which they cause alternations." (p. 140). And, as already mentioned in the previous section, certain phonological phenomena, such as sentence stress in 'pace', seem to contribute to the temporal aspect of fluency as well.

Phonological fluency refers to the various kinds of phonological processes occurring while fluent speech is produced, sometimes equated to 'connected speech'. These processes include: assimilation, linking, elision, contraction, stress, rhythm, and intonation, among others. Hieke (1984) attributes the motivation of these processes to 'ease-of-effort' principles. "It is the generation of maximally fluid temporal, and minimally complex articulatory speech which furnishes the rationale for such phenomena, and the whole variety of adjustments observed appears orchestrated to meet speech dynamic necessities and constraints." (p. 345). The main interest

here resides in finding out whether these phonological processes play a role in facilitating fluency in speech production, and hence, be considered as a marker of fluent speech. In his study, Hieke (1984) finds evidence for consonant attraction acting as a facilitator of fluency. He measured both the actual realizations and the potential ones in NS and NNS speech. Findings show a consistent use of consonant attraction by NS of approximately 80% of the possible cases, while for NNSs the percentage is reduced to little above 50%. These results are interpreted as evidence for this measure being a reliable indicator of NNS' fluent speech production, and suggest it may be used as a valid measure in fluency assessment. Several years later, Hieke (1989) takes up the issue of the study of phonological processes, but from a relatively different standpoint, by contrasting what he calls 'pre-dynamic citation form strings, with 'dynamic running speech'. He claims that "[the transition from one form to the other] is not possible without a host of absorption processes that alter segmental sequences through assimilation, reduction, loss and similar leveling features" (p. 197).

In a study designed to explore the adequacy of measures of sentence stress placement to determine difficulty in speaking and listening, Vanderplank (1993) elaborates on the notions of 'pacing' (speed or tempo at which stressed words are spoken) and 'spacing' (proportion of stressed words to total number of words on average). The participants proved to have serious difficulties in imitating the tempo of a NS's speech, being unable to fit the exact number of words in the same length of time. The problem was mainly due to the lack of reduced forms in their speech and the non-mastery of stress placement. Vanderplank (1993) concludes that "using the notions of pacing and spacing, it should be possible to grade passages of spoken English more accurately than simply using syllable-per-minute speech rates." (p. 123) Even though his research was exploratory in nature, Vanderplank's findings are very suggestive. In her (2004) study, Kormos

and Dénes provided further evidence for the adequacy of 'pacing' to measure L2 speech.

Wennerstrom (2000) conducted a study on the influence of intonation in the perception of fluency. The speech of 10 NNS was analyzed in terms of intonation patterns, and then, rated for fluency by two NS of English. A detailed description of the intonation patterns produced by a group of most fluent speakers, and one of less fluent speakers is provided. Most fluent speakers differ from less fluent ones basically in the assignment of different pitch levels according to the role of the words in the information structure of the discourse. Less fluent speakers tend to produce all words with relatively equal pitch. Wennerstrom finishes her chapter by suggesting the use of intonation (especially pitch on lexical items and at boundaries) in fluency assessments.

The small number of studies presented above is representative of the lack of research in this area of phonological fluency. However, the findings are very suggestive and promising for an under-researched area of fluency with great relevance to the field. More detailed analysis, with more participants, and for different processes should be undertaken in order to advance the understanding of fluency, and speech production in general.

In an attempt to shed new light on the area of phonological fluency, which has been and still is surprisingly largely under-researched, we examine the acquisition of L2 rhythm and its relationship with fluency. This is an area which will be scrutinized in our study, in a further attempt to explore and expand the notion of fluency in the measurement of progress during SA. The next chapter deals specifically with this phonological process, its learning and its potential role in L2 fluency.

4

Rhythm in Speech Production

As referred to in the previous chapter, opposite to what happens with fluency, research on other L2 suprasegmentals is still scarce in SLA. In the last decades, however, several researchers have claimed that prosody and suprasegmentals are important for L2 intelligibility (Hahn, 2004; Field, 2005; Anderson-Hsieh et al., 1992; Derwing, Munro & Wiebe, 1998) comprehensibility and foreign accent (FA) (White & Mattys, 2007b; Kang, 2010a, 2010b; Kang, Rubin & Pickering, 2010; Derwing & Munro, 1997; Derwing et al., 1998; and Munro & Derwing, 1999, 2001). Most of these studies sought to identify which prosodic factors contribute to listener's perceptions. For instance, Munro and Derwing (2001) report on two experiments they conducted to gain insight into the role of speech rate in listener's judgements of FA. They found that the rate of speech in the L2 influenced listeners' ratings of comprehensibility and FA, and that the optimal speed was somehow higher than the normal L2 speaking rate of their Mandarin participants.

In a study whose main objective was to look at effects of experience on the production of suprasegmentals (fluency based and melody based), Trofimovich & Baker (2006) confirmed that, independently of experience, the suprasegmentals they tested (stress timing, peak alignment, speech rate, pause frequency, and pause duration) contributed to FA. Nevertheless, fluency based suprasegmentals were found to contribute more to the perception of FA than the melody based ones, at least when ratings on low-pass-filtered speech were used. Apart from identifying the specific prosodic contributors to listeners' judgements, Kang (2010a, 2010b) sought to quantify the relative weights of temporal and prosodic factors in relation to FA judgements. In these studies she analysed the speech of 8 and 11

international teacher assistants (ITAs), acoustically and perceptually. In the 2010a study she found that overall pitch range, stress measures and mean length of pauses exerted significant effects on the ratings of the NSs, explaining approximately 42% of the variance. Similarly, for the 2010b study, accent ratings were better predicted by pitch range (24% of variance) and stress measures, while comprehensibility scores where better predicted by speaking rates.

The term suprasegmentals, introduced in the aforementioned studies, encompasses a large number of different features, such as rate, pitch, stress, rhythm, intonation, etc. Chun (2002:3) defines suprasegmentals as "phonological units that stand in contrast to so-called segmental features or simple sounds", and goes on to specify that "a suprasegmental feature typically extends over more than one sound segment in an utterance, over longer stretches of speech." This definition provides an adequate frame for gauging rhythm as a suprasegmental; it works at utterance level, and it is determined by the regularity of syllable prominence in sentences. Yet, very few studies have included rhythm in their research agenda, at least not as the central phenomenon under investigation. For example, Derwing, Munro and Wiebe (1998) report on an experiment where rhythm was part (together with intonation and stress) of instruction based on interactive discourse and which proved to be effective to improve comprehensibility, FA and fluency. However, no direct inference was made of which of the instructed factors (rhythm, intonation or stress) was responsible for the gains.

The following sections in this chapter provide a detailed account of what is meant by rhythm and how research in this area has evolved, especially in the last years. We start by defining the central concept of rhythm, and summarizing the assumptions made by research on L1 rhythm, with a special section on vowel reduction, as one of the core features of English rhythm and one on the most commonly used metrics. Then, a thorough review of research on L2 rhythm is provided.

4.1. L1 Rhythm

A plausible reason for not having many studies in SLA devoted to the study of rhythm might be the lack of consensus among researchers on what the term rhythm itself conveys and/or which rhythmic metrics can better discriminate between languages, or varieties within languages. Encouragingly, in the last years, research on rhythmic metrics has been more productive and findings seem to provide more consistent evidence for the use of certain metrics compared to others.

The term rhythm is widely used in relation to both music and spoken language. Even though the two domains are closely related, and both imply the regular or quasi-regular patterns of either sounds or words, spoken language rhythm conveys specific nuances, such as frequency variability or rhythm typology among languages. The first entry for rhythm in the Merriam Webster Dictionary (online, 2010) is "an ordered recurrent alternation of strong and weak elements in the flow of sound and silence in speech". Within the field of linguistics, Dauer (1987) characterizes rhythm as a "total effect involving phonetic and phonological as well as segmental and prosodic phenomena". This concept of rhythm has been the most widely followed by researchers in the last twenty years with minor changes. Barry (2007) proposes an understanding of rhythm based on the "conceptual *grouping* of a number of structural correlates, possibly already established at other levels of description. Ideally, these structural properties should have identifiable phonetic exponents,"(p.104). However, during this period, the key issue in speech rhythm research has been rhythmic variability as a mean to understand which acoustic correlates better define the perceptually based rhythm typologies.

The categorical distinction between languages as being either syllable-timed or stress-timed, together with the isochrony hypothesis (Pike, 1945; Abercrombie, 1967), which assumed the equal duration of speech intervals, have long been abandoned. The lack of evidence found for this isochrony in early studies led to a

review of the issue. Dauer (1983, 1987) suggested that languages could be placed along a rhythm continuum depending on some of their phonological features; languages such as Spanish and English would be placed at opposite ends of this continuum. Languages such as Catalan and Polish, or even dialects within a language (e.g. Singaporean English), have been reported to fall in intermediate positions on this continuum (Nespor 1990; Low, Grabe & Nolan, 2000; Grabe & Low, 2002). Nevertheless, the results on these intermediate languages are far from conclusive, since several other studies claim that Catalan should be classified as syllable-timed, mainly due to its phonological properties (Ramus, Nespor, & Mehler, 1999; Prieto, Vanrell, Astruc, Payne, & Post, *in press*).[9]

Dauer (1983) observed three phonological and phonetic properties which differentiate rhythm classes: a) syllable structure, b) vowel reduction and, c) word stress. Languages with a more stress-timed rhythm have more complex phonotactic rules, allowing onsets and codas with three and four consonant segments. On the other hand, syllable-timed languages permit a very limited array of combinations within syllables, with open syllables of the CV type prevailing. English, for instance, allows syllables with a CCC– structure in the syllable onset and a –CCCC in the coda (e.g. spray, worlds) (Chela-Flores, 2006b; Toledo, 2010). Opposite to this, Spanish phonotactics limit both onset and codas to the CC– and –CC structure (e.g. trans-con-ti-nen-tal) with the simple structure CV representing 55,81% of the total in Spanish, and CC– accounting only for 3,14% and –CC for 0,13% (Guerra, 1983, cited in Alfano, 2009). Catalan, likewise, allows a maximum CC– in onsets, but –CCC in codas (e.g. bruscs) (Recasens, 1993; Prieto, 2004). Despite its greater syllable complexity and proportion of closed syllables than Spanish, Catalan predominant vowel structure remains CV (Prieto, Vanrell, Astruc, Payne & Post, 2010). For English, the CV structure respresents only a 25,33% of the possible combinations (Gut & Milde, 2002 cited in Toledo, 2010). Hence, the

[9] Most participants in this study are Catalan/Spanish bilinguals learning English as an L3, hence the special reference to these languages.

amply different phonotactic properties of the languages under study provides an interesting case for the study of rhythmic variability.

Apart from these differences in syllable structure, stress-timed languages make use of vowel reduction in order to comply with temporal requirements of the language, in which stressed syllables occur at fairly regular intervals. The variation in length between vowels in stressed or non-stressed syllables is much greater for stress-timed than for syllable-timed languages. The latter seldom have vowel reduction, or if they do, as for example Catalan, it is a reduction in vocalic quality, not in length (Gavaldà-Ferré, 2007; Ortega-Llebaria & Prieto, 2011) (see next section on phonetic vowel reduction).

4.1.1. Phonetic Vowel reduction

As we have seen above, vowel reduction is one of the features which most influence the categorization of languages along the rhythmicity continuum. The term is defined by *A Dictionary of phonetics and phonology* as: "Any phonological process in connected speech which makes a vowel shorter, less loud, lower in pitch or more central in quality, or which neutralizes some vowel contrasts in unstressed syllables." (Trask, 1996:384). However, in the more specialized literature, a distinction is often made between phonological and phonetic vowel reduction. According to Fourakis (1991), the phonological process is independent of tempo, and dependent only on assigned degree of stress. On the other hand, phonetic vowel reduction is affected by several factors such as consonantal context, de-stressing, and rate of speech. Gómez-Lacabex (2009) describes vowel reduction as "a universal phenomenon through which vowels may lose some of their durational, acoustic or articulatory properties" (p. 41). For the purpose of the present study, and with the intention to relate it to L2 fluency, the phonetic meaning is used. In this sense, it can be argued that vowel reduction is a phonetic process occurring due to the effects of segmental and suprasegmental contexts. Gimson (1970, cited in Farnetani & Recasens, 2010) explains the occurrence of vowel reduction due to

"the pressures of its sound environment or of the accentual or rhythmic group of which it forms part," (pag. 287) and the speed of the utterance. In the phonetic-informational content of speech signals, Harris (2003) described vowel reduction as the suppression of phonetic information in the speech signal. Lindblom (1963) refered to vowel reduction as a correlate of stress in languages with heavy stress such as English and claimed that "vowel reduction is associated with stress, rate of utterance and contextual influence." (p. 1774). His findings reflect that a reduction in duration seems to be determinant for vowels to undergo contextual assimilation, hence a change in quality, and ascribes the results to physical limitations in the articulators.

As mentioned above, vowel reduction implies various changes in the phonetic quality of vowels (duration, pitch, articulation). Harrington (2010) distinguishes two kinds of vowel reduction: centralization and coarticulation. The first is defined as paradigmatic (vowel shrinks towards the center) and the latter as syntagmatic (shifts in vowels attributed to the preceding and following contexts). Here we focus on the paradigmatic form in which vowels centralize in unstressed positions. Vowel features in such positions become more obscure, with the neutralization of acoustic distinctions and the tendency towards the realization of schwa-like quality vowels.

Apart from the changes in vowel quality, and in relation to those, vowel reduction often entails a decrease in duration, especially in stress-timed languages. In these types of languages, vowels in unstressed positions tend to be shortened, in contrast to the full quality and length of vowels in stressed positions. In a review of the theoretical accounts of coarticulation, Farnetani & Recasens (2010) refer to Lidbloms's studies as main representatives of coarticulation as speech economy emphasizing that "reduction is the automatic response of the motor system to an increase of the rate of the motor commands" (p. 330). Lee, Guion & Harada (2006) report on a ratio of unstressed to stressed vowels to compute duration-

based vowel reduction for which they found that English NSs' unstressed vowels were produced in almost half the duration of the stressed vowels (ratio= .45).

Several studies have shown an association between vowel reduction and rate of speech (eg. Fourakis, 1991; Turner et al., 1995; Weismer et al., 2000). Miller (1981) affirms that "most of the change in total utterance duration that occurs with rate actually takes place in the vocalic segment. (pag. 49)". Although speech rate has been demonstrated to affect vowel duration, not all vowels in all position experience the same degree of reduction. A detailed account of the vowel reduction processes occurring in relation to their position in the syllable and/or phrase falls outside the scope of this study (but see Farnetani & Recasens (2010) and Harrington (2010) for an account on the issue). Tsao, Weismer & Iqbal (2006) conducted a study on the strength and frequency of relationships between vowel duration and formant frequencies (or vowel space). They found that vowel space was not influenced by speaking rate, therefore, suggesting that "the vowel space – at least its extreme limits– is a rather stable phonetic characteristic of English" (p. 1079). Interestingly, they additionally found that vowel duration changed significantly as a function of speech rate, with the fast-speaking group producing shorter vowel duration than the slow-speaking group.

These studies on speech rate and vowel reduction have been mainly undertaken with English data. Tsao et al. (2006) question whether the same results would be found for other languages with "relatively sparse vowel spaces" since their articulatory demands would be lower than those of English. Equally important is the fact that English falls within the stress-timed end of the rhythmic continuum, where variation between stressed and unstressed syllables is very high and noticeably affects vowel reduction. In this case, vowels in unstressed positions are shorter than vowels in stressed syllables in order to comply with temporal demands. On the other hand, languages at the more syllable-timed end of the continuum tend to present much lower variation (or almost null –in the case of

languages such as Spanish) between stressed and unstressed syllables, resulting in practically inexistent vowel reduction, at least as a distinctive feature.

Spanish is representative of a typical syllable-timed language, with almost inexistent vowel reduction and a non-complex syllable structure. By comparison, Catalan has been often referred to as a language taking a middle position on the rhythmic continuum, since the syllable structure is comparable to that of Spanish, but it has vowel reduction (Nespor, 1990). Nevertheless, as already mentioned, it has also been argued that Catalan vowel reduction is more of a qualitative than a durational nature (Ramus, Dupoux, & Mehler, 2003; Gavaldà-Ferré, 2007; Prieto et al. 2010; Ortega-Llebaria & Prieto, 2011) and stress only affects vowel duration marginally compared to other factors (Aguilar, Giménez, Machuca, Marín & Riera, 1997). Furthermore, Cantín & Ríos (1991 citing Bertinetto, 1984) list the influence of tempo on syllabic duration as a factor explaining different rhythm types. They contend that in stress-timed languages faster speech is accomplished at the expense of unstressed syllables, which are much shorter than stressed syllables. Conversely, syllable-timed languages tend to administer a proportional reduction to both syllable types.

In a study especially devoted to the analysis of Catalan vowel reduction and rhythm, Gavaldà-Ferré (2007) found that results on the vocalic rhythmic metrics could lead to quite different conclusions depending on the metrics used. Whereas %V seemed to place Catalan within the range of the syllable-timed languages, nPVI fell somewhere in between the usual range of stress-timed and syllable-timed languages.

Few studies have been undertaken on the second language acquisition of English phonetic vowel reduction (Flege & Bohn, 1989; Lee, Guion & Harada, 2006; Gómez-Lacabex, 2009). The L1s of speakers in the three studies (Spanish for the first and third and, Korean and Japanese for the second) are reported by the researchers not to have vowel reduction. The two studies with L1-Spanish speakers

showed L2 learners' failure to appropiately reduce English unstressed vowels, both in quality and duration, when no instruction was given. When instruction on vowel reduction was provided, Gómez-Lacabex (2009) found that participants showed (limited) improvement on vowel reduction perception and production. In Lee et al. (2006), Japanese speakers produced durational differences between stressed and non-stressed vowels more accurately than the Korean speakers, and non-significantly different from English NSs. The authors attribute this finding to the characteristics of the Japanese language, which uses duration to signal phonological differences such as moraic length.

In this section we have acknowledged the importance of vowel reduction in speech, especially for rhythmically differentiating types of languages. Additionally, we have also learnt that English L2 learners whose L1 does not reduce vowels (as is the case with Catalan and Spanish), at least in a duration-like manner, often fail to appropriately accommodate to English NS standards. Since vowel reduction is an essential part of English rhythm, we assume that this process should be noticeable in the metrics used to assess rhythm. The next section is devoted to review these metrics.

4.1.2. Rhythm Metrics

In the last two decades a number of rhythm metrics which adopt objective criteria to place languages along the rhythm continuum have been proposed (Ramus et al., 1999; Grabe & Low, 2002; Dellwo, 2006). These metrics are based on the duration of vocalic and intervocalic intervals as means of capturing the rhythmic properties of a language. Ramus et al., 1999 introduced a set of measures which they found discriminated well between rhythm types when tested on eight different languages. These measures are all based on durations of vocalic and consonantal intervals: percentage over which speech is vocalic (%V), and standard deviation of consonantal (ΔC) and vocalic (ΔV) intervals. Grabe, Post, Nolan, & Farrar (2000) and Grabe & Low (2002) introduced two new measures based on the durations of

vowels and intervals between vowels; the raw and normalised Pairwise Variability Indices (nPVI and rPVI). Dellwo & Wagner (2003) analysed the effect of speech rate on Ramus et al. (1999) rhythm measures. They found that ΔC was especially affected by the speed of delivery, so Dellwo (2006) introduced a modification to Ramus ΔC by calculating a variation coefficient (varcoΔC), which enabled the monitoring of speech rate differences across languages. The same was done for ΔV by Barry et al. (2003), for which they introduced the varcoV as the normalized version.

In a more recent study, Dellwo et al. (2007) and Fourcin & Dellwo (2009) introduce the idea of using voiced and voiceless intervals instead of the commonly used vocalic and consonantal ones. They adapt some of the former measures (%V, ΔC, nPVI-V and rPVI-C) creating a %VO (percentage over which speech is voiced), ΔUV (standard deviation of unvoiced intervals), and nPVI-VO / rPVI-UV (the average difference between consecutive voiced and unvoiced intervals, respectively). The appropriateness of these two new measures to classify languages in different rhythmic classes is proved for the data used in the study. The authors claim that their use is of advantage for researchers due to its easiness and reliability of method. They go on to argue that voicing contrasts possibly offer a plausible perceptual cue in the distinction of rhythm classes by infants, but further research in this direction is needed to confirm its validity.

An interesting novel approach to the study of rhythmic classes which challenges the commonly accepted properties of rhythm (syllable structure and vowel reduction) was proposed by Prieto et al. (2010) by introducing phrasal timing phenomena (durational marking of prosodic heads and prosodic edges). Findings were reported on the limited role of syllable structure in predicting rhythmic behaviour, suggesting that phonotactic differences (especially those measured by varcoV and nPVI-V) could not fully explain the distinction of rhythmic classes and that other prosody factors could further explain it. The factors examined (durational marking of prosodic heads and prosodic edges) provided evidence that

cues to prosodic properties could partly add to the distinction of rhythmic classes. However, further research is called for to support these findings and to determine how language-specific durational variability could be related to rhythm.

In the present study, a set of widely applied measures in the field was first chosen for the categorization of rhythmic productions at the three data collection times. This choice was based on the robustness of the measures demonstrated by previous research and for comparison purposes with these previous studies. Subsequently, due to the longitudinal design of the study, and the semispontaneous nature of the data, a subset of 3 measures was selected for the analysis of rhythm as a fluency component and its development during a SA period. The selection criteria were mainly based on the influence of rate on rhythm reasures, since it was previously acknowledged that speech rate in our participants' productions was significantly different before (T2) and after (T3) the SA period, as confirmed by some of the temporal fluency measures. Hence, measures demonstrated not to be affected by rate, %V, varcoC, nPVI were chosen. This will be further developed in the methods and results chapters.

4.2. L2 Rhythm

Based on the previous studies on L1 rhythm, in the recent past, several researchers have started to look at L2 rhythm. These studies have mainly focused on the evaluation of rhythm metrics, the influence of L1 on L2 rhythm, or its importance for L2 oral instruction. The latter has been mainly undertaken within research on prosody instruction. Two books have been devoted in the last decade to bridge the gap between theoretically-oriented research on prosody and second language teaching (Chun, 2002; Trouvain & Gut, 2007). The emphasis of both books is on the acquisition of intonation in the L2, however, they also cover other suprasegmentals such as stress and rhythm. The conclusion is straightforward for the authors; even though they provide some examples and recommendations on how to teach suprasegmentals, further research is called for to better understand

suprasegmentals language-specifically and cross-linguistically, as well as a much closer cooperation among researchers and teachers. In an article which exemplified this connection between research and teaching, Hahn (2004) examined the impact NNSs misplaced primary stress had on the inteligibility of NSs' listeners and provided some strategies for its teaching in EFL classrooms, such as including perception exercises or using longer pieces of discourse. In a study which focused on the fundamental features that differentiate English and Spanish rhythm and intonation, Chela-Flores (2004) sought to provide a thorough description of these features in order to optimize its teaching in EFL programs. She introduced a model which "forces students to focus on the phonological patterns without the normal segments and sequences with which they co-occur in language" (p.14) in which durational differences between syllables was established as a priority for the teaching of rhythm to Spanish EFL learners. The literature on suprasegmentals' instruction seems to provide evidence for the benefits of instruction, although limited, and to emphasize the difficulty of teaching certain suprasegmentals such as rhythm (Barry, 2007), and the need for large amounts of practice, especially in extemporaneous speech, to automatize pronunciation patterns (Chela-Flores, 2004). Further and more focalised research on rhythm is called for by these authors.

As previously mentioned, another central aspect for L2 researchers has been the evaluation of rhythm metrics, mostly undertaken by L. White and his colleagues (Wiget et al., 2010; White & Mattys, 2007a & 2007b; etc.) In these studies, the authors have tested the metrics used in L1 rhythm research with the intention of finding out how appropriate each of them are in predicting the position of a language along the rhythmic continuum. White & Mattys (2007a) reported %V, VarcoV and nPVI-V as the metrics that best discriminated between rhythm classes, and further claimed that %V (which is not subject to variation with speech rate) provided support for the gradient distinctions in rhythmicity (as seen with the English-Dutch results). Based on these findings they suggested the use of %V with either VarcoV or nPVI-V to overcome limitations of using a single metric.

White & Mattys (2007b) went a bit further and correlated the acoustic measures with perception of foreign accent. They used foreign accent as the perceptual variable since rhythm *per se* is very difficult to rate perceptually, despite having been attested to be part of the prosodic elements which influence foreign accent perception. A group of 12 native English speakers rated the utterances produced by Standard Southern British English (SSBE), Dutch and Spanish speakers. A significant main effect of accent group on ratings was found, with Dutch English being closer to SSBE than Spanish English. When correlated with the acoustic measures, %V, VarcoV and nPVI-V yielded strong significant correlations. As expected, speech rate was inversely correlated with FA ratings as well. VarcoV was found to be the best single predictor of the foreign accent ratings obtained. When coupled with speech rate, these two measures accounted for a greater proportion of FA ratings, but no other metrics did. The authors suggest that these findings support the idea of a close relationship between gradient rhythmic distinctions and prosodic timing processes.

In a study which investigated both the rhythm metrics of an L2 and the influence of L1 on L2 rhythm, Tortel and Hirst (2010) examined the rhythmic parameters in the production of French learners of English. They found a set of rhythmic parameters (ΔC, cvC (comparable to varcoC)) which distinguished NSs from NNSs, as well as different groups of NNSs. With these same parameters, they also observed a 3-level gradation: from French non-experienced at one end, to French experienced in the middle, to British English at the other end. However, they acknowledged that the results may only apply to French speakers of English and not to other L1 speakers, so further research on other L1s is encouraged, as well as the application of these measures to spontaneous speech.

Other studies have focused on the sole influence of L1 rhythm on L2 productions (Gut, 2003; Carter, 2005; Lin & Wang, 2005; White & Mattys, 2007b; Grenon & White, 2008). The main finding being that L2 productions usually stand somewhere between L1 and L2 productions. Studies on Spanish EFL learners

provide an accurate picture of this finding. Gutierrez-Diez (2001) offered a contrastive approach to the analysis of rhythm. Using judges to determine stress placement, he found that the Spanish EFL learners' mean syllable duration ratio fell between the native Spanish and English ratios. Later studies have used acoustic measures to capture L2 learners' position in the rhythmic scale. Carter (2005), within a variationist perspective, analysed PVI values in the performance of four native Spanish speakers of English, to observe that PVI values were in an intermediate position between the typical low Spanish and high English values. Similar results were found by White & Mattys (2007b) reporting on VarcoV scores. Spanish speakers of English seemed to accommodate towards the shorter unstressed vowels of the L2, and produced longer stressed vowels than in their L1. However, the distinction between stressed and unstressed vowels never reached native English speakers' values.

Taking into account the findings the previous literature on rhythm has produced and the areas in which further research has been called for, we intend to contribute to the understanding of how L2 rhythm develops over time and under a specific learning condition, SA. To our knowledge, no studies to date have looked at the acquisition of L2 rhythm in a SA context, nor at the development of L2 rhythm at two different points in time. From a more methodological point of view, two of the rhythmic metrics are considered as measures which can contribute to better explain oral fluency. In addition, this study introduces the use of extemporaneous speech, since rhythm has been mainly studied using read-language in laboratory conditions.

PART II. THE STUDY

5

Objectives and Research Questions

In this chapter, a recapitulation of Part I is provided (section 5.1.), followed by the objectives (section 5.2.) and research questions which have motivated this study (section 5.3.).

5.1. Recapitulation

In the first chapter a thorough description of the learning contexts investigated in SLA was provided. These learning contexts were characterized by a series of factors which are central to the study of SLA: input, interaction, output and practice. SA has been described as a rich context in terms of opportunities to obtain meaningful input, to become involved in varied type of interactions and to enhance the quantity and quality of practice. Previous research has depicted SA as a context which especially benefits oral-aural skills, most especifically fluency. However, the study of fluency has been mainly undertaken from an utterance fluency perspective, leaving perceived fluency largely under-researched.

A further issue which has been shown to have received little attention until the last years is that of an initial threshold level. Several authors have now claimed its role in learners' linguistic development during SA periods (Golonka, 2006; Collentine, 2009), but more empirical research is needed. Our study aims to provide a better understanding of whether and how the initial fluency level within a general advanced level group can still have an impact on the SA outcomes in terms of fluency and rhythm.

Attention is drawn to specific factors in L2 speech production which are central for the development of fluency, automaticity, output, practice and memory. All of them are core features in the larger construct of fluency in L2 speech production to which this thesis is devoted, together with the phenomenon of rhythm presented in the last chapter of the theoretical overview.

Few studies have been conducted to examine the relationship between perceived fluency and utterance fluency in the speech of advanced learners (cf. Cucchiarini et al., 2002). Additionally, only two early studies have undertaken this type of research within a SA context (Lennon, 1990 and Freed, 1995) and these were limited in terms of number of participants and use of statistical analyses, in addition to the difference in participants' proficiency level being lower than intermediate.

The few studies which have examined the influence of listener type on fluency ratings have reported different findings, with Kormos & Dénes (2004) observing no differences between ratings from NSs and those of NNSs, and Rossiter (2009) finding that novice NSs gave higher fluency ratings than NS experts, and at the same time, these gave higher ratings than NNSs. The 69 listeners in the present study, differing in native language and experience with Catalan/Spanish accented speech, will provide further evidence about the impact that listener type may have on fluency ratings.

There seems to be a direct relationship between many suprasegmental phenomena and speed of delivery. In order to accommodate to the temporal demands of connected speech, certain changes need to occur in speech both at segmental and suprasegmental level.

As previously mentioned, in stress-timed languages, rhythm is highly dependent on several processes which need to take place in order to

comply with language-specific rhythmic patterns. The core concept of English rhythm lies in the assigment of stress to the right words in a sentence. By allocating the main stress to one syllable of the 'focus' word, all other syllables in the utterance become de-emphasized to different degrees. This phenomenon triggers most reduction processes, especially when conditioned by time constraints in, for instance, fast speech, with vowel reduction as one of the most significant features, together with assimilation, elision, blending, etc.

In the last decade, research on rhythm has focused on the duration of vocalic and consonantal intervals to assign languages a place in the rhythm continnum depending on whether they are more stress-timed or syllable-timed. Vocalic-based rhythm metrics have provided robust results, especially in L2 rhythm research, probably due to their heavy dependance on vowel reduction processes. Spanish, commonly used as the typical example of syllable-timed language, does not have vowel reduction in unstressed positions, so the total duration of vowels in L2 English utterances would be expected to be longer, in general, than those produced by native English counterparts.

As the participants in this study are bilingual Catalan/Spanish, and Spanish has been clearly classified as a syllable-timed language and Catalan, depending on the metrics used, as either a syllable timed or intermediate-language (Ramus et al., 2003; Gavaldà-Ferré, 2007; Prieto et al., 2010; Ortega-Llebaria & Prieto, 2011), it seems plausible to assume that participants' L1 rhythm is totally different from English rhythm.

The present study aims to shed new light into the development of both fluency and rhythm during a 3-month study abroad period, since as we have seen in the background of the study, research on perceived fluency and rhythm is very scarce in SA learning contexts. More especifically, the

present research will try to contribute to the current body of research on fluency in SA settings by providing a more comprehensive assessment of the development of fluency during SA periods by focusing on two aspects of fluency: utterance fluency and perceived fluency.

The study examines a group of 30 EFL learners during a period of 15 months comprising two learning contexts, FI and SA. The SA period lasted 3 months and students undertook it as a requirement for their undergraduate Translation studies in Barcelona. They all followed the regular curriculum which included 6 months of FI at home and a 3-month SA period.

5.2. Objectives

The present study examines the development of L2 fluency and rhythm in the extemporaneous speech of advanced EFL learners during a study abroad period, and the relationship between utterance fluency and rhythm and listeners' judgments. We also intend to assess the impact of listeners' background on perceived fluency ratings, and of *initial fluency level* (pre-SA fluency level) on SA outcomes in both fluency and rhythm.

Taking into account that participants' oral fluency improved significantly during the period abroad, as captured by gains in utterance fluency measures (Valls-Ferrer, 2008), we further explore the fluency construct by: a) providing non-trained listeners' ratings to examine which of the utterance fluency measures better predict perceived fluency scores, b) analyzing how rhythm impinges on such development with a set of well established measures, and c) examining the effect of *initial fluency level* on both fluency and rhythm. The research questions are presented in the following section.

5.3. Research Questions

Due to its broad scope, the three research questions are subdivided into two subquestions each.

Research Question 1 (RQ1).

Will there be an effect of time and learning context on the fluency and rhythm of L2 speakers?

> *RQ.1.1.* To what extent will listeners perceive gains in participants' oral fluency after the study abroad period? That is, when asked to rate L2 speech samples, will listeners assign more target-like scores to T3 than to T2 speech samples?
>
> > *RQ.1.1.a.* When grouped by L1 and experience, will any differences in fluency ratings between the two listener groups arise?
>
> *RQ.1.2.* Will there be any change in rhythm before and after the period spent in each learning context as measured by rhythm metrics? If so, to what degree will changes represent significant gains in each of the learning contexts?

Research Question 2 (RQ2).

To what extent are hypothesized changes in perceived fluency related to the temporal characteristics of non-native speakers' utterances? That is, to what extent will utterance fluency and rhythm be related to perceived fluency ratings?

> *RQ.2.1.* Will utterance fluency scores be related to perceived fluency ratings and result as good predictors of perceived fluency?

RQ.2.2. How and to what extent will rhythm be related to perceived fluency ratings and contribute to explaining variance in perceived fluency scores?

Research Question 3 (RQ3).

To what extent will *initial fluency level*, indexed as a temporal composite fluency measure, be a factor affecting SA outcomes on fluency and rhythm?

RQ.3.1. How and to what extent will *initial fluency level* be related to fluency and rhythm after the SA? What's more, will it have any effect on the amount of gains?

RQ.3.2. After the period abroad, will there be differences in fluency and rhythm between a high fluency and a low fluency group assembled by *initial fluency level*? Moreover, will the two groups obtain different gains during the period abroad?

6

Methods

In this chapter, the empirical study is presented. In order to answer the research questions, two closely related tasks were conducted, one production task (semi-guided interview task) and one perception task (rating task). The production task was used to examine utterance fluency and rhythm and the perception task perceived fluency[10]. This chapter starts with a general contextualization of the empirical study (section 6.1.), followed by a section on the comprehensive design of the study (section 6.2). Afterwards, the two tasks are introduced: 1) production task (including utterance fluency and rhythm) (section 6.3.) and 2) perception task (perceived fluency) (section 6.4.). At the end of the chapter, an overall analysis section is included to describe how the different tasks have been analyzed in relation to one another (section 6.5).

6.1. Contextualization of the study

The context within which the study is developed responds to a complex sociolinguistic situation, common to many tertiary educational settings around the world, especially Europe, where multilingualism is becoming the norm.

The study takes place at the Pompeu Fabra University (UPF), in Barcelona. Barcelona is a bilingual city, where both Catalan and Spanish

[10] The analysis of utterance fluency was conducted as part of the author's DEA [Diploma D'Estudins Avançats, equivalent to a master thesis] (Valls-Ferrer, 2008).

are spoken on a daily basis. Most people living in Barcelona are bilingual speakers –to different degrees. Linguistically, we find ourselves in a context where bilingualism is the norm, confirming what some researchers (Crystal, 1987; Sharwood-Smith, 1994) announced years ago; monolingual speakers are becoming fewer in number, and bilingualism (or even multilingualism) is becoming the norm.

Catalan being a minority language[11], both government and Universities have the duty to promote the knowledge and usage of the language among all members of the community. In Catalonia, Universities have long contributed to the process of language normalization instigated by the Catalan Government in the 80s. University lectures and seminars are mainly conducted in Catalan. However, at present, with the new European Higher Education Area and the EU promotion of students' mobility, some Universities are adopting new language policies, fostering a new multilingual classroom reality.

In 2005, the Council of the European Union adopted a policy on multilingualism with three main aims: "to encourage language learning and promoting linguistic diversity in society, to promote a healthy multilingual economy, and to give citizens access to European Union legislation, procedures and information in their own languages." (p.3). The Commission's long-term objective is to "increase individual multilingualism until every citizen has practical skills in at least two languages in addition to his or her mother tongue" (see Pérez-Vidal, 2002 for a more comprehensive view).

In 2007, the Pompeu Fabra University launched a Plan of Action for

[11] Even though more than 9 million speakers speak Catalan, we consider it a *minority language* in Catalonia since only 36% of the population use it regularly, compared to the 46% which uses Spanish on a daily basis. Source: Institut d'Estadística de Catalunya (Idescat, census 2008).

Multilingualism, which introduces English as a working language for all studies offered, together with the already used Catalan and Spanish. (see www.upf.edu/llengues/). The official bilingualism characteristic of the Catalan society and its universities is giving way to an imminent trilingualism in the university context.

This trilingualism is boosted by the ERASMUS program, funded by the EU with the intention of helping promote the learning of European languages within the university context (see chapter 1). Coleman (1998) refers to the ERASMUS as having "made European cooperation between universities into the norm rather than the exception." (p. 170). The experience helps to create cultural and linguistic links among countries and individuals. The main goal of students participating in this mobility program is very often to become more proficient in the foreign language. Hence, the study of second language acquisition at university level, and within the study abroad context, becomes highly relevant in this sociolinguistic context.

The profile of Spanish students participating in the ERASMUS program depicts university students aged 18 to 30 (with the highest proportion between 20 and 25), with women representing almost two thirds of the students going abroad, and doing it in the second half of their studies. The duration of the stay ranges from 3 to 12 months, with an average of 7 months[12].

The European actions towards multilingualism, the universities promotion of foreign languages, and the current number of students spending time abroad define the reality of the context where the study takes place. SA programs offer the opportunity to benefit from "informal out-of-class

[12] Source: "Movilidad de Estudiantes" Agencia Nacional Erasmus (year 2005/2006)

exposure to the TL [target language]" (Huebner, 1998;2) and it represents a real turn towards personal growth (Murphy-Lejeune, 2002:30; Pellegrino-Aveni, 2005).

6.2. Design

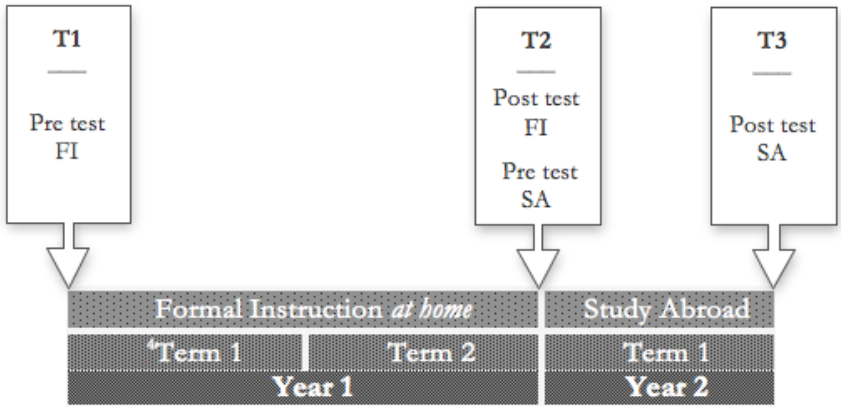
This study is part of a larger state-funded research project, *Stay Abroad and Language Acquisition* (SALA[13]), which intends to uncover the effects of a SA period, following a FI period, on the development of target-like proficiency in English, both in oral and written production and receptive skills. A battery of oral and written tests was used to measure gains in linguistic competence over time. In addition, a set of questionnaires gathered information about participants' linguistic profile, attitude and motivation, and contact with the target language during their SA (see Pérez-Vidal, 2011 for further details).

The SALA project follows a pre-test post-test design. The data collection took place at four different times over a 15-month period. First, data was collected upon students' entrance at university (T1), and after a 6-month period of formal instruction (FI) in the home university (T2). Subsequently, the third data collection (T3) was conducted upon students' return from a 3-month SA in an English speaking country, and one year after their return (T4). For the present study, only the first three data collection times are used (see figure 6.1.). As shown in figure 6.1., the data collected at T2 functions both as pre- and post-test; T2 is the post-test of T1 (defining the end of FI) and at the same time the pre-test of T3 (defining the beginning of the SA).

---

[13] Stay Abroad and Language Acquisition. Research project at the Universitat Pompeu Fabra funded by the Spanish Ministry of Education [HUM2004-05442-C02-01, 2004-2007; HUM2007-66053-C02-02, 2007-2010; FFI2010-21483-C02-01, 2010-2013].

Figure 6.1. Design with data collection times for the present study (adapted from the SALA project design).



The design of the project responds to a longitudinal type of research, described by Menard (cited in Dörnyei, 2007) as "research in which a) data are collected at two or more distinct time periods; b) subjects analyzed are the same from one period to the next; and c) the analysis involves some comparison of data between periods." (Dörnyei, 2007). This type of design also corresponds with a within subjects design, very common in psychology, and highly efficient to control for participants variability, but sometimes at the expense of internal validity. As Højen (2004) describes it:

> "Often in longitudinal studies in the humanities and social and psychological sciences, any obtained difference between different times of testing may be assumed to be caused by the effect of the independent variable (Riney & Flege, 1998; Snijders, 2001). In other words, scores would not be expected to differ across repeated times of testing, if no "treatment" had been given. This kind of scenario warrants for a direct comparison of scores for each subject group across testing times, i.e., a within-subjects design." (p. 84)

Data obtained through the current design qualify for a longitudinal analysis of participants' linguistic development, as well as the analysis of acquisition in different learning contexts, FI at home and SA. However, the threat to internal validity comes when comparing the same group of participants in the two different contexts where "carry over effects" can occur from the

FI period to the SA.

In this respect, Milton & Meara (1995) already advanced such a methodological problem that SA research would have to face in relation to the studies' design. Ideally, a SA study should include two groups of students which "should have roughly the same degree of exposure to the L2 over a similar time span." (p. 18). The same authors acknowledged that it is almost impossible for this situation to occur. In this same study, Milton & Meara used an unconventional design in the field consisting of the use of the same participants as their own matched pairs, using a within subjects design for one of the three sections of their study. A decade later, in a short review of methodological approaches used in SA research, Rees & Klapper (2008) concluded that: "given the vast number of variables at play in this type of research, it is unlikely that any study will be able to secure genuine "control" groups, but researchers should at least attempt to make the best comparisons possible." (p.98).

It is with the intention to use the most appropriate methodology to provide accurate comparisons that the Milton & Meara (1995) design was followed in the production task. Hence, participants were tested under two different treatment conditions, FI and SA. Subsequently, gains obtained in the FI, that is, between T1 and T2, were then compared to gains obtained during the SA, that is, between T2 and T3. Participants received no specific practice on pronunciation or oral skills during the first period (FI). Previous studies on the role of instruction in phonetics and phonology had reported no gains in the absence of instruction in FI settings (Aliaga-Garcia & Mora, 2009, Chela-Flores, 2001). Therefore, we decided to check this assumption. However, the main focus still was the development of fluency and rhythm during the SA, since the nature of the learning context (with increased opportunities for input, output, and all kinds of practice) seemed ideal to boost the development of a suprasegmental such as

rhythm, even when no previous instruction had been provided. Nevertheless, the instruction received during this period, conducted entirely in English, could have had an effect on participants' linguistic knowledge. This will have to be taken into account when reporting gains at T3 (after the SA) and certain statistical analyses will need to be used to neutralize this effect. Notwithstanding, the design used is thought to be the best to control for participants' homogeneity, very often ignored by SA researchers and crucial in this type of studies due to important potential group differences in motivation, attitude and proficiency levels between participants willing to go abroad and those who are not.

For the perception task, exclusively data from T2 and T3 were used. The rationale for using these two times only, corresponding to the SA period, responds to two basic observations: first, the non-significant differences in the preliminary study on utterance fluency between T1 and T2 led us to assume that a similar pattern was bound to emerge for perceived fluency. So, it was deemed not necessary to ask listeners to assess differences on perceived fluency on the basis of speech samples for which an analysis in terms of utterance fluency had found no differences on any of the measures. After having conducted the study on utterance fluency and observing significant gains during SA only, our main interest focused on checking whether these acoustically observed changes could also be subjectively perceived by listeners.

The methodological approach taken here allows us to comply with the aforementioned objectives of the study; observing gains in linguistic competence over time, focusing on the SA context. Two different tasks were undertaken (production and perception tasks) to provide an answer to the research questions. The production task was centered on the participants' oral production and enabled the assessment of utterance fluency and rhythm. The speech samples collected in the production task

served as the instrument for the perception task, which examined perceived fluency assessment by different groups of listeners.

6.3. Production Task: Semi-guided Interview Task

For each task section, participants, instruments and procedures, and analyses are presented. For the production task, the analysis is divided into two subsections, one on utterance fluency and the other on rhythm. This task was intended to gather data for the analyses of fluency and rhythm in the participants' speech production.

6.3.1. Participants

Participants (*n=30*) were selected from a larger pool of students studying Translation and Interpreting at the UPF in Barcelona. All subjects were bilingual (90% Catalan/Spanish and 10% Basque/Spanish), with ages between 17 and 25 (*mean*=18,36) and an advanced level of proficiency in English[14].

Participants were tested under two different treatment conditions, FI at home and SA. For the sake of clarity, we have assigned each treatment a group name, and they will be referred to as:

FI period. Participants attended formal English lessons for a period of 6 months at the home University.

SA period. Participants went on a SA to an English speaking country for a period of 3 months.

[14] Participants general level was established according to two factors: participants score on a comprehensive university entrance exam, which includes an English language test, and an internal SALA test on grammar and writing skills in English.

Participants had studied English through formal instruction within the Catalan educational system; during primary education students receive 420 hours of instruction in English, and 630h during secondary education (2,5/3 hours weekly)[15]. As a complement to the official education, it is quite common that students attend extra-school private English lessons. Many of the participants in this study (86,6%) reported having attended private language lessons for approximately three hours a week at some time during their secondary education. Previous experience abroad, in an English speaking country, was reported to have been no longer than one month (10% of the participants had been abroad for 1 month, 80% had visited an English speaking country for tourism (between 4 to 10 days), and 10% had never set foot in any English speaking country).

Before entering University, participants had successfully completed primary and secondary education and taken a university entrance examination (Prova d'Accés a la Universitat: PAU). A cut-off mark is required to enroll in the Translation and Interpreting degree (a minimum of 6,73 out of 10; *mean* = 7,54 in 2005[16], when participants in the present study entered University), as well as passing a test measuring students' linguistic competence in their mother tongue and target language (English).

Translation and Interpreting undergraduate courses at the UPF last 4 academic years, and each year is divided into three terms. During the first two terms of the first year, participants attended English lectures for a total of 80 hours. These lectures were all taught through English and they covered lexical, syntactic and morphological aspects of the language, as well as practicing some reading and writing skills. No explicit training was

[15] Source: Generalitat de Catalunya. Departament d'Educació
http://phobos.xtec.cat/edubib/intranet/index.php?module=Pages&func=display&pageid=2
[16] Source: Universitat Pompeu Fabra webpage:
http://www.upf.edu/universitat/upf_xifres/estudis/tra.html [20/01/2011]

provided in listening or any kind of oral production during the course.

Participants were required to spend the first term (mid-September to mid-December) of the second academic year abroad, in an English speaking country (most of them went to the UK or Ireland (88%), and the rest to the USA, Australia or Canada (12%)). This was accomplished through the ERASMUS scheme. During this three-month period, students attended lectures and seminars at the host university. No official requirements regarding the number of lectures students had to attend while abroad were specified, so the number varied depending on the host university regulations. However, all students took a minimum of 2 courses while abroad (about 23 hours of English language instruction and 59 hours of content subjects taught through English). The host university occasionally provided accommodation arrangements, but this was not always the case. Participants lived in university dorms (76%), shared apartments (16%) or with host families (8%).

After the SA term, students went back to the UPF and followed the regular curriculum for the Translation and Interpreting Studies, which did not include any other subject on the English language for the remaining two and a half years.

Baseline data was collected from a group of 10 highly comparable participants, both in age (*mean*=20) and education. They were all native speakers of English (NSs) and, at the time of the data collection, they were participating in an international exchange program at the Universitat de les Illes Balears in Mallorca (Spain). The use of native speakers here provided a comparable set of data to contrast L1 and L2 productions for the same task. "SLA research can justifiably use native speakers' language as one perspective on the language of L2 learners, provided it does not make native speakers' language the measure of final achievement in the L2."

(Cook, 1999:190). These NSs were asked to perform the same task as the experimental group so the type of language they produced would be highly comparable. In this sense, by having NS baseline data, "any discerned differences between native and nonnative performance could more confidently be ascribed to the additional processing demands of a second and imperfectly known language." (Foster & Tavakoli, 2009:867).

6.3.2. Instruments and Procedures

Participants' L2 productions were elicited by means of an oral interview performed in pairs (student A and student B). This task was deemed appropriate to elicit spontaneous speech that would more closely reflect real-life interactions, compared to other tasks used in previous research such as oral narratives. However, since the priority was the comparison between NNSs' productions at different data collection times, as well as NSs', semi-spontaneous speech was elicited instead. This was accomplished by providing participants with the questions they had to ask, which proved very effective since participants used extemporaneous speech, but the type of language was controlled, since the topic was common to all speakers, including the native English ones.

Furthermore, having participants interact with each other, and not the researcher, favored much balanced interactions and a more relaxed atmosphere. This last condition allowed the creation of a scenario relatively close to an informal interaction, so the participants' productions closely resembled language used in real-life conversations. A researcher was present as an observer and to solve any technical problem which might have arisen, but never intervened in the conversation.

A battery of seven questions was given to each participant (see appendix A1.1.). The questions were all different but shared the same topic,

university life. The main rationale for choosing the topic was familiarity of content. As previous research by Foster and colleagues has shown (Foster & Skehan, 1996; Foster & Tavakoli, 2009; Skehan & Foster, 1997; Tavakoli & Foster, 2008), cognitive complexity, operationalized as familiar and unfamiliar task content can have an impact on task performance, especially on fluency and accuracy. In a study specially addressing differences in fluency ratings across task types, Derwing et al. (2004) found that perception of L2 speakers' fluency varied across tasks, speech samples from a monologue and a conversation received significantly higher ratings than those from a narrative. Since our main interest was on eliciting semi-spontaneous speech as close to real-life speech as possible, having a low-level cognitive complexity task, that is, a topic familiar to participants, was a priority. In addition, the familiarity of the topic ensured that participants would not run out of ideas, so productions were kept to a fair length. This same instrument held good for the native speakers' group as well.

Data was collected according to the following procedure. First, student A formulated one question at a time out of seven to student B, who answered after each question. Once finished, they changed roles so student B formulated the questions and student A responded. Time allowance for this task was five minutes, and no participant went overtime. The recordings took place inside an interpreting booth in the presence of a researcher, who gave the instructions. The same proceedings were followed at the three data collection times. For T1 and T2, speech samples were recorded on tape and then converted into 16 bit computer audio files, for T3 minidisks were used instead but the same digitalization procedure was followed.

At a later stage participants were asked to collaborate a last time with the project. On that occasion they were asked to perform the interview task in

their native language. They could choose between Catalan and Spanish depending on which of the two they felt was their strongest language. 10 participants performed the interview task in their L1 (7 in Spanish, 3 in Catalan). In the present study, the Spanish recordings are used to provide the L1 baseline speech materials for appropriateness of rhythm metrics in distinguishing between stress-timed and syllable-timed productions.

6.3.3 Analyses

Participants' oral productions were transcribed and coded for fluency of speech and rhythm. Speech samples were segmented using a digital audio editor (Goldwave Audio Editor, version 5.19) and only responses (interviewees' productions) were considered for analysis[17]. Different analyses were undertaken for each of the phenomena under study. The levels of analysis are defined according to EAGLES (1996) categories based on a compilation of events and labels commonly used in the transcription of spoken language. For fluency production, the analysis was at a syllabic, suprasegmental level, labeling and coding syllable boundaries, speed and pausing phenomena, and semi-lexical (e.g. oops, er) and non-lexical (e.g. laugh, cough) vocalized events. For rhythm, the analysis was at the segmental, syllabic and suprasegmental level, adding the segmentation and labeling of vocalic and consonantal elements to the previous analysis.

6.3.3.a. Temporal Fluency Measures and Analysis.

Speech samples were orthographically transcribed[18] in Word documents for each participant and at the three different data collection times using

[17] Apart from the interviewer asking the question and interviewee answering it, for this task any further spontaneous interaction (i.e. interruptions) between interviewer and interviewee was almost inexistent. Hence, the use of the interviewers' productions was disregarded.

[18] CLAN conventions were used and adapted for the purpose of this study.

Goldwave Audio Editor. By means of this program dysfluent pauses occurring between and at clause boundaries were detected and measured in milliseconds. The cut-off point for pauses was established at 0.4 sec. The 0.4 sec. criterion was selected as the defining criterion of a dysfluent silent pause, following Riggenbach, who suggested that "pauses shorter than 400 ms are within the range of normal or fluent speech and do not reflect dysfluency" (Riggenbach, 1991:426).

Other studies using the same or similar temporal measures among L2 learners have established the cut-off point for pauses at different lengths. The 0.25 cut-off point has been largely applied, at least in early studies (see Grosjean & Deschamps, 1975; Raupach, 1987). However, Raupach (1980) used the 0.3 sec. criterion to measure silent pauses of German and French learners and Towell et al. (1996) applied the 0.28 sec[19]. criterion to measure the French productions of English speakers who spent a period of time in France. Riggenbach (1991) distinguished between different type of pauses: 'micropauses' of 0.2 sec. or less, 'hesitations' of 0.3 to 0.4 sec. and 'unfilled pauses' of 0.5 sec. or greater. Freed (1995b) established the cut-off point at 0.4 sec. after native judges' fluency evaluations of the performance of two L2 learners' groups, one group that spent a period abroad and the other that remained at home. This 0.4 sec. criterion accounts for dysfluent pauses produced by L2 learners. In the present study the 0.4 sec. criterion was also adopted, since we were interested in measuring dysfluent pauses, that is, pauses which break the fluent run of speech, and that it could also be compared with the Freed's (1995b) study which measured fluency gains in study abroad.

The transcription system used was adapted from the CLAN conventions (see appendix A1.2.). Specific decisions were made on the segmentation and labeling, e.g. inclusion (or not) of repetitions for the syllable count,

[19] The use of .28 sec. was an artifact of the data collection procedure.

status of contractions as one or two words, etc., which are explained below together with the description of the measures.

In this task, the analysis of oral fluency was based on objective temporal measures and hesitation phenomena. Previous research on SLA has extensively used this temporal approach to oral fluency providing a good account of how L2 fluency develops under specific conditions (Lennon, 1990; Raupach, 1987; Towell et al., 1996; Freed et al., 2004). Still, Kormos & Dénes (2004), after a thorough review of the most commonly used temporal measures in SLA, suggest that a cluster of temporal variables could explain a high proportion of the variation in perceived fluency scores, hesitation phenomena accounting for much less variation in the perception of fluency. In contrast, Rossiter (2009) claimed that fluency ratings by three different groups of judges were highly influenced by pausing and hesitations in the speech of NNSs, accounting for approximately half of the reported impressions. This, together with speech rate, accounted for over three quarters of the temporal fluency impressions provided by judges, further supporting the above cited early studies. On the basis of the aforementioned research the following measures were adopted in this study:

As for the temporal fluency measures, Tavakoli & Skehan (2005) proposed a categorization of fluency measures in three sub-dimensions: 1) speed fluency, 2) breakdown fluency (or silence), and 3) repair fluency. In accordance, the measures used in our study, included in each sub-category, are described here.

Speed fluency: this dimension is based on the speed with which language is produced. It includes 4 measures:

Speech rate (SR): total number of syllables produced in a given speech

sample divided by the amount of total time required to produce the speech sample (including pause time) expressed in seconds. This figure is then multiplied by sixty to obtain syllables per minute. For the present study we decided to use a pruned speech rate measure, for which any repetitions, false starts or repairs were excluded from the counts. We did so because several studies had demonstrated that it was a more reliable measure of speed (Derwing et al., 2004; Gilabert, 2005; Ortega, 1999; etc.). As de Jong, Steinel, Florijn, Schoonen & Hulstijn (*in press*) point out, this measure has often been taken as representative of speed fluency, however, since it includes pause time, it is also representative of breakdown fluency.

Articulation Rate (AR): total number of syllables produced in a given speech sample divided by the amount of time taken to produce them (excluding pause time) in seconds. This figure has been calculated following Kormos & Dénes (2004). As in the previous measure, repeated words were not included in the calculation. In recent studies, this measure has been taken as a more pure reflection of speed than speech rate, since it is not affected by the duration of pauses.

Phonation-time Ratio (PhonRat): "percentage of time spent speaking as a percentage proportion of the time taken to produce the speech sample." (Towell et al., 1996:91).

Mean Length of Runs (MLoR): Average number of syllables produced in utterances between pauses of 0.4 sec. and above (refer to the beginning of this section for the argumentation on the selection of 0.4 sec. criterion). Even though Tavakoli & Skehan (2005) considered this measure as representative of speed fluency, here we argue that, similarly to speech rate, this measure could be considered a hybrid between speed and breakdown fluency, since it represents the speed

that can be achieved in an utterance, but conditioned by frequency of pauses.

Breakdown fluency: this dimension is based on the pausing behavior (frequency and duration of silent and filled pauses). It includes 3 measures.

Pause Frequency (PauseFreq): total number of pauses divided by the total amount of time expressed in seconds and multiplied by 60. Only pauses of 0.4 sec. and above have been used for the calculations.

Pause duration ratio (PauseDur): average duration of pauses calculated by dividing the total length of pauses above 0.4 seconds by the total number of pauses of above 0.4 seconds.

Internal Pause duration ratio (Pause i Dur): average duration of internal pauses calculated by dividing the total length of internal pauses above 0.4 seconds by the total number of internal pauses of above 0.4 seconds.

Repair fluency: this dimension is based on the number of repair elements (repetitions, false-starts, and repairs) in the productions. It includes only 1 measure.

Dysfluencies per minute (DysRat): total number of dysfluencies (repetitions, restarts and repairs) divided by the total amount of time expressed in seconds and multiplied by 60.

For the statistical analyses, one-way repeated measures ANOVAs were performed to capture the development of fluency over time. Then, in order to measure the incidence of the two learning contexts on the development of fluency, t-tests were performed comparing the size of gains obtained during the two learning contexts (FI vs. SA). As gains obtained during the SA period may have been influenced by improvement

during the previous FI period, we performed a hierarchical regression analysis, controlling for gains during FI. Subsequently, these results were compared to native speakers' performance on the same task by means of t-tests.

6.3.3.b. Rhythm Analyses and Metrics

Rhythm was analyzed acoustically using wide band spectrograms in Praat (Boersma & Weenink, 2007). First, the speech samples were transcribed using the SAMPA (Speech Assessment Methods Phonetic Alphabet)[20] conventions in a Texgrid Editor. Then, the vocalic and consonantal intervals were delimited and labeled in a second tier. Segmentations were based on the visual characteristic of speech sounds as represented in the spectrograms and the researcher's judgments from the audio files to adjust it to the maximum precision (see appendix A1.3. for examples of segmentation and labeling).

Decisions on how to segment and label speech sounds, particularly speech reductions present in connected speech were informed by previous research (Peterson & Lehiste, 1960; Grabe & Low, 2002; Prieto, 2004; Varden, 2006). Vocalic intervals comprised the stretches between vowel onset (F2) and vowel offset, including adjacent vowels, diphthongs, and vowels brought together by co-articulation processes. Intervals between vowel offset and vowel onset were considered intervocalic (or consonantal) intervals, and could include both single consonants and consonant clusters. Vowel transitions were considered, from the beginning of voicing, as part of the vocalic interval. Following Grabe & Low (2002) glides were labeled as consonants only when changes in formant structure

[20] SAMPA is a computer readable phonetic alphabet designed as a mapping of the IPA (International Phonetic Alphabet) codes. It was developed in several stages by an international group of phoneticians, and applied to 27 different languages. For further information see: www.phon.ucl.ac.uk/home/sampa

and/or amplitude were clearly observable; otherwise they were treated as vocalic intervals. Working with L2 speech complicated the segmentation procedures, and several assumptions had to be made for the coding of certain segments. Even though participants' productions were in English, the quality of vowels was often closer to the Catalan inventory than the English one, as well as the VOT durations.

Scripts written in Praat were used at different stages of segmentation and labeling (i.e. script to automatically correct the manually labeled data, script to combine consecutive consonant and vowel intervals, etc.). Moreover, a final script was used to extract vocalic and consonantal intervals to process all rhythm measurements (%V, nPVIs, etc.) for further statistical analysis[21]. Following Dellwo (2010), each measure was calculated for an interval of speech between naturally occurring pauses (inter-pause-interval), so no pause time was included in the calculations.

As for the rhythm metrics, well-established rhythm metrics proposed by other researchers in previous studies examining L2 rhythm were used here. White & Mattys (2007a, 2007b) suggest varcoV and %V as the measures with the most discriminative power to assess L2 rhythm. In contrast, Tortel and Hirst (2010) find that the combination of ΔC and cvC (comparable to varcoC) are best to predict the rhythmic tendencies of productions by French learners of English. Yet, Carter (2005b) suggests nPVI-V as a good measure to assess the productions of Spanish learners of English, but without testing any other measure. Since findings from previous research are far from conclusive, and none of the aforementioned studies have used extemporaneous speech, it was deemed appropriate to include as many measures as possible to observe which

[21] We are most grateful to Volker Dellwo who kindly provided us with the Praat scripts for the rhythm analysis and who supervised the researcher's work during her 3-month academic research stay at the University College London towards the completion of the European Mention to the Doctorate.

ones could best explain changes in the participants' performance, as well as discriminate among groups (NNS-T2, NNS-T3, NS) and place them along a syllable-timed – stress-timed rhythm continuum. A definition and description of how the measures were calculated is given next:

Interval measures: these are three raw measures introduced by Ramus et al. (1999) and two rate normalized measures by Dellwo (2006) and Barry et al. (2003).

percentage V (%V): the proportion of vocalic intervals within the utterance. It is calculated as the sum of the duration of vocalic intervals (*100) divided by the total duration (CV) of the utterance (Ramus et al., 1999). This measure captures the balance between vocalic and consonantal intervals and has been reported as not being affected by speech rate (Dellwo & Wagner, 2003; White & Mattys, 2007b). Indeed, it has largely been considered a good measure to discriminate between stress-timed and syllable-timed languages, since it detects the reduction (or non-reduction) of vowel duration for unstressed vowels, one of the key properties that differentiate languages along the rhythm continuum. Hence, stress-timed languages have significantly lower %V values than syllable-timed languages (Ramus et al., 1999; White & Mattys, 2007b).

deltaC (ΔC): the standard deviation of the duration of consonantal intervals within each utterance (Ramus et al. 1999). It represents the distance of each consonantal interval from the mean. Highly influenced by speech rate (Dellwo & Wagner, 2003; White & Mattys, 2007b). Stress-timed languages are predicted to have larger ΔC than syllable-timed languages due to phonotactic differences (Ramus et al., 1999; White & Mattys, 2007b).

deltaV (ΔV): the standard deviation of the duration of vocalic intervals

within each sentence (Ramus et al., 1999). This metric is analogous to the previous one but with the vocalic intervals. Also influenced by speech rate (Dellwo & Wagner, 2003; White & Mattys, 2007b) and with larger figures for stress-timed languages.

varcoC: percentage of the standard deviation of the consonantal interval duration divided by the average duration of consonantal intervals. This is a rate-normalized measure for the ΔC (Dellwo, 2006). Figures are predicted to be larger for stress-timed than syllable timed languages.

varcoV: percentage of the standard deviation of the vocalic interval duration divided by the average duration of vocalic intervals. This is a rate-normalized measure for the ΔV (Barry et al., 2003; White & Mattys, 2007a) with larger figures for stress-timed languages.

Pairwise Variability Indexes: these are three metrics introduced by Grabe & Low (2002); and Low et al. (2000).

raw Pairwise Variability Index V (rPVI-V): average difference between consecutive vocalic interval durations. (Grabe & Low, 2002). This metric is non-normalized for rate (see next entry for the normalized version)

normalized Pairwise Variability Index V (nPVI-V): average of relative differences between consecutive vocalic interval durations. Obtained by calculating the absolute value of the difference in duration between each pair of successive measurements and dividing it by the mean duration of the pair. Then, the differences are summed and divided by the number of differences. (Low et al., 2000; Grabe & Low, 2002). This metric is a rate-normalized measure for rate. Figures are predicted to be higher for stress-timed than syllable-timed languages due to the larger variability of stressed-unstressed vowels in the former group.

raw Pairwise Variability Index C (rPVI-C): average difference between consecutive consonantal interval durations. (Grabe & Low, 2002). This is a non-normalized for rate. Lower figures are predicted for syllable-timed languages due to the basic syllable structure.

Firstly, the rhythm metrics were evaluated in terms of their adequacy in the characterization of rhythm with extemporaneous speech data. In order to do so, first, the influence of speech rate on these metrics was assessed. Then, a second analysis was carried out in order to assess the extent to which the various rhythm metrics could effectively discriminate among speaker groups (NSs vs. NNSs) and among different data collection times for NNSs. For this analysis, native and non-native speakers were considered, and for non-natives, each data collection time was treated as a separate group, with T1 and T2 representing non-experienced speakers, and T3 experienced speakers. First, a 3 level principal component analysis was performed to determine to which group the speakers belonged to according to the rhythmic metrics used. Then, a discriminant analysis was used to check the strength of the predictions for the group distinctions. Finally, after having checked which of the metrics used were not influenced by rate and which could discriminate between rhythm types, hence capturing the features of NSs and NNSs' speech, the scores from each measure were submitted to a series of t-tests to test for differences between data collection times.

6.4. Perception Task: Rating Task

This task was intended to gauge the perception of L2 fluency by different groups of listeners. The section is divided into three subsections which present the participants (listeners), instruments and procedures and analyses undertaken to answer the research questions.

6.4.1. Listeners

Listeners (*n=69*) in the perception task were assembled in 5 groups of raters differing in L1 background (Catalan/Spanish vs. English) and degree of experience with Catalan/Spanish-accented English. The following groups were made: G1) Native English EFL (English as a Foreign Language) teachers in Britain teaching English to learners of various L1 backgrounds, G2) Native English SFL (Spanish as a Foreign Language) students in Britain (at university level), G3) Native English EFL teachers in Spain teaching English to Catalan/Spanish learners of English, G4) Native Catalan/Spanish EFL teachers in Spain teaching English to Catalan/Spanish learners of English, and G5) Native Catalan/Spanish EFL students in Spain (at university level). (see Table 6.1.). Listeners' degree of experience (and comprehensibility) with Catalan/Spanish-accented English was assessed by means of a questionnaire.

Table 6.1. Listeners' profiles.

Raters Groups	Age	Native Language	Teacher/Student	Country of residence
Group 1	48.3	English	EFL T	UK
Group 2	20	English	SFL S	UK
Group 3	43.4	English	EFL T	Spain
Group 4	46.5	Cat/Sp	EFL T	Spain
Group 5	21	Cat/Sp	EFL S	Spain

The 5 groups of listeners were further arranged according to 2 different dimensions: a) experienced vs. inexperienced, and b) L1 English vs. L1 Catalan/Spanish. Experienced listeners were those familiar with Catalan/Spanish-accented English (G3, G4, G5). Conversely, inexperienced listeners (G1 and G2) had not been regularly in contact with

Catalan/Spanish-accented English[22]. As for L1, G1, G2 and G3 were all composed of English native speakers, whereas G4 and G5 of Catalan/Spanish native speakers (see Table 6.2.).

Table 6.2. Listeners' L1 and experience.

	English	**Cat/Sp**
Experienced	G3	G4 – G5
Inexperienced	G1 - G2	N/A

Note: N/A= no applicable.

6.4.2. Instruments and Procedures

The first 20-second excerpts from the interviews in the production task at T2 and T3 were selected (initial pauses were not included). They were converted into 16-bit computer audio files, edited and presented binaurally and in randomized order using the DmDx display software (Forster & Forster, 2003) to run the task and obtain responses and response times (RTs). The decision to use 20 sec. speech samples instead of longer excerpts was taken considering previous research, which reported that this length is appropriate for fluency evaluations by judges. Longer samples might only be detrimental to listeners since it lengthens the duration of the tasks, and they have to hold the string of information longer in memory before producing an evaluation (Derwing et al., 2009; 541).

The ratings were collected individually in Salford (UK) for G1 and G2,

[22] The two inexperienced groups of listeners (G1 and G2) were to a certain extent familiar with Catalan/Spanish accented English due to contact with Spanish learners at some point in their teaching history (G1), or awareness of the Spanish pronunciation features taught in Spanish lessons (G2). However, in both cases the experience was very limited compared to the other three groups who lived in constant contact with the Catalan/Spanish accented English.

and in Barcelona (Spain) for G3, G4 and G5.[23] 65 speech samples were presented to raters over headphones with self-adjusted volume level; 60 target items (produced by L2 speakers at two different data collection times) and 5 control English native speaker samples. All speech samples were normalized for peak and mean intensity. The instructions were pre-recorded and presented visually in written form and orally to listeners on a computer before starting the task. A few practice trials were included in order to familiarize the raters with the type of speech samples to be evaluated, the task procedures and the use of the 5-point rating scale. The scale went from 1= extremely fluent to 5= extremely dysfluent. No definition of fluency was given to complete the task[24]. This was to avoid any kind of influence on the listeners. A few practice items were first presented so participants would familiarize themselves with the task. Once the perception task started, the items were automatically presented 1 sec. after response and the time-out for response was 6 sec. The judges were instructed to press a number key from 1 to 5 on the computer keyboard to rate each speech sample. They were asked to use the whole scale to rate the speech samples for fluency, and to rate the speech samples as accurately and fast as possible[25].

6.4.3. Analyses

Listeners' ratings were extracted from the DmDx outputs and transferred to an SPSS data editor. NS samples were removed from the data after verifying that all judges had assigned a 1 to most of these samples.

[23] We are most indebted to professor Richard Towell for allowing us access to the groups of listeners at the at University of Salford, to Joan C. Mora for the groups at Universitat de Barcelona, and to Carmen Pérez Vidal for the groups at the Pompeu Fabra University.
[24] However, prior to starting the experiment listeners were informed that they would be asked to rate the same speech samples for fluency, foreign accent, and comprehensibility, one at a time and in two different sessions (5 days apart).
[25] The RT data were not used for the present study.

Statistical analyses were performed using the SPSS package. The alpha level of significance was set at .05 for all analyses. At this point, interrater reliability was computed to assess level of agreement among judgments. First, all listeners ratings were included in the calculation, which gave a high reliability of Cronbach's $\alpha = .98$. However, taking into account that the value of α depends on the number of items on the scale, its value increasing with large sets of data (Field, 2009), we decided to compute the reliability within each group of raters and between the five groups. Table 6.3. summarizes the results for all the reliability analysis performed evidencing very high α values both between and within groups, thus, the reliability of the scores.

Table 6.3. Inter-rater reliability

		Cronbach' α
Individual ratings		.98
Between-groups		.95
Within-groups	G 1	.89
	G 2	.91
	G 3	.88
	G 4	.93
	G 5	.94

Once the inter-rater reliability was verified, the data was submitted to further statistical analysis. First, Pearson correlations were used to determine the direction and strength of the relationship between perceived and utterance fluency. Then, in order to assess whether judges perceived any differences in fluency between T2 and T3, *t-tests* were performed. Further analysis to determine the effect of the listeners' group variables were undertaken by means of ANOVAs.

6.5. Overall Analyses

In this chapter, the type of analyses undertaken and instruments used for

each individual aspect under study have been described including utterance fluency, perceived fluency and rhythm. It was also mentioned that in the study of utterance fluency, the possible effects of the previous FI period were controlled for. Regarding rhythm, a comparison of the results on rhythmic metrics before and after FI and SA has been proposed after having checked for the appropriatenes of the rhythm metrics used for extemporaneous speech data. Finally, the task on perceived fluency has been described comparing the scores given by listeners for the speech samples collected before and after the SA. Comparisons between and within listeners' groups are undertaken.

In addition to the longitudinal analyses for each L2 speech aspect, several other analyses have been proposed interrelating the different aspects of fluency and rhythm considered in this study. First, utterance fluency is compared to perceived fluency before and after the SA. Pearson's *r* Correlations are used to check for the existence of a relationship between the two dimensions and its strength. Then, after checking for multicollinearity, in order to explain how much variance in the perceived fluency ratings could be explained by the utterance fluency measures, a series of regression analyses are carried out. In a second stage, rhythm is correlated with both utterance fluency and perceived fluency to explore the hypothesized relationship between the two domains of L2 speech. Finally we compute a hierarchical multiregression analysis to check whether rhythm scores can explain any variance in perceived fluency after the SA, and hence, be considered one of the components of fluency.

Confidence intervals are set at 95% throughout the study. Apart from testing for the null hypothesis and reporting the p-value, we also include the effect sizes in this study. Effect sizes measure the magnitude of a treatment effect, so they help us to provide a more accurate interpretation of significant findings.

7

Results

The results for the analyses undertaken in the present study to provide an answer to the three research questions formulated in chapter 5 are reported here, with the main objective of gaining a better understanding of the effect of context of learning and experience on L2 fluency and rhythm.

The chapter has been organized according to the three main research questions. Accordingly, within each section, the research question and subquestions are dealt with. First, in section 7.1., the results on the effect of experience, operationalized as time and learning context, on L2 fluency and rhythm are reported. These correspond to the first research question:

Research Question 1

Will there be an effect of time and learning context on the fluency and rhythm of L2 speakers?

In this section results on the effect of time and learning context on fluency and rhythm are presented. We begin with utterance fluency where a summary of the results of the preliminary study of Valls-Ferrer (2008) on temporal fluency measures is provided. These are core values for the analysis of fluency in the present study and are compared to results on perceived fluency and rhythm (section 7.2.). After covering utterance fluency, a section on gains in perceived fluency during the SA period is

presented (section 7.1.2). Finally, results on the changes in rhythm performance in the two learning contexts are shown (section 7.1.3.).

Second, section 7.2. presents the results on the role of utterance features in the perception of L2 fluency, corresponding to the second research question:

Research Question 2

To what extent are hypothesized changes in perceived fluency related to the temporal characteristics of non-native speakers' utterances? That is, to what extent will utterance fluency and rhythm be related to perceived fluency ratings?

Within this section, first, a subsection on the relationship between temporal measures of utterance fluency and perceived fluency is introduced (section 7.2.1.). This relationship is further explored with a multiple regression analysis performed to examine the predictive power of the temporal measures on perceived fluency. Parallel subdivisions are used for the relationship between perceived fluency and rhythm, with section 7.2.2. reporting results on the correlation and the multiple regression analyses.

The last section, 7.3. considers *initial fluency level* as a crucial factor in the acquisition of L2 speech relating it to temporal measures and rhythm metrics. The research question reads:

Research Question 3

To what extent will *initial fluency level*, indexed as a temporal composite fluency measure, be a factor affecting SA outcomes on fluency and rhythm?

Two different subsections report on this third research question. The first one (section 7.3.1.) tackles the issue of the relationship between initial fluency level and SA outcomes in fluency and rhythm. The second subsection, 7.3.2., presents the results of SA outcomes according to two different groups difined by initial fluency level.

7.1. L2 speech development as a function of time and learning context.

This first section is subdivided into three subsections. In the first subsection, 7.1.1., the results from the temporal and hesitation phenomena are presented as they are core for the ensuing comparisons with perceived fluency and rhtyhm scores. Subsequently, section 7.1.2. is devoted to the analysis of perceived fluency. First, results on the effect of time and learning context on listeners' perceptions are presented, followed by the impact of listeners' type on these results. The last subsection deals with how rhythm metrics can capture changes in participants' rhythmic patterns (section 7.1.3). This subsection includes a thorough analysis of the rhythm metrics used in the study to assess their adequacy for the type of data used i.e. extemporanous speech.

7.1.1. Utterance Fluency

This subsection deals with *RQ.1.1.* together with subsection 7.1.2. The research question reads as follows:

RQ.1.1. To what extent will listeners perceive gains in participants' oral fluency after the study abroad period? That is, when asked to rate L2 speech samples, will listeners assign more target-like scores to T3 than to T2 speech samples?

Utterance fluency was assessed through a set of temporal fluency measures. Improvement in these measures during the 15-month period under study was first examined, followed by the comparison of gains in each learning context, FI vs. SA. Baseline data for a group of English NSs was provided for comparison purposes.

First, tests for normality were performed on all utterance fluency measures. A brief inspection of the normal Q-Q Plots suggested that most measures were normally distributed. In order to check this first impression, a Kolmogorov-Smirnov test was run for all variables. The p-values were larger than .05 for all measures but MLoR, indicating that the assumption of normality had been violated by this measure. When observing the histogram, the MLoR data was slightly positively skewed. We then proceeded to transform the data for this measure using a log transformation. Once the transformation was done, the Kolmogorov-Smirnov test was computed on the transformed variable and this time the p-value was larger than .05.

The data was then tested for homogeneity of variance using Levene's test. The p-values for all measures were larger than .05 indicating that the variances were equal.

The general pattern for the temporal fluency measures was an increase in scores over the 15-month period under study for the speed fluency measures and a decrease for the breakdown and repair fluency measures. NNSs scores never reached NSs values in any of the temporal fluency measures (see table 7.1).

Table 7.1. Mean fluency scores at T1, T2 and T3 for NNSs (*N*=30) and NSs (*N*=10) (*SD* in parenthesis)

Temporal Fluency Measures	**NNSs**			**NSs**
	T1	**T2**	**T3**	
SR	155.72 (27.97)	157.28 (27.90)	176.12 (29.40)	242.04 (34.82)
AR	206.64 (30.80)	213.46 (28.62)	226.92 (28.05)	280.12 (24.77)
PhonRat	75.21 (5.63)	73.41 (6.31)	77.38 (6.05)	86.02 (7.23)
MLoR	7.74 (2.24)	7.42 (2.20)	9.30 (2.43)	16.54 (3.89)
DysRat	8.25 (3.76)	7.51 (3.58)	6.22 (3.41)	1.54 (1.83)
PauseFreq	10.43 (2.90)	11.50 (2.98)	8.44 (2.83)	3.51 (1.34)
PauseDur	24.79 (5.63)	26.59 (6.31)	22.62 (6.05)	13.98 (7.23)
Pause_i_Dur	12.14 (3.89)	13.57 (4.64)	9.44 (3.76)	3.89 (2.29)
FLIndx.	0.47 (0.16)	0.44 (0.19)	0.62 (0.18)	1.0 (0.16)

Source: adapted from Valls-Ferrer (2008) and Mora & Valls-Ferrer (submitted)

*SR= speech rate; MLoR= mean length of run; PhonRat =phonation time ratio; AR= articulation rate; DysRat= dysfluency ratio; PauseFreq= pause frequency; PauseDur= internal pause duration ratio; Pause_i_Dur= pause duration ratio; FLIndx=fluency index.

Fluency scores for the whole period under study were submitted to a One-Way Repeated Measures ANOVA with time (T1, T2, T3) as the within-subjects factor and the temporal fluency measures as the dependent variables. Bonferroni-adjusted pairwise comparisons followed to assess where significant improvement occurred. The ANOVA yielded a significant main effect of *time* and pairwise comparisons revealed that whereas all T3 fluency scores were significantly different from T2 (and T1) scores, none of the T1-T2 differences in fluency scores reached significance (see Table 7.2.); speed fluency scores were higher at T3 than at T1 and breakdown and repair fluency scores were lower at T3 than at T1.

Table 7.2. Results of one-way RM-ANOVAs with Time (T1, T2, T3) and Bonferroni-adjusted pairwise comparisons T1-T2 (FI) and T2-T3 (SA) (all Fs = [2, 28]).

RM-ANOVAs	*F*	η^2	*p*	FI	SA
SR	10.3	.424	<.001	1.0	<.001
AR	12.9	.481	<.001	.35	.017
PhonRat	6.8	.328	.004	.48	.002
MLoR	14.2	.504	<.001	1.0	<.001
DysRat	12.4	.470	<.001	.63	.029
PauseFreq	15.0	.517	<.001	.47	.001
PauseDur	6.8	.328	.004	.48	.002
Pause_i_Dur	14.9	.516	<.001	.31	<.001
FLIndx.	17.7	.559	<.001	1.0	<.001

Because significant changes over time were observed for the whole period under study, further pairwise comparisons (Bonferroni adjusted) were conducted to examine where changes occurred, if at home, abroad, or in both learning contexts. No significant changes were found for any of the temporal fluency measures during the FI period (between T1 and T2). By comparison, statistically significant changes were found for all measures during the 3-month SA period (between T2 and T3). **During this period abroad, scores in speed fluency measures increased, and breakdown and repair fluency scores decreased, pointing towards NS standards in both cases.**

Lastly, to assess differences in the size of the gains obtained in each learning context, amount of gains in temporal fluency was calculated by substracting T1 scores from T2 scores (gains during FI) and T2 scores from T3 scores (gains during SA). Marginal gains were found during the FI period for three of the temporal fluency measures (SR, AR and DysRat); all other measures remained practically the same. Conversely, gains during the SA period were large for all temporal fluency measures. Differences in gains between the two learning contexts were significant for

all measures but AR and DysRat (see table 7.3.). Effect sizes were moderate to large for all measures. **These results indicate that the time spent in different learning contexts had an effect on the amount of gains obtained from each context, with the 3-month SA period exerting a significantly greater influence than the previous FI period at home.**

Table 7.3. Amount of gains in oral fluency (T1-T2 and T2-T3) (SDs in parentheses). Gains are indicated by a + sign, negative values (-) indicate loss between testing times. Dysfluency, pause frequency and duration have been reversed. Results of dependent-sample T-Tests comparing amount of gains.

Fluency	T1-T2	T2-T3	*t(29)*	*p*	*r*
SR	+1.56 (23.4)	+18.84 (26.0)	-2.25	.032	.39
AR	+6.82 (23.0)	+13.46 (0.18)	-0.85	.401	.15
PhonRat	-1.80 (6.81)	+3.97 (5.79)	-2.95	.006	.48
MLoR	-0.24 (1.52)	+1.58 (1.57)	-3.84	.001	.58
DysRat	+0.74 (3.20)	+1.29 (2.54)	0.56	.580	.10
PauseFreq	-1.07 (4.05)	+3.06 (3.94)	-2.97	.006	.48
PauseDur	-1.80 (6.81)	+3.97 (5.79)	3.44	.006	.54
Pause_i_Dur	-1.43 (4.67)	+4.13 (4.77)	2.95	.002	.48
FLIndx.	-0.03 (0.17)	+0.18 (0.18)	-3.64	.001	.56

In short, **participants' utterance fluency developed as a function of time and learning context as indicated by temporal fluency measures. Temporal fluency scores after the SA (T3) are closer to NSs' standards than scores at both T1 and T2. Time spent abroad prompted significantly larger gains than time spent in the FI context at home.**

7.1.2. Perceived Fluency.

In this subsection, results from the listeners' perceptions of fluency in the participants' speech samples analysed in the previous section are presented. The first part deals with the extent to which listeners perceive improvement in participants' fluency during a 3-month SA period. Note that this section focuses on the SA period only, hence, comparing results

from T2 and T3. This is followed by a subsection on the impact of listener characteristics (L1 and experience) on fluency judgements.

To begin with, the data was explored and checked for assumptions to perform parametric tests. First, histograms were inspected visually to check whether scores were normally distributed. This subjective impression was then quantified by running a Kormogorov-Smirnov test. The mean perceived fluency scores at T2, $D(69) = 0.10$, $p > 0.05$, and the mean perceived fluency scores at T3, $D(69) = 0.07$, $p > 0.05$, were both normally distributed. As for homogeneity of variance, Levene's tests indicated that for both mean fluency scores at T2 and mean fluency scores at T3, the variances were equal for NSs and NNSs, $F(1, 67) = 0.09$, $p > 0.05$ and $F(1, 67) = 0.06$, $p > 0.05$.

Once the normality and homogeneity of variance had been tested, statistical analyses were performed on these data to check whether the following hypotheses could be confirmed. The first hypothesis predicted that listeners would perceive a difference in participants' oral fluency as a function of testing time: before vs. after the SA period. More precisely, this would hold true if listeners gave T3 speech samples more target like scores than those at T2 when asked to rate speech samples for fluency. The 69 listeners were grouped together and paired-sample t-tests were used to test whether the pooled listeners' scores for fluency at T2 and at T3 were different. On average, listeners rated T3 speech samples (M = 2.78, SD = .40) as significantly more fluent than T2 speech samples (M = 3.06, SD = .42), $t(68) = 5.540$, $p < .05$, $r = .56$. Not only was the difference significant, but also the effect size was large, as illustrated by the r larger than .5.

A mixed between-within subjects ANOVA with Time as the within-subjects factor (T2, T3) and Rater group (G1, G2, G3, G4, G5) as the

between-groups factor was performed to test whether differences between perceived fluency at the two data collection times were significantly different (main effect for time), while also comparing the ability of the groups of listeners to differently perceive fluency (main effect for group). In this way it was possible to determine whether changes in perceived fluency over time were different for the 5 groups of listeners (interaction effect). The mean ratings in Table 7.4. seem to suggest that **listeners perceived the speech samples as being more fluent at T3 (after the SA) than at T2 (before the SA)** and also that **different groups of listeners differed in the fluency scores they assigned to the speech samples.** However, the mixed between-within subjects ANOVA indicated that not all differences were statistically significant. There was a statistically significant main effect for time [$F(1,64) = 30.32$, $p < .05$], suggesting that there was a change in perceived fluency between the two data collection times. As previously observed in the T-test analysis, the effect size was large (partial eta squared .322). There was no significant effect of listeners' groups, indicating that ratings from the 5 groups of listeners were in general the same [$F(4,64) = 2.07$, $p = .09$]. Finally, the non-interaction effect did not reach statistical significance either [$F(4,64) = 1.43$, $p = .08$].

RQ.1.1.a. When grouped by L1 and experience, will any differences in fluency ratings between the two listener groups arise?

Regarding the impact of listener type in the perception of L2 fluency, when the results were analyzed by listeners' group, not all 5 groups perceived differences between the two data collection times in the same manner. First, we started by examining which group/s of listeners were more sensitive to differences between the two data collection times (see table 7.4.). Listeners in G1 (NS Teachers in the UK) perceived significantly

larger differences between speech samples at T2 and T3 (M = .45, SD = .32, Z = -3.20, p < .05)[26]. On the contrary, listeners in G2 (NS Students in the UK) were the ones who perceived less of a difference between T2 and T3 speech samples (M = .10, SD = .31), not reaching significance. Perceived differences between T2 and T3 for the two groups (G1 and G2) as measured by a T-test revealed that the degree to which these differed was significant, $t(27) = 3$, $p < .05$, $r = .50$. No other significant differences emerged among any of the other groups of listeners.

Table 7.4. Perceived differences (amount) in fluency between Time 2 and Time 3 by individual and pooled groups of listeners.

Listeners	**Difference btw T2 & T3**	
	M	*SD*
Inexperienced NS Teachers (UK)	.46	.32
Inexperienced NS Students (UK)	.10	.32
Experienced NS Teachers (Spain)	.25	.37
Experienced NNS Teachers (Spain)	.27	.44
Experienced NNS Students (Spain)	.31	.52
Pooled listeners	.27	.41

Overall differences in perceived fluency as a function of listener group were explored by grouping T2 and T3 rating scores. A one-way between groups ANOVA was conducted to explore the impact of the listeners' grouping on perceived fluency as a whole, without differentiating between data collection times. The dependent variable was fluency scores, for which both T2 and T3 fluency scores were put together, and the independent variable was *listener group* (G1, G2, G3, G4, G5). There was a statistically significant effect of *listener group* on perceived fluency, F(4, 133)

[26] For this comparison we used the Wilcoxon Signed Rank non-parametric test because the distribution was not normal and the total population low (n=14).

= 2.75, $p < .05$, omega squared[27] (ω^2) = .05. Post-hoc comparisons using the Bonferroni test indicated that the differences between **mean perceived fluency scores were only statistically significant between G2 (inexperienced NS Students in the UK) [M = 2.80 SD = .39] and G5 (experienced NNS Students in Spain) [M = 3.12 SD = .47]**. No other significant differences were found among the other groups.

Then, comparisons among groups were conducted separately for perceived fluency scores given at each data collection time (T2 and T3). See table 7.5 for the mean fluency ratings and SD by individual and pooled groups of listeners. The group of NNSs students in Barcelona (G5) gave the lowest fluency ratings to the speech samples at both T2 and T3. On the other hand, NS students in the UK (G2), followed by NNS teachers living in Barcelona (G4) were the ones who gave the highest fluency ratings both at T2 and T3. The differences between ratings given by the different groups of listeners were only significant between G2 (M = 2.85 SD = .40) and G5 (M = 3.27 SD = .41) at T2, but any difference among these groups vanished for the T3 speech samples.

[27] For one-way between groups ANOVA, we used omega squared (ω^2) instead of eta squared (η^2) based on Field 2009:389 recomendation who claims that η^2 is a biased measure of population variance explained (overestimating the sample), and ω^2 corrects this bias by using the variance explained by the model and the error variance. This correction results in ω^2 being always smaller than η^2. The effect sizes for ω^2 are considered small (.01), medium (.06) and large (.14), following Kirk (1996).

Table 7.5. Mean fluency ratings of L2 productions at Time 2 and Time 3 by individual and pooled groups of listeners.

Listeners	**Time 2**		**Time 3**	
	M	*SD*	*M*	*SD*
Inexperienced NS Teachers (UK)	3.15	.36	2.69	.36
Inexperienced NS Students (UK)	2.85	.40	2.75	.39
Experienced NS Teachers (Spain)	3.01	.46	2.76	.47
Experienced NNS Teachers (Spain)	2.97	.36	2.70	.28
Experienced NNS Students (Spain)	3.28	.41	2.97	.48
Pooled listeners	3.06	.42	2.78	.41

Note: 1 = extremely fluent, 5 = extremely dysfluent.

Subsequently, listeners were grouped by L1 (English NS = G1 + G2 + G3; English NNS = G4 + G5) and experience (Inexperienced = G1 + G2; Experienced = G3 + G4 + G5) in order to explore whether these two factors would affect perceived fluency. When these new groupings were made, none of the group differences reached significance, only differences in L1 showed a tendency [$F(1, 136) = 3.15$, $p = .07$] for NSs ($M = 2.86$ $SD = .42$) to assign higher fluency ratings than NNSs ($M = 3$ $SD = .44$). However, this tendency disappeared when the speech samples were grouped by data collection time, instead of being analyzed all together.

To sum up, the development of oral fluency as a perceptual phenomenon over time and learning context has been tackled so far. Overall, the listeners perceived a significant change between speech samples produced before (T1) and after (T2) the 3-month SA period. Participants were perceived as more fluent after the SA, as illustrated by the scores obtained at T3, which are closer to NSs' mean fluency scores. Differences in perceived fluency scores as a function of listeners' L1 and experience were not statistically significant, suggesting that they focused on similar aspects of fluency when rating the speech samples. When amount of perceived gains was assessed, differences between the group of native teachers in the

UK, who perceived the largest gains, and native students in the UK, who perceived the narrower gains, were statistically significant.

7.1.3. Rhythm

This section and further subsections 7.1.3. a, b and c deal with *RQ1.2.*

RQ.1.2. Will there be any change in rhythm before and after the period spent in each learning context as measured by rhythm metrics? If so, to what degree will changes represent significant gains in each of the learning contexts?

The whole section is devoted to the assessment of rhythm. First of all, a wide array of the most commonly used rhythmic metrics was evaluated on their appropriateness for use with the oral production samples under study. The reason for performing this analysis instead of exclusively basing the decision on findings from previous research is mainly due to the nature of the data. Most studies testing rhythm metrics have been based on highly controlled read speech samples (at least for length and syllable structure). However, the data collected from the present study is semi-spontaneous speech, hence, it could not be controlled for either syllable structure or duration. Due to the higher variability that this type of data may contain, compared to read speech, especially in the lengthening, combination and duration of phonetic segments, it was deemed necessary to evaluate the metrics before doing any comparisons.

To begin with, we checked whether any of the measures were affected by speed. Knowing beforehand that there were significant differences in speech rate between samples at different data collection times, particularly between T2 and T3, we wanted to make sure that only those metrics unaffected by speech rate were used. Previous research has provided quite conclusive results about the dependence of ΔC and ΔV, as well as rPVIs

on speech rate. However, we decided to compute all the metrics on our data, including ΔC, ΔV and rPVIs, to check whether these would correlate with speech rate.

Next, the rhythm metrics were assessed in order to find out how well they could discriminate Spanish from English according to the auditory impression of languages sounding more stress-timed or syllable-timed. The main reason for performing this analysis was to check how the metrics responded to extemporaneous speech, hence their adequacy to assess rhythm with this type of data could be evaluated in relation to previous findings for read speech. First, scatterplots were drawn to assess the distribution of speech sample types (native Spanish, native English and non-native English) based on several combination of measures (%V-VarcoV, nPVI, etc.) and compare it to previous findings. Then, a series of ANOVAs were performed to see whether the differences between languages or varieties within languages (the case of NNS speech was treated as a variety of English) were significant.

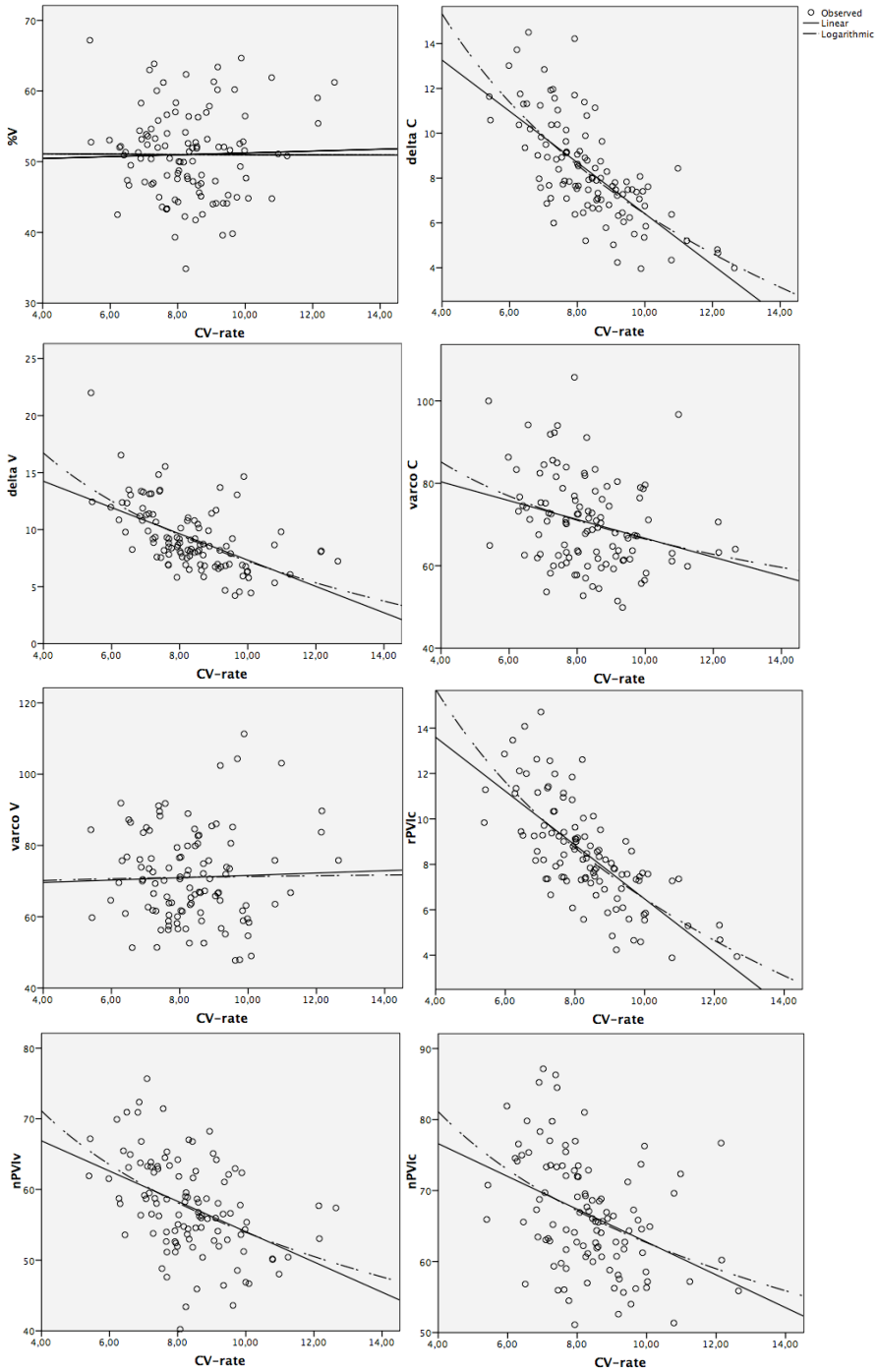
The findings from the analyses just mentioned were then used to decide on the metrics which were going to be included in the developmental analysis in which language gains in rhythm were assessed as a function of time and learning context. First, the development of rhythm towards TL norms for the entire period of time (15 months) was considered. Subsequently, the analysis for the two separate periods of time (FI and SA) was undertaken to assess gains in each context, as well as the size of these gains. Finally a hierarchical regression analysis was used to compare gains between the two learning contexts. All of this is presented in the following subsections.

7.1.3.a. Rhythm metrics and the influence of speech rate.

The influence of speech rate on rhythm is analysed here. Dellwo's (2010) procedures were followed to assess this influence on rhythm measures. To begin with, scatterplots were used to check whether there was a relationship between commonly used rhythm measures and rate of speech. The variable speech rate (pruned), was operationalized as the duration of C + V intervals (without pausing time) and it was displayed in the horizontal axis. A different plot was drawn for each rhythm metric (%V, ΔC, ΔV, varcoC, varcoV, rPVI, nPVI) displayed in the y-axis and plotted against the independent variable.

The scatterplots with a fitted linear and logarithmic line are illustrated in Figure 7.1. They show no relation between speech rate and the variables %V and varcoV. In contrast, a strong relation can be observed for ΔC, ΔV, rPVI and rate. For VarcoC and nPVI the picture is not so clear, so the R^2 results had to be further examined.

Figure 7.1. Scatterplots relating speech in the horizontal axis to rhythm metrics in the vertical axis.



For %V/CV-rate and varcoV/CV-rate, values for the R^2 were very low. In both cases, %V and varcoV, the linear regression was $R^2 = .001$ ($p>.05$) and the logarithmic $R^2 = .000$ ($p>.05$), attesting that these two measures had not been influenced by rate. On the other hand, a strong relationship between CV-rate and non-normalized measures, ΔC, ΔV and rPVI was found as illustrated in figure 7.1. The linear fit for ΔC resulted in $R^2 = .49$ and the logarithmic in $R^2 = .50$ ($p<.05$ in both cases). Similar results were found for ΔV with $R^2 = .30$ for the linear curve and $R^2 = .34$ for the logarithmic ($p<.05$), and rPVIc with both the linear fit and the logarithmic in $R^2 = .52$ ($p<.05$). Hence, ΔC, ΔV and rPVI seemed to be highly affected by the variability in CV-rate.

The relationship between the normalized measures varcoC, nPVIc and nPVIv and CV-rate could not be interpreted straightforwardly from the graphs, but they didn't seem to be strongly related. In all cases, the linear and logarithmic curve *R* values were quite low (varcoC: linear $R^2 = .07$ and log. $R^2 = .08$, nPVIc: linear $R^2 = .15$ and log. $R^2 = .16$ and nPVIv: linear $R^2 = .19$ and log. $R^2 = .20$), however, the *p-values* were significant for all the measures. These results suggested that there might be a relationship between these normalized measures and CV-rate, but, if so, it would be very low; rate would only be responsible for a very low amount of variance in these rhythm metrics.

The results in the present study corroborate Dellwo & Wagner's (2003), Barry et al.'s (2003), White & Mattys' (2007a) and Dellwo's (2010) findings as regards the influence of speech rate on ΔC, ΔV and rPVI and its lack of influence on %V and varcoV. Regarding varcoC, nPVIc and nPVIv, results differ slightly from what White & Mattys (2007a) and Dellwo (2010) found for their data. The former did not find significant correlations between varcoC and speech rate, and the latter reported lower R^2 values than those found here. However, because the general pattern

was the same, it could be that the slight differences might have been caused by the type of data under analysis, since those studies used read-speech in contrast with the extemporaneous speech used in the present study.

The present analysis suggests that the ΔC, ΔV and rPVI measures are not adequate to be used with extemporaneous speech data where time, number of words and type of syllables and phonemes cannot be controlled for. Nevertheless, in the next section, where the distribution of languages along the rhythm continuum is assessed, all measures were used so as to compare the results with previous research.

7.1.3.b. Distribution of languages and language varieties: L1 stress-timed (Eng) vs. syllable-timed (Sp) vs. L2 (Eng).

The above mentioned metrics were further evaluated to find out whether they could discriminate between languages hypothesized to be rhythmically different, as well as to distinguish between first and second language rhythm. Hence, the aim of this section was to assess which rhythm metrics could better distinguish the syllable-timed rhythm of Spanish from the stress-timed rhythm of English in our corpus of semi-spontaneous speech productions. Such rhythm metrics would be better able to capture changes in rhythm patterns towards a target-like stress-timed rhythm. Table 7.6 presents the mean and standard error for each of the measures grouped according to native language (Spanish and English) and non-native productions of English (by Spanish speakers).

Table 7.6. Means (standard errors) of rhythm metrics for Spanish and English NS and Spanish NNS of English.

	Native Speakers		Non-Native Speakers		
	Sp	**Eng**	***Eng*T1**	***Eng*T2**	***Eng*T3**
Interval					
ΔC	04.4 (0.1)	07.9 (0.7)	08.7 (0.4)	08.9 (0.3)	08.4 (0.3)
ΔV	10.3 (1.1)	07.2 (0.7)	09.5 (0.4)	10.3 (0.6)	08.1 (0.3)
%V	60.9 (1.1)	45.9 (1.6)	51.6 (0.7)	54.2 (1.0)	46.2 (0.7)
VarcoC	60.7 (2.3)	64.6 (3.1)	71.7 (1.8)	75.7 (2.1)	67.4 (1.7)
VarcoV	89.2 (5.0)	66.1 (2.8)	68.9 (1.9)	69.0 (2.3)	70.4 (1.9)
Pairwise var.					
*r*PVIc	04.5 (0.2)	08.4 (0.7)	08.8 (0.4)	08.9 (0.3)	08.7 (0.3)
nPVIc	59.1 (3.2)	66.1 (2.6)	67.2 (1.6)	69.0 (1.5)	65.6 (1.0)
nPVIv	55.9 (1.5)	55.0 (1.5)	58.9 (1.1)	57.9 (1.2)	57.3 (1.4)

The distribution of the languages according to the different metrics can be observed in Figures 7.2 to 7.6. Both first languages (English & Spanish) and second language English (at 3 data collection times) are represented in each figure.

Figure 7.2. Distribution of English, Spanish L1 and English L2 over the ΔC, %V plane. Bars represent the standard error around the means.

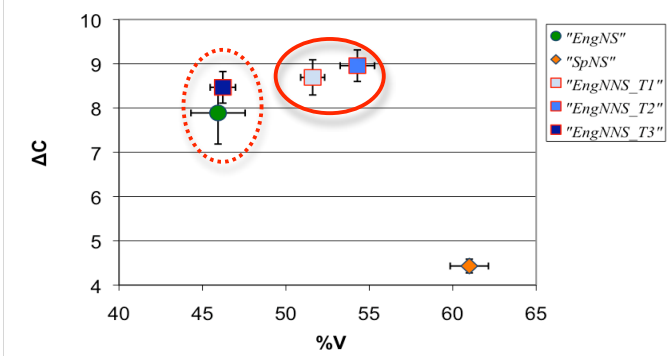


Figure 7.3. Distribution of English, Spanish L1 and English L2 over the ΔV, %V plane. Bars represent the standard error around the means.

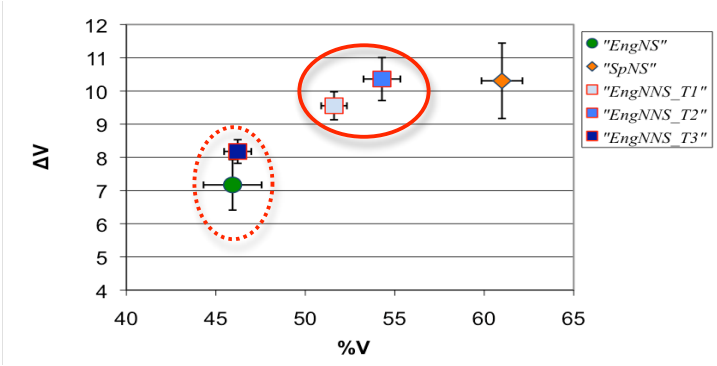


Figure 7.4. Distribution of English, Spanish L1 and English L2 over the varcoC, %V plane. Bars represent the standard error around the means.

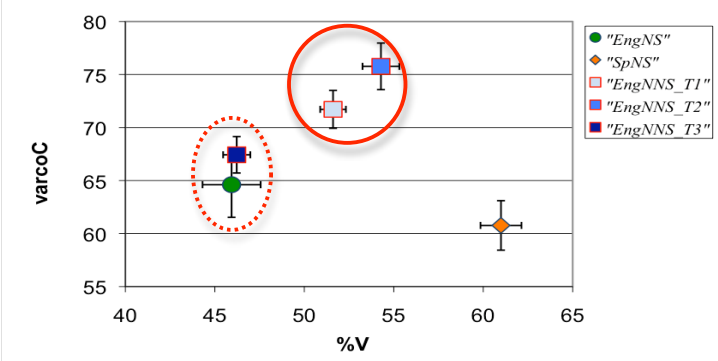


Figure 7.5. Distribution of English, Spanish L1 and English L2 over the varcoV, %V plane. Bars represent the standard error around the means.

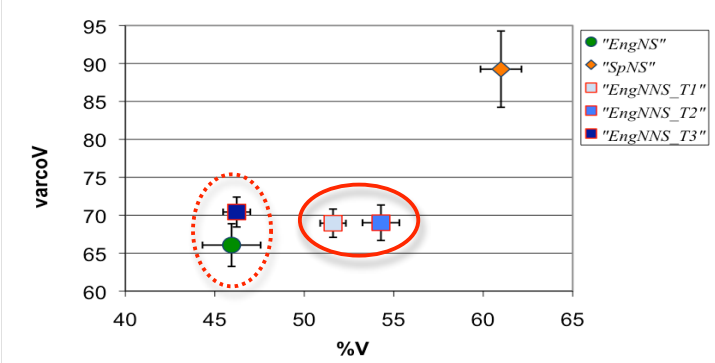
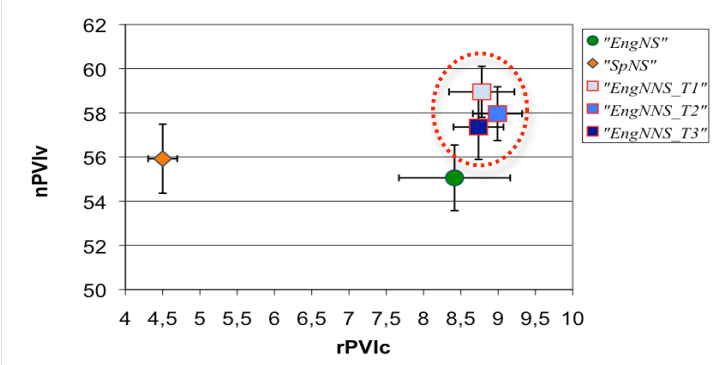


Figure 7.6. Distribution of English, Spanish L1 and English L2 over the nPVIv, %V plane. Bars represent the standard error around the means.



Comparisons among the different languages were made for each variable. For all interval measures, ΔC, ΔV, %V, VarcoC, VarcoV, there was a significant effect of language, but not for the normalised pairwise variability indices, except one (rPVIc). Effect sizes were large for all measures with a significant main effect of language (except for ΔV which showed a medium effect size). See Table 7.7 for a summary of results.

Table 7.7. Results from the analysis of variance testing for effects of language and effect sizes.

Variable	df_M	df_R	F	$Sig.$	Effect size (ω^2)
ΔC	4	102	7.78	$p < 0.01$	.20
ΔV	4	102	4.28	$p < 0.01$	.10
%V	4	102	23.26	$p < 0.01$	.45
varcoC	4	102	5.20	$p < 0.01$	.13
varcoV	4	102	5.51	$p < 0.01$	.14
rPVIc	4	102	7.64	$p < 0.01$	.20
nPVIc	4	102	2.47	$p = 0.05$	.05
nPVIv	4	102	0.79	$p > 0.05$	0

Because statistically significant main effects were found for language, post-hoc[28] tests were performed to check which pair/s of languages or language varieties differed from each other for the eight rhythm metrics tested. Only comparisons between first languages (*Eng*L1 vs *Sp*L1) and first and second languages (*Eng*L1 & *Sp*L1 vs. *Eng*L2_T1, *Eng*L2_T2, *Eng*L2_T3) are reported here. Comparisons within the *Eng*L2 groups are dealt with in the next section (7.1.3.c). For %V, the mean score for *Eng*L1 was significantly lower than for all the other languages[29] [vs *Sp*L1, $p < .001$; vs *Eng*L2_T1 $p < .05$; vs *Eng*L2_T2, $p < .01$], but not statistically different for *Eng*L2_T3, even though the mean score for *Eng*L2_T3 was higher than *Eng*L1. *Sp*L1 had a %V mean score significantly higher than all the other languages [vs *Eng*L1, $p < .001$; vs *Eng*L2_T1 $p < .001$; vs *Eng*L2_T2, $p < .005$; vs *Eng*L2_T3, $p < .001$]. So, the %V mean for the L2 productions was higher than the *Eng*L1 mean and lower than the *Sp*L1 mean, as predicted.

For ΔC, the mean score for *Eng*L1 was significantly higher than for *Sp*L1 [$p < .005$], and the mean for *Sp*L1 was significantly lower than all other languages [vs *Eng*L1, $p < .005$; vs *Eng*L2_T1 $p < .001$; vs *Eng*L2_T2, $p < .001$; vs *Eng*L2_T3, $p < .001$]. Here we see that *Eng*L2 obtained higher scores than *Sp*L1, and even higher scores than *Eng*L1 (although non-significantly different) against expectations. For ΔV, the mean score for *Eng*L1 was lower than all other language varieties, but the difference was only significant for the *Eng*L2_T2 [$p < 0.05$]. There were no significant

---

[28] The *Games-Howell* procedure was used instead of the commonly-used *Tukey's* test due to the difference in group sizes. This test is recommended when the group sizes and population variances are very different. However, a quick examination of a *Tukey's* test performed revealed highly similar results to the Games-Howell's.

[29] In this section, for the comparisons between L1s and the L2, the three different times of the *Eng*L2 are treated as varieties within a language.

differences between *Sp*L1 and any of the English varieties, even though they all had lower values (except for *Eng*L2_T2 which had the same mean, but different SD). These results are in the opposite direction than expected.

VarcoC had a higher mean score for *Eng*L1 than *Sp*L1, but lower than all the *Eng*L2 varieties. However, none of the differences were significant. On the contrary, *Sp*L1 had a significantly lower mean than *Eng*L2_T1 [$p < .05$] and *Eng*L2_T2 [$p < .001$]. In this case, the *Eng*L2 scores are higher than the *Sp*L1 as expected, but they seemed to reach *Eng*L1, contrary to the slightly lower values predicted. For VarcoV, the mean score for *Eng*L1 was lower than any other language varieties, but significantly different only from *Sp*L1 [$p < .05$]. On the other hand, *Sp*L1's varcoV mean was significantly higher than that of any of the English varieties [vs *Eng*L1, $p < .05$; vs *Eng*L2_T1 $p < .05$; vs *Eng*L2_T2, $p < .05$; vs *Eng*L2_T3, $p < .05$]. In this case, the values for *Eng*L2 were between the values of the *Eng*L1 and *Sp*L1, but the L1s values were right in the opposite direction to what was expected.

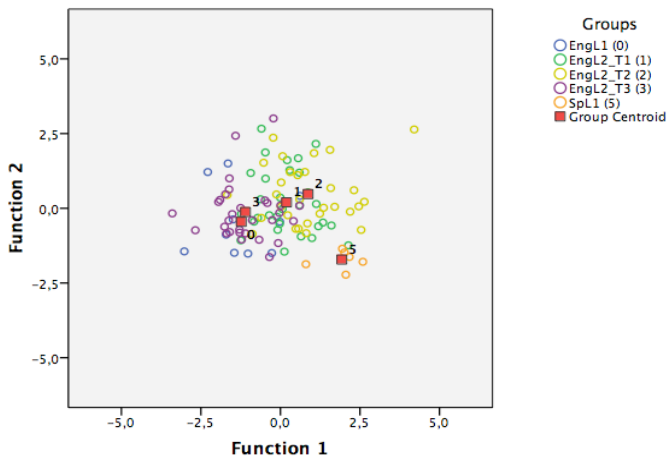
The same pattern resulted for rPVIc, with a mean score of *Eng*L1 significantly higher than *Sp*L1 [$p < .005$], and *Sp*L1 significantly lower than any other languages [vs *Eng*L1, $p < .005$; vs *Eng*L2_T1 $p < .001$; vs *Eng*L2_T2, $p < .001$; vs *Eng*L2_T3, $p < .001$]. For the normalized *Pairwise Variability Indices* (nPVIc and nPVIv) no significant differences between any pair of languages or language varieties were found.

When comparing the grouping patterns found among the rhythm metrics with findings from previous studies on read speech, several similarities and differences were observed. The following metrics discriminated among L1 rhythm in the expected direction: %V, ΔC, varcoC, rPVIc. The other metrics, either followed the opposite pattern to that expected (ΔV,

varcoV) or did not show any differences between the languages (nPVIc and nPVIv). Regarding the position of the *Eng*L2 productions with reference to the L1 productions, only %V followed the expected pattern with *Eng*L2 falling in between *Eng*L1 and *Sp*L1 positions. However, we deal with the non-native productions in the following section.

A discriminant analysis was performed to test the accuracy in predicting group membership of the 4 variables reported in the previous paragraph as good discriminators between the L1s in our data, and their consistency with findings from previous studies in the field. A five-way classification of the participants into five groups –with group centroids– is presented in figure 7.7. The *Sp*L1 group (orange) was clearly positioned on the lowest more right-hand side part of the figure, at a considerable distance from the other groups. Right above it, the *Eng*L2_T1 and *Eng*L2_T2 were hardly distinguishable from each other. And left from these two groups, the *Eng*L2_T3 and the *Eng*L1 were practically overlapping.

Figure 7.7. Canonical Discriminant Functions for group membership based on %V, ΔC, varcoC, rPVIc.



According to the discriminant analysis including the 4 aforementioned variables (%V, ΔC, varcoC, rPVIc), 51.4% of the original grouped cases were correctly classified. This result is considered more than acceptable taking into account the variability within the speech samples analysed due to the extemporaneous nature of the data. The *Sp*L1 group was well distinguished with a correct prediction of 100%. Opposite to this finding, and contrary to our expectations, the *Eng*L1 group could not be distinguished from the *Eng*L2_T3 group, and no values were correctly predicted, 80% being placed under the *Eng*L2_T3 range. For this latter group, 73.3% of the values were predicted correctly. For EngL2_T2 and EngL2_T1, 46.7% and 40% respectively were correctly classified. Most of the remaining values were placed within the other group's range. Therefore, if the two groups had been taken together, an approximate 85% of the values would have been correctly predicted.

To summarize, the two previous sections provided ample evidence of the adequacy and limitations of certain measures to accommodate to the extemporaneous speech composing the data set in the present study. In the first section we saw that some measures were affected by speed (ΔC, ΔV and rPVI), and in the second, that other measures did not discriminate among different types of languages (ΔV, varcoV, nPVIc and nPVIv). **These results suggest that two rhythm metrics, %V and varcoC are the most adequate ones to examine rhythm in extemporaneous speech.** This is further elaborated in the next section with comparisons on L2 data from the same participants at three different points in time.

7.1.3.c. Longitudinal L2 rhythm: comparison between data collection times (FI vs SA).

Once comparisons had been made to establish how the tested languages (L1s & L2) were distributed according to the rhythm metrics, further analyses were undertaken with the L2 data. As previously explained, oral productions in L2 English were collected longitudinally at three different points in time. These data not only represent a longitudinal approach to the analysis of L2 rhythm (based on time), but they also enclose two different learning conditions, which are mainly characterized by the type and amount of exposure to the TL that students receive.

This third subsection deals with RQ1.2. which asked whether there would be an effect of time and learning context on the rhythm of L2 speakers. In order to test if (and how) these changes occurred, we look at changes in rhythm outcomes from pre- to post-test in each learning context and the amount of gains obtained. To begin with, an analysis of the longitudinal data was performed to detect changes in participants' productions over the 15 month period under study. First, each of the eight rhythm metrics were entered into the repeated measures ANOVA analysis, to be followed by post-hoc tests to asses where the significant changes occurred. However, because some of the measures were affected by speech rate, and previous work with this set of data had demonstrated that speech rate increased in participants productions over time, we focused on the measures which had shown not to be affected by speech rate, even though results for all measures are reported.

The assumption of sphericity[30] for repeated-measures ANOVA was assessed through the Mauchly's test for all metrics to find that it was not violated in any of the cases.

Table 7.8. Results of the repeated-measures ANOVA

		Mean (SD)					
	N	T1	T2	T3	*F*	$\eta^2_{partial}$	*Sig.*
ΔC	30	8.69 (2.1)	8.95 (1.9)	8.46 (1.9)	0.53	.04	.594
ΔV	30	9.55 (2.3)	10.36(3.5)	8.17 (2.0)	05.40	.28	.010*
%V	30	51.60(3.9)	54.28(5.6)	46.22(4.1)	30.18	.68	.001*
varcoC	30	71.71(9.7)	75.77(12)	67.43(9.4)	05.53	.28	.009*
varcoV	30	68.94(10.2)	69(12.8)	70.41(10.7)	0.15	.01	.860
rPVIc	30	8.77 (2.4)	8.99 (1.8)	8.73 (1.8)	0.19	.01	.829
nPVIc	30	67.24(8.8)	69.03(8.1)	65.66(5.8)	1.88	.12	.170
nPVIv	30	58.95(6.3)	57.96(8.0)	57.35(8.0)	0.40	.03	.672

The overall pattern of development for the rhythm measures under analysis showed that for the interval measures, there was a change in participants' rhythm over time. This change was sometimes in the expected direction (%V, varcoV) and sometimes in the opposite direction (ΔV, ΔC and varcoC). The differences that emerged were significant for most of the interval measures (ΔV, %V, varcoC), but not for ΔC and varcoV, for which scores at the three data collection times remained practically the same. For the pairwise variability indices a stable pattern was found, with no significant differences emerging for any of the measures (Table 7.8). The effect sizes were quite small for all the measures except for %V with 68% of explained variance on data collection times, and ΔV and varcoC with a 28% each. The fact that varcoC changed in the opposite direction to that expected can probably be explained by the abnormally high values in this measure at T1 and T2, much higher than

[30] This assumption holds that the variation within the two treatment conditions is not significantly different.

NS values. Indeed, NNSs' values at T3 were comparable to NSs' values, and no significant differences were found between them.

Post-hoc analyses were performed to detect where the significant changes occurred. For most of the measures, a similar pattern emerged, with hardly any changes occurring during the FI period, and few changes during the SA period (see Table 7.9.). These changes were only significantly different for ΔV, varco C and %V, and between T2 and T3 (after the SA). The main effect of time for %V before and after the SA was not only significant, but it also represented a large effect size of r =.79. ΔV and varcoC effect sizes were large as well, .50 and .52 respectively. However, as previously mentioned, it should be noticed that the changes in ΔV and varcoC were not int the expected direction, with lower values at T3 than at T2, but getting closer to NS standards.

Table 7.9. Post-hoc comparisons for the repeated measures ANOVA.

	Context	*Mean diff*	*Std. Error*	*Sig.*	*Effect Size (r)*[31]
ΔC	FI (T1-T2)	-0.26	.40	1	.12
	SA (T2-T3)	0.49	.46	.910	.19
ΔV	FI (T1-T2)	-0.81	.56	.492	.26
	SA (T2-T3)	2.18	.69	.012*	.50
%V	FI (T1-T2)	-2.67	1.12	.068	.41
	SA (T2-T3)	8.06	1.15	.001*	.79
varcoC	FI (T1-T2)	-4.05	2.61	.392	.27
	SA (T2-T3)	8.34	2.52	.008*	.52
varcoV	FI (T1-T2)	-0.06	2.11	1	.00
	SA (T2-T3)	-1.41	3.10	1	.08
rPVIc	FI (T1-T2)	-0.21	.46	1	.08
	SA (T2-T3)	0.25	.42	1	.11
nPVIc	FI (T1-T2)	-1.79	2.10	1	.15
	SA (T2-T3)	3.37	1.78	.205	.33
nPVIv	FI (T1-T2)	0.98	1.72	1	.10
	SA (T2-T3)	0.60	1.65	1	.06

[31] The coefficient r was used following Fields (2009) recommendation to calculate the effect size of individual comparisons in repeated-measures ANOVA.

The boxplots in Figures 7.8 to 7.10 illustrate the position of the NNSs scores in relation to the rhythm metrics at the three data collection times. NSs scores are also included for comparison purposes. They show how close the non-native speakers at T3 were to the NSs, even in the cases where the developmental pattern did not follow the expected direction.

Figure 7.8. Boxplot for %V for the NNS at the 3 data collection times and the NS baseline data.

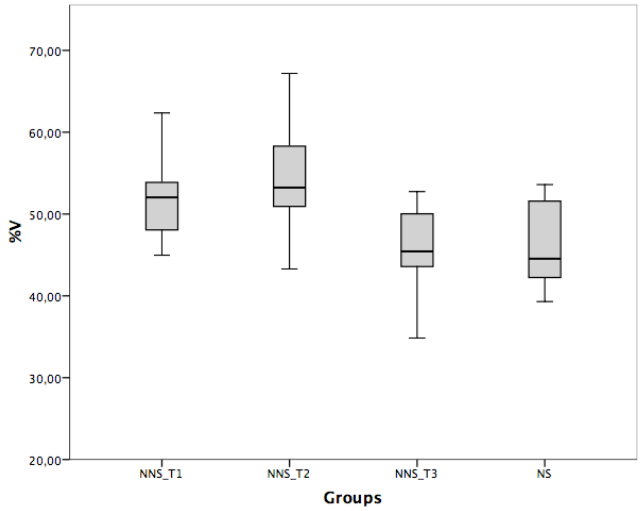


Figure 7.9. Boxplot for ΔV for the NNS at the 3 data collection times and the NS as base line data.

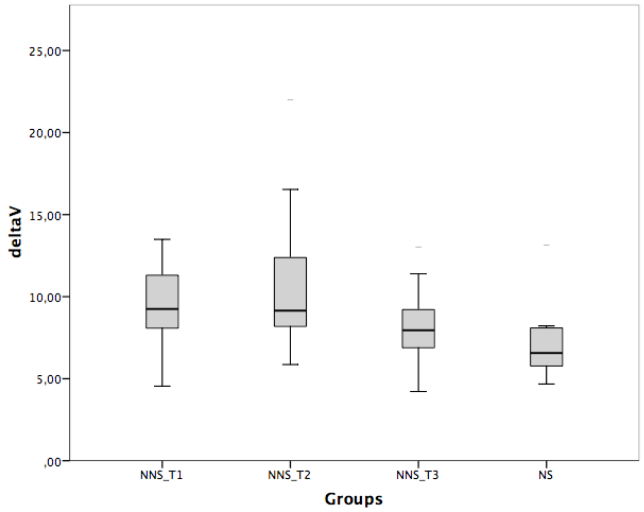
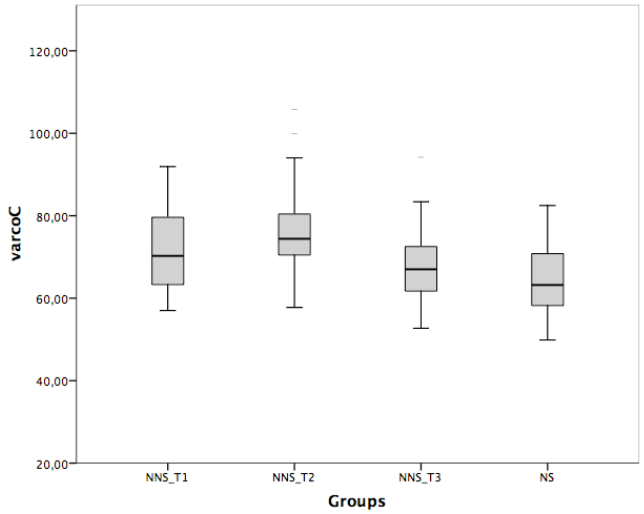


Figure 7.10. Boxplot for varcoC for the NNS at the 3 data collection times and the NS as base line data.



The size of gains between T1-T2 and T2-T3 were analysed. The t-tests revealed that significantly larger gains were obtained during the period

abroad (T2-T3) than while at home (T1-T2). These differences were significant for %V, ΔV and varcoC (see Table 7.10.).

Table 7.10. Results for a T-test comparison between gains in rhythm (T1-T2 vs. T2-T3) indicated by a + sign (loss by a – sign), and effect sizes.

	Gains T1-T2	**Gains T2-T3**	***p***	***r***
%V	-2.68 (6.08)	+8.06 (6.33)	.0001	.69
ΔV	+.81 (3.09)	-2.18 (3.81)	.016	.43
ΔC	+.26 (2.21)	-.49 (2.55)	ns	.17
varcoC	+4.05 (14.29)	-8.34 (13.82)	.012	.44
varcoV	+.06 (11.59)	+1.41 (17.01)	ns	.05
rPVIc	+.21 (2.52)	-.25 (2.29)	ns	.11
nPVIc	+1.79 (11.54)	-3.37 (9.75)	ns	.26
nPVIv	-.98 (9.41)	-.60 (9.06)	ns	.02

A Hierarchical Multiple Regression (HMR) analysis was performed on %V, ΔV and varcoC to check how much of the variance at T3 could be accounted for by T2 scores after the initial T1 fluency level was controlled for.

To begin with, %V was analysed, and %V_T3 (%V after the SA) was selected as the dependent variable. Then, we had two independent variables, %V_T1 and %V_T2. The first variable to be introduced into the model was %V_T1, and the second variable was %V_T2. The output generated by the HMR showed that %V_T1 alone accounted for 11.2% of the variance on T3 scores. When %V_T2 was introduced into the model, the percentage of total variance explained on T3 scores was increased to 12.6%. This means that participants' rhythm before the FI (T1) already accounted for 11.2% of the variance in rhythm after the SA (T3). When rhythm after FI (T2) was included in the model, the value explained a 12.6% of the variance in rhythm after the SA. Therefore, if rhythm before FI accounted for 11.2%, rhythm after the FI only accounted for an additional 1.4%, corroborating the above findings that FI hardly made any difference in participants' rhythm, and that the gains could be mainly attributed to the SA period.

The same analysis was applied for the ΔV measure, introducing ΔV_T3 as the dependent variable and ΔV_T1 first and ΔV_T2 second as independent variables in the HRM analysis. Results follow the same pattern as %V, with ΔV_T1 accounting for most of the variance compared to ΔV_T2. However, the model is less strong, with an $R^2 = .058$ for ΔV_T1 and marginally increasing to $R^2 = .059$ when the second variable (ΔV_T2) was included. Highly comparable results were found for varcoC, with an $R^2 = .052$ in the first step, and an $R^2 = .076$ when varcoC_T2 was added. For these two last measures (ΔV and varcoC), none of the models account for a significant amount of variance in the dependent variable (T3 rhythm scores).

To sum up, the analysis in this last subsection concerning *RQ.1.2.* showed that L2 rhythm changed over time as illustrated by some of the rhythm metrics. However, changes were not always significant and/or occurred in the expected direction. Rate-independent measures distinguishing L1 from L2 speech samples (%V and VarcoC) were found to vary significantly as a function of time. Post-hoc tests provided evidence that changes occurred during the SA period only. As already mentioned, the changes in varcoC were not in the expected direction, with values getting lower over time. However, values at T1 and T2 were much higher than NSs' values, and values at T3 were comparable to NSs'. This might be due to the length of consonants at T1 and T2 being longer than target language standards. Subsequently, the size of gains between T1-T2 and T2-T3 were compared. **Results showed significantly larger gains obtained during the SA period for both measures.** Effect sizes for %V were large in both analyses, and medium to large for varcoC. A final analysis was undertaken to find out the amount of variance in rhythm gains which could be accounted for during the SA if gains during FI were controlled for. The percentage of variance explained by results in the two variables at T3 was

practically the same at T1 than at T2, hence indicating that FI did not significantly improve participants' rhythm.

Section 7.3. has been entirely devoted to the study of rhythm. The various analyses undertaken aimed at gaining an understanding of the adequacy of the rhythm metrics used to analyse the type of data in the present study, as well as the development of L2 rhythm over time. The first two subsections dealt with the examination of the suitability of the rhythm metrics for the type of data under study, and the third subsection with changes over time and learning context. Findings were conclusive regarding the effect that speech rate had on some of the metrics used (ΔC, ΔV and rPVI) but not on the others. Hence, the results on these three measures for the subsequent analyses could not be taken into account since the nature of the data we had here, being extemporaneous, and not controlled for length or speed, would have made it impossible to discern between the effects of speed and those of rhythm.

The subsequent analysis was undertaken to assess whether the rhythm metrics used could discriminate between language types in the rhythm continuum; not only first, but also second languages. Two of the metrics (%V and varcoC) turned out to be good discriminators among languages, while the others were not (varcoV, nPVIc and nPVIv). They could discriminate between *Eng*L1 and *Sp*L1 productions, positioning them at the stress- and syllable-timed ends of the rhythm continuum respectively, and the three variants of the L2 in between L1 productions.

Finally, in the third part of this section, **the longitudinal analysis gave good account of how L2 rhythm experienced changes over time towards TL norms. The pattern for %V followed the expected pattern with values decreasing over time to approximate the low *Eng*L1 figures.** VarcoC, on the other hand, showed a slightly

controversial pattern, since the values decreased over time instead of increasing as expected. However, the oddity was for the figures at T1 and T2 which were much higher than normal, falling down to near native values at T3.

7.2. The role of utterance features –temporal and hesitation phenomena and rhythm phenomena– in the perception of L2 fluency.

This section deals with *RQ.2*. Before performing any comparison between utterance fluency and perceived fluency, we checked whether the data were comparable. For the utterance fluency analysis, participants' entire speech productions (*Mean* =115 sec.) were used, whereas for the perceived fluency analysis 20 sec. excerpts were used. We were confronted with the decision to either calculate the temporal fluency counts for the 20 sec. only, or test whether these excerpts were comparable to the entire productions. If so, the temporal fluency counts from the entire productions could be used to compare them with perceived fluency scores. We opted for the last option, so we performed a set of analyses on these measures to check whether these excerpts were representative of the entire productions.

A randomly selected sample of speech productions (10%) of the population at each data collection time was used. We took the 20 sec excerpts used in the perception experiment and calculated the temporal measures for utterance fluency. We then correlated the results of the 20 sec. excerpts with the results of the total production using Spearman's Rank Order Correlations, due to the low number of speech samples analyzed. There were strong correlations between equal variables for the

two sets of data (AR: $r = .83$, $n = 12$, $p < 0.05$; SR: $r = .94$, $n = 12$, $p < 0.05$; PhonRat: $r = .94$, $n = 12$, $p < 0.05$).

A non-parametric test (Mann-Whitney U-test) was used to check whether differences between the two sets of data were significant. As shown in table 7.11. this test revealed non-significant differences between the data sets for all fluency variables measured. Taking this into account, the decision to use the results of the entire speech productions to correlate with perceived fluency scores was made.

Table 7.11. Temporal fluency counts and comparison between the entire productions and the 20 sec. excerpts.

Variables	Set	N	Mean	Sd	Median	*Z*	*p*	*r*
AR	*A*	6	3.76	.64	3.52	-.16	.87	-.05
	P	6	3.66	.56	3.61			
SR	*A*	6	135.1	32.11	122.16	-.48	.63	-.14
	P	6	127.5	26.88	124.5			
PhonRat	*A*	6	75.17	5.18	73.87	-.96	.33	-.28
	P	6	78.02	7.46	76.35			
PauseFreq	*A*	6	8.17	3.64	7.82	-.64	.52	-.19
	P	6	5.50	5.16	6.00			
Pause_i_Dur	*A*	6	9.57	3.96	9.72	-.80	.42	-.23
	P	6	6.19	5.74	6.35			

Set: *A* = all (whole production); *P* = partial (20 sec. excerpts)

7.2.1. Utterance Fluency vs. Perceived Fluency

This subsection deals with *RQ.2.1. Will utterance fluency scores be related to perceived fluency ratings and result as good predictors of perceived fluency?* The relationship between perceived fluency scores and objective temporal fluency measures was assessed through correlation analyses. Moderate to strong significant correlations were found between most temporal measures and fluency scores assigned by the raters (see Table 7.12). Scores for perceived fluency were reversed so that high scores indicated a high level of fluency. Positive correlations were found between speed fluency

measures (SR, AR, MLoR, PhonRat) and perceived fluency ratings; the higher the speed fluency, the higher the fluency ratings. Conversely, negative correlations were found for the breakdown fluency measures (PauseFreq, PauseDur, Pause_i_Dur) and the perceived fluency ratings; the lower the frequency and duration of pauses, the more target like the scores. The only measure which showed no relation with fluency ratings was the dysfluency ratio. Correlation coefficients between speed fluency scores and mean fluency ratings were relatively high (all above .5), hence, representing a relatively strong correlation.

For breakdown fluency, pause duration (as a whole and internal) also showed high correlation coefficients ($r > .5$), however, pause frequency presented a more moderate correlation ($r = .36$), maybe due to the cut off point being set a 400 ms. As already mentioned, no relationship emerged between the dysfluency ratio and perceived fluency.

Table 7.12. Pearson Correlations of temporal fluency measures and mean fluency scores assigned by raters.

	SR	AR	$Phon_{rat}$	MLoR	P_{freq}	P_{dur}	P_{dur_int}	Dys_{rat}
FL score	,619**	,519**	,524**	,650**	-,362**	-,524**	-,510**	,065
Sig.	,000	,000	,000	,000	,004	,000	,000	,619
N	60	60	60	60	60	60	60	60

**. Correlation significant at 0.01 level (2-tailed).

The strongest relationship between perceived fluency scores and temporal fluency measures was found for two of the speed fluency measures, SR and MLoR, both with a *r*-value over .6. Subsequently, correlations were run between the fluency index (FLIndx) and the perceived fluency scores. As expected the correlation was large ($r = -.633$, $p < .01$) and significant between the two variables. The higher the figure in the fluency index (higher speed and lower breakdown fluency), the more TL the perceived fluency score.

We then proceeded to check whether the correlations between temporal fluency measures and perceived fluency scores at T2 and T3 provided different results when analyzed separately. Overall, the correlations between the temporal fluency measures and perceived fluency scores given at T3 (after the SA) were lower than the correlations between the temporal fluency measures and the perceived fluency scores given at T2 (before the SA). However, when the correlation coefficients were compared for each of the measures between the two data collection times, converting the coefficients to z_r, no statistically significant differences emerged.

Temporal fluency measures and perceived fluency scores for the 5 different groups of listeners were also highly intercorrelated (see table 7.13). Perceived fluency scores correlated most strongly with MLoR and SR —the two measures which encompass both speed and breakdown fluency— for all groups of listeners independent of experience and L1. On the contrary, the dysfluency ratio did not correlate with the ratings of any of the groups. Nevertheless, if we focus on the speed and breakdown fluency measures, two slightly different patterns emerged: perceived fluency scores given by inexperienced listeners (G1 and G2) correlated more strongly with pause duration measures than speed measures. On the other hand, experienced listeners' (G3, G4 and G5) ratings correlated more strongly with speed fluency measures than pause duration phenomena. This suggests that inexperienced listeners attended more to pausing than speed phenomena, whereas experienced listeners attended more to speed of delivery. However, this statement must be taken with caution since, when the correlations' *r*s were compared among groups of raters, no significant differences were found. That is, when rating NNSs' fluency, all listeners, irrespective of L1 and experience, paid attention to practically the same utterance fluency features.

Table 7.13. Pearson Correlations of temporal fluency measures and perceived fluency scores by listeners' groups.

Listeners' Groups	MLoR	SR	AR	$Phon_{rat}$	P_{dur}	P_{dur_int}	P_{freq}	Dys_{rat}
Group 1	,641**	,574**	,467**	,507**	-,507**	-,489**	-,357**	,036
Sig.	,000	,000	,000	,000	,000	,000	,005	,784
Group 2	,558**	,522**	,422**	,474**	-,474**	-,444**	-,302**	,082
Sig.	,000	,000	,001	,000	,000	,000	,019	,536
Group 3	,605**	,577**	,487**	,475**	-,475**	-,461**	-,327**	,045
Sig.	,000	,000	,000	,000	,000	,000	,011	,735
Group 4	,619**	,595**	,528**	,456**	-,456**	-,493**	-,363**	,43
Sig.	,000	,000	,000	,000	,000	,000	,004	,746
Group 5	,659**	,663**	,557**	,562**	-,562**	-,529**	-,369**	,096
Sig.	,000	,000	,000	,000	,000	,000	,004	,464

**. Correlation significant at 0.01 level (2-tailed).

Regression analyses were performed to assess the predictive power of the temporal fluency measures on perceived fluency scores. First, simple regressions were run to assess the single contribution of each temporal fluency measure to perceived fluency scores. Table 7.14 shows a summary of the results for each variable.

Table 7.14. Simple regressions between temporal fluency variables and perceived fluency scores.

	R^2	*adj* R^2	β	*F*	Sig.
MLoR	.422	.412	-.650	42.38	$p < .001$
SR	.383	.372	-.619	35.97	$p < .001$
PhonRat	.274	.262	-.524	21.90	$p < .001$
PauseDur	.274	.262	.524	21.90	$p < .001$
AR	.269	.257	-.519	21.36	$p < .001$
Pause_i_Dur	.260	.247	.510	20.39	$p < .001$
PauseFreq	.131	.116	.362	8.74	$p < .001$
DysRat	.004	-.013	-.065	*0.25*	$p = .619$
FLIndx	.401	.391	-.633	*38.81*	$p < .001$

The R^2 value for each variable represents the percentage of variation in perceived fluency scores that that particular variable, when entered as a single factor, could explain. As observed from Table 7.14, MLoR was the variable with the greatest predictive power, since this variable by itself accounted for 42.2% of the variance in perceived fluency scores, almost half of the variance. This left 57.8% of the variance in perceived fluency to be explained by other variables. The *sig.* value tells us that this single variable model (MLoR) results in a significant degree of prediction of perceived fluency.

Several other single variables accounted for relatively large amounts of variance on perceived fluency scores (SR, PhonRat, PauseDur, AR, Pause_i_Dur) when assessed individually. On the other hand, one variable (PauseFreq) accounted for a fairly low percentage of variance, and another (DysRat) didn't seem to be related to perceived fluency scores.

The FLIndx (a composite measure including both speed fluency and breakdown fluency) turned out to be a good predictor of perceived fluency scores as was expected, accounting for 40% of the variance. However, from these results we could interpret that with a single variable (MLoR) we could get better predictability than with any other variable, including a composite one which included both types of fluency (speed and breakdown).

As for the regression analysis, in order to determine whether a combination of variables could predict a higher amount of variance than using single variables, the data were submitted to a multiple regression analysis. First, we checked for multicollinearity between the predictor variables since we were interested in assessing the individual importance of the predictors. High correlations between some of the temporal fluency measures indicated that the multicollinearity problem would arise with our

data if all the measures were introduced in the regression model (see appendix A2.1. for the correlation matrix of temporal fluency measures). The correlation coefficients were higher than .7 in many cases, and the variance inflation factor (VIF) diagnostic showed collinearity among some of the measures.

This problem with multicollinearity came as no surprise since, as already mentioned in the methodology chapter, some of the measures used were confound measures which included both speed and breakdown fluency. For the regression analyses, a collinearity problem was considered and avoided by only using variables that would prevent this problem.

In order to decide which measures would be included in the analysis to establish the temporal variables which could better predict fluency ratings, the correlations between these measures and the fluency ratings were calculated (refer to Table 7.13.). As mentioned above, MLoR presented the highest correlation with fluency ratings, closely followed by SR (both with an $r > .6$). This is not surprising since we had seen that these two measures were a confound for speed and breakdown fluency; the main difference between them being that MLoR, apart from AR and Pause duration, also includes Pause Frequency, whereas SR does not. Excluding these two confound measures from the multiple regression analysis, the next variables with high correlations were PauseDur, AR and Pause_i_Dur (all above .5). Hence, we decided to introduce these three variables as predictors of fluency ratings.

Results showed that predictive power was not enhanced by combining these temporal fluency measures (adjusted $R^2 = .38$, $p < .0001$), on the contrary, it was a bit lower than the model with MLoR as a single predictor. Nevertheless, this multiple regression model provides the opportunity to observe which specific temporal elements of fluency exert

a higher influence on the perceived fluency ratings. In this case, speed (represented by AR) was the variable with higher influence (adjusted R^2 = .26), followed by PauseDur (adjunted R^2 = .12). Pause_i_Dur was excluded from the model since its contribution was non-significant.

To sum up, regarding *RQ.2.1.* **results showed a strong correlation between speed fluency measures and perceived fluency scores, moderate to strong for the breakdown fluency measures**, and low non-significant correlations for repair fluency. **Listeners seemed to focus both on the speed and the pausing phenomena at the same time, as correlations with MLoRs**, a measure including speed of articulation and pause frequency and duration, were the strongest. This variable was also found to be the single best predictor of perceived fluency. No significant differences were found among listeners' groups on the variables they seemed to attend to when producing their fluency judgments.

7.2.2. Rhythm vs. Perceived Fluency

This subsection deals with results related to *RQ.2.2. How and to what extent will rhythm be related to perceived fluency ratings and contribute to explaining variance in perceived fluency scores?* The relationship between rhythm, as an utterance feature which could be related to perceived fluency was examined in this section. Two rhythm metrics (%V and varcoC) are used here as representative of rhythm. The reasoning behind this decision derives from the previous section (see section 7.1.3a and b for a full account).

Correlations between rhythm scores and perceived fluency ratings were run to test whether these two features were related. The results for %V and varcoC are presented in Table 7.15. The complete correlation matrix with all the rhythm metrics is included in appendix A2.2.

Table 7.15. Pearson-r correlations between rhythm metrics (%V & varcoC) and perceived fluency scores by listeners' groupings.

Perceived Fluency	%V	varcoC
Group 1	.351**	.409**
Group 2	.179	.277*
Group 3	.227	.362**
Group 4	.321*	.330**
Group 5	.321*	.365**
Inexperienced G.	.278*	.358**
Experienced G.	.301*	.365**
Native G.	.265*	.365**
Non-native G.	.328*	.354**
Pooled_listeners	.298*	.366**

Significance: *$p<.05$; **$p<.01$

Correlations between perceived fluency and rhythm were positive and moderate when the ratings from all listeners were pooled together. The lower the scores in %V and varcoC, the more fluent participants were perceived to be. When divided by L1 and experience, correlations for the %V were always a bit lower than for varcoC, but were mainly mantained at a moderate strengh and significant level.

Correlations between scores on rhythm metrics affected by speed and perceived fluency ratings were much stronger (see appendix A2.2. for the full matrix). Listeners were asked to rate the speech samples according to their perception of how fluent they were. In section 7.1.2., we saw that listeners' perceptions were highly affected by speed, as attested by the fluency scores obtained. Therefore, if listeners attended to fluency to provide their ratings, and a large part of their perceptions were based on speed fluency, then, high correlations between rhythm metrics affected by speed and perceived fluency could be expected. However, because we were interested in disentangling speed from the counts of rhythm, these measures were not further considered in the analysis of the relation between rhythm and perceived fluency. **To sum up, significant correlations were found between rhythm and perceived fluency. However, when the measures affected by speed were discarded, the**

correlations between perceived fluency and %V and varcoC were only moderate.

With the intention to broaden the understanding of the relationship between fluency and rhythm, correlations between rhythm metrics and temporal fluency measures were run as well. As expected, very strong correlations between rhythm metrics reported to be affected by speech rate and speed fluency measures were observed (see appendix A2.3). Here we only report the correlations between the objective fluency measures and %V and varcoC.

Table 7.16. Pearson-r correlations between rhythm metrics (%V & varcoC) and utterance FL measures.

FL measures	%V	varcoC
SR	-,227	-,249
MLoR	-,276*	-,330**
AR	-,200	-,196
PhonRat	-,199	-,253
PauseFreq	,298*	,382**
Pause_i_Dur	,408**	,363**
PauseDur	,199	,253
DysRat	-,028	,089

Significance: *$p<.05$; **$p<.01$

Correlations between temporal fluency measures and rhythm metrics were significant and moderate for some of the measures. No correlations were found between pure speed fluency measures (AR), nor for speech rate. Interestingly, correlations were found between rhythm metrics and MLoR, pause_freq and pause_i-Dur. **Correlations between MLoR and rhythm were positive, that is, the higher the MLoR, the lower the %V and varcoC.** On the other hand, **correlations between pause frequency and pause internal duration were negative, the higher the pause frequency and duration, especially within clause**

boundaries, the higher the %V and varcoC. That is, the more target-like the scores in the three temporal fluency measures (MLoR, PauseFreq and Pause_i_Dur), the more target-like the scores in the two rhythm metrics (%V and varcoC). Those participants who were perceived to be more fluent were those whose MLoR was higher and had a low %V and varcoC, approximating target language norms. Hence, these results suggest that rhythm is highly affected by pausing phenomena, probably more than speed. Rhythm is compromised by a type of speech full of lengthy and frequent pauses, since the rhythm pattern, whether more stress-timed or syllable-timed will suffer from breaks in the speech stream.

With regards to the significant positive relationship between MLoR and rhythm, but not for more speed based measures such us SR or AR, several observations can be made. First, as previously mentioned, MLoR and SR are complex measures which include characteristics of both speed and breakdown fluency, while AR is purely a speed measure. However, there is an important difference between MLoR and SR which makes them different. SR does not take into account the distribution of pauses while MLoR does. In this sense, MLoR gives account of the speed at which the speech is produced within pause boundaries. It is a more refined measure than SR, which does not give an account of phrases produced without fluency breakdowns. When related to rhythm, it is not surprising to find that only MLoR is significantly related and not SR or AR. This is because the runs accounted for with the MLoR measure are fluent runs, from which any breakdown fluency phenomena has been eradicated. Hence, rhythm can only come into play when fluent strings of speech occur.

The results from the above correlations between perceived fluency and rhythm metrics were used to decide which variables were introduced in the regression analysis to examine if the rhythm metrics could explain any variance in perceived fluency ratings. A multiple linear regression analysis

was used to help determine which of the rhythm metrics could be used to predict perceived fluency ratings. The measures chosen to be included in the analysis were %V, varcoC and nPVIc. These were the three metrics, apart from the ones affected by speech rate, which showed significant correlations with perceived fluency. The correlations were moderate for all three measures [varcoC $r = .37$, $p < .01$; %V $r = .30$, $p < .05$; nPVIc $r = .48$, $p < .01$], so they were deemed adequate to enter in the regression analysis simultaneously.

Assumptions were checked for the model. No multicollinearity problems were found for the variables under investigation. They all correlated substantially with perceived fluency (above .30) and correlations between variables were not too high (below .50). A collinearity diagnosis was performed and the VIF values were all well below 10, providing further evidence that multicollinearity was not a problem with these data. Mahalanobis distances were inspected for outliers in the data and none were found. Homoscedasticity was examined through scatterplots and the spread was considered reasonable.

Since there were no a priori preferences to determine the order of entry of the metrics, a standard multiple regression analysis was performed. The three rhythm metrics produced an R^2 of .31 (adjusted $R^2 = .27$) [$F(3,59) = 8.407$ $p < .001$] for the prediction of perceived fluency ratings. Together these three predictors shared 19% of explained variance and uniquely predicted 12% of the variance[32]. The most influential predictor was nPVIc ($\beta = .39$, $p<.01$), followed by %V ($\beta = .23$, $p<.05$) and varcoC ($\beta = .16$, ns.).

[32] sr^2 indicates the unique variance that each independent variable predicts. Based on Neill (2007) recommendation to report MLR results, the uniquely explained variance and the shared variance were computed.

Table 7.17. Standard multiple regression of rhythm metrics on perceived fluency ratings.

Variables	*B*	*β*	sr^2
nPVIc	.032*	.388	.12
percentV	.022*	.232	.05
varcoC	.008	.155	.02

* p<.05; *B* = unstandarized beta; *β* = standarized beta; sr^2 = squared semipartial correlations.

To sum up, regarding the analysis of results related to *RQ.2.2.*, **rhythm was found to be a variable which helped to explain some variance in perceived fluency scores. From these findings we suggest that rhythm could be included as a feature of utterance fluency which would complement the temporal fluency phenomena**.

7.3. *Initial Fluency Level* as a factor affecting the development of L2 fluency and rhythm.

This section deals with the results related to RQ.3. The analyses in the previous two sections captured how fluency and rhythm developed over time in two different learning situations. As a general pattern, the SA was shown to be the time when most improvement occured. This section examines how the development during the SA period can be conditioned by the fluency level of students before the SA (T2) (from now on *initial fluency level*). That is, how the fluency level participants have before going abroad (T2) can affect the development of both fluency and rhythm so that it can be observed in the results obtained after the SA.

Firstly, correlations between initial fluency level scores and post-SA scores in perceived fluency, temporal fluency and rhythm were performed. Subsequently, participants were divided into two groups according to their initial fluency level scores (T2) and comparisons were performed in

relation to fluency and rhythm after the SA (T3). Additional analyses tested the difference for amount of gains between the two groups.

7.3.1. *Initial Fluency Level* and SA outcomes in fluency and rhythm.

This section deals with *RQ.3.1. How and to what extent will initial fluency level be related to fluency and rhythm after the SA? What's more, will it have any effect on amount of gains?* Correlations were run between *initial fluency level* and perceived fluency ratings after the SA. Perceived fluency scores were reversed so that high scores indicated high fluency. There was a positive significant relationship between the initial fluency scores and the fluency ratings given by the listeners after the SA period, $r = .489$, $p < .01$. That is, the higher the fluency scores before the SA, the higher the perceived fluency ratings after the SA. Subsequently, correlations were run between initial fluency scores and the temporal fluency measures after the SA (see Table 7.18).

Table 7.18. Pearson Correlations between initial fluency level and objective fluency measures after the SA (*N*=30).

Correlations

		SR	MLoR	AR	$Phon_{rat}$	P_{freq}	P_{dur_int}	P_{dur}	Dys_{rat}
Initial FL level	Pearson	,573**	,593**	,453*	,513**	-,044	-,278	-,513**	,125
	Sig.	,001	,001	,012	,004	,818	,137	,004	,512

** Correlation significant at 0.01 level (2-tailed).
* Correlation significant at 0.05 level (2-tailed).

A relationship between *initial fluency level* and scores in most of the temporal fluency measures after the SA (but PauseFreq, Pause_i_Dur and DysRat) **was found. The relationship between *initial fluency level* and pauses was negative. The higher the starting fluency level, the less participants paused after the SA. With all the other fluency**

measures the relationship was positive, the higher the fluency level, the more (and faster) speech produced.

Then, correlations between initial fluency level and rhythm metrics after the SA were performed. No relations were found between *initial fluency level* and outcomes in any of the rhythm metrics after the SA.

Regarding *initial fluency level* and gains in fluency and rhythm, a marginally significant negative relation was found between *initial fluency level* and amount of gains in perceived fluency during the SA, r = -.348, p = .059 (approaching significance). The higher the starting fluency level, the narrower the amount of gains perceived by the listeners after the SA period. Likewise, the temporal fluency scores showed a similar pattern (see Table 7.19), with gains in PauseFreq, r = .537 and Pause_i_Dur, r = .587 reaching significance (ps < .01). The higher the *initial fluency level*, the lower the size of the gains in these fluency measures. These results suggest that fluency level before a SA period is a factor that needs to be taken into account when examining gains in fluency during a SA period.

Table 7.19. Pearson correlations between initial fluency level and amount of gains between T2 and T3.

Correlations

		SR	MLoR	AR	$Phon_{rat}$	P_{freq}	P_{dur_int}	P_{dur}	Dys_{rat}
Initial FL level	Pearson	-,334	-,284	-,347	-,297	-,537**	-,587**	-,297	-,005
	Sig.	,072	,128	,061	,112	,002	,001	,112	,981

** Correlation significant at 0.01 level (2-tailed).

Subsequently, correlations between *initial fluency level* and amount of gains in rhythm metrics during the SA were performed. *Initial fluency level* was significantly correlated with gains in ΔV, r = .543, p < .01, varcoV, r = .426, p < .05, %V, r = .418, p < .05, and nPVIc, r = .433, p < .05 . The higher the initial fluency level, the narrower the gains in these measures.

7.3.2. SA outcomes in relation to High and Low *Initial Fluency Level* groups.

This subsection deals with *RQ.3.2. After the period abroad, will there be differences in fluency and rhythm between a high fluency and a low fluency group assembled by the initial fluency level? Moreover, will the two groups obtain different gains during the period abroad?* Participants were then divided into two groups according to their *initial fluency level* before going abroad. The high fluency group (*H-FL*) was formed by the 10 participants with the highest fluency scores on the Fluency Index at T2. Similarly, the 10 participants with the lowest fluency scores before going abroad were grouped together in the low fluency group (*L-FL*). The 10 participants in the middle were not included in the analyses. Differences in the FLIndx at T2 –representing the *initial fluency level*— between these two groups (*H-FL*, *Mdn* = .40; *L-FL*, *Mdn* = .17) were highly significant [W_s = 55.00, z = -3.78, p < .001, r = -0.85][33].

Perceived fluency ratings were significantly higher for the *H-FL* (*Mdn* =2.42) than for the *L-FL* (*Mdn* =2.89) group [W_s = 73.50, z = -2.38, p < .05, r = -0.53]. The same occurred with SR and MLoR. The *H-FL* group (SR, *Mdn* =184.45; MLoR, *Mdn* =11.42) obtained significantly higher results after the SA than the *L-FL* group (SR, *Mdn* =155.44; MLoR, *Mdn* =7.01) in both measures [W_s = 79, z = -1.96, p < .05, r = -0.44; W_s = 72, z = -2.49, p < .05, r = -0.55, respectively]. However, no statistically significant differences were found between the two groups for the rhythm metrics after the SA. Notwithstanding, there was a tendency towards more

[33] Non-parametric tests were used for the comparisons between the *H-FL* and the *L-FL* groups since the population was low (*N*=10 per group). The Wilcoxon's test was reported here, equivalent to the parametric T-Test. The median was reported instead of the mean for this non-parametric test, following Field's (2009) suggestion.

native-like scores for the *H-FL* group than the *L-FL* one in some of the rhythm metrics (%V and rPVIc, p = .08, and nPVIc, p = .06). **This suggests that the *initial fluency level* had an impact on L2 speech during the SA period, especially for temporal fluency, and to a lesser extent for rhythm.**

Furthermore, differences between the two groups were tested for amount of gains during the SA period. **The *L-FL* group was predicted to obtain more gains during the SA period than the *H-FL* group both for fluency and rhythm.** Regarding fluency ratings, listeners perceived *H-FL* group speech samples with less difference between T2 and T3 than the samples from the *L-FL* group (*H-FL*, *Mdn* = .18; *L-FL*, *Mdn* =.30), but the difference was non-significant. For the temporal fluency measures, the *L-FL* group obtained larger gains than the *H-FL* group in all measures (see table 7.20 below). These **differences were significant for most measures, indicating that the *L-FL* group obtained larger benefits in fluency than the *H-FL* group during the 3-month SA period.**

Table 7.20. Median scores for the *L-FL* and *H-FL* groups in temporal fluency measures. Wilcoxon's test comparing gains between *H-FL* and *L-FL* groups.

Variables	Group	Gains (Median)	Wilcoxon W	Z	Sig. (2-tailed)
SR-gains	1	35.46	71	-2.57	.010
	2	8.33			
MLoR-gains	1	2.83	95	-.756	.450
	2	2.12			
AR-gains	1	0.44	72.5	-2.46	.014
	2	0.19			
PhonRat-gains	1	5.92	78	-2.04	.041
	2	0.60			
PauseFreq-gains	1	5.63	73	-2.42	.016
	2	1.81			
Pause_i_Dur-gains	1	7.92	70	-2.65	.008
	2	1.74			
Pause_Dur-gains	1	5.92	78	-2.04	.041
	2	0.60			
DysRat-gains	1	1.06	100	-.38	.705
	2	1.25			

Note: 1 = *L-FL* group (*N*=10); 2 = *H-FL* group (*N* = 10)

A similar pattern was found for the rhythm metrics. However, even though gains were larger for the *L-FL* than the *H-FL* group, the differences between the two groups were only significant for ΔV (*H-FL*, *Mdn* = .15; *L-FL*, *Mdn* = 4.09) [W_s = 72, z = -2.49, $p < .05$, r = -0.55]. In addition, two other measures, %V (*H-FL*, *Mdn* = 4.68; *L-FL*, *Mdn* = 8.46) and nPVIc (*H-FL*, *Mdn* = 1.41; *L-FL*, *Mdn* = 9.66) approached significance [%V: W_s = 82, z = -1.74, p = .08, r = -0.39, and nPVIc: W_s = 81, z = -1.81, p = .07, r = -0.40] suggesting that *initial fluency level* is important to determine the amount of gains that learners will obtain from a period of study abroad. There is more room of improvement for those learners starting with a lower utterance fluency level.

Summarizing, results reported in this section related to *RQ.3* indicate that *initial fluency level* has an effect on fluency outcomes after the SA both in utterance and perceived fluency. Moreover, a negative relationship between starting fluency level and size of gains during the SA has also

been observed. On the other hand, the relation between *initial fluency level* and rhythm did not prove to be so straightforward. *Initial fluency level* exerted an influence on amount of gains in rhythm during the SA period in the predicted direction, whereas, even though some trends existed, no significant relationship was found between *initial fluency level* and rhythm outcomes after the SA.

The differences in fluency and rhythm after the SA between the groups of *H-FL* and *L-FL* participants were shown to significantly differentiate the two groups for fluency in most cases, but not for rhythm (for which only tendencies were observed). However, when the amount of gains during the 3-months SA were analysed, *initial fluency level* seemed to have an important effect. The *L-FL* group obtained larger gains in fluency and rhythm than the *H-FL* group (significant in the case of the temporal fluency measures and approaching significance for perceived fluency and most of the rhythm metrics).

8

Discussion

The results from the analyses of utterance fluency, perceived fluency, and rhythm undertaken in the current study were presented in the preceding chapter. The analyses were performed on the speech samples produced by 30 non-native speakers of English at three different points in time comprising two different learning contexts when carrying out a short semi-guided interview between peers. The effect of time, learning context and *initial fluency level* was examined, as well as the factors affecting perceived fluency ratings. Concurrently, relationships among the three aforementioned speech phenomena (utterance fluency, perceived fluency and rhythm) were explored. The present chapter builds on the preceding one and discusses the results in relation to the research questions. The discussion is organized dealing with each question and its subquestions in turn.

Research Question 1

> Will there be an effect of time and learning context on the fluency and rhythm of L2 speakers?

Results in the previous chapter confirm that there was an effect of time and learning context on both the fluency and rhythm of advanced L2 speakers of English. Significant differences were found for fluency and rhythm during the SA period, but not during the FI period. The discussion

on the subquestions of *RQ.1.* attempts to provide a better understanding of how and why these changes occurred.

The subsequent research questions that this *RQ.1.* raised provide a thorough account of the complex issue of progress in fluency made by learners in different contexts of learning, one of them being SA. We begin with *RQ.1.1.*

RQ.1.1.

To what extent will listeners perceive gains in participants' oral fluency after the study abroad period? That is, when asked to rate L2 speech samples, will listeners assign more target-like scores to T3 than to T2 speech samples?

It is worth reminding the circumstances in which this research subquestion was raised. Preliminary research (Valls-Ferrer, 2008), analyzing the development of utterance fluency through a set of temporal fluency measures had found that the participants' productions remained practically the same from T1 to T2, but were significantly different at T2 and T3, fluency at T3 being closer to NSs' standards than at T2. From those findings, it could not be inferred whether these differences were large enough for them to be perceptible to listeners. We were interested in finding out whether listeners could perceive any changes in the participants' utterance fluency since, at the end of the day, L2 learners need to achieve a certain level of fluency to be able to communicate effectively with their interlocutors. Listeners' fluency ratings provide an assessment of this perceptual dimension of non-native speech (Trofimovich & Baker, 2006; Rossiter, 2009). In order to gain new insights into this matter, a task in which the speech samples of the participants' oral production interview were rated by different types of listeners was conducted.

Before starting with the discussion, we briefly review the specificities of the speech samples used as stimuli in the perception experiment. They were collected from a group of advanced EFL learners, undergraduate students at a university in Barcelona. The samples consisted of 20-second excerpts taken from longer interviews on *University Life* performed in pairs at 3 data collection times, at the beginning of FI (T1), and before (T2) and after (T3) a 3-month SA period (total 90 samples). Five groups of non-expert listeners were asked to rate these speech samples for fluency at T2 and T3 using a 5-point Likert-scale. The scale went from 1 = extremely fluent to 5 = extremely dysfluent.

Hence, the first research question aimed to uncover whether this change in fluency, captured by the temporal fluency measures, during a 3-month SA period could also be perceived by a variety of untrained listeners. To our knowledge, only two other studies have approached this issue from a longitudinal perspective, examining speech samples from a group of participants at two different points in time (Lennon, 1990; Derwing, Thomson & Munro, 2006). Still, Lennon (1990) was an early small-scale study (*N*=4), and Derwing et al., (2006) analysed perceived fluency in speech samples produced by non-native immigrant L2 speakers with beginners level acquiring their L2 in a naturalistic setting, hence, a different population than the one here. In both studies listeners were native speakers of English.

Results from the T-Test comparing listeners' ratings before and after the SA show that there was a significant difference between ratings assigned to T2 and T3 speech samples by the groups of untrained listeners, and that ratings at T3 were higher than those at T2. This confirms the hypothesis that a 3-month SA period providing massive exposure to input and numerous opportunities for practice, produced a positive effect on participants' fluency to the extent that changes were detectable even by

non-expert listeners as opposed to listeners trained in phonetics or as exam evaluators.

These findings are in line with those of Lennon (1990) in which, by analyzing the responses of 9 NS listeners on the fluency of 4 participants' speech samples, he found that the judges perceived participants to be more fluent after a 6-month SA period (however, no statistical analyses were reported for these data). Contrary to our findings, in a study which also used listeners to judge the speech of a group of learners who went abroad, Freed (1995) found no differences in listeners' ratings before and after the SA. Even though her interest was not on examining listeners' ratings at the pre- and post-test, and accordingly, did not report any statistics regarding this comparison, she provided the raw scores for each participant. We took these scores and ran a T-Test to see whether her judges had perceived differences between pre- and pot-test productions. No significant differences were perceived by the listeners in Freed's study between pre- and post-SA scores on fluency [$t(14) = -1{,}78$, $p = {,}094$].

Since no other studies have been conducted on SA using listeners' ratings, we turn to studies conducted on naturalistic settings, which are to certain extent comparable. Derwing et al. (2006) found improvement in fluency over time for one of their groups under study (Slavic speakers), but not the other (Mandarin speakers). In that study, the Slavic speakers were reported to have a lower level at the beginning of the study than the other group, so they suggested that Slavic speakers might have had more room for improvement than the Mandarin speakers who might be at a ceiling. However, this hypothesis of the Mandarin speakers being at ceiling seems somehow unlikely since the mean fluency ratings listeners assigned was quite low. An alternative explanation is that group differences in amount of input and interaction explain group differences in fluency gains.

The results for fluency development in the present study to some extent contradict those found by Rossiter (2009). In her study, the raters did not perceive an improvement in fluency over the 10 weeks the study lasted, which is a period of time highly comparable to our 12 weeks. However, the type of participants, and the length of time they had been living in the TL country before the first data collection time differ from the present study. The author offers several tentative explanations for the lack of improvement, one of them being that 10 weeks is too short a period of time. The results in the present study show that it is possible for fluency to improve during a 12-week period, or even shorter (see Llanes, 2010). The other two explanations seem more plausible to us, either that participants had already attained a threshold in speaking fluency, or that they did not get enough exposure outside classroom to make a difference in perceived fluency. Rossiter also points to the fact that participants did not receive enough instruction on oral communication skills in class as one of the reasons why they might have not improved. However, our participants did not receive any instruction on oral communication skills and nevertheless they were perceived to have improved in fluency.

Both Derwing et al. (2006) and Freed (1995) suggested that one of the reasons why differences between speech samples before and after the SA could not be found in their studies might have been that the speakers' initial proficiency level was already advanced, and fluency might have needed more time to manifest an improvement even in a SA context. As indicated above, the present study provides evidence that this assertion cannot be generalizable since participants in the present study had an advanced proficiency level, and conversely, they were found to significantly improve between the two data collection times (cf. Lennon, 1990).

The divergence in findings between the studies mentioned and the present work could be attributed to either methodological differences (participants' profiles, the number of participants and listeners included in the analyses, the type of listeners) or to the treatment *per se* (the nature of the SA experience). For instance, participants in Derwing et al.'s (2006) study were immigrants who went to live to Canada, with a mean age of 34 years, a high-beginner English level, who mainly lived within their ethno-cultural groups, and whose L1 was typologically very different from English. Hence, not only were there important differences in participants' age, L1 background and L2 proficiency level, but the kind of immersion experience participants underwent was also very different. On the other hand, Freed's (1995) participants were more comparable, undergraduate students in a SA program who were spending a semester abroad. However, neither the L1 (English) nor the L2 (French) were the same as those in the present study. In addition, participants proficiency level in that study was lower (mainly intermediate). Moreover, the relatively low number of participants in the SA group (N=15) and the NS judges (N=6) suggest that findings on the perception of improvement are not generalizable. A third factor which might have affected the results in these studies, as well as the present study, is that the L1 inter-subject differences in fluency might have carried over to the L2. No L1 data is provided in any of the studies. However, this does not invalidate any of the results because fluency is investigated longitudinally for the same group of subjects.

In the present study, the perception of differences in fluency between pre- and post-SA speech samples by 5 groups of listeners differing in L1 background and experience was probably due to, a) the (large) size of changes undergone by participants L2 fluency during the 3-month experience in the TL community, and b) the nature of these changes (e.g. increased MLoR), which were prone to be perceived by listeners. The type

of experience undergone must have had some influence in the improvement of fluency, at least in part, but also the nature of the changes in the participants' speech. We discuss the type of experience here. The nature of the changes will be discussed in relation to *RQ.2*.

Firstly, there are several possible explanations for the results in our study. One possible explanation for this improvement might be the increased amount of out-of-class contact with the TL during the SA period. Brecht & Robinson (1993) and Meara (1994) suggest that time-on task and interaction were crucial for obtaining gains while abroad. In Brecht & Robinson's study, students who used the L2 more were the ones who obtained larger gains. In Meara's study, time spent speaking in the target language while abroad was a good predictor of linguistic gains. In research on naturalistic settings, Flege & Liu (2001) suggested that the conflicting evidence about the benefits of length of residence (LOR) (with some studies finding evidence for LOR as a predictor of pronunciation gains, and others not) were mainly caused because the quantity and quality of input had not been taken into account. Højen (2004) observed that LOR *per se* did not correlate with improvement in pronunciation but a composite measure of LOR and language use did. He concludes that "Language use affects the degree to which LOR correlates with improvement in L2 performance." (p. 108). Similarly, Juan-Garau & Pérez-Vidal (2007) found that gains in oral fluency were related to type of contact with the TL during the SA period. On the contrary, Segalowitz & Freed (2004) found only a weak indirect impact of out-of-class contact on oral gains.

In the present study, during the 3 months spent abroad, participants received a considerable higher amount of input in the TL than what they had normally received at home. Not only was the input larger in quantity but also larger in quality, since they took different roles, using different

registers and in different settings, from university lessons where the language was very formal, to informal gatherings where less formal language was used, short exchanges with staff in supermarkets, shops and cafeterias, etc. This language-rich context, not only in quantity but also in quality, might have played an important role in the development of L2 speech.

Furthermore, the input they received was supplemented by the output they produced, which at the same time was quantitatively larger and qualitatively more varied than that produced during the FI period at home. In naturalistic research, this has been operationalized as language use, and it has been one of the factors shown to influence L2 pronunciation (Flege, Munro & MacKay, 1995; Piske, MacKay & Flege, 2001; Flege, 2009). At least several basic exchanges occurred during the SA period for which interaction between participants and NSs (or NNSs with different degrees of proficiency) took place. These basic exchanges happened in everyday situations such as having to go to the bank to withdraw money, buying a mobile phone, requesting a students' travel pass, asking for directions, etc. (as reported in the participants diaries). Even though these tasks could be simulated in a language class, the difference here is that the situation was real; they really need to get to a certain place, or to obtain the money from the cashier. The attention paid to both the interlocutors' language, and one's own message to get it through is much more intense than any class simulations. Nevertheless, we are not saying that class role-plays simulating these everyday exchanges through meaningful tasks are not important, on the contrary. We deem these tasks necessary and appropriate to provide the learners with the practice needed to enhance their speaking skills in the real world. These in-class interactions are very useful to focus on meaning, in addition to the most commonly used form-focused activities mostly used in FI, since in the real world interactions the focus on meaning is so prevalent that much of the time no attentional resources can be devoted to

form-related aspects in the speech production process. During these interactions participants were requested not only to produce language, but also to accommodate to the interlocutor's demands. Participants' fluency might have been enhanced through this increased interaction in quantity and variety. It is clear that, as already pointed out by some researchers (Freed, 1995; Højen, 2004; Segalowitz & Freed, 2004; DeKeyser, 2007), not all learners engage in language interactions to the same extent, or take advantage of the benefits that the learning context offers in terms of language learning. In general, participants in this study, being translation majors, were highly motivated to improve their linguistic skills and their understanding of the TL culture (as confirmed by findings in Valls-Ferrer, 2008). Motivation has often been considered a factor which plays a role in shaping the amount of contact with the target language students who go abroad avail to (DuFon, 2006; Kinginger, 2008; Hernandez, 2010). The attitude and motivation towards language learning for participants in this study might have promoted their willingness to seek opportunities to interact in the TL, hence, increasing practice which might have promoted the development of L2 fluency.

Second, a further element which might have contributed to perceived fluency gains through increased and varied interactions is the development of a specific awareness leading participants to become more experienced L2 users and capable of avoiding troublesome language they do not feel comfortable with, hence, avoiding communication breakdowns which might have been otherwise present. According to Lennon (1990), a fluent speaker must be able to maintain his/her interlocutor's attention and interest in the conversation. In this case, extensive interactional practice might have provided participants with better skills at holding the interlocutor's attention after the SA than before.

A third explanation might be that the nature of the task used to elicit speech might have also contributed to the perceived increase in fluency after the SA period. The nature of the task has been claimed by several researchers (Ejzenberg, 2000; Cucchiarini et al., 2002; Towell, 2002) to play a very important role on fluency. As described in the methods chapter, the task consisted of a personal information exchange task, essentially monologic. The familiarity with the task topic –*University life*–, not being cognitively demanding, might have allowed participants to have greater control of the content, and rely on formulaic language acquired during the SA period. It might well be that the questions used favored a type of language participants had, often practised during the 3 months abroad.

A fourth reason for the progress found in fluency might be related to output practice. As it has already been mentioned, the SA context seems ideal for the development of fluency, especially due to the enhanced amount of practice in a rich variety of situations; at university, at home, in the street, in shops, with friends, with professors, etc. Output practice is important for the automatization of knowledge, since only through practice can learners reach the automatization stage (see Towell et al., 1996; DeKeyser, 1991, 2001, 2007). However, as pointed out by DeKeyser (2007) students do not always benefit from the opportunities for practice the SA context offers them. The decision on the degree of involvement in practicing the language while abroad depends exclusively on each participant.

Finally, a fifth explanation might be, although of a slightly different nature, as follows. At the beginning of this section we suggested that a possible reason for changes in fluency being noticed by listeners might have been the nature of the changes in fluency in participants' productions. Although we will deal with this in greater depth later in *RQ.2.*, we want to point to

some factors here. Knowing that, according to results from the utterance fluency measures, speech samples were produced faster at T3 than at T2, one of the possibilities for judges perceiving T3 productions to be more native like than those at T2 might have been processing difficulties. Munro & Derwing (2001) suggest that native listeners might prefer speech that is at a slightly lower rate than the common NSs speech, mainly due to L2 speech being generally more difficult to process. However, they also claimed that if speech were too slow, the ratings would also suffer because listeners would have more time to focus on other deviant linguistic features present in the speech signal such as phonological errors. In this sense, and according to their findings, speech produced at a slightly faster speed might be easier to process, since listeners would not have time to focus as much on phonological errors and, "very slow speech may also be difficult (or even tedious) to process because listeners are required to keep information in short-term memory for a longer period of time." (p. 466).

Participants in our study may have benefited from this increase in speech rate from T2 to T3 as far as ratings from listeners is concerned. Valls-Ferrer (2008) found that participants' productions at T3 generally approximated NSs' productions, however, never reaching NSs standards. Hence, as the T3 speech samples were only slightly slower than those of NSs performing the same task, this might have practically eliminated the processing difficulties present for listeners when faced with T2 speech samples, such as having to retain information in the short-term memory for a longer period of time, or the noticing of phonological errors, hence, listeners' ratings being more reliable.

In the end, it is probably the combination of the different factors converging in the SA context —increased amount of input, interaction, output and practice— and the variety of the language encountered, as well as the nature of the task used to elicit the speech samples, which can

explain the participants' progress in fluency and why they are perceived as significantly more fluent after the SA period than before.

We would like to round up the discussion of *RQ.1.1.* by emphasizing that the effect that the previous six months of FI might have had on participants' gains in fluency cannot be understimated. Brecht et al. (1993) found that grammar and reading achievement scores before the SA were significant predictors of gains in speaking proficiency, listening and reading after the stay abroad. They suggest that grammar practice during the first years of learning a language results in advances in speaking and listening skills at upper-intermediate and advanced levels. DeKeyser (2007) suggests that students' preparation at home is crucial to shape the quality of the students' learning experience abroad. In explicit reference to the improvement of fluency, he points out that "it is important that the students be ready for fluency improvement in the sense of automatization." (p.217). Pérez-Vidal, Juan-Garau & Mora (2011), inspired by DeKeyser's proposal, also suggest that during the SA learners might put into practice the explicit knowledge obtained in the FI context previous to the SA. In a multifaceted study with the same SALA[34] population, they found that oral accuracy and general written ability benefited from both, a FI period followed by a SA period. The authors state that:

> "… although benefits only became apparent abroad, it can be assumed that the AH and the SA contexts have both supported gains in oral accuracy and general written ability, albeit differentially- the former by providing learners with a solid grammatical and lexical foundation and academic written practice that would bear fruit later on, and the latter by allowing learners to participate in abundant complex communicative

[34] The present dissertation is ascribed to the *Study Abroad and Language Acquisition* (SALA) project, as introduced in the methods section. This is a six-year research project which seeks to address the acquisition of a wide range of English language skills by Catalan/Spanish bilinguals as a result of a SA period.

> situations that helped them in most cases both boost their oral performance," (p. 124-125)

During the first period under research, corresponding to the FI at home, students in the present study attended English lessons, which provided explicit declarative knowledge with the purpose of increasing their linguistic knowledge. During the FI period, and for some students (especially those who started with a higher language proficiency level), some of this knowledge had started to proceduralize, while for others it might have remained in declarative form. This knowledge, either in declarative or procedural form, accompanied students when they went on SA, where it could be further stretched. While abroad, students received large amounts of input in the TL and the opportunity to practice their linguistic skills. Even though different levels of interaction with the target culture may have occurred, in all cases exposure to language and practice would have been more intensive than at home. This SA context seems to offer the ideal opportunity for students to proceduralize the declarative knowledge acquired during the FI, or even automatize some of the knowledge they had already proceduralized at home. However, one needs to take into account DeKeyser's (2007) observation that:

> "They [students] try to do the best they can to communicate, i.e., transmit reasonably accurate and relevant information at an acceptable speed, and that leaves them no time to draw on their hard-to-access declarative knowledge. As a result, the automatization that eventually takes place as a result of many encounters of this kind leads to automatic use of formulas only, not automaticity of rule use." (p. 212).

This idea of students getting more fluent due to the use of formulas and pre-fabricated chunks has also been suggested by other researchers (Towell et al., 1996; Chambers, 1997; Juan-Garau & Pérez-Vidal, 2007). We now tackle *RQ.1.1.a.*

RQ.1.1.a

When grouped by L1 and experience, will any differences in fluency ratings between the two listeners' groups arise?

The results of our analyses show that the inter-rater reliabilities were very high for fluency ratings of the different groups of untrained listeners (Cronbach's α varied between 0.88 and 0.94). These high values were comparable to those in Cucchiarini et al. (2002) for phonetically trained raters and Derwing et al. (2004) for untrained raters, but much higher than Lennon (1990), Riggenbach (1991), Freed (1995) and Kormos & Dénes (2004). This suggests that fluency can be reliably assessed even by groups of untrained listeners, irrespective of their L1 (being NSs or NNSs of the TL) or experience with the TL under investigation. In fact, the kind of interlocutors one finds in everyday life.

As already mentioned in the theoretical background chapter, very few studies have used listeners to assess fluency development in SA research. Most of the studies using listeners have been undertaken to asses foreign accent, intelligibility and comprehensibility (Derwing et al., 1998; Munro & Derwing, 2001; Hanh, 2004; Trofimovich & Baker, 2006; Kang, 2010). In this sense, our study has been one of the few studies (cf. Lennon, 1990; Freed, 1995) to use listeners to directly assess the development in oral fluency of students who have spent a period abroad. Previous studies measuring the effects of this learning context on fluency have mostly been based on utterance fluency and used temporal fluency measures (Towell et al., 1996; Freed et al., 2004; Segalowitz et al., 2004; among others). As noted above, Freed (1995) is an exception since she uses both listeners and utterance fluency measures. However, we should bear in mind that, as explained earlier in this dissertation, her study does not focus as much on listeners' assessments to reflect a development in fluency from pre- to

post- SA results, but on the comparison of these ratings at the end of the semester between the SA group and the control group which remained AH. Neither is the aim of her study to focus on the type of listeners (all NSs) and its relation with the ratings provided.

Although not within the SA research, other SLA studies have focused on determining the influence of listeners' characteristics on the fluency ratings they provide. Derwing & Munro (2001) found that experience with Mandarin-accented speech made a difference in 'ideal rate' preferences for a group of native Mandarin listeners and a group of ESL listeners with mixed L1s. Mandarin listeners 'ideal rate' for both the Mandarin-accented speech and the native English speech was practically the same, while the mix-L1 group listeners preferred the Mandarin-accented English to be spoken at a slower rate than the normal English NSs' rate. They suggest that the Mandarin-accented English probably did not present an extra processing cost for the native Mandarin listeners but did so for the listeners with a variety of L1s.

The results of the present study were not in line with the above mentioned research since here no effect was found for experience on L2 fluency ratings between the two groups of listeners, those who had experience with the Catalan/Spanish accented English and those who had not. Nor were differences found between judgments made by English NSs and NNSs. Nevertheless, a tendency for NSs to assign higher fluency ratings than NNSs was found.

When differences in ratings assignments were observed for the five groups of listeners separately, statistically significant differences were found between the inexperienced NS students in the UK and the experienced NNS students in Spain, with NSs assigning higher fluency ratings than the NNSs. These two groups differed not only in their native language but also in the experience they had with Catalan/Spanish-accented speech.

One could argue that it was the L1 which made a difference, but one could also argue that it was the experience, or both. If we compare these results to results for the other group where members were experienced NSs of English (EFL teachers in Barcelona), hence the difference being nativeness only, non-significant differences were found between this group and the experienced NNS group regarding rates assigned. This seems to point to group differences in fluency ratings being due to listeners' degree of experience with Catalan/Spanish-accented speech rather than to L1. However, a group of non-experienced NNS listeners had not been recruited to confirm this possibility.

Studies in which NS and NNS listeners had been used to rate L2 speech samples found no differences in listener groups' ratings between listener groups (Kormos & Dénes, 2004). However, Rossiter (2009) found that ratings of novice NSs were higher than ratings of NNSs, but ratings from expert NS did not differ from those of the other two groups (novice NSs and NNSs). The author attributes this difference between novice NSs and NNSs ratings to a reminiscence of findings in studies on error gravity, for which NNSs evaluated L2 productions harsher than NSs. In fact, this seems to have been the general pattern followed by NNSs in our study as well.

Further analyses were undertaken to find out which group/s of listeners perceived the largest differences between speech samples before and after the SA. No differences were found between groups differing in L1s and experience with Catalan/Spanish-accented speech. However, one group of listeners seemed to be particularly sensitive to perceiving changes in fluency from the pre- to the post test. This group was made up of NS teachers in the UK. When compared to the other groups, they perceived significantly larger differences between the pre-SA and post-SA speech samples than the group of NS students in the UK, which hardly perceived

any differences between speech samples at the two data collection times. Thus, what makes these two groups different is not their L1, nor their experience with Catalan/Spanish-accented speech, but probably their overall experience with NNSs accented speech, irrespective of their L1. Even though the NS teachers reported not to be especially sensitive or to have been largely exposed to Catalan/Spanish-accented speech, their experience with accented speech from different L1 backgrounds probably made them more sensitive to the perception of differences in L2 speech.

Overall, no main differences were found in listeners' ratings regarding L1 and experience with Catalan/Spanish-accented speech, only a tendency for NSs listeners to assign higher ratings than NNSs. Yet, differences were found between the NSs group who had extensive contact with accented speech from different L1 backgrounds and the other groups. This group perceived the largest amount of gains in comparison to the other four groups of listeners. We now turn to *RQ.1.2*.

RQ.1.2.

> Will there be any change in rhythm before and after the period spent in each learning context as measured by rhythm metrics? If so, to what degree will changes represent significant gains in each of the learning contexts?

Results indicate that there was a change in the participants' L2 rhythm during the 15-month period under study. However, the rate at which these changes occurred differed depending on the period of time and learning context. Thus, no significant differences were found between rhythm scores at T1 (before FI) and at T2 (after FI): rhythm was stable during these first 6 months in which participants received 60 hours of syntactico-discursive formal instruction through English. No specific instruction on pronunciation skills was provided, nor speaking skills practiced during this

period of time. On the other hand, statistically significant differences were found between rhythm scores at T2 (before SA) and at T3 (after SA): participants' rhythm experienced a change during this 3-month SA period. The changes in rhythm were towards TL norms and the size of gains during the SA period were significantly larger than during the FI period at home. No specific instruction on pronunciation skills was given during this period of time either.

These results mirror those found in the previous section on fluency development, revealing that spending a relatively short period of time in the TL country is highly beneficial for the improvement of rhythm. To date, no other studies have investigated the development of rhythm over time in a SA context.

However, we can compare these findings to the learning of other suprasegmentals under similar conditions. For instance, Højen (2004) examined the L2 speech of Danish participants who undertook a study abroad in southern England. He found that improvement in foreign accent ratings did not occur at the segmental level, and suggested that changes might have occurred at the suprasegmental level instead. In addition, we can also compare it to studies examining the effect of experience on prosody which have been conducted in naturalistic settings. Trofimovich & Baker (2006) examined the role of experience on the acquisition of fluency and suprasegmentals by 30 Korean learners of English. They found that amount of experience influenced one of the suprasegmentals (stress timing) and suggested that the exact nature of the L2 experience might have determined the success and rate of L2 suprasegmental learning. The other few studies available have found that foreign accent ratings are affected by prosodic factors such as intonation (Munro & Derwing, 1995), overall prosody (Anderson-Hsieh et al., 1992) and pitch range and word stress (Kang, 2010).

Studies on vowel reduction, one of the basic elements of rhythm, have found that L2 learners generally failed to appropriately reduce unstressed vowels, both in quantity and duration, when no instruction was given (Flege & Bohn, 1989; Gómez-Lacabex, 2009). However, when instruction on vowel reduction was provided, participants showed limited improvement on vowel reduction perception and production (Gómez-Lacabex, 2009). Accordingly, studies on the influence of instruction on suprasegmentals have provided evidence for the benefit of instruction, although limited to some suprasegmentals such as rhythm, which is claimed to be very difficult to teach (Barry, 2007) and needs large amounts of practice to develop, especially in extemporaneous speech (Chela-Flores, 2004).

The results of the present study suggest that when formal instruction is not available but the type of linguistic experience participants' undergo is appropriate, with a large and varied amount of L2 input, interaction and output, as well as motivation from the part of learners to improve their L2 speech, an improvement in rhythm can occur. In such a case, the learning might have followed a bottom up pattern which, in the absence of rules, might have drawn on data and memory driven processes (Robinson, 1997). Participants in our study might have benefited from both the quality and variety of the input received (Flege & Liu, 2001; Højen, 2004), as well as the quality and large amount of practice in extemporaneous speech that they had the opportunity to be involved in (Chela-Flores, 2004; DeKeyser, 2007). In a review of research on L2 speech, Munro & Ocke-Schwen (2007) include quantity and quality of exposure to the L2 and its use over time within a list of variables that previous research on L2 speech learning has found to affect phonetic learning.

Flege & Liu (2001) conducted a study in which the effect of experience on the acquisition of L2 speech in a naturalistic context was tested. They

found that amount of experience, operationalized as LOR was not the unique factor which contributed to the improvement of L2 speech, the quantity and quality of the input received during that experience was as important as LOR for improvement to take place. They claimed that adult learners' L2 performance would improve over time only if the experience was as similar as possible to that of "input-rich L2 environment that is typical of children who immigrate to North America" (Flege & Liu, 2001:547). In this sense, our findings would suggest that the experience abroad the participants in our study underwent was of the type needed for the acquisition of L2 speech as identified by Flege & Liu (2001).

As discussed in *RQ.1.1.*, apart from the input-rich context that the SA offers, gains in rhythm might have been the result of large and varied practice in the production of L2 speech. Participants were highly motivated not only to learn the language but also to learn about the TL culture as future professionals who would be working with that language. This high integrative motivation might have led them to seek opportunities to interact with native speakers. Hernandez (2010) found that students integrative motivation was a predictor of students' amount of interaction with the L2, which at the same time was a predictor itself of improvement on the SOPI interview evaluating speaking abilities. Similarly, Isabelli-Garcia (2006) found that the social networks students developed while abroad and the motivation and attitudes towards the language and the culture were important factors affecting learners' linguistic gains.

Studies on L2 speech conducted in naturalistic settings have pointed out that the differences in findings among studies focusing on experience, measured as LOR in the TL community, might be due to the fact that most changes in speech seem to occur during the first year of communicative contact with the TL (Flege & Fletcher, 1992; Piske et al.,

2001; Højen, 2004), and many of these studies take place when participants have already been living in the TL community for some months or even years. The linguistic changes occurring during the first months abroad are largely under-researched, and it has been complemented by findings in SA research (Towell et al. 1996; Isabelli-Garcia, 2004; Segalowitz & Freed, 2004; Mora & Valls-Ferrer, *submitted*). Højen (2004) demonstrated that an average of 7 months of immersion was sufficient for learners to make considerable progress in L2 speech. Our study provides evidence that much learning takes place during the first 3-month SA period already. However, we don't have information about what would have happened had our participants stayed longer.

In addition, the speech samples analyzed in the present study corresponded to the first three months that the participants had been in contact with the TL environment. One of the conditions to take part in the study was not to have lived abroad prior to the study. As suggested by Flege & Fletcher (1992) and Højen (2004), most of the improvement in L2 speech takes place within the first year of extensive exposure to conversational L2. So, the large gains observed in fluency and rhythm during the SA might also be in part an effect of the benefits of the first extensive exposure to the L2.

The 3-month SA period in the TL country seems to have benefited participants L2 speech in terms of fluency, both at the temporal and phonological level. Participants were perceived as more fluent after the period abroad by several groups of listeners differing in L1 and experience with Catalan/Spanish-accented speech. Not only were they perceived more fluent, but also utterance fluency measures and rhythm metrics confirmed that their performance had significantly changed from the pre- to the post-SA tests. Gains during the SA period were significantly larger than gains during the FI period at home for both temporal measures and

rhythm metrics. These gains might have been the result of a combination of factors favored by the 3-month SA: a) increased amount of L2 input and highly varied L2 input, b) increased opportunities to interact with NSs and NNSs in the TL and varied type of interactions, c) increased amount of output produced by the participants in different real-life situations, d) increased amount of opportunities to practice the L2, and in a variety of situations, e) participants proficiency level, which might have been at an optimal moment to automatize the previously acquired linguistic knowledge, f) participants intrinsic motivation to learn the language and to be part of the TL culture. All these factors might have contributed to the gains in both fluency and rhythm during the 3-month SA period in the TL country.

Research Question 2

> To what extent are hypothesized changes in perceived fluency related to the temporal characteristics of non-native speakers' utterances? That is, to what extent will utterance fluency and rhythm be related to perceived fluency ratings?

In chapter 7, we reported results showing that both utterance fluency and rhythm scores were related to perceived fluency ratings. However, the strength of these relationships were of varying degrees depending on the comparisons; utterance fluency and perceived fluency were more strongly related to one another than rhythm and perceived fluency. This indicates that when listeners are exposed to L2 speech they mostly attend to utterance fluency components and rhythm makes less of a contribution to the perception of fluency than temporal fluency measures. In order to thoroughly examine the underlying changes in the temporal dimension of speech production which listeners perceive, we turn to the sub-questions of this second research question.

RQ.2.1.

Will utterance fluency scores be related to perceived fluency ratings and result as good predictors of perceived fluency?

As we have just mentioned, results provide strong evidence that perceived fluency ratings are related to utterance fluency scores. Yet, the degree of the relationships is not the same for all utterance fluency measures, with SR and MLoR presenting the strongest correlations and PauseFreq the weakest correlation. Interestingly, from the wide range of temporal fluency measures used, only one measure did not correlate with perceived fluency scores: the dysfluency ratio. This result was somehow predictable since no previous study had found any relationship between dysfluencies and listeners' judgments on fluency (Cucchiarini et al., 2002; Kormos & Dénes, 2004; Derwing et al., 2004). It seems that listeners do not consider elements such as repetitions, repairs or false-starts to be dysfluency markers, and hence, they do not pay much attention to them when listening to non-native speech. Similar results were reported by Lennon (1990). It might be that listeners consider these speech markers to be part of both native and non-native speech and its presence goes unnoticed, even though the occurrence is significantly higher in NNSs than in NSs' speech as evidenced in previous research.

Interestingly, the correlation coefficients between temporal fluency measures and perceived fluency ratings found in our study are comparable to those obtained by Cucchiarini et al. (2002) for spontaneous speech, and Derwing et al. (2009) for one of the groups (the Mandarin speakers' group) and the ones in Rossiter (2009), and much lower than those obtained by Kormos & Dénes (2004). In fact, values in the latter study are comparatively high, similar to those obtained for read speech in Chucchiarini et al. (2000).

The difference between the correlation coefficients among the studies is difficult to explain. One possible difference might be due to the cut-off point for pauses. Cucchiarini et al. (2002) and Kormos & Dénes (2004) used 0.20 sec. and 0.25 sec. respectively, whereas Derwing et al. (2009) and Rossiter (2009) used 0.40 sec., as we did in the present study. The correlation coefficients, however, are closer for Cucchiarini et al. (2002) and one of the groups in Derwing et al. (2009) and the present study than within studies using the same cut-off point. It may also be the case that the extremely high values found in Kormos & Dénes (2004) are due to the proficiency level of participants. From their report, it is unclear whether for the correlation analyses, productions from the two proficiency groups (advanced and low-intermediate) were pooled together. If this is so, this would mean that when speech samples were presented to listeners, the differences in fluency were likely to be very large, so listeners could easily differentiate between what they perceived as fluent or dysfluent. On the other hand, in the present study, participants had an advanced level, and even though there were differences in fluency between them, they were still quite homogeneous. Hence, their speech samples might have been more difficult to rate. From the studies mentioned here, the most comparable to the present study regarding the type of speech samples produced, the cut-off point for pauses and the type of listeners are Rossiter (2009) and Derwing et al. (2009), even though the participants were quite different (immigrant population in these two and SA students in the present study); the correlation coefficients among these three studies are quite comparable, with Rossiters' (2009) study and the Slavic group in Derwing et al. (2009) being a bit higher.

Nevertheless, the fluency ratings given may be quite different depending on the starting proficiency level of the population in each study, so that what raters attend to when providing the ratings would be very much related to the homogeneity of the proficiency level at the beginning of the

study. When asked to rate relatively fluent speech samples, listeners probably focus on aspects of the speech signal which they would have not focussed on had the speech samples been more dysfluent, since speech was not constantly interrupted by pauses. Hence, this should be taken into account when comparing findings from different studies. Yet, it might also be that there are other features in our participants' speech apart from merely temporal aspects of fluency which have influenced listeners' ratings. In the discussion of the findings relating to *RQ2.2* we relate L2 rhythm to listeners' perception of fluency.

Unexpectedly, we found that measures such as AR and PauseDur presented moderate to strong positive correlations with perceived fluency scores. Previous research in which these measures had been used had found no relationship between articulation rate and perceived fluency scores (Cucchiarini et al., 2002; Kormos & Dénes, 2004) and a weak relation with pause duration (Cucchiarini et al, 2002; Kormos & Dénes, 2004). As Chambers' (1997) puts it, "Becoming fluent therefore is not about speaking faster (articulation rate), but about pausing less often and pausing at the appropriate junctures in an utterance" (p. 540). Cucchiarini et al. (2002) suggested that the importance of articulation rate decreased in perceived fluency judgments when pauses were more frequent. In the present study, listeners seem to have paid attention to both pausing and speed of delivery. Given the fact that participants had an advanced level of English when tested, and that their fluency was judged to be relatively high, the number of pauses might have not been an obstacle for listeners, so they might have also focused on speed.

Regarding SR and MLoR, they were the two variables with the highest correlations with perceived fluency ratings, coming as no surprise since previous research had also emphasized their importance in fluency judgments (Lennon, 1990: Kormos & Dénes, 2004; Riggenbach, 1991 and

Freed, 1995 only for SR). Having argued that these two variables are confound measures including both features of speed and breakdown fluency, it seems reasonable for them to be highly correlated with perceptions of fluency. Most probably, listeners cannot disentangle the two aspects of fluency, focusing on speed or pausing phenomena at a time. Both aspects of temporal fluency are closely interconnected within the utterance; not only the speed at which speech is produced is important, but also the frequency with which this speech gets interrupted by pauses, the length of these pauses and the place where they occur.

Overall, the results in the present study suggest that listeners largely attended to similar features as the ones represented by the utterance fluency measures, especially those of speed and breakdown fluency. We now turn to examine which of these measures work as better predictors of perceived fluency.

Some of the utterance fluency measures analyzed in this study turned out to be good predictors of perceived fluency. MLoR was found to be the best single predictor of listener's perceived fluency, accounting for 42.2% of the variance in perceived fluency scores. Explained variance was not increased by any other variable when added to the model in a second place.

Towell et al. (1996) and Towell (2002) suggest that MLoR is a fluency measure which helps explain the changes in fluency. In his (2002) study Towell argues:

> "A priori an increased mean length of run indicates that the speaker is able to process more language within a single time span and could therefore indicate greater proceduralization of knowledge.
> […] if the subject manages to produce longer runs without at the same time pausing more, it can plausibly be argued that there must be some change in the way the language is stored in order for the speaker to be able to produce more of it in a single run between unfilled pauses." (p. 121).

In the present study, MLoR increased and pause length decreased, fulfilling the conditions for proceduralization formulated by Towell et al. (1996) and Towell (2002). This suggests that participants in this study were ready to proceduralize much of the knowledge they had gained previous to the SA. Once abroad, and probably due to the increased amount of practice (DeKeyser, 2007a) much of this knowledge was proceduralized as suggested by the increase in MLoR and the decrease in pause duration.

Cucchiarini et al. (2000) suggest that pause frequency is more relevant than pause duration for perceived fluency. However, one has to take into account that the cut-off point for pauses in their study was 0.20 sec. and, that it was based on read speech. Interestingly, a second study by the same authors, Cucchiarini et al. (2002), comparing read and spontaneous speech shed new light on this issue. For spontaneous speech, and the intermediate level group, the variable which explains the greatest amount of variance in perceived fluency was MLoR, with exactly the same amount of variance explained as in the present study (R^2= 0.422). They then add pause frequency, but the extra variance explained by this measure is quite marginal. AR is not included in the model since the correlation the authors find between perceived fluency and this measure are relatively low (contrary to what we find in our study). When discussing the importance of MLoR to perceived fluency ratings the authors argue that "the importance of this variable seems to suggest that pauses are tolerated, provided that sufficiently long uninterrupted stretches of speech are produced." (p. 2871).

Our results support Cucchiarini et al.'s (2002) findings of MLoR being the most important correlate between listeners' ratings and objective temporal measures for intermediate to advance learners. Most previous studies had suggested that SR was the most appropriate global measure to correlate with perceived fluency. However, most of these studies were conducted

on participants with a low or intermediate proficiency level (Freed, 1995; Derwing et al., 2004; Derwing et al., 2009). Cucchiarini et al. (2002) found that indeed for the beginner level, SR was the best predictor of fluency ratings, whereas for advanced level students, MLoR showed the strongest relation with fluency ratings.

As mentioned above, within SA research, only Lennon (1990) and Freed (1995) have examined perceived fluency to assess the development of fluency during a SA. Freed reported speech rate was the variable that better discriminated between the SA and AH groups in relation to listeners' ratings. Towell et al. (1996) and Raupach (1987) reported mean length of runs as the main factor contributing to improvement in fluency after spending time abroad to learn French. However, their claim was not based on the relationship between this objective measure and perceived fluency but on a thorough analysis of participants' speech.

Since MLoR is neither a single speed fluency nor breakdown fluency measure, but it includes both speed and pausing, it is not surprising that it reflects perceived fluency well. The nature of this measure and its high power of predictability can make it the most appropriate variable to use as a measure of fluency for studies with advanced learners. To date, most studies in SLA, when using a single variable of fluency to relate it to other linguistic phenomena, have mainly chosen SR. Based on previous research (Freed, 1995; Derwing et al., 2009; Cucchiarini et al., 2002) and on the results in the present study, the participants' proficiency level should always be taken into account when choosing the most adequate measure to capture changes in L2 fluency. From the present study it can be suggested that when participants have an advanced level of the language, MLoR is the measure that most accurately captures differences in (perceived) fluency.

In addition, the findings in this section suggest that the temporal aspects of utterance fluency, which can explain a high proportion of the variation in perceived fluency ratings, define fluency as mainly a temporal phenomenon. However, there may be other factors which listeners attend to when evaluating L2 speech since approximately half of the variance could not be explained by the temporal fluency measures. It might be that listeners were attending to things such as 'richness of vocabulary', 'accent', 'confidence in speaking', or 'rhythm of the phrases' (Freed, 1995:143) among other things. This will be further considered in following section dealing with *RQ.2.2.*

RQ.2.2.

How and to what extent will rhythm be related to perceived fluency ratings and contribute to explaining variance in perceived fluency scores?

The discussion on *RQ.2.1.* has provided a better understanding of how utterance fluency scores are related to perceived fluency ratings. In brief, listeners seem to largely attend to temporal and pausing phenomena when asked to rate non-native speech for fluency. However, apart from temporal fluency phenomena, other factors seem to play a role in perceived fluency. Phonological factors, such as consonant attraction and linking (Hieke, 1984), intonation (Wennerstrom, 2000) and sentence stress placement (Vanderplank, 1993) have been found to be indicators of English non-native fluency. In this study we build on this previous phonological research and expand it by introducing rhythm as a possible indicator of perceived fluency.

Moderate positive correlations were found between rhythm scores and perceived fluency ratings. The closer participants' rhythm scores were to TL norms, the more fluent they were perceived to be. This result suggests

that, to a certain extent, listeners' fluency judgments were affected by rhythm.

After providing fluency judgements, follow-up discussions on the aspects listeners themselves believed to have affected their fluency judgments have been promoted by some researchers (Lennon, 1990; Freed, 1995; Kormos & Dénes, 2004; Rossiter, 2009), and phonological and linguistic aspects (degree of perceived foreign accent, intelligibility and comprehensibility, as well as lexical and grammatical accuracy and complexity) have often been considered by the same listeners to have affected their fluency ratings. Specifically, Freed (1995) suggested that one of the aspects that listeners might attend to when asked to judge NNSs' speech could be 'rhythm of the phrase'. Results in the present study indicate that indeed rhythm played a role as a suprasegmental phonological feature affecting perceived fluency ratings.

Derwing, Munro, & Wiebe (1998) found that NS listeners' perceived improvement in fluency and comprehensibility in the speech of L2 learners who had received instruction in global pronunciation (based on prosodic factors: word and sentence stress, intonation and rhythm, projection, and speech rate), but did not perceive it in the speech of a group who received segmental pronunciation instruction. In a follow up study using the same population and treatment, Derwing & Rossiter (2003) found the same results for judgments of fluency and comprehensibility, with improvement over time for the global pronunciation group only. Then they analyzed which specific factors contributed to the impression of improvement. They were classified as either phonological, morphological, syntactic, semantic, filled pauses, repetitions and prosodic (stress, intonation, vowel length). When temporal measures where calculated, SR was found to increase significantly over time but between-group differences were non-significant. They also asked

judges for their overall impressions. The prosodic impressions were significantly different between groups, with the global pronunciation group receiving less negative impressions at T2 than T1 and the reverse pattern for the segmental group. So, the global pronunciation group spoke faster, with less filled pauses and the impressions on prosody improved over time. The researchers suggested that activities undertaken by the global pronunciation group might have promoted automaticity, which resulted in the improvement of fluency and "the release of attentional resources for other purposes." (Derwing & Rossiter, 2003:13). So, impressions of better prosody were related to higher fluency scores.

In our study, when listeners were grouped by L1 and experience, very similar correlations for all groups were found. So native and non-native listeners seemed not to differ substantially in the strength of the relationship between rhythm and perceived fluency. Interestingly, the group of English NSs teachers living in the UK was the one presenting higher correlations between rhythm scores and their perceived fluency ratings. However, the differences in the correlation coefficients between this group and all the other groups were not significant.

We are unaware of any previous studies investigating the relationship between perceived fluency and rhythm. In a study which assessed the relationship between rhythm and FA, White & Mattys (2007b) correlated the acoustic rhythm metrics with perception of foreign accent (FA) for three groups of speakers, English NSs, and Dutch and Spanish NNSs of English. Significant correlations between %V, VarcoV, and nPVIv and degree of perceived FA were found. Speech rate was inversely correlated with FA ratings. VarcoV was found to be the strongest single predictor of the FA ratings obtained. Only when speech rate was added did the two variables account for a greater amount variation in FA ratings, but no other rhythm metric did. The authors suggested that a close relationship

between rhythmic distinctions and prosodic timing processes exist. Results from the present study also confirm that there is a close relationship between some rhythm measures and timing processes.

In another study in which the relation between prosodic features (some of them closely related to rhythm) and FA was examined, Kang (2010) found that accent ratings where best predicted by pitch range, word stress measures and mean length of pauses. The less pitch variation, the more stressed syllables, and the longer the pauses in the speech samples the more heavily accented they were perceived to be.

The correlation analysis between utterance fluency measures and rhythm metrics provides an extra element for a better understanding of the relationship between rhythm and perceived fluency. That is, rhythm scores were significantly related to scores in three utterance fluency measures: MLoR, PauseFreq and Pause_i_Dur. The longer the runs produced without interruptions, the fewer pauses, and the shorter the pause durations within clauses, the more the scores on rhythm approximated target language norms. This seems an obvious relation since the only way a fairly regular rhythm patter can occur is with long enough intervals of speech without dysfluent pauses. Then, it comes as no surprise that those participants with more target-like rhythm were those with longer MLoRs, fewer pauses and of a shorter duration within clauses, and also those who obtained higher fluency ratings.

The results of the present study contribute to previous research on the importance of prosody to the perception of fluency (Derwing & Rossiter, 2003; Derwing et al., 2004; Wennerstrom, 2000) by adding to the understanding of how L2 rhythm is acquired, and its relatioship with perceived fluency.

On the basis of the findings in previous research regarding other aspects of fluency related to perceived fluency, apart from temporal and pausing phenomena, in the present study we did not expect that anything more than a small proportion of the variance in perceived fluency could be explained by rhythm variation.

Interestingly, we found that rhythm explained 27% of variance in perceived fluency. This means that listeners' judgements were partly atributable to the varying rhythmic patterns in participants' speech productions. None of the aforementioned studies including phonological factors in the study of fluency (Hieke, 1984; Vanderplank, 1993; Wennerstrom, 2000; Kormos & Dénes, 2004) provided information on the percentage of variance in perceived fluency that these factors could explain.

All in all, this study provides evidence that there exists a relationship between rhythm scores and perceived fluency ratings. The speech of those participants presenting a more stress-timed rhythm was judged as being more fluent than that of those with a more syllable-timed rhythm. Then, it is not only the ability to produce fast speech without dysfluent pauses which contributes to the perception of fluency, but also speakers' ability to use target-like rhythmic patterns in their speech. We now turn to the last research question.

Research question 3

> To what extent will *initial fluency level*, indexed as a temporal composite fluency measure, be a factor affecting SA outcomes on fluency and rhythm?

SA outcomes on fluency and rhythm were partially affected by the fluency level participants had prior to the SA (*initial fluency level*). Previous research

had observed an influence of initial proficiency level and cognitive abilities on several aspects of L2 speech (Brecht et al., 1993; Towell, 2002; Segalowitz & Freed, 2004). Furthermore, DeKeyser (2007) maintains that to maximally benefit from the experience abroad, students should have acquired "functional knowledge of the grammar that is assumed to be known at an intermediate level" (p.217) so that the process of proceduralization can be completed and make progress towards automatization. In this study we do not focus on proficiency in general, neither on the general grammatical knowledge, but on the initial fluency level. In the next two subquestions the impact of this factor on post-SA scores is discussed.

RQ.3.1.

How and to what extent will *initial fluency level* be related to fluency and rhythm after the SA? What's more, will it have any effect on amount of gains?

A general tendency for findings in SA research has been towards reporting higher gains for students who went abroad with a low proficiency level (Brecht et al, 1993; Freed, 1995; Lapkin et al., 1995; Segalotwitz & Freed, 2004). In the present study, this same pattern was found. Within an advanced level of competence in the language, those students who started with a relatively lower fluency level obtained larger gains in fluency during the period abroad. In a sense, this finding confirms what other researchers had found about the benefits of the SA for low to intermediate level students. However, participants in this study went abroad with an advanced level in the target language and still the findings were similar.

Brecht et al. (1993) reported that for reading and listening, the higher the initial level, the less the gain. A similar finding was supported by the speaking results on the OPI test. The authors suggested that this happened

as a function of the learning curve and the nature of the scale used, which might have not been able to capture effects at higher proficiency levels. However, they also found that learners at a certain threshold level (roughly corresponding to upper intermediate) obtained larger significant gains, compared to leaners with a lower proficiency level. Freed (1995) found that students who started with a lower initial fluency level obtained most gains during the semester abroad. Likewise, Towell (2002) demonstrated that learners who had started with a lower fluency level increased their fluency the most during study abroad, mainly by changing their pausing behaviour. On the other hand, learners who had started with a higher fluency level did not increase their temporal fluency as much, in spite of improving their pausing performance and their syntax becoming more complex.

Segalowitz & Freed (2004) suggested that initial oral performace levels were related to learners' fluency scores after the SA through their predisposition to make use of communicative opportunities. This was based on results showing that longest turn length correlated significantly with indexes of extracurricular contact with the L2 through reading and listening. They interpreted this correlation as an indication that short pre-test length of runs might have signalled an inability to process long messages, and this predisposed learners to avoid activities such as reading and writing, which involved the processing of long messages. In addition, they also found a relationship between pre-test levels in cognitive skills (processing speed and efficiency in lexical access) and gains in oral performance. They suggest that "oral gains may depend, to some extent, on the cognitive readiness to benefit from the learning opportunities available" (p.194).

We may conclude by stating that learners with a high fluency at the beginning of the stay can benefit more from the opportunities that the

context can provide them. On the other hand, learners with a very low initial fluency level at the beginning might not be capable of benefiting from such opportunities.

RQ.3.2.

After the period abroad, will there be differences in fluency and rhythm between a high fluency and a low fluency group assembled by the *initial fluency level*? Moreover, will the two groups obtain different gains during the period abroad?

Participants were assigned either to the *H-FL* group or the *L-FL* group depending on the score on the composite initial fluency index. Results revealed that after the 3-month SA period, participants in the *H-FL* group were assigned significantly higher fluency ratings than participants in the *L-FL* group. That is, the group which started with a high fluency level before going abroad was perceived to be significantly more fluent after the SA than the group which had started with a lower fluency level.

On the other hand, participants in the L-FL group were the ones who obtained more gains during the SA period as perceived by untrained listeners. Similar findings have been reported by several researchers (Freed, 1995; Towell, 2002), with low proficiency students obtaining larger gains during the SA than high proficiency students.

In his study, Towell (2002) observed that the initially low-scoring group had a much higher percentage of improvement than the high-scoring group. However, even though both groups increased their scores, the initially low-scoring group never caught up in absolute terms with the high group in terms of MLoR, PhonRat and PauseDur. By qualitatively analysing the data from an initially high and an initially low performer, he suggests that the differences are mainly in that the low performer has been

able to proceduralize basic syntactic patterns and the high performer gained knowledge of more complex syntax and learned how to use it in context.

In terms of rhythm, participants in the *H-FL* group showed a tendency to present more TL scores in rhythm than the *L-FL* group after the SA. That is, participants whose fluency level was high before SA presented a more native like rhythm than participants who went abroad with a lower fluency level. Regarding gains, the *L-FL* group obtained significantly larger gains in rhythm during the SA than the *H-FL* group, as instanced by three of the rhythm metrics.

These findings suggest that the pre-SA fluency level had an effect not only on the fluency level students reached during SA, but also on their rhythm. As participants in the *H-FL* group were producing L2 speech with a more TL rhythm, as demonstrated by certain rhythm metrics (%V, nPVIc), we suggest that rhythm metrics should be included as a component of fluency in English at advanced levels. That is, when the participants under assessment are advanced learners, a measure of rhythm can help to better discriminate between the most fluent and the less fluent participants. It would be interesting to explore the possibility of there being a threshold in fluency at which speakers must arrive before improvement in rhythm can occur. As we saw above, native-like rhythm is very difficult to attain when the speech is continuously broken by dysfluent pauses.

We now turn to the conclusion of our study in the final chapter.

9

Conclusions

The present study has tackled four major topics in SLA research today: learning context (or L2 experience), L2 fluency development (utterance and perceived fluency), L2 rhythm performance and *initial fluency level*. The four issues have been interrelated to provide a detailed account of how two different aspects of L2 speech, fluency and rhythm develop during a 15-month period containing a 6-month FI period and a 3-month SA period. The effect of the two learning contexts shapes the development of these two aspects of L2 speech, as well as the fluency level that participants had before the SA period.

This study aimed at finding out how two different learning contexts had an effect on the acquisition of L2 fluency and rhythm. Two senses of fluency were examined: utterance fluency and perceived fluency. The relationship between these two fluency domains was assessed, as well as the relationship between rhythm and perceived fluency. The initial fluency level was assessed as a factor influencing SA outcomes in fluency and rhythm.

The findings in this dissertation contribute to the understanding of four central areas in the study of SLA by providing empirical evidence on each of them. In the area of linguistic progress in a specific learning context, SA, this study has found that a 3-month SA period brings about significant benefits in the learners' oral fluency.

On methodological grounds, MLoR, a measure which encompases both speed and breakdown fluency, and which has been claimed to capture changes in the proceduralization process (Towell et al., 1996), has proven to be the most robust measure in relation to listeners' perceptions of the speech of a population of advanced L2 learners.

In the area of L2 fluency development, gains in utterance fluency were found to take place during SA and such gains were found to be perceivable by various groups of listeners. The non-experienced native speakers listener group was found to be more sensitive to perceived changes in fluency than the other listener groups. These are comparable to the kind of interlocutors that learners are more likely to find when going abroad, and in their future life.

As regards rhythm performance, our findings support two claims: 1) rhythm can be understood as part of what listeners' attend to when judging fluency and constitutes an important component that needs to be included in the assessment of L2 fluency, and 2) in relation to linguistic development, it has been proven that L2 rhythm develops over time as a function of learning context. The analysis of rhythm in the present study constitutes ground-breaking work in the analysis of the phonological aspects of L2 fluency and has important implications for the assessment of L2 fluency.

Finally, in relation to participants' variables, with the *initial fluency level* variable we have added to the ongoing debate on the existence of a threshold level for certain language skills (fluency) and linguistic phenomena (rhythm) by providing empirical support in an area where empirical findings are still scarce. We have been able to confirm that within a range of advanced learners, those with a lower *initial fluency level* improved the most, without ever reaching NSs' standards.

There are two practical implications that stand out in considering these findings. Firstly, our learners seem to have improved both quantitatively and qualitatively in a way that enables an untrained interlocutor, that is, the lay person with whom they would have been in contact while abroad, and in the future, to understand what they are saying. Having said this, the second implication is more directed towards the formal learning in that the students we have analysed lack specific instruction on both fluency and prosodical aspects of the language. When abroad, they would have probably benefited more from the interactions with the target language community, had they been instructed on how to interact with the TL community (Dudley, 2007) or had they been required by the home institution to accomplish a number of activities which would have guaranteed a regular contact with the TL community.

We would like to conclude with a quote from Ortega (2009): "what matters in the linguistic environment is not simply 'what's out there' physically or even socially surrounding learners, but rather what learners make of it, how they process (or not) the linguistic data and how they live and experience that environment." (p. 78).

Further Research

In light of the findings and limitations found while carrying out this dissertation, several considerations follow that will need to be taken into account in further research.

Firstly, the study here is limited in the sense that it is restricted to the L1 Spanish – L2 English pair. In the area of rhythm, it seems plausible to think that the importance of reaching a more TL rhythm is especially relevant when we have languages which stand at opposite ends of the syllable-timed vs. stress-timed rhythm continuum. Moreover, the

relationship between fluency and rhythm seems to be important in this respect, speakers of a syllable-timed L1 aiming at mastering a L2 stress-timed language. The relationship between fluency and rhythm would probably not be the same depending on whether the language is stress-timed or syllable-timed. As suggested by Derwing et al. (2009) it might be that some language pairings are closer than others, so the inclusion of different L1s is necessary for the full understanding of aspects of L2 development such as fluency and prosody.

Secondly, it would be interesting to analyze L1 productions of these same learners for fluency on the same task in order to find out which aspects of the L2 fluency are attributable to the command of the second language and which belong to L1 fluency characteristics such as the personal speaking style of the learner (as recommended by Segalowitz, 2010 and de Jong et al., *in press*).

Thirdly, we have found that a 3-month SA period has been enough time for the L2 speech of learners to develop significantly. The size of gains obtained by learners, especially in fluency, is already very high. One might consider these results very exceptional. However, when interpreted in the light of research on L2 speech, it is in fact not so. Studies on the effect of length of residence on other aspects of L2 speech learning have shown that the first year of exposure to authentic conversational speech is probably the period when most gains are found, compared to the much lower rate in benefits in subsequent years. It would be interesting to expand this analysis to an analogous group of learners who had spent a longer period abroad to see if fluency and rhythm keep changing at the same rate, or if a certain time threshold is reached at which the rate of improvement stagnates.

Finally, future research might benefit from a comprehensive and detailed account of amount of L1 use while abroad. Nowadays, new technologies (e.g. Skype, Facebook, Twitter, etc.) are widely available to larger segments of the population than years ago, which allow learners to keep a closer contact with the L1 communtiy at home. It would be interesting to compare if L1 use while abroad makes a difference in linguistic gains. New studies should control this variable if reliable comparisons between learning contexts are to be made. The impact of these new technologies should not be underestimated.

References

Abercrombie, D. (1967). *Elements of General Phonetics*, Edinburgh University Press.

Aguilar, L., Gimenez, J.A., Machuca, M., Marin, R., & Riera, M. (1997). Catalan vowel duration, In *EUROSPEECH*, 771-774.

Alfano, I. (2009). *La percepción del acento léxico en Italiano y español como lenguas extranjeras.* (Proyecto de investigación de tercer ciclo). Universitat Autònoma de Barcelona.

Aliaga-García, C. & Mora, J. C. (2009). Assessing the effects of phonetic training on L2 sound perception and production. In M. A. Watkins, A. S. Rauber, y B. O. Baptista (eds) *Recent Research in Second Language Phonetics/Phonology: Perception and Production.* Newcastle upon Tyne: Cambridge Scholars Publishing. 2-31.

Anderson, J. R. (1983). A Spreading Activation Theory of Memory. *Journal of Verbal Learning and Verbal Behavior, 22.*

Anderson, J. R. (1993). *Rules of the mind.* Hillsdale, NJ: Lawrence Erlbaum.

Anderson-Hsieh, J., Johnson, R., & Koehler, K. (1992). The Relationship Between Native Speaker Judgment of Nonnative Pronunciation and Deviance in Segmentals, Prosody, and Syllable Structure. *Language Learning, 42*(4), 529-555.

Barry, W. J. (2007). Rhythm as an L2 problem: How prosodic is it? In Jürgen Trouvain and Ulrike Gut (eds.), *Non-Native Prosody. Phonetic Description and Teaching Practice.* Berlin, New York: Mouton de Gruyter, 97-120

Barry, W. J., Andreeva, B., Russo, M., Dimitrova, S., & Kostadinova, T. (2003). Do rhythm measures tell us anything about language type? *Proceedings 15th Int. Congr. Phonet. Sci.*, Barcelona 2003, pp. 2693–2696.

Batstone, R. (2002). Contexts of engagement: A discourse perspective on "intake" and "pushed output." *System* 30, 1-14.

Beattie, J. (2008). Evaluating the acquisition of English listening comprehension skills in the framework of the SALA project. *Monografías de Aula Abierta32, Estudios de desarrollo del lenguaje y educación*. Oviedo: Universidad de Oviedo, 301-309.

Bialystok, E. (1978). A Theoretical Model of Second Language Learning. *Language Learning.* 28(1). 69-83.

Boersma, P., & Weenink, D. (2007). Praat (Version 4.5.25) [Software]. Latest version available for download from www.praat.org.

Brecht, R. D., & Robinson, J.L. (1995). On the value of formal instruction in study abroad: Student reactions in context. In B.F. Freed (Ed.), *Second Language Acquisition in a Study Abroad Context*. Amsterdam and Philadephia: John Benjamins. 317-334.

Brecht, R., Davidson, D., & Ginsberg, R. (1993). *Predictors of foreign language gain during study abroad.* Washington D.C.: National Foreign Language Center.

Brecht, R., Davidson, D., & Ginsberg, R. (1995). Predictors of Foreign Language Gain during Study Abroad. In B. F. Freed (Ed.), *Second Language Acquisition in a Study Abroad Context*. Amsterdam and Philadephia: John Benjamins. pag.

Brown, J. (2003). Promoting fluency in EFL classrooms. Paper presented at the JALT Pan-SIG Conference Proceedings.

Bygate, M. (2006). Areas of research that influence L2 speaking instruction. In E. Usó-Juan, & A. Martínez-Flor (Ed.), *Current Trends in the Development and Teaching of the four Language Skills*. Berlin, New York: Mouton de Gruyter.

Byram, M., & Feng, A. (2006). *Living and Studying Abroad: Research and Practice.* Clevedon: Multilingual Matters.

Cantín M. & Ríos A. (1991). Análisis experimental del ritmo de la lengua catalana, Anuario del Seminario de Filología Vasca Julio de Urquijo / *International Journal of Basque Linguistics and Philology (ASJU)*, 25, 487-514.

Carter, P. M. (2005a). Prosodic variation in SLA: Rhythm in an urban North Carolina Hispanic community. *Penn Working Papers in Linguistics, 11*, 59-71.

Carter, P. M. (2005b). Quantifying Rhythmic Differences between Spanish, English, and Hispanic English. *Amsterdam Studies in the Theory and History of Linguistic Science Series 4, 272*, 63-76.

Cenoz, J., & Perales, J. (2000). Las variedades contextuales y el efecto de la instrucción en la adquisición de segundas lenguas. In C. Muñoz (Ed.), *Segundas lenguas. Adquisición en el aula.* Barcelona: Ariel Lingüística.

Chambers, F. (1997). What do we mean by fluency? *System,* 25, 535–544.

Chela-Flores, B. (2001). Pronunciation and language learning: An integrative approach. *IRAL*, 39: 85–101

Chela-Flores, B. (2004). Optimizing the Teaching of English Suprasegmentals. *BELLS: Barcelona English Language and Literature Studies, 12*.

Cheng, P. W. (1985). Restructuring Versus Automaticity: Alternative Accounts of Skill Acquisition. *Psychological Review,* Vol. 92, No.3, 414-423.

Chomsky, N. (1986) *Knowledge of language: Its nature, origin, and use*. New York; Prague.

Chun, D. M. (2002). *Discourse Intonation in L2: from theory and research to practice*.: John Benjamins Publishing Company.

Churchill, E. DuFON, M. A. (2006). Evolving Threads in Study Abroad Research. In E. Churchill, M.A. DuFon (Ed.), *Language Learners in Study Abroad Contexts* (pp. 1-27). Clevedon: Multilingual Matters.

Coleman, J. A. (1997). Residence abroad within language study. *Language Teaching*, 30, 1-20.

Coleman, J. A. (1998). Language Learning and Study Abroad: the European Perspective. *Frontiers, 4*, 167-203.

Collentine, J. (2004). The effects of learning contexts on morphosyntactic and lexical development. *Studies in Second Language Acquisition, 26*, 227-248.

Collentine, J. (2009). Study abroad research: Findings, implications and future directions. In C. Doughty, & M. Long (Eds.), Handbook of language teaching (pp. 218-233). Malden, MA: Blackwell.

Collentine, J., & Freed, B. F. (2004). Learning contexts and its effects on second language acquisition: Introduction. *Studies in Second Language Acquisition*, 26, 153–171.

Cook, V. J. (1999). Going beyond the native speaker in language teaching, *TESOL Quarterly* 33, 2, 185-209.

Costa, A., Miozzo, M., & Caramazza, A. (1999). Lexical selection in bilinguals: Do words in the bilingual's two lexicons compete for selection? *Journal of Memory and Language, 41,* 365-397.

Crookes, G. (1991). Second Language Speech Production Research: A Methodologically Oriented Review. *Studies in Secon Language Acquisition, 13*(2), 113-132.

Crystal, D. (1987). *The Cambridge encyclopedia of language*. Cambridge: Cambridge University Press.

Cucchiarini, C., Strik, H., & Boves, L. (2000). Quantitative assessment of second language learners' fluency by means of automatic speech recognition technology. *Journal of the Acoustical Society of America, 107*(2), 989-999.

Cucchiarini, C., Strik, H., & Boves, L. (2002). Quantitative assessment of second language learners' fluency: Comparisons between read and spontaneous speech. *Journal of the Acoustical Society of America, 111*(6), 2862-2873.

Dauer, R. M. (1983). Stress-timing and Syllable-timing Reanalysed. *Journal of Phonetics*, 11, 51-62.

De Bot, K. (1992). A bilingual production model: Levelt's "Speaking" model adapted. *Applied Linguistics, 13*(1), 1-24.

De Bot, K. (1996). The psycholinguistics of the output hypothesis. *Language Learning, 46*(3), 529-555.

de Jong, N. H. Wempe, T. (2007). Automatic measurement of speech rate in spoken Dutch. *ACLC Working papers, 2*(2), 49-58.

de Jong, N. H., Steinel, M.P., Florijn, A., Schoonen, R., & Hulstijn, J.H. (2007). The Effects of Tasks Complexity on Fluency and Functional Adequacy of Speaking Performance. In S. Van Daele, Housen, A.., Pierrard, M., Kuiken, F. and Vedder, I. (Ed.), *Complexity, Accuracy and Fluency in second language Use, Learning and Teaching.* (pp. 53-63). Brussels: Koninklijke Vlaamse Academie van België voor Wetenschappen en Kunsten.

de Jong, N. H., Steinel, M. P, Florijn, A., Schoonen, R., & Hulstijn, J.H. (*in press*). Facets of speaking proficiency. *Studies in Second Language Acquisition.*

Dechert, H. W. (1980). Pauses and intonation as indicators of verbal planning in second-language speech productions: Two examples from a case study. In H. W. Dechert & M. Raupach (Eds.), *Temporal variables in speech* (pp. 271-285). The Hague: Mouton.

Dechert, H. W., & Raupach, M. (Ed.). (1980). *Temporal Variables in Speech.* Mouton: The Hague.

De Jong, N. (2008). Second language learning of grammar: Output matters too. In T. P. M. Young-Scholten (Ed.), *Input Matters in SLA* (pp. 95-115). Bristol, UK: Multilingual Matters.

DeKeyser, R. (1991). Foreign language development during a semester abroad. In B. F. Freed (Ed.), *Foreign Language Acquisition: Research and the Classroom* (pp. 104-119). Lexington, MA: D. C. Heath.

DeKeyser, R. (1997). Beyond explicit rule learning: Automatizing second language morpho-syntax. *Studies in Second Language Acquisition, 19*, 195-221.

DeKeyser, R. (2001). Automaticity and automatization. In P. Robinson (Ed.), *Cognition and second language instruction*. Cambridge: Cambridge University Press.

DeKeyser, R. (Ed.). (2007a). *Practicing in a second language: Perspectives from applied linguistics and cognitive psychology*. New York: Cambridge University Press.

DeKeyser, R. (2007b). Situating the concept of practice. In R. DeKeyser (Ed.), *Practicing in a second language: Perspectives from applied linguistics and cognitive psychology* (pp. 1-18). New York: Cambridge University Press.

DeKeyser, R. (2007c). Study abroad as foreign language practice. In R. DeKeyser (Ed.), *Practicing in a second language: Perspectives from applied linguistics and cognitive psychology* (pp. 208-226). New York: Cambridge University Press.

DeKeyser, R. (2007d). The future of practice. In R. DeKeyser (Ed.), *Practicing in a second language: Perspectives from applied linguistics and cognitive psychology* (pp. 287-304). New York: Cambridge University Press.

DeKeyser, R. & Sokalski, K.J. (1996). The differential role of comprehension and productive practice. *Language Learning, 46*(4), 613-642.

DeKeyser, R. (2010). Monitoring processes in Spanish as a second language during a study abroad program. Foreign Language Annals, 43 (1), 80-92.

Dell, G. S. (1986). A spreading-activation model of retrieval in sentence production. *Psychological Review*, 93, 231-241.

Dellwo, V. (2006). *Rhythm and Speech Rate: A Variation Coeficient for ΔC.* Paper presented at the 38th Linguistics Colloquium, Piliscsaba.

Dellwo, V. (2010). Choosing the right rate normalization method for measurements of speech rhythm. In S. Schmidt, Schwarzenbach, M. & Studer, D. (Ed.), *La dimensione temporale del parlato.* (pp. 13-32). Torriana: EDK editore.

Dellwo, V., Wagner, P. (2003). *Relations between language rhythm and speech rate.* Paper presented at the International Congress of Phonetics Science, Barcelona.

Dellwo, V., Fourcin, A., & Abberton, E. . (2007). *Rhythmical classification of languages based on voice parameters.* Paper presented at the ICPhS XVI, Saarbrücken.

Derwing, T. M., & Munro, M. J. (1997). Accent, Intelligibility, and Comprehensibility. *Studies in Secon Language Acquisition, 20*, 1-16.

Derwing, T. M., Munro, M. J. & Wiebe, G. E. (1998) Evidence in favor of a broad framework for pronunciation instruction. *Language Learning, 48*, 393-410.

Derwing, T. M., & Munro, M.J. (2001). What speaking rates do nonnative listeners prefer? *Applied Linguistics,* 22(3), 324–37.

Derwing, T. M. & Rossiter, M. J. (2003). The effects of pronunciation instruction on the accuracy, fluency and complexity of L2 accented speech. *Applied Language Learning, 13*, 1-18.

Derwing, T. M., Rossiter, M. J., Munro, M. J., & Thomson, R. I. (2004). Second Language Fluency: Judgments on different tasks. *Language Learning, 54*, 655-679.

Derwing, T. M., Thomson, R. I., & Munro, M. J. (2006). English pronunciation and fl uency development in Mandarin and Slavic speakers. System, 34, 183–193 .

Derwing, T. M., Munro, M. J., & Thomson, R. I. (2008). A longitudinal study of ESL learners' fluency and comprehensibility development. *Applied Linguistics,* 29, 359–380 .

Derwing, T. M., Munro, M. M., Thomson, R. I., & Rossiter, M. J. (2009). The relationship between L1 fluency and L2 fluency development. *Studies in Second Language Acquisition, 31(4),* 533-557.

Dewey, D. P. (2004). A comparison of reading development by learners of Japanese in intensive domestic immersion and study abroad contexts. *Studies in Second Language Acquisition, 26*(2), 301-327.

Diaz-Campos, M. (2004). Context of learning in the acquisition of Spanish second language phonology. *Studies in Second Language Acquisition*, 26(2), 249-273.

Dörnyei, Z. (2007). *Research methods in applied linguistics : quantitative, qualitative, and mixed methodologies.* Oxford: Oxford University Press.

Dörnyei, Z., & Scott, M.L. (1997). Communication Strategies in a Second Language: Definitions and Taxonomies. *Language Learning, 47*(1), 173-210.

Doughty, C. (2001). Cognitive underpinnings of focus on form. In Robinson, P. (ed.), *Cognition and second language instruction* (pp. 206-257). New York: Cambridge University Press.

Doughty, C., & Long, M. (Ed.). (2003). *Handbook of second language acquisition*. Oxford: Blackwell.

DuFon, M., & Churchill, E. (2006). *Language learners in study abroad contexts*. Clevedon, UK: Multilingual Matters.

Ejzenberg, R. (1992). *Understanding nonnative oral fluency: The role of task structure and discourse variability*. State University of New York, Albany.

Ejzenberg, R. (2000). The juggling act of oral fluency: a psycho-sociolinguistic metaphor. In H. Riggenbach (Ed.), *Perspectives on Fluency* (pp. 287-313). Ann Arbor: The University of Michigan Press.

Ellis, N. C. (2003). Constructions, chunking, and connectionism: The emergence of second language structure. In C. J. Doughty & M. H. Long (Eds.) *The handbook of second language acquisition* (pp. 63-103). Malden, MA: Blackwell.

Ellis, R. (1993). Naturally simplified input, comprehension and second language acquisition. In M. Tickoo (Ed.), *Simplification theory and application*. Singapore: SEAMEO Regional Language Centre.

Ellis, R. (2000). *Second Language Acquisition*. Oxford University Press.

Ellis, R. (2003). Tasks, production, and language acquisition. In R. Ellis (Ed.), *Task-based language learning and teaching* (pp. 103-139): Oxford University Press.

Ellis, R. (2008). *The study of second language acquisition*. Oxford: Oxford University Press.

Farnetani, E. & Recasens, D. (2010). Coarticulation and connected speech processes, Hardcastle, W.J., Laver, J. & Gibbon, F.E. (Eds.) *The Handbook of Phonetic Sciences*, Wiley-Blackwell, Oxford, 316-352.

Field, J. (2005). Intelligibility and the listener: The role of lexical stress. *TESOL Quarterly*, 39, 399-423.

Field, A. (2009). *Discovering Statistics using SPSS*. London: Sage.

Fillmore, C. (1979). On fluency. In C. Fillmore, D. Kempler, & W. Wang (Eds.), *Individual Differences in Language Ability and Language Behavior* (pp. 85–101). New York: Academic Press.

Flege, J. E. (2009). Give input a chance! In T. Piske and M. Young-Scholten (Eds.) *Input Matters in SLA*. Bristol: Multilingual Matters, Pp. 175-190.

Flege, J. E., & Bohn, O.S. (1989). An instrumental study of vowel reduction and stress placement in Spanish-accented English. *Studies in Second Language Acquisition* 11, 35-62.

Flege, J. E., & Fletcher, K. L. (1992). Talker and Listener Effects on Degree of Perceived Foreign Accent. *Journal of the Acoustical Society of America, 91*(1), 370-389.

Flege, J. E., Munro, M., & MacKay, I. (1995). Factors affecting strength of perceived foreign accent in a second language. *Journal of the Acoustical Society of America*, *97*, 3125-3134.

Forster, K. I., & Forster, J. C. (2003). DmDx: A window display program with millisecond accuracy. *Behavior Research Methods, Instruments & Computers, 35*(1), 116-124.

Foster, P. & Skehan, P. (1996). The influence of planning on performance in task-based learning. *Studies in Second language Acquisition*, 18, 299-324.

Foster, P. Tonkyn, A., & Wigglesworth, G. (2000). Measuring spoken language: a unit for all reasons. *Applied Linguistics, 21*, 354-375.

Foster, P. & Tavakoli, P. (2009). Native speakers and task performance: comparing task effects on complexity, fluency and lexical diversity. *Language Learning*, 59 (4), 866-896.

Fourcin, A., & Dellwo, V. (2009). *Rhythmic classification of languages based on voice timing. Working Papers*. London: Speech, Hearing and Phonetic Sciences UCL.

Fourakis, M. (1991). Tempo, Stress and Vowel Reduction in American English. *Journal of the Acoustical Society of America* 90:1816-1827.

Freed, B. F. (Ed.). (1995a). *Second Language Acquisition in a Study Abroad Context*. Amsterdam and Philadelphia: John Benjamins.

Freed, B. F. (1995b). What makes us think that students who study abroad become fluent? In B. F. Freed (Ed.), *Second Language Acquisition in a Study Abroad Context* (pp. 123-148). Amsterdam: Benjamins.

Freed, B. F., So, S., & Lazar, N.A. (2003). Language learning abroad: how do gains in written fluency compare with oral fluency in French as a second language? *ADFL Bulletin, 43*(3), 34-40.

Freed, B.F., Dewey, D.P., Segalowitz, N., & Halter, R. (2004). The Language Contact Profile. *Studies in Second Language Acquisition, 26*, 349-356.

Freed, B. F., Segalowitz, N., and Dewey, D. P. (2004). Context of learning and second language fluency in French. *Studies in Second Language Acquisition, 26*, 275-301.

Fullana, N., & MacKay, I.R.A. (2010). L2 fluency in the production of English extemporaneous speech by Catalan/Spanish bilinguals. *Proceedings of the 6th International Symposium on the Acquisition of Second Language Speech*, Poznan. 125-130.

Gass, S., & Mackey, A. (2006). Input, Interaction and Output. *AILA Review, 19*, 3-17.

Gavaldà-Ferré, N. (2007). *Vowel reduction and Catalan speech rhythm.* Unpublished MA Thesis. University College London.

Gilabert, R. (2005). Task Complexity and L2 Narrative Oral Production (Doctoral Dissertation).

Ginsberg, R. B., & Miller, L. (2000). What Do They Do?: Activities of Students during Study Abroad. In Lambert, R. D. & E. Shohamy (Eds.), *Language Policy and Pedagogy*, 237-261

Golonka, E. (2006). Predictors revised: Linguistic knowledge and metalinguistic awareness in second language gain in Russian. *Modem Language Journal* 90, 496-505.

Gómez-Lacabex, E. (2009). *Perception and production of vowel reduction in second language acquisition*. Unpublished PhD thesis, UPV/EHU.

Gómez-Lacabex, E., García-Lecumberri, M.L., & Cooke, M. (2009). Training and generalizing effects of English vowel reduction for Spanish listeners. In A. S. R. B. O. B. M. A. Watkins (Ed.), *Recent Research in Second Language Phonetics/Phonology: Perception and Production.* (pp. 32-42). Newcastle: Cambridge Scholars Publising.

Grabe, E., Post, B., Nolan, F., & Farrar, K. (2000). Pitch accent realisation in four varieties of British English. *Journal of Phonetics* 28.

Grabe, E., & Low, E.L. (2002). Durational Variability in Speech and the Rhythm Class Hypothesis. In C. W. Gussenhover, N. (Ed.), *Papers in Laboratory Phonology 7* (pp. 515-546). Berlin: Mouton de Gruyter.

Grenon, I., & White, L. (2008). *Acquiring Rhythm: A Comparison of L1 and L2 Speakers of Canadian English and Japanese.* Paper presented at the 32nd Boston University Conference on Language Development (BUCLD).

Grosjean, F. (1980). Temporal variables within and between languages. In H. Dechert, & M. Raupach (Eds.), *Towards a Cross-Linguistic Assessment of Speech Production*. Bern: Peter Lang.

Gut , U. (2003). Non-Native Speech Rhythm in German. *Proceedings of the 15th International Congress of Phonetic Sciences*, Barcelona, 2437-2440.

Gutiérrez-Díez, F. (2001). The Acquisition of English Syllable Timing by Native Spanish Speakers Learners of English. An Empirical Study. *International Journal of English Studies, 1*(1), 93-113.

Hahn, L. D. (2004). Primary stress and intelligibility: Research to motivate the teaching of suprasegmentals. *Tesol Quarterly*, 38(2), 201-223.

Harrington, J. (2010). *The phonetic analysis of speech corpora.* Chichester: Wiley-Blackwell.

Harris, J. (2005). Vowel reduction as information loss. In P. Carr, J. Durand & C. J. Ewen (Eds.), *Headhood, elements, specification and contrastivity*, pp. 119-132. Amsterdam: Benjamins.

Hernandez, T. A. (2010) Promoting Speaking Proficiency through Motivation and Interaction: The Study Abroad and Classroom Learning Contexts. *Foreign Language Annals,* 43-4 650-670.

Hieke, A. E. (1984). Linking as a marker of fluent speech. *Language and Speech, 27*(4), 343-354.

Hieke, A. E. (1985). A Componential Approach to Oral Fluency Evaluation. *The Modern Language Journal, 69*, 135-142.

Hieke, A. E. (1989). Spoken Language Phonotactics Implications for the ESL/EFL Classroom in Speech Production and Perception. *Language Sciences, 11*(2), 197-213.

Hilton, H. (2009). Annotation and Analyses of Temporal Aspects of Spoken Fluency. *CALICO Journal, 26*(3), 644-661.

Højen, A. (2004). *Second language speech perception and production in adult learners before and after short-term immersion*. Unpublished doctoral dissertation, University of Aarhus.

Howard, M. (2001). The effects of study abroad on the L2 learner's structural skills: Evidence from advanced learners of French. *Eurosla Yearbook, I*, 123-141.

Howard, M. (2005). On the role of context in the development of learner language: Insights from study abroad research. *ITL International Journal of Applied Linguistics*, 148, 1_20.

Huebner, T. (1995). The effects of overseas language programs: Report on a case study of an intensive Japanese course. In B. F. Freed (Ed.), *Second Language Acquisition in a Study Abroad Context* (pp. 171-193). Amsterdam: Benjamins.

Huebner, T. (1998). Methodological Considerations in Data Collection for Language Learning in a Study Abroad Context. *Frontiers: The Interdisciplinary Journal of Study Abroad, 4*.

Ife, A., Vives Boix, G., & Meara, P. (2000). The impact of study abroad on the vocabulary development of different proficiency groups. *Spanish Applied Linguistics* 4, 55-84.

Isabelli-García, C. (2006). Study abroad social networks, motivation and attitudes: Implications for second language acquisition. In E. Churchill & M. DuFon (Eds.), *Language learners in study abroad contexts* (pp. 231-258). Clevedon, UK: Multilingual Matters.

Izumi, S. (2002). Output, input enhancement and the noticing hypothesis: An experimental study on ESL relativization. *Studies in Second Language Acquisition, 24*(4), 541-577.

Juan-Garau, M., & Pérez-Vidal, C. (2007). The effect of context and contact on oral performance in students who go on a stay abroad. *Vigo International Journal of Applied Linguistic, 4*, 117-134.

Kang, O. (2010a). Relative salience of suprasegmental features on judgments of L2 comprehensibility and accentedness. *System,* 38(2), 301-315.

Kang, O. (2010b). *Salient Prosodic Features on Judgments of Second Language Accent.* Paper presented at the Speech Prosody 2010 100016, Chicago.

Kang, O., Rubin, D., & Pickering, L. (2010). Suprasegmental measures of accentedness and judgments of English language learner proficiency in oral English, *Modern Language Journal, 94,* 554-566.

Kinginger, C. (2007). Research on language education through programmes of study and residence abroad: Recent publications. *International Journal of Bilingual Education and Bilingualism*, 10, 104_112.

Kinginger, C. (2008). Language Learning in Study Abroad: Case Studies of Americans in France. *Modern Language Journal, 92*, 1-124.

Kinginger, C. (2009). *Language learning and study abroad: a critical reading of research.* Basingstoke, UK: Palgrave Macmillan.

Kormos, J. (1999) Monitoring and self-repair. *Language Learning* 49: 303-342.

Kormos, J. (2006). *Speech Production and Second Language Acquisition.* Mahwah, New Jersey: Lawrence Earlbaum.

Kormos, J. & Dénes, M. (2004) Exploring measures and perceptions of fluency in the speech of second language learners. *System* 32, 145-164.

Krashen, S.D. (1985). *The Input Hipothesis: Issues and Implications.* New York: Longman.

Lafford, B. A. (2004). The effect of context of learning on the use of communication strategies by learners of Spanish as a second language. *Studies in Secon Language Acquisition, 26*, 201-226.

Lapkin, S., Hart, D., & Swain, M. (1995). A Canadian Interprovincial Exchange: Evaluating the Linguistic Impact of a Three-Month Stay in Quebec. In B. F. Freed (Ed.), *Second Language Acquisition in a Study Abroad Context* (pp. 67-94). Amsterdam /Philadelphia: John Benjamins Publishing Company.

Lee, B., Guion, S.G., & Harada, T. (2006). Acoustic analysis of the production of unstressed English vowels by early and late Korean and Japanese bilinguals. *Studies in Second Language Acquisition* 28, 487-513.

Lennon, P. (1990). Investigating fluency in EFL: a quantitative approach. *Language Learning, 40*(3), 387-417.

Lennon, P. (2000). The lexical element in spoken second language fluency. In H. Riggenbach (Ed.), *Perspectives on fluency* (pp. 25-42). Ann Arbor: University of Michigan Press.

Levelt, W. J. M. (1989). *Speaking: From intention to articulation.* Cambridge, MA: MIT Press.

Levelt, W. J. M. (1993). *Speaking: from intention to articulation* Cambridge, Mass.: MIT Press.

Lindblom, B. (1963). Spectrographic study of vowel reduction. *Journal of the Acoustical Society of America* 35: 1773-1781.

Llanes, À. (2010). *Children and adults learning English in a study abroad contex.* Unpublished doctoral dissertation. Universitat de Barcelona.

Llanes, À. (2011). The many faces of study abroad: an update on the research on L2 gains emerged during a study abroad experience. *International Journal of Multilingualism.*

Llanes, À. & Muñoz, C. (2009). A short stay abroad: Does it make a difference? *System*, 37 (3), 353-365.

Logan, G. D. (1988). Toward an Instance Theory of Automatization. *Psychological Review, 95*(4), 492-527.

Long, M. H. (1981). Input, interaction and second language. In H. Winitz (Ed.), *Native language and foreign language acquisition* (pp. 259-278). New York: Annals of the New York Academy of Sciences.

Long, M.H. (1983). Native speaker / non-native speaker conversation and the negotiation of comprehensible input. *Applied Linguistics*, 4(2), 126-41.

Long, M. H. (1996). The role of linguistic environment in second language acquisition. In T. K. B. W.C. Ritchie (Ed.), *Handbook of second language acquistion* (pp. 413-468). New York: Academic Press.

Long, M., S. Inagaki & L. Ortega (1998). The Role of Implicit Negative Feedback in SLA: Models and Recasts in Japanese and Spanish. *The Modern Language Journal* 82:357-71.

Long, M. H., & Doughty, C. J. (Eds) (2009). *Handbook of Language Teaching*. Hoboken, NJ, USA: Wiley-Blackwell,. p 218

Lord, G. (2006). *Defining the Indefinable: Study Abroad and Phonological Memory Abilities*. Paper presented at the 7th Conference on the Acquisition of Spanish and Portuguese as First and Second Languages, Somerville, MA.

Low, E. L., Grabe, E., & Nolan, F. (2000). Quantitative characterizations of speech rhythm: syllable-timing in Singapore English. *Language & Speech* 43, 377-401.

MacKay, D. G. (1982). The problems of flexibility, fluency and speech-accuracy tradeoff in skilled behavior. *Psychological Review, 89*, 483-506.

Magnan, S., & Back, M. (2007). Social interaction and linguistic gain during study abroad. *Foreign Language Annals 40*, 43-61.

Meara, P. (1994) Second language acquisition: Lexis. In Asher, R.E. (Ed.) *The encyclopedia of language and linguistics. Oxford*: Pergamon. Vol. 7. 3276-28.

Miller, J. (1981). Effects of speaking rate on segmental distinctions. In P. Eimas & J. Miller (Eds.), *Perspectives on the Study of Speech*. pp 39-74, Lawrence Erlbaum, New Jersey.

Miller, L., & Ginsberg, R.B. (1995). Folklinguistic theories of language learning. In B. F. Freed (Ed.), *Second Language Acquisition in a Study Abroad Context*. Amsterdam and Philadephia: John Benjamins.

Milton, J., & Meara, P. (1995). How periods abroad affect vocabulary growth in a foreign language. *Review of Applied Linguistics, 107/108*, 17-34.

Mizera, G. J. (2005). *Working Memory and L2 Oral Fluency*. University of Pittsburgh, Pittsburgh.

Möhle, D. (1984). A comparison of the second language speech production of different native speakers. In H. W. Dechert, D. Möhle, & M. Raupach (Eds.), *Second language productions* (pp. 26-49). Tubingen: Gunter Narr Verlag.

Mora, J. C. (2006). Age effects on oral fluency development. In C. Muñoz (Ed.), *Age and the Rate of Foreign Language Learning* (pp. 65-88). Clevedon: Multilingual Matters.

Mora, J. C. (2008). Learning context effects on the acquisition of a second language phonology. In C. Pérez-Vidal, M. Juan-Garau, & A. Bel (Eds.), *A portrait of the young in the new multilingual Spain*. Clevedon: Multilingual Matters.

Mora, J. C., & Valls-Ferrer, M. (*submitted*). Study Abroad Effects on the Oral Production of Advanced L2 Learners: the Interaction between Fluency, Accuracy and Complexity.

Moyer, A. (2004). *Age, Accent and Experience in Secong Language Acquisition. An Integrated Approach to Critical Period Inquiry*. Clevedon: Multilingual Matters.

Moyer, A. (2008). Input as a critical means to an end: quantity and quality of experience in L2 phonological attainment. In T. Piske, & M. Young-Scholten (Eds.) *Input Matters in SLA*. (pp. 159-174) Bristol: Multilingual Matters.

Munro, M., & Derwing, T. M. (1995). Foreign accent, comprehensibility, and intelligibility in the speech of second language learners. *Language Learning* , 45, 73-97.

Munro, M. J., & Derwing, T. M. (1999). Foreign accent, comprehensibility, and intelligibility in the speech of second language learners. *Language Learning, 49*, 285-310.

Munro, M. J., & Derwing, T. M. (2001). Modeling Perceptions of the Accentedness and Comprehensibility of L2 Speech. The Role of Speaking Rate. *Studies in Second Language Acquisition, 23*(04), 451-468.

Munro, M. J., & Ocke-Schwen, B. (2007). The study of second language speech learning: A brief overview. In B. Ocke-Schwen, & M. J. Munro (Eds.) *Language Experience in Second Language Speech Learning,* 3–11.

Muñoz, C. (2007). CLIL: Some thoughts on its psycholinguistic principles. In Modelos Y Prácticas En Eicle. *Revista Española de Lingüística Aplicada (RESLA)*.

Muranoi, H. (2000). Focus on form through interaction enhancement: Integrating formal instruction into a communicative task in EFL classrooms. *Language Learning, 50*(4), 617-673.

Muranoi, H. (2007). Output practice in the L2 classroom. In R. DeKeyser (Ed.), *Practicing in a second language: Perspectives from applied linguistics and cognitive psychology*. New York: Cambridge University Press.

Murphy-Lejeune, E. (2002). *Student Mobility and Narrative in Europe. The New Strangers*. London: Routledge.

Nespor, M. (1990). On the rhythm parameter in phonology. In I. Roca (Ed.) *Logical Issues in Language Acquisition*. Dordrecht: Foris, 157-175.

O'Brien, I., Segalowitz, N., Collentine, J., & Freed, B. (2006). Phonological memory and lexical, narrative, and grammatical skills in second language oral production by adult learners. *Applied Psycholinguistics, 27*(3), 377-402.

O'Brien, I., Segalowitz, N., Freed, B., & Collentine, J. (2007). Phonological memory predicts second language oral fluency gains in adults. *Studies in Second Language Acquisition, 29*(4), 557-581.

Ortega, L. (1999). Planning and focus on form in L2 oral performance. Studies in Second Language Acquisition, 21, 109-48.

Ortega, L. (2009). *Understanding second language acquisition*. London: Hodder Arnold.

Ortega, L., & Iberri-Shea, G. (2005). Longitudinal Research in Second Language Acquisition: Recent Trends and Future Directions. *Annual Review of Applied Linguistics, 25*, 26-45.

Ortega-Llebaria, M., & Prieto, P. (2007). Disentangling stress from accent in Spanish: production patterns of the stress contrast in deaccented syllables. In J. M. P. Prieto, and M.J. Solé (Ed.), *Segmental and Prosodic Issues in Romance Phonology* (pp. 155-175). Amsterdam/Philadelphia: John Benjamins.

Ortega-Llebaria, M., & Prieto, P. (2011). Acoustic Correlates of Stress in Central Catalan and Castilian Spanish. *Language and Speech, 54*(1), 1-25.

Pellegrino Aveni, V. A. (2005). *Study Abroad and Language Use: Constructing the Self*. Cambridge: Cambridge University Press.

Pérez-Vidal, C. (2007). The need for focus on form in content and language integrated approaches: an exploratory study. In Modelos Y Prácticas En Eicle. *Revista Española de Lingüística Aplicada (RESLA)*.

Pérez-Vidal, C. (2008). La enseñanza bilingüe: El Enfoque Integrado de Contenidos y Lenguas en Europa. *In Aula de Innovación Educativa*. Editorial GRAÓ

Pérez-Vidal, C. (2011). Language acquisition in three different contexts of learning: formal instruction, stay abroad, and semi-immersion (CLIL). In Ruiz de Zarobe, Y., Sierra, J.M. and Gallardo, F.

Content and Foreign Language Integrated Learning: Contributions to Multilingualism in European Contexts. Spain: Peter Lang Publications.

Pérez-Vidal, C. & Escobar Urmeneta, C. (2002) Spain. In M. Grenfell (Ed.) *Modern Languages across the curriculum*. London: Routledge.

Pérez-Vidal, C. & Juan-Garau, M. (2010). To CLIL or not to CLIL? From bilingualism to multilingualism in Catalan/Spanish communities in Spain. In Y. Ruiz de Zarobe, & D. Lasagabaster (Eds.), *CLIL in Spain: Implementation, Results and Teacher Training*. CUP: Cambridge Scholars.

Pérez-Vidal, C., & Juan-Garau, M. (2009). The effect of Study Abroad (SA) on written performance. *Eurosla Yearbook, 9*, 269-295.

Pérez-Vidal, C. & Juan-Garau, M. (2011). The effect of context and input conditions on oral and written development: A Study Abroad perspective. *International Review of Applied Linguistics in Language Teaching*. 49, 2, 157–185.

Peterson, G. E., & Lehiste, I. (1960). Duration of syllable nuclei in Enghsh, *J. Acoust Soc Am* 32, 693-703

Pike, K. L. (1945). *The Intonation of American English*. Ann Arbor. University of Michigan Press.

Piske, T., MacKay, I.R.A., & Flege, J.E. (2001). Factors affecting degree of foreign accent in an L2: a review. *Journal of Phonetics, 29*, 191-215.

Piske, T., & M. Young-Scholten (Eds.). (2009). *Input matters in SLA*. Bristol: Multilingual Matters.

Prieto, P. (2004). *Fonètica i fonologia. Els sons del català*. Edicions de la Universitat Oberta de Catalunya: Barcelona.

Prieto, P., & Ortega-Llebaria, M. (2006). *Stress and Accent in Catalan and Spanish: Patterns of duration, vowel quality, overall intensity, and spectral balance*. Paper presented at the Speech Prosody, Dresden.

Prieto, P., Vanrell, MM., Astruc, Ll., Payne, E., & Post, B. (2010). *Speech rhythm as durational marking of prosodic heads and edges. Evidence from Catalan, English, and Spanish*. Paper presented at the Speech Prosody 2010, Chicago.

Prieto, P., Vanrell, MM., Astruc, Ll., Payne, E., Post, B. (*in press*). Phonotactic and phrasal properties of speech rhythm. Evidence from Catalan, English, and Spanish.

Poulisse, N., & Bongaerts, T. (1994). First language use in second language production. Applied Linguistics, 15, 36–57.

Ramus, F., Nespor, M., & Mehler, J. (1999). Correlates of linguistic rhythm in the speech signal. *Cognition, 75*(1), AD3-AD30.

Ramus, F., Dupoux, E. & Mehler, J. (2003). The psychological reality of rhythm classes: Perceptual studies. *Proceedings of the 15th International Congress of Phonetic Sciences*, Barcelona, 337-342.

Rast, R. (2008). *Foreign Language Input: Initial Processing*. Clevedon: Multilingual Matters.

Raupach, M. (1980). Temporal variables in first and second language speech production. In H. W. Dechert. & M. Raupach (Eds.), *Temporal Variables in Speech* (pp. 271-285). The Hague: Mouton.

Raupach, M. (1987). Procedural knowledge in advanced learners of a foreign language. In J. Coleman, & R. Towell (Eds.), *The Advanced Language Learner* (pp. 123-157). London: AFSL/CILT.

Recasens, D. (1993). *Fonètica i fonologia*, Enciclopèdia Catalana, Barcelona.

Rees, J., & Klapper, J. (2008). Issues in the quantitative longitudinal measurement of second language progress in the study abroad context. In L. Ortega & H. Byrnes (Eds.), *The longitudinal study of advanced L2 capacities* (pp. 89–105). London: Routledge.

Regan, V. (1995). The acquisition of sociolinguistics native speech norms. Effects of a year abroad on second language learners of French. In B. Freed (Ed.), *Second Language Acquisition in a Study Abroad Context* (pp. 245–267). Amsterdam: Benjamins.

Regan, V. (1998). Sociolinguistics and Language Learning in a Study Abroad Context. *Frontiers: The Interdisciplinary Journal of Study Abroad, 4*, 61-91.

Regan, V., Howard, M., & Lemée, I. (2009). *The acquisition of sociolinguistic competence in a study abroad context*. Buffalo: Multilingual Matters.

Richards, J., Platt, J. & Weber, H. (1985). *Longman dictionary of applied linguistics*. Harlow, Essex, England: Longman.

Riggenbach, H. (1991). Toward an understanding of fluency: a microanalysis of nonnative speaker conversations. *Discourse Processes, 14*, 423-441.

Riggenbach, H. (Ed.). (2000). *Perspectives on Fluency*. Ann Harbor: University of Michigan Press.

Robinson, P. (1995). Task complexity and second language narrative discourse. *Language Learning*, 45, 99-140.

Robinson, P. (1997). Generalizability and Automaticity of Second Language Learning Under Implicit, Incidental, Enhanced, and Instructed Conditions. *Studies in Second Language Acquisition, 19*, 223-247.

Roquet i Pugès, H. (2011) *A Study of the Acquisition of English as a Foreign Language: Integrating Content and Language in Mainstream Education in Barcelona*. Unpublished Doctoral Dissertation. Universitat Pompeu Fabra.

Rossiter, M. J. (2009). Perceptions of L2 fluency by native and non-native speakers of English. *The Canadian Modern Language Review*, 65, 395-412

Sasaki, M. (2007). Effects of Study-Abroad experiences on EFL writers: A multiple-data analysis. The Modern Language Journal, 9, 602-620.

Schachter, J. (1986). Three approaches to the study of input. *Language Learning, 36*(2), 221-225.

Schmidt, R. (1990). The role of consciousness in second language learning. *Applied Linguistics*, 12(2), 17-46.

Schmidt, R. (1992). Psychological mechanisms underlying second language fluency. *Studies in Second Language Acquisition, 14*, 357-385.

Schmidt, R. (1995). Consciousness and foreign language learning: A tutorial on the role of attention and awareness. In R. Schmidt (Ed.), *Attention and awareness in foreign language teaching and learning,* (pp. 1-64). Honolulu: University of Hawai'i at Manoa.

Schmidt, R. (2001). Attention. In Robinson, P. *Cognition and Second Language Instruction*. New York, NY: Cambridge University Press, 3-32.

Schmidt, R., & Frota, S. (1986). Developing basic conversational ability in a second language. A case study of an adult learner of Portuguese. In R. Day (Ed.), *Talking to learn: Conversation in second language acquisition*, (pp. 237-326) Rowley, MA: Newbury House.

Segalowitz, N. (2000). Automacity and attentional skill in fluent performance. In H. Riggenbach (Ed.), *Perspectives on fluency* (pp. 200-219). Ann Arbor:: University of Michigan Press.

Segalowitz, N. (2003) Automaticity and second languages. In C. Doughty & M. Long (Eds.), *The handbook of second language acquisition* (pp. 382-408). Oxford: Blackwell Publishers.

Segalowitz, N. (2010). *Cognitive Basis of Second Language Fluency*. New York: Routledge.

Segalowitz, N., & Segalowitz, S. J. (1993). Skilled performance, practice, and the differentiation of speed-up from automatization effects: Evidence from second language word recognition. *Applied Psycholinguistics*, 14, 369-385.

Segalowitz, N., Segalowitz, S. J., & Wood, A. (1998). Assessing the development of automaticity in second language word recognition. *Applied Psycholinguistics*, 19, 53-67.

Segalowitz, N., & Freed, B. F. (2004). Context, contact, and cognition in oral fluency acquisition. *Studies in Second Language Acquisition, 26*, 173-199.

Segalowitz, N., Freed, B., Collentine, J., Lafford, B., Lazar, N. & Díaz-Campos, M. (2004). A Comparison of Spanish Second Language Acquisition in Two Different Learning Contexts: Study Abroad and the Domestic Classroom. *Frontiers, 10*, 1-18.

Sharwood-Smith, M. (1994). *Second Language Learning: Theoretical Foundations*. New York: Longman.

Shehadeh, A. (1999). Non-native speakers' production of modified comprehensible output and second language learning. *Language Learning, 49*(4), 627-675.

Skehan, P. (1989) *Individual differences in second-language learning*. Edward Arnold.

Skehan, P. (1998). *A cognitive approach to language learning*. Oxford: Oxford University Press.

Skehan, P. (2002) Theorising and updating aptitude. In P. Robinson (Ed.) *Individual differences and instructed language learning*, (pp. 69-93(. Amsterdam/Philadelphia: John Benjamins.

Skehan, P., & Foster, P. (1997). Task type and task processing conditions as influences on foreign language performance. *Language Teaching Research*, 1 (3), 185-211.

Sunderman, G. & Kroll, J.F. (2009). When study abroad experience fails to deliver: The internal resources threshold effect. *Applied Psycholinguistics*, 30, 1-21.

Swain, M. (1985). Communicative competence: Some roles of comprehensible input and comprehensible input and comprehensible output in its development. In S. Gass, & C. Madden, *Input in Second Language Adquisition*, 235-53. Rowley, MA: Newbury House.

Swain, M. (1995). Three functions of output in second language learning. In G. Cook & B. Seidlhofer (Eds.), *Principle and Practice in Applied*

Linguistics: Oxford University Press.

Swain, M. (1998). Focus on form through conscious reflection. In C. Doughty & J. Williams (Eds.), *Focus on form in classroom second language acquisition* (pp. 64-81). New York, NY: Cambridge University Press.

Swain, M., & Lapkin, S. (1995). Problems in output and the cognitive processes they generate: A step towards second language learning. *Applied Linguistics, 16*(3), 371-391.

Tauroza, S., & Allison, D. (1990). Speech Rates in British English. *Applied Linguistics, 11*(1), 90-105.

Tavakoli, P. & Skehan., P. (2005). Strategic planning, task structure and performance testing. In R. Ellis (Ed.), *Planning and task performance in a second language* (pp. 239-277). Amsterdam: John Benjamins.

Tavakoli, P., & Foster, P. (2008). Task design and second language performance: the effect of narrative type on learner output. *Language Learning*, 58 (2), 439-473.

Toledo, G. (2010). Métricas rítmicas en discursos peninsulares. *Boletín de Linguística,* vol. XXII, núm. 33, pp. 88-113

Tortel, A., & Hirst, D. (2010). *Rhythm metrics and the production of English L1/L2*. Paper presented at the Speech Prosody 2010 100959, Chicago.

Towell, R., & Hawkins, R. (1994). *Approaches to Second Language Acquisition*, Multilingual Matters, Clevedon.

Towell, R., Hawkins, R., & Bazergui, N. (1996). The development of fluency in advanced learners of French. *Applied Linguistics, 17*(1), 84-119.

Towell, R. (2002). Relative degrees of fluency: A comparative case study of advanced learners of French. *International Review of Applied Linguistics, 40*, 117-150.

Towell, R., & Dewaele, J.M. (2005). The role of psycholinguistic factors in the development of fluency amongst advanced learners of French. In J.-M. Dewaele (Ed.), *Focus on French as a foreign language: Multidisciplinary approaches* (pp. 210–239). Clevedon, UK: Multilingual Matters

Trask, R.L. (1996). *A Dictionary of Phonetics and Phonology*, Routledge.

Trenchs-Parera, M. (2009). Effects of formal instruction and a stay abroad on the acquisition of native-like oral fluency. *The Canadian Modern Language Review*, 65 (3), 365-393.

Trofimovich P., & Baker, W. (2006). Learning Second Language Suprasegmentals: Effect of L2 Experience on Prosody and Fluency Characteristics of Speech. *Studies in Second Language Acquisition, 28*, 1-30.

Trofimovich, P., & Baker, W. (2007). Learning Prosody and fluency characteristics of second language speech: The effect of experience on child learners' acquisition of five suprasegmentals. *Applied Psycholinguistics, 28*, 251-276.

Trouvain, J., & Gut, U. (Eds) (2007). *Non-Native Prosody. Phonetic Description and Teaching Practice* (Trends in Linguistics. Studies and Monographs [TiLSM] 186). Berlin/New York: Mouton de Gruyter.

Tsao Y. C., Weismer, G., Iqbal, K. (2006). The effect of intertalker speech rate variation on acoustic vowel space. *Journal of the Acoustical Society of America*. 2006 Feb;119(2):1074-82.

Turner, C. W., Souza, P. E., and Forget, L. N. (1995). Use of temporal envelope cues in speech recognition by normal and hearing-impaired listeners, *J. Acoust. Soc. Am.* 97, 2568–2576.

Valls-Ferrer, M. (2008) Language acquisition during a stay abroad period following formal instruction: Temporal effects on oral fluency development. Unpublished master's thesis. Barcelona: Universitat Pompeu Fabra.

Valls-Ferrer, M., Roquet-Pugès, H., & Pérez-Vidal, C. (2010). The effect of practice in different contexts of language acquistion: Contrasting FI, SA and CBLC. *Proceedings XXVII Congreso Nacional de Lingüística Aplicada 2009, Ciudad Real.*

Vanderplank, R. (1993). 'Pacing' and 'spacing' as predictors of difficulty in speaking and understanding English. *ELT Journal*, Vol. 47/2, 117-125, Oxford: Oxford University Press

VanPatten, B., & Cadierno, T. (1993). Explicit instruction and input processing. *Studies in Second Language Acquisition*, 15, 225–243.

VanPatten, B. (2004). Input processing in second language acquisition. In B. VanPatten (Ed.), *Processing Instruction: Theory, research, and commentary*. Mahwah, NJ: Lawrence Erlbaum.

Varden, J. K. (2006). Visualizing English Speech Reductions Using the Free Phonetic Software Package WASP. In J.D. Brown & K. Kondo-Brown (Eds.). *Perspectives on Teaching Connected Speech to Second Language Speakers,* NFLRC, University of Hawaii Press.

Weismer, G., Laures, J.S., Jeng, J-Y., Kent, R.D., & Kent, J.F. (2000). Effect of speaking rate manipulations on acoustic and perceptual aspects of the dysarthria in amyotrophic lateral sclerosis. *Folia Phoniatr. Logop.* 52, 201–219.

Wennerstrom, A. (2000). The Role of Intonation in Second Language Fluency, in: H. Riggenbach (Ed.), *Perspectives on Fluency*. Ann Arbor, MI, University of Michigan Press.

White, L., & Mattys, S. L. (2007a). Calibrating rhythm: First language and second language studies. *Journal of Phonetics, 35*(4), 501-522.

White, L., & Mattys, S. L. (2007b). Rhythmic typology and variation in first and second languages. In P. Prieto, J. Mascaró, & M-J. Solé (Eds.), *Segmental and prosodic issues in Romance phonology* (pp. 237-257): John Benjamins.

Wiget, L., White, L., Schuppler, B., & Grenon, I. (2010). How stable are acoustic metrics of contrastive speech rhythm? *Journal of the Acoustical Society of America, 127*(3), 1559-1569.

Wood, D. (2006). Uses and Functions of Formulaic Sequences in Second Language Speech: An Exploration of the Foundations of Fluency. *The Canadian Modern Language Review, 63*(1), 13-33.

List of Appendices

A.1.1. Instrument: Interview questions (sample).

INTERVIEW QUESTIONS

1. Have you met any good friends at the university yet? If yes, what are they like? if not, why do you think you haven't?

2. How different is your university life from your high school life?

3. How do you like the library? how often do you go there? and what resources do you usually use?

4. What advantages and disadvantages do you see in the location of the Ramblas building?

5. What do you normally do for lunch during school days?

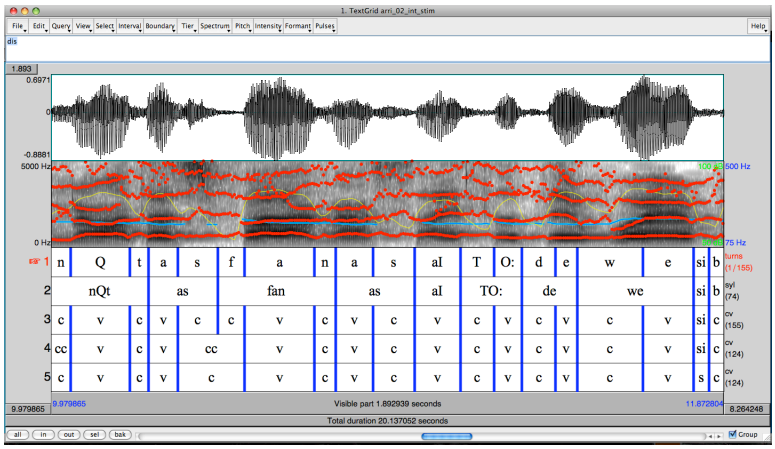
6. What do your friends and family think of your future profession as a translator and/or interpreter?

7. What do you think of the requirement at UPF of having to study abroad?

A.1.2. Transcription Conventions for Utterance Fluency Analysis.

Standards used	Description of Standards
A	Student A
B	Student B
R	Researcher
#	Unfilled Pause (Short pause approx. 1 sec)
##	Unfilled Pause (Longer pause approx. 2-3 sec.)
#4	Unfilled Pause (More than 3 sec, indicating sec.)
.	Period. End of an unmarked (declarative) utterance.
?	Question Mark. End of a question.
!	Exclamation Point. End of an imperative or emphatic utterance.
+...	Trailing Off. Incomplete, but not interrupted, utterance.
+,	Self-Completion. Completion of an utterance after an interruption.
+/.	Interruption by another speaker.
++	Latching. Completion of another speaker's utterance.
underlined	Speech in Spanish or Catalan
[*text]	Ungrammatical word
xxx	Unintelligible Speech
&	Phonological Fragment
:	Lengthening of a sound or syllable
<fragment>	String of words modified by the following symbol.
[!]	Stressing. Preceding word or fragment is stressed.
[=! text]	Paralinguistic Events produced while speaking
&=	Simple Events. Sounds produced by the speaker not being words.
[^text]	Complex Local Event. Description of a non linguistic event.
[>]	Overlap Follows. Text said at the same time as other speaker.
[<]	Overlap Precedes. Text said at the same time by preceding speaker.
[>N] [<N]	Overlap Enumeration. When there are several overlapping overlaps, to mark which pieces are overlapped.
[/]	Retracing Without Correction. Repetition of early material without change.
[//]	Retracing With Correction. Repetition of the basic phrase, changing the syntax but maintaining the idea.

A1.3. Transcription rhythm: example in Praat.



Waveform, spectrogram, and labeling used for the speech sample '…not as fun as I thought they were…' produced by a NNS participant.

A.2.1. Pearson r correlations between temporal fluency measures.

	1	2	3	4	5	6	7	8	9
1.SR	1	,897**	,740**	-,489**	-,647**	-,740**	-,027	,792**	-,619**
2.AR	,897**	1	,373**	-,334**	-,430**	-,373**	-,141	,537**	-,519**
3.Phon_Rat	,740**	,373**	1	-,537**	-,730**	-1,000**	,170	,854**	-,524**
4.Pause_F	-,489**	-,334**	-,537**	1	,886**	,537**	,133	-,615**	,362**
5.P_i_Dur	-,647**	-,430**	-,730**	,886**	1	,730**	,089	-,722**	,510**
6.P_Dur	-,740**	-,373**	-1,000E0	,537**	,730**	1	-,170	-,854**	,524**
7.Dys_Rat	-,027	-,141	,170	,133	,089	-,170	1	,098	-,065
8.MLoR	,792**	,537**	,854**	-,615**	-,722**	-,854**	,098	1	-,650**
9.Total_Mean	-,619**	-,519**	-,524**	,362**	,510**	,524**	-,065	-,650**	1

**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).

A.2.2. Pearson-r correlations between rhythm metrics and perceived fluency scores by listeners' groupings

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1.deltaV	1	,332**	,696**	,386**	,628**	,324*	,358**	,471**	,553**	,467**	,520**	,545**	,552**	,559**	,530**	,560**	,534**	,555**
2.deltaC	,332**	1	,224	,771**	-,202	,874**	,451**	,100	,396**	,319*	,431**	,336**	,402**	,402**	,371**	,375**	,397**	,393**
3.varcoV	,696**	,224	1	,249	,169	,188	,228	,280*	,213	,239	,235	,187	,185	,210	,234	,190	,238	,220
4.varcoC	,386**	,771**	,249	1	,189	,545**	,431**	,037	,409**	,277*	,362**	,330**	,365**	,365**	,358**	,354**	,365**	,366**
5.percentV	,628**	-,202	,169	,189	1	-,237	,094	,338**	,351**	,179	,227	,321*	,321*	,301*	,278*	,328*	,265*	,298*
6.rPVIc	,324*	,874**	,188	,545**	-,237	1	,725**	,164	,466**	,412**	,491**	,442**	,499**	,494**	,456**	,479**	,475**	,485**
7.nPVIc	,358**	,451**	,228	,431**	,094	,725**	1	,208	,492**	,383**	,439**	,465**	,478**	,478**	,455**	,481**	,457**	,476**
8.nPVIv	,471**	,100	,280*	,037	,338**	,164	,208	1	,158	,133	,182	,216	,243	,222	,151	,233	,164	,197
9.G1_Mean	,553**	,396**	,213	,409**	,351**	,466**	,492**	,158	1	,866**	,929**	,889**	,898**	,939**	,969**	,913**	,969**	,961**
10.G2_Mean	,467**	,319*	,239	,277*	,179	,412**	,383**	,133	,866**	1	,877**	,847**	,852**	,891**	,963**	,868**	,947**	,932**
11.G3_Mean	,520**	,431**	,235	,362**	,227	,491**	,439**	,182	,929**	,877**	1	,873**	,890**	,954**	,936**	,900**	,972**	,953**
12.G4_Mean	,545**	,336**	,187	,330**	,321*	,442**	,465**	,216	,889**	,847**	,873**	1	,915**	,967**	,899**	,981**	,903**	,952**
13.G5_Mean	,552**	,402**	,185	,365**	,321*	,499**	,478**	,243	,898**	,852**	,890**	,915**	1	,969**	,907**	,976**	,914**	,961**
14.Expcd_J.	,559**	,402**	,210	,365**	,301*	,494**	,478**	,222	,939**	,891**	,954**	,967**	,969**	1	,948**	,989**	,964**	,991**
15.Inexpcd_J.	,530**	,371**	,234	,358**	,278*	,456**	,455**	,151	,969**	,963**	,936**	,899**	,907**	,948**	1	,923**	,993**	,981**
16.NonNattive_J.	,560**	,375**	,190	,354**	,328*	,479**	,481**	,233	,913**	,868**	,900**	,981**	,976**	,989**	,923**	1	,928**	,977**
17.Native_J.	,534**	,397**	,238	,365**	,265*	,475**	,457**	,164	,969**	,947**	,972**	,903**	,914**	,964**	,993**	,928**	1	,986**
18.Total_Mean	,555**	,393**	,220	,366**	,298*	,485**	,476**	,197	,961**	,932**	,953**	,952**	,961**	,991**	,981**	,977**	,986**	1

A.2.3. Pearson-r correlations between rhythm metrics and utterance fluency measures.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1.deltaV	1	,332**	,696**	,386**	,628**	,324*	,358**	,471**	-,545**	-,508**	-,512**	-,398**	,367**	,436**	,398**	,089
2.deltaC	,332**	1	,224	,771**	-,202	,874**	,451**	,100	-,397**	-,281*	-,411**	-,212	,212	,333**	,212	-,063
3.varcoV	,696**	,224	1	,249	,169	,188	,228	,280*	-,142	-,210	-,111	-,145	,090	,111	,145	,083
4.varcoC	,386**	,771**	,249	1	,189	,545**	,431**	,037	-,227	-,276*	-,200	-,199	,298*	,408**	,199	-,028
5.percentV	,628**	-,202	,169	,189	1	-,237	,094	,338**	-,249	-,330**	-,196	-,253	,382**	,363**	,253	,089
6.rPVIc	,324*	,874**	,188	,545**	-,237	1	,725**	,164	-,574**	-,456**	-,557**	-,368**	,282*	,412**	,368**	-,086
7.nPVIc	,358**	,451**	,228	,431**	,094	,725**	1	,208	-,499**	-,541**	-,439**	-,402**	,310*	,418**	,402**	-,063
8.nPVIv	,471**	,100	,280*	,037	,338**	,164	,208	1	-,312*	-,184	-,322*	-,191	,175	,148	,191	-,009
9.SR_S	-,545**	-,397**	-,142	-,227	-,249	-,574**	-,499**	-,312*	1	,792**	,897**	,740**	-,489**	-,647**	-,740**	-,027
10.MLoR	-,508**	-,281*	-,210	-,276*	-,330**	-,456**	-,541**	-,184	,792**	1	,537**	,854**	-,615**	-,722**	-,854**	,098
11.AR	-,512**	-,411**	-,111	-,200	-,196	-,557**	-,439**	-,322*	,897**	,537**	1	,373**	-,334**	-,430**	-,373**	-,141
12.Phon_Rat	-,398**	-,212	-,145	-,199	-,253	-,368**	-,402**	-,191	,740**	,854**	,373**	1	-,537**	-,730**	-1,00**	,170
13.Pause_F	,367**	,212	,090	,298*	,382**	,282*	,310*	,175	-,489**	-,615**	-,334**	-,537**	1	,886**	,537**	,133
14.P_i_Dur	,436**	,333**	,111	,408**	,363**	,412**	,418**	,148	-,647**	-,722**	-,430**	-,730**	,886**	1	,730**	,089
15.P_Dur	,398**	,212	,145	,199	,253	,368**	,402**	,191	-,740**	-,854**	-,373**	-1,00**	,537**	,730**	1	-,170
16.Dys_Rat	,089	-,063	,083	-,028	,089	-,086	-,063	-,009	-,027	,098	-,141	,170	,133	,089	-,170	1

**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).