UNIVERSIDADE FEDERAL DE MINAS GERAIS

FACULDADE DE LETRAS

PROGRAMA DE PÓS-GRADUAÇÃO EM ESTUDOS LINGUÍSTICOS

JESIEL SOARES SILVA

EXPLORING A VOCABULARY TEST AND A JUDGMENT TASK AS DIAGNOSES OF EARLY AND LATE BILINGUALS' L2 PROFICIENCY

BELO HORIZONTE

FEVEREIRO, 2016

JESIEL SOARES SILVA

EXPLORING A VOCABULARY TEST AND A JUDGMENT TASK AS DIAGNOSES OF EARLY AND LATE BILINGUALS' L2 PROFICIENCY

Tese apresentada ao Programa de Pós-Graduação em Estudos Linguísticos da Faculdade de Letras da Universidade Federal de Minas Gerais, como requisito parcial para obtenção do título de DOUTOR em Linguística Teórica e Descritiva.

Área de Concentração: Linguística Teórica e Descritiva

Linha de pesquisa: Processamento da Linguagem

Orientador: Prof. Dr. Ricardo Augusto de Souza

Co-Orientadora: Profa. Dra. Jordana Grazziela Alves Coelho dos Reis (Instituto de Pesquisas René Rachou-FIOCRUZ/FAMINAS-BH)

BELO HORIZONTE

FACULDADE DE LETRAS DA UFMG

2016

Ficha catalográfica elaborada pelos Bibliotecários da Biblioteca FALE/UFMG

S586e

Silva, Jesiel Soares.

Exploring a vocabulary test and a judgment task as diagnoses of early and late bilinguals' L2 proficiency [manuscrito] / Jesiel Soares Silva . – 2016.

143 f., enc. : il., grafs. tabs., p&b.

Orientador: Ricardo Augusto de Souza.

Coorientadora: Jordana Grazziela Alves Coelho dos Reis.

Área de concentração: Linguística Teórica e Descritiva.

1. Língua inglesa – Estudo e ensino – Falantes estrangeiros – Teses. 2. Psicolínguistica – Teses. 3. Aquisição da segunda linguagem – Teses. 4. Língua inglesa – Testes de aptidão – Teses. 5. Língua inglesa – Vocabulário – Teses. I. Souza, Ricardo Augusto de. II. Reis, Jordana Grazziela Alves Coelho dos. III. Universidade Federal de Minas Gerais. Faculdade de Letras. IV. Título.

CDD : 428.07

UNIVERSIDADE FEDERAL DE MINAS GERAIS

PROGRAMA DE PÓS-GRADUAÇÃO EM ESTUDOS LINGUÍSTICOS

PosLin

FOLHA DE APROVAÇÃO

Exploring a Vocabulary Test and a Judgment Task as Diagnoses of Early and Late Bilinguals' L2 Proficiency

JESIEL SOARES SILVA

Tese submetida à Banca Examinadora designada pelo Colegiado do Programa de Pós-Graduação em ESTUDOS LINGUÍSTICOS, como requisito para obtenção do grau de Doutor em ESTUDOS LINGUÍSTICOS, área de concentração LINGUÍSTICA TEÓRICA E DESCRITIVA, linha de pesquisa Processamento da Linguagem.

Aprovada em 01 de abril de 2016, pela banca constituída pelos membros:

Prof(a). Ricardo Augusto de Souza - Orientador
UFMG

Prof(a). Jordana Granziela Alves Coelho dos Reis
FIOCRUZ

Prof(a). Natália Carolina Alencar de Resende
CEFET/MG

Marisa Mendonça Carneiro
Prof(a). Marisa Mendonça Carneiro

Deise Prina Dutra
Prof(a). Deise Prina Dutra
UFMG

Prof(a). Maria Luiza Gonçalves Aragão da Cunha Lima
UFMG

Belo Horizonte, 1 de abril de 2016.

To Valdivino da Silva (in memorian). Daddy, looks like I made it!

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to:

My advisor Dr. Ricardo Augusto de Souza for the unceasing support of my PhD study and related research, for his patience, motivation, and immense knowledge.

Dr. Jordana Reis, who revealed me new possibilities of working with linguistic data.

Dr. Ingrid Finger, Dr. Deise Dutra, and Dr. Fernando de Oliveira for the insightful comments and encouragement in the primary version of this text, but also for the hard questions which incented me to widen my research from various perspectives.

Dr. Eva Fernández for all the attention, help and patience during my time in New York. Without her precious support it would not be possible to conduct this research.

Dr. Obler and Eve Higby, who provided me an opportunity to join their team as a visiting student, and who gave me access to the Neurolinguistics Laboratory and research facilities in the Graduate Center (CUNY).

My fellow lab mates of the Laboratório de Psicolinguística/UFMG for the stimulating discussions, for the sleepless nights we were working together before deadlines (Alexandre), and for all the fun we have had in the last four years.

CAPES (Coordenação de Aperfeiçoamento de Pessoal de nível Superior) for supporting this research from the beginning, including my time in New York.

Programa de Ps-Graduação em Estudos Linguísticos (POSLIN/UFMG).

Special thanks...

To daddy, who is the reason of all the work, effort and time I spent in this research.

To mommy and sis (Gilda), who are always there to remind me I can go further.

To Gute (Rute), who is the reason I love Belo Horizonte! My wife/girlfriend/friend for the rest of my days. We're getting there together, baby!

To Rogério and Alexandre, the dearest brothers.

Colleen Gibson, for the friendship always followed by great pun moments, and for the help with my text edition. Keep the Bryant Park ready, I'll be back!

ABSTRACT

This psycholinguistics study delved into the adequacy of a measure of vocabulary size – the Vocabulary Levels Test (VLT) – as a predictor of Brazilian Portuguese-English speakers' and Spanish Heritage Language (HL) speakers' capacity to access grammatical representations when using their non-dominant language. Considering the scenario with two types of bilinguals (L2 learners and HL), two types of knowledge (implicit and explicit), and two types of proficiency measures (vocabulary size and sentence judgment), we conducted an exploratory study designed in two experiments. In the first experiment, we compared participants' performances in the VLT (English version), an overall abilities test (Oxford Placement Test – OPT), and a Speeded Acceptability Judgment (SAJ) task in English, in which a ceiling of 8 seconds was set for each judgment call. In the second experiment, we compared participants' performances in the VLT (Spanish version), the Spanish Placement Test (SPT), a SAJ task in Spanish, in which a ceiling of 6 seconds was set for each judgment call, and a self-assessment test on participants' language abilities. Moreover in the second experiment, we employed a Bilingual History Questionnaire (BHQ) to operationalize HL speakers' language use/dominance. For both experiments, our SAJ task stimuli were composed of 56 sentences, and 16 of them contained grammatical violations. There were two types of sentence violations applied to 8 sentences each: argument structure realization violations involving unergative verbs in transitive syntax, and explicit morpho-syntactic violations involving long-distance dependencies (Wh-movement) and subject-verb agreement. In both experiments, the VLT, the OPT/SPT and the SAJ tasks were submitted to the Receiver Operating Characteristic (ROC) curve procedures for the assessment of their sensitivity and specificity as diagnostic tests of language proficiency. Results for both experiments show that only those participants who classified as high proficiency in the VLT and the OPT/SPT were capable of detecting grammatical violations. We interpret the results as indicating that a measure of vocabulary size is a predictor of both fluency in lexical access and fluency in grammatical knowledge access of Second Language (L2) and HL speakers. Moreover, ROC curve analysis from both experiments revealed that the VLT and the SAJ task are adequate instruments for language proficiency diagnosis, since they are able to differentiate two groups of proficiency.

Keywords: Bilingualism; Vocabulary knowledge; Proficiency measurement.

RESUMO

Este estudo psicolinguístico investigou a adequação de uma medida de tamanho de vocabulário – o *Vocabulary Levels Test (VLT)* – como preditor da capacidade de acesso às representações gramaticais de falantes de português brasileiro e inglês, bem como de falantes de espanhol como língua de herança (HL) no momento em que estão usando sua língua não-dominante. Considerando o cenário com dois tipos de bilíngues (falantes de L2 e falantes de HL), dois tipos de conhecimento (implícito e explícito) e dois tipos de medidas de proficiência (tamanho de vocabulário e julgamento de sentença), nós conduzimos um estudo exploratório desenhado como dois experimentos. No primeiro experimento, comparamos o desempenho dos participantes no VLT (versão em inglês), um teste de habilidades gerais (*Oxford Placement Test – OPT*) e uma tarefa de julgamento temporalizado (SAJ) em inglês, na qual um teto temporal de 8 segundos foi estabelecido para cada julgamento. No Segundo experimento, nós comparamos o desempenho dos participantes no VLT (versão em espanhol), na versão em espanhol do teste de habilidades gerais (SPT), a SAJ em espanhol, na qual um teto temporal de 6 segundos foi estabelecido para cada julgamento e um teste de auto-avaliação sobre habilidades linguísticas. Além disso, no segundo experimento, um questionário de histórico bilingue (o BHQ) foi usado na operacionalização da classificação dos HL pelo uso/dominância linguística. Em ambos os experimentos, os nossos estímulos da SAJ foram compostas de 56 sentenças, 16 delas continham violações gramaticais. Havia dois tipos de violações de sentenças: violações de realização de estrutura argumental envolvendo verbos inergativos transitivizados e violações morfosintáticas explícitas envolvendo dependências de longa distância (movimento WH) e concordância sujeito-verbo. Em ambos os experimentos, o VLT, o OPT / SPT e as tarefas SAJ foram submetidos à curva ROC (*Receiver Operating Characteristic*) para a avaliação da sensibilidade e especificidade como diagnóstico adequado de proficiência linguística. Os resultados para ambos os experimentos mostram que somente os participantes que foram classificados como de alta proficiência no VLT e OPT / SPT foram capazes de detectar violações gramaticais. Nós interpretamos os resultados como indicAdores de que a medida do tamanho do vocabulário é um preditor tanto de fluência em acesso lexical quanto de fluência no acesso ao conhecimento gramatical de falantes de L2 e HL. Além disso, análises com a curva ROC nos dois experimentos revelaram que o VLT e a SAJ são instrumentos adequados para o diagnóstico de proficiência da língua, uma vez que são capazes de diferenciar dois grupos de proficiência.

Palavras-chave: Bilinguismo; conhecimento de vocabulário; medidas de proficiência.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

CHAPTER 1

INTRODUCTION

1.1 Preliminaries

When he opened a book chapter by stating "everyone is bilingual," Edwards (2006, p.7) did not overestimate his appraisal, since his definition of bilingualism encompasses the knowledge of at least one word from a language that is different than the mother tongue. Edwards' description is one among several definitions of bilingualism that reach many aspects from linguistic to political issues (Edwards, 2012). The last decades have shown the necessity of inserting the bilingualism issue into the psychological, political, and social debate, because the discussion on bilingualism has played a crucial role in constructs, such as ethnicity, communities, minority groups (Edwards, 2012).

In this study, we define bilingualism under a psychological aspect rather than social or cultural. We consider bilinguals to be those who operate in two languages, regardless of their level of proficiency in either language (Grosjean, 1998, 2013). Grosjean (2013, p. 5) characterizes bilingualism (and multilingualism) as "the use of two or more languages (or dialects) in everyday life," and bilinguals as those who "use the two languages—separately or together—for different purposes, in different domains of life, with different people" (Grosjean, 2008, p. 14). We align ourselves with Grosjean (2008, p. 13) on the idea that "the bilingual is not the sum of two complete or incomplete monolinguals; rather, he or she has a unique and specific linguistic configuration."

Although the proficiency level does not determine whether someone is bilingual or not, an important question to be considered is how to measure one's degree of language competence. Edwards (2012) mentions types of measurement such as rating scales, tests of speaking fluency, and self-assessment. The author points out that the major part of these measurements is the ability to provide information about a set of one's abilities, but not about all of the facets in which a bilingual is involved. There are several 'labels' to define the highly proficient bilingual, such as balanced bilingual, ambilingual, or equilingual. However, this idea of equilibrium has been overcome, since bilinguals seem

to be those ones who operate two languages in two different ways (Edwards, 2012, Grosjean, 1998).

In this work, we have two types of bilinguals: the Second Language (L2) speakers and the Heritage Language (HL) speakers. For our purpose, in this study, we define L1 as the first language acquired in the childhood in a context in which such language is dominant. On the other hand, L2 is the second language learned, normally after puberty. In this work, we will use L2 to represent the non-dominant language and not necessarily the second in a scale. For that reason, for English native speakers who learned Spanish after puberty (Spanish L2 learners) and for HL speakers who have Spanish as the non-dominant language, we will name Spanish as L2. Although the term 'L2' seems to be unappropriated for the second group, we are adopting it not as matter of sequentially, but as a matter of different-from-the-dominant language.

1.2 Statement of the purpose

It is broadly known that L2 proficiency pairs with language dominance as the fundamental constructs in bilingualism studies (Alderson, 2005, Bedore et al., 2012, Birdsong, 2006). The former construct entails specifications of observable L2 ability and fluency (Harley et al. 1990, Hulstijn, 2010), whereas the latter entails patterns and preferences of language use in everyday life (Dunn & Fox Tree, 2009, Gertken et al. 2014). In studies with L2 speakers, participants' L2 proficiency level is frequently taken as an independent variable, and researchers often take language ability profiles as a screening factor for between-subject comparisons (Souza & Oliveira, 2014, Souza & Soares-Silva, 2015).

Although employment of some sort of measurement of L2 proficiency and/or language dominance has been a common practice in the scientific investigation of bilingualism, such a procedure has been criticized for recurrently being inconsistent and lacking sufficiently powerful generalizability (Grosjean, 1998; Hulstijn, 2012). This concern has led several researchers to consider the study of measures of language dominance (e.g. Dunn & Fox Tree, 2009; Bedore et al., 2012), and measures of specific

dimensions of L2 proficiency (e.g. Alderson, 2005; Souza, Duarte & Berg, 2015, Hulstijn, 2010) as a primary goal in the L2 research agenda.

From a cognitive perspective, it is crucial to outline the nature of knowledge being measured in L2 proficiency tests. According to the literature, linguistic knowledge can be implicit or explicit (Ellis, R. 1995, Ellis, R. 2005). Implicit knowledge is intuitive, unconscious, and inherent to the learner. In other words, implicit knowledge means *to know the language*. An example of implicit knowledge that entails most use of the first language is the First Language (L1). As he speaks, an L1 speaker is not aware of rules being used and sometimes he would not even be able to explain the reason he is using such structures. Contrarily, explicit knowledge is conscious and it means *know about the language.* Learners are able to verbalize the language structures and even to explain some of the rules.

Both implicit and explicit knowledge are constructs derived from models of memory (Paradis, 1994). The association of implicit knowledge with procedural memory and explicit memory with declarative knowledge supports several studies on the nature of L2 knowledge (Ellis R, 1994, 2005, Ellis N, 2005, Bowles, 2011). Bilingualism studies that depart from experimental data have considered this distinction and the interaction among both types of knowledge in L2 learning/acquisition process (Ellis N, 2005, 2007).

Another question worth raising is what type of bilingual will be investigated. Although the major part of bilingual studies deals with L2 speakers, notably those who acquire a second language consequently after the L1 in the last decades, studies with a different kind of bilingual have increased considerably: Heritage Language (HL) speakers. HL speakers are individuals who were born in a given linguistic context and, still as a toddler, move to a different linguistic environment where they are raised. In these cases, the language of the new context becomes the dominant as time goes by (Montrul, 2005, 2010).

For instance, a girl who was born in Peru (and her parents are Spanish L1 speakers) moves to the United States of America by the age of 1 year. Before going to school, she keeps Spanish as the only language while interacting with parents, relatives, and/or nannies. As she grows up and goes to school, she gets to speak in English while studying and interacting with teachers and colleagues. When she becomes an adult, English is her

first language and Spanish becomes the second (as a matter of use). Due to this inversion, a great endeavor of bilingualism studies is to measure these individuals' proficiency in both languages, as well as the nature of their knowledge in the native language.

A measurable aspect of L2 proficiency that has been amply studied is L2 lexical knowledge (Nation, 1990; Meara, 1996; Laufer & Nation, 1999; Read, 2000; De Groot, 2011). Tests explicitly measuring L2 lexical knowledge may focus on two concepts of vocabulary: breadth and depth (Schmitt , 2014). Vocabulary breadth represents the amount of words a speaker is able to acknowledge, understand, and associate to a meaning. In a nutshell, it is the size of someone's vocabulary. Differently, vocabulary depth is the extent of representational detail and connectedness in the mental lexicon. By knowing a word deeply, an individual is capable of internalizing its collocations, morphological restrictions, and pragmatic adequacy (Meara, 1996; Read, 2000; Milton, 2010).

According to Meara & Alcoy (2010), one of the most widely accepted instruments to measure vocabulary size is Nation's (1990) *Vocabulary Levels Test* –the VLT. The VLT is a 5-part vocabulary test concerning word frequency, and consists of associating words with meanings. Because the VLT deals with word association, we assume that the knowledge being measured is explicit.

Another possible measure of L2 knowledge used in bilingualism studies is the acceptability judgment of sentences (Ellis R, 2015, Souza & Soares-Silva, 2015). This is a behavioral task in which participants are asked to emit a judgment on the licitness of a set of sentences according to their first impression. An alternative to this task is the application of a time limit on the task. Based on statistical measures with L1 speakers judging their first language, a time ceiling is delimited to L2 learners to do judgment calls in their L2 (Souza et al., 2015).

Considering this scenario with two types of bilinguals (L2 learners and HL), two types of knowledge (implicit and explicit), and two types of proficiency measures (vocabulary size and sentence judgment), we propose a psycholinguistic study. The general goal is to validate the proficiency measures as precise diagnostic tests, with both types of bilinguals, considering their explicit and implicit knowledge. Such endeavor will be guided by the following research questions:

1.3 Research questions

• Are the VLT (as a diagnostic test for explicit knowledge of L2 vocabulary size) and the Acceptability Judgment Task (as a diagnostic measure for implicit knowledge of L2 representation) able to discriminate L2 proficiency profiles that correlate among L2 learners and HL?

• Assuming that the increase in the proficiency level reflects on the type of memory that learners rely on: Do highly proficient L2 learners and HL speakers perform better than low ones in the speeded acceptability judgment task?

In order to answer these questions, we designed two experiments with one major objective each. Experiment 1 was conducted in Brazil, in the Psycholinguistics Lab of the Universidade Federal de Minas Gerais. Thirty participants (age mean=25.6) took part on the first experiment. The second experiment took place in New York, in the Psycholinguistics Lab of the City University of New York (CUNY). Two groups formed the participants for the second experiment. The first group was composed of 20 English native speakers who were Spanish L2 learners. The second group was composed of 40 individuals (age mean=23). All of them were born in a Spanish language context—some in a Spanish language country, and some in New York in a Spanish-speaking family.

1.4 Research objectives

• Objective 1 (Experiment 1)

The main goal is to investigate the ability of the English version of VLT to discriminate L2 English proficiency profiles in relation to an objective grammatical test (Oxford Placement Test - OPT), and to a Speeded Acceptability Judgment task among L1 Brazilian-Portuguese speakers who are learners of English. The OPT scores roughly meet a broadly accepted framework of L2 proficiency: the *Common European Reference Framework* (CERF) (Council of Europe, 2001). In the Speeded Acceptability Judgment task, both grammatical and ungrammatical sentences in English were displayed as stimuli. Acceptability judgment tasks may vary in stimuli presentation mode and task requirements. Nevertheless, it is worth noting that there is evidence that more than one of

its varieties can yield data that reflects differences in the state of L2 representations (Ellis, 2005; Souza & Oliveira, 2014). Therefore, the acceptability judgment is a behavioral off-line task that is likely to be sensitive to differential profiles in L2 proficiency.

As a secondary goal, we aimed at estimating the minimum time frame in which bilinguals (Portuguese/English) could make accurate judgment calls about a sentence in English. To achieve this specific goal, we mainly replicated the design of the study reported by Souza et al. (2014), in which the authors have established the mean least it took monolinguals of both English and Portuguese with a college-level education to accurately judge the grammaticality of sentences in their L1.

• Objective 2 (Experiment 2)

The second goal of this study is to investigate the ability of the Spanish version of the VLT to discriminate L2 Spanish proficiency profiles in relation to an objective grammatical test (Spanish Placement Test - SPT) and to a Speeded Acceptability Judgment task among Spanish HL speakers who lived in the U.S, and Spanish L2 learners. The scores of SPT roughly meet a broadly accepted framework of L2 proficiency: the *Common European Reference Framework* (CERF) (Council of Europe, 2001). In the Speeded Acceptability Judgment task, both grammatical and ungrammatical sentences in Spanish were displayed as stimuli.

In order to analyze the data, we applied statistics tests such as the Kolmogorov-Smirnov test for normality, to guarantee that the population of this study is normally distributed, which allows us to use parametric tests. In order to test the variance between groups across the levels of proficiency, we apply an analysis of variance. In order to assess the ability of the proficiency tests accurately producing a diagnosis for proficiency, we applied the *ROC* (*Receiver Operating Characteristic*) curve. These statistic tests and procedures for analysis will be presented in the details of the methodological chapter. For this study we stablished the following hypothesis:

1.5 Research hypotheses

• Hypothesis 1: The VLT scores correlate with the placement test and with the SAJ task scores regardless the bilingual type.

• Hypothesis 2: The VLT and the AJ task are adequate diagnostic instruments for language proficiency in L2 English and Spanish HL.

• Hypothesis 3: High proficient individuals have a higher level of automaticity than low proficient ones in L2 English and Spanish HL processing, since they perform better in the speeded SAJ task.

In Table 1 below, the experiments are summarized:

Table 1: Summary of the research

	Experiment 1	Experiment 2
Objectives	• To investigate the ability of VLT to discriminate L2 English proficiency profiles in relation to the OPT and the AJ task; •To estimate the minimum timeframe for accurate sentence judgments.	• To investigate the ability of the Spanish VLT to discriminate L2 Spanish proficiency profiles in relation to the SPT and the AJ task
tests/tasks	•VLT, OPT, AJ task	VLT, SPT, AJ task, Self-Assessment, Bilingual Questionnaire
Participants	30 Brazilian Portuguese-English bilinguals	• 20 Spanish L2 speakers (English native speakers) • 40 Spanish HL speakers

Following this introduction, a theoretical background chapter, a methodological chapter, a data analysis chapter, and a conclusion will be presented. In the forthcoming chapter, we will present two models for the global architecture of grammar, highlighting the role of lexical items in their configurations. After that, we will present three models that contemplate lexical access/production in both L1 and L2. Following, we will define the construct of proficiency under the distinction of implicit and explicit knowledge in models of declarative/procedural memory. Finally, we will present the vocabulary size as a measure of overall L2 proficiency. Afterward, we continue on to describing the methods and the analyses of data we compiled in the present study. We finish with a presentation

of our interpretation of our findings, and a brief discussion of what we believe to be the ensuing steps to be taken in our research agenda.

CHAPTER 2

THEORETICAL BACKGROUND

In this chapter, we will provide a theoretical discussion on the syntax-lexical sematic interface, followed by the principles of the *construction grammar*. After that, we present models that proposes mechanisms in which the lexical item is formulated, accessed and processed. First, Levelts's (1989) *Speech Production Model*; then *Hartsuiker et al's model*, followed by the *Modular Online Growth and Use of Language* (MOGUL) framework. After that, we discuss the construct of proficiency under models of memory notions, pervading the concepts of declarative/procedural memory, implicit/explicit knowledge and automaticity. Consequently, we present an overview of vocabulary as a global measure of language proficiency and its correlation with other types of tests. Finally, we present some aspects and studies of the HL speakers.

2.1 The syntax-lexical semantics interface

The syntax-lexical semantics interface is a field of linguistic investigation in which it is assumed that there is interrelation between lexis and syntax as an effort to understand the parallelism of semantic (or thematic) relations with syntactic positions of expressions. Through this interface, the argument structure of the predicates plays a role in the syntactic configuration. According to Souza (2011, p. 155):

> The linguistic realization of arguments may be regarded as the transition between mental representations of concepts and the manifestations that emerge from them in morphosyntactic structures. Therefore, the semantics of argument realization is of crucial importance, and argument structure should be understood as a component of grammar in which there is an unquestionable interface between semantics and syntax.

An enlightening example of the semantic configuration playing a role in the syntactic arrangement is the arguments of the verbs *persuade* and *promise* in English[1] in

[1] For a detailed discussion on the thematic characteristics of these types of verbs, see Culicover & Jackendoff (2005, p. 419/420).

1(a) and 1(b). In both cases, the verb requires an NP and a CP (infinitival clause, in this particular case) as complements, thus the syntactic structure appears to be equivalent. However, the agentive value of the infinitival clause is determined by the semantic of the main verb (Culicover & Jackendoff, 2005, p. 419):

1(a). John$_i$ persuaded Sarah$_j$ to $_j$/*$_i$dance.
(b). John$_i$ promised Sarah$_j$ to $_i$/*$_j$dance.

According to Cullicover & Jackendoff (2005), this is a strong argument against the centrality of syntax in determining the control. In both 1(a) and 1(b), although the syntactic configuration seems to be similar, the controller is different. In 1(a), we notice that *Sarah* is the agent (controller) of the infinitival form "*to dance.*" Differently, in 1(b) the controller is *John.* In this example, the controller position is not dependent on the syntactic arrangement, since the difference between 1(a) and 1(b) relies on the verb meaning. Considering the conceptual structure level, the semantics of the verbs itself is able to affect the controller regardless the syntactic configuration.

Another example of syntactic restrictions imposed by verbal semantics is in psychological verbs. According to Cançado (1996) and Hsin e Lee (2009), psychological verbs convey the meanings of emotional states and present obligatorily the thematic role of *experiencer* as argument. However, psychological verbs differ on the syntactic position (subject/object) of the *experiencer* regarding the semantic of the verb. In *fear*, for example, the *experiencer* has to appear in the subject position (2a). If this condition is not satisfied, the sentence will be inadequate (2b). With *frighten,* on the other hand, the *experiencer* must appear in the object position (2c), since if it placed in the subject position, the sentence will be also inadequate (2d):

2(a) Jimmy feared thunderstorms.
(b)*Thunderstorms feared Jimmy.
(c) Thunderstorms frightened Jimmy.
(d) *Jimmy frightened thunderstorms.
(Souza & Oliveira, 2011, p. 107)

According Souza & Oliveira (2011, p. 107), in 2(a) and 2(b), sentences are ungrammatical due to the fact that "the entities referred to by respectively the grammatical subject and the grammatical object fail to match the semantic configuration specified by the verbs." These verbs are examples of the interface between a configuration placed by the meaning of the lexical item and its syntactic structural disposition. These particularities of the argument structures echoes in a theoretical framework conception that considers argument structures to be a subclass of constructions: the *Construction Grammar* (CG) (Goldberg, 1995, Souza & Mello, 2007).

Goldberg (1995, p.1) concisely describes constructions as the "form -meaning correspondences that exist independently of particular verbs". Such definition implies the idea that construction conveys meaning indifferent from the meaning each word itself carries. The two main ideas of the construction grammar are that (1) the construction should be considered the basic units of syntactic operation, not the lexical units themselves, and (2) these constructions are systematized in the speaker's mind as a conceptual entanglement. For example, in English, a speaker certainly understands the meaning of the verb "miss" and the noun "boat", however, the meaning of the expression "miss the boat" (idiom) cannot be obtained through the integration of both meaning of "miss" and "boat" separately.

In *Constructions*, Goldberg (1995) admits that the main argument encompasses a criticism towards the GG, especially on the lexicalist conception of language. According to the author, this 'bottom-up' perception, in which the lexical entity determines the syntactic organization into which it is inserted, fails to explain structures that present the same verb into quite different syntactic configuration. For instance, in this bottom-up view, the meaning of a sentence could be understood from the analysis of the particularity of the verbs. In this way, it would be necessary to add a new sense for a verb in every new syntactic structure such verb appears. Therefore, Goldberg (1995, p. 5) claims that constructions have to be conceived as a theoretical instance *per se: "*The collection of constructions is not assumed to consist of an unstructured set of independent entities, but instead it is taken to constitute a highly structured lattice of interrelated information."

The comparison of argument structures of bilinguals' languages has been an important line of investigation on bilingualism studies, since it is informative about bilinguals' representational state in both languages (Souza 2011, Souza & Mello, 2007).

In construction grammar, argument structure patterns are considered as construction (Souza, 2012). The specificities of argument structure patterns differ between languages and these differences play a role in bilingual's L2 representational state. According to White (2003, p. 206):

> The L2 learner must arrive at a representation for lexical items in the second language and must map from argument structure to syntax. Since there are crosslinguistic differences in argument structures, there will be cases where the L1 and L2 realize argument structure somewhat differently. In some cases, there is a potential for overgeneralization from the L1 to the L2.

According to Goldberg (1995), some argument structures such as ditransitive[2], resultative[3] and caused-motion[4] should be looked at, since "in these argument structures, the dissimilarities found in relation to the meaning with the same verb is credited to specific constructions" (Goldberg, 1995, p. 3). In terms of L2 studies, induced-movement (Souza, 2012; Guimarães, 2012), resultatives (Oliveira, 2013) and ditransitive (Souza, Soares-Silva & Silva, 2016) have been fruitful to the understanding of how bilinguals process such in the L1 and L2. Moreover, they are informative about the learnability of such structures. In this work, we call the attention to the induced-movement alternation in Brazilian Portuguese and English.

The argument structure of some verbs in English and Brazilian Portuguese do not share the same configuration, for example, the pair *run/correr* (unergatives). Unergatives and unnacusatives are types of intransitive verbs, the former has an agentive subject who deliberately acts (for instance, *run, walk)*, as opposed to the latter, in which the action played by the subject is not necessarily deliberate (for instance, *hear*, *die, fall)* (Souza, 2012).

In English, *run* requires only one argument (agent) when it is in the intransitive form (*John ran*). Alternatively, it can be transitivized by receiving accusativity, thus requiring two arguments in a process called induced movement alternation (Levin, 1993; Levin & Rappaport-Hovav, 2005) (*John ran the guinea pigs through the maze*). The

[2] "Joe promised Bob a car." (Goldberg, 1995, p. 75)
[3] "Pat hammered the metal flat." (Goldberg, 1995, p. 81)
[4] "She allowed him into the room." (Goldberg, 1995, p. 84)

equivalent *correr* in Brazilian Portuguese only accepts the intransitive version. This way, a sentence such as *João correu as cobaias pelo labirinto* is not licensed.

There are several empirical studies investigating the influence of these differences in bilinguals' sentence processing in both L1 and L2. For example, Souza (2011), through a psycholinguistic study on bilingual sentence processing, demonstrated the importance of this relation by corroborating the effects of linguistic transference in the bilingual's mental representation of the induced-movement alternation. Four groups took part in the experiment: native speakers of Brazilian Portuguese who had low-proficiency (N=19) and high-proficiency (N=20) in English (L2), Brazilian Portuguese monolinguals (N=10), and English monolinguals (N=10).

All participants took an acceptability judgment task in order to assess three types of grammatical structures with the same verbs. The target structure was sentences with induced movement alternation (*The psychologists ran the rat through the maze*). The structure was also displayed sentences with a prepositional phrase as adjunct (*The psychologists ran with the rat through the maze*), and constructions with the verb *make* (*The psychologists made the rat run through the maze*). The result showed that proficiency level plays a role in the process of identifying the induced movement alternation structures in L2. English L2 speakers with higher proficiency were significantly better in recognizing the grammaticality of these sentences. It must reflect the influence of L1 configuration L2 representation access; moreover it suggests the learnability of the structure.

In other study investigating the influence of L2 argument structure in L1 online processing, Souza & Oliveira (2011) administered a self-paced reading with grammatical sentences in English with induced-movement alternation verbs (*The researcher ran the mice through the maze)* and the equivalent structure in Portuguese forcing the accusativity (*O pesquisador correu os ratos pelo labirinto*). Three groups took part on the experiment: Brazilian-Portuguese monolinguals reading sentences in Portuguese (N=9); bilinguals (Brazilian-Portuguese/English) reading sentences in English (N=9), and bilinguals (Brazilian-Portuguese/English) reading sentences in Portuguese (N=9). Results suggested that bilinguals reading sentences in Portuguese tend to accept the forged accusativity of intransitive verbs (even these sentences are ungrammatical) more than the monolinguals. This may suggest an interference of L2 configuration in L1 processing.

In order to better understand this interface between the lexical semantic with the syntax, it is crucial to comprehend the lexical item since its formulation to its articulation, especially the notion of *lemma.* Usually, models for lexical access/production encompass a design of a combinatorial-nodes processing machine in which the first stage of lexical access involves only *lemma* selection, the phonetic-shaping stage is not activated yet. This definition of lemma is an echo of Levelts's (1989) Speech Production Model. Another model comprising this modular notion and based on Levelt's postulations, but which is elaborated to structure bilinguals' lexical access, is the Hartsuiker et al's. (2004) Model. Along the same lines, but proposing a broader definition of the linguistic arrangements as part of a general cognitive system covering constructs such as memory, consciousness, attentional focus, and access competition, the Modular Online Growth and Use of Language (MOGUL) is a framework that attempts to merge linguistics theory with other cognitive domains. Following, the three models will be presented.

2.2 Models for lexical access and production

2.2.1 Speech Production Model

The vast literature on language acquisition has provided studies with several models for lexical production and processing. According to the main lexical access models, the access to the lexis is a process that must take place in two stages. In the first stage, the lexical item related to the dimension of meaning is activated/selected from a group of lexical nodes belonging to the same semantic field. In the second stage, the phonological properties shape the word, and then it leads to word articulation. Such models are known to be discrete since the stages are separated.

An example is Levelt's (1989) proposal. Through a model of lexical processing and production, Levelt postulates that speech production consists of three general modules: Conceptualizer, Formulator, and Articulator. The Conceptualizer is responsible for managing the ideas to be transmitted, and arranges them within a formal system. The Formulator is more complex and has two primary components: a grammar encoder, which generates the sentence pattern, and a phonological encoder, which produces a phonetic plan (generating output). The Formulator accesses the lexicon containing every word

known by the speaker. Such words are shaped as *lemmas* (essentially, a set of knowledge about the meaning of a given word) and *forms* (containing similar knowledge of morphology and phonology of the word). The final module is the articulator, which transforms the phonological instance into something explicit (audible) in order to be accessed and assessed (Figure 1). Due to its discreetness, Levelt's model is considered serial and incremental (De Bot, 1992).

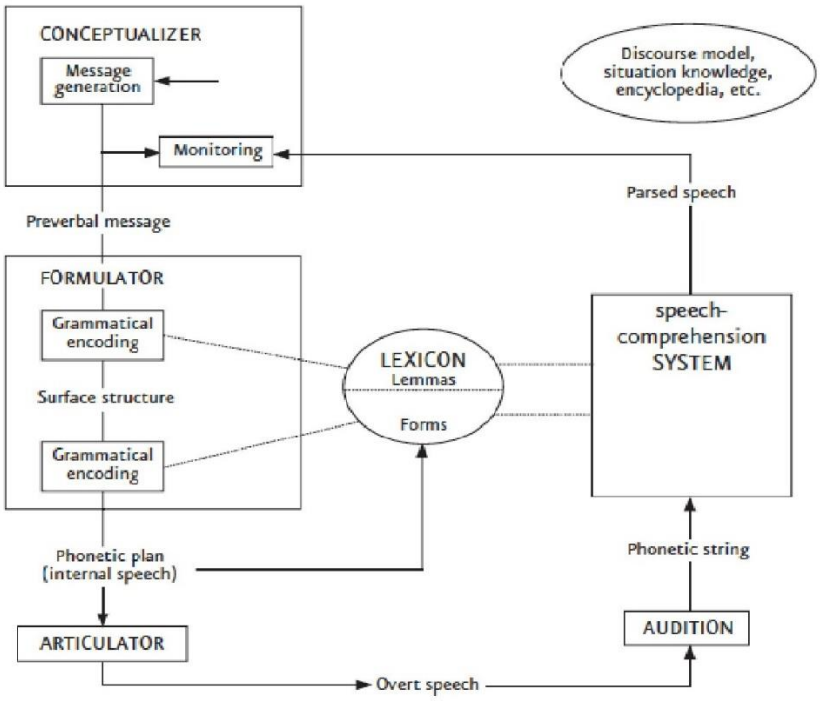


Figure 1 – Blueprint of Levelt's model for the speaker (LEVELT, 1989, p. 9)

A remarkable concept from Levelt's model is the binarity of lemmas. A lemma is the word part that indicates basic meaning, and establishes both syntactic category and argument structure. Levelt (1989, p. 191) exemplifies with the lemma *give* which has the conceptual specifications: CAUSE (X GOposs (Y, (FROM/TO (X,Z)))); Conceptual arguments: (X, Y, Z); the syntactic category: Verb; the grammatical functions: Subject, Direct Object, Indirect Object; the relation to complementizers: none in this case; and the diacritic parameters: tense, aspect, mood, person, number, and pitch accent. To Levelt (1989, p. 181), the grammatical and phonological encodings are mediated by the lexical entries, which is why the entire process is considered lexically driven:

> [...] grammatical and phonological encodings are mediated by lexical entries. The preverbal message triggers lexical items into activity. The syntactic, morphological, and phonological properties of an activated lexical item trigger, in turn, the grammatical, morphological and

> phonological encoding procedures underlying the generation of an utterance.

Ten years after the first version of the model, Levelt et al. (1999) elaborated a broader overview of the model, including a conceptual network in which nodes are selected in the conceptual level. A remarkable enhancement to the model is that Levelt et al. (1999) attribute more value to the semantic role in the node selection. Therefore, as a conceptual node is activated, several others belonging to the same semantic field are equally selected in a process called spread activation:

> A core feature of the theory is that lexical selection is conceived of as selecting the syntactic word. What the speaker selects from the mental lexicon is an item that is just sufficiently specified to function in the developing syntax. To generate fluent speech incrementally, the first bit of lexical information needed is the word's syntax. Accessing word form information is less urgent in the process. (Levelt et al. 1999, p. 14)

According to the conceptual network (Figure 2), once e*scort* is selected, it is also established that it has transitivity and two arguments. The syntactic properties are available automatically together with lemma selection. In the lemma stratum, the diacritic parameters (person, number, tense, and aspect) are set up. Finally, in the form stratum the word is shaped with the phonological and articulatory features.

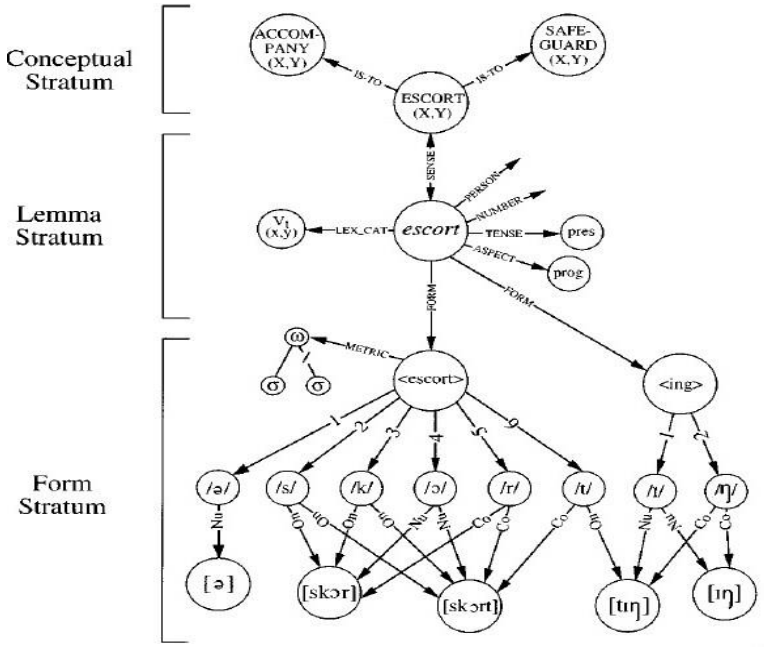


Figure 2 - Fragment of the lexical network underlying lexical access of e*scort.* (Levelt et al., 1999, p. 4)

Although Levelt's model is clearly a description of L1 production, its scheme can be applied to L2 acquisition studies. De Bot (1992) proposed the version of Levelt's model as an attempt to structure the speaking production of bilinguals, what the author himself claimed to be a necessary endeavor back then. De Bot (1992) proposes that in bilinguals' speech production, differently from Levelt's notion, the deep level of the conceptualizer is not language specific. In the formulator level, a single lexicon is in charge of storing information of both languages. An important adaptation proposed by De Bot (1992) is the fact that a lemma can be linked to various form characteristics in both languages. Another model that covers the bilinguals' processing and which is based on Levelt's speech production is Hartsuiker et al.'s model (2004).

2.2.2 Hartsuiker et al.'s Model

The Hartsuiker et al.'s Model (Hartsuiker et al. 2004) comprises the notion that convergent syntactic structures in L1 and L2 are shared in a single representation. A similarity between this model and De Bot's (1992) proposal (and consequently Levelt's model) is the definition of lemma. According to Bernolet et al. (2013, p. 288), in this model "the lemma stratum thus contains lemma nodes (corresponding to the base forms of words) from both languages which are connected to language nodes."

Based on cross-linguistic **s**yntactic priming, Hartsuiker et al. (2004) tested Spanish/English bilinguals describing Figures. Twenty-four Spanish/English bilinguals (who lived in an English language country) took the task. Results revealed that when participants heard a sentence in a given language, they described the next picture in the other language, however using the same structure. In Figure 3, the blueprint of the model is displayed, highlighting the shared syntactic representation.

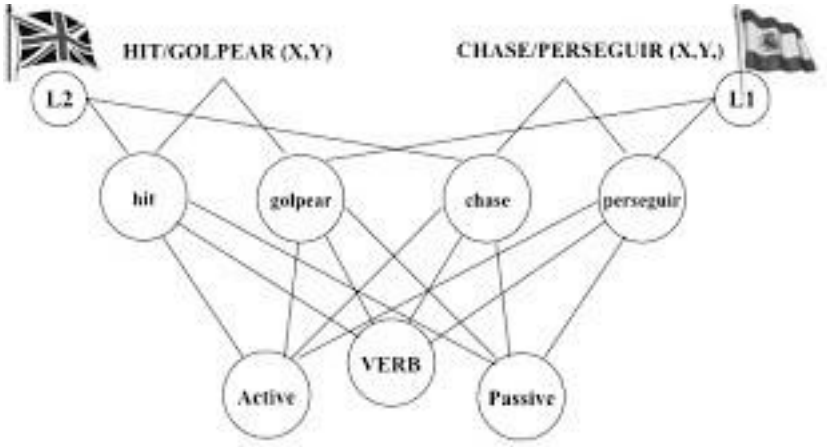


Figure 3- Example of lexical entries for ‘‘to chase’’ and ‘‘to hit’’
(Hartsuiker et al.’s (2004)

As presented in the model’s blueprint (Figure 3), there is a connection among lemma node, conceptual node, category node, combinatorial nodes, and language node. Agreeing to Levelt’s model, Hartsuiker et al.’s model is lexically driven, although it presents combinatorial nodes that are not dependent of a specific language, since “the activation of a grammatical structure in itself does not determine the language of an utterance. Instead, the language of the utterance is dependent on the choice of lexical items” (Bernolet , 2008, p. 22). Hartsuiker et al.’s model claims that the assumption of shared syntax can explain some interesting bilingual situations:

> According to the shared-syntax account, rules that are the same in the two languages are represented once. […] Even if there are some grammatical differences between the languages (such as the presence or absence of a preposition), the bilingual could represent the shared aspects of the construction once, and store additional language-specific information as necessary. (Hartsuiker et al. 2004, p. 409)

Levelt’s model (1989), De Bot’s adaptation to bilingual production (1992), and Hartsuiker et al.’s (2004) model appear to corroborate the notion of a non-language specific lexicon, in which the lemma activation also trigger its syntactic features to be accounted when the lexical item is selected. Moreover, in both models, the lexical item appears the locus where the interface among the elements takes place, agreeing with the *PA* assumptions.

2.2.3 The Modular Online Growth and Use of Language (MOGUL)

The MOGUL (Truscott & Sharwood Smith, 2004; Truscott, 2015) is a framework for language processing that adopts the notion of modularity. The MOGUL is fundamentally based on the *Acquisition by Processing Theory* (Truscott & Sharwood Smith, 2004), which assumes that linguistic modules are created as a natural consequence of processing. The main goal of this theory is "to build a cross-disciplinary platform which can bring together research on linguistic structure and research on general cognition" (Truscott & Sharwood Smith, 2004, p. 1).

According to MOGUL's principles, language system is understood as a domain of consciousness, as it fits into the cognitive system as a whole. This assumption appears to be an attempt of merging different theoretical apparatus from linguistic theory, which are employed to explain the language system, with constructs from cognitive and psycholinguistic constructs such as memory, competition, and activation levels. Different from other approaches that conceive the linguistic system as non-modular, in other words "nothing especial" (Truscott, 2015, p. 414), MOGUL adopts a modular notion to cover the combinatoriality of the system's components, conceiving them as a processing chain. The possibility of autonomy for each module, as well as a non-hierarchical interface among them, echoes Jackendoff's (1997, 2002) conception of linguistic representation. In Truscott's (2015, p. 419) words:

> The MOGUL framework assumes a modular architecture [...] consisting of processor–store pairs and interfaces connecting stores. Each such pair is a module in that it serves a specific function, based on innate constraints. The processors construct representations on their stores, by combining representations already present there.

MOGUL offers grant for SLA research on the shared-processing assumption. As stated in MOGUL's definitions, two knowledges are constantly striving for engaging in a single processing system in bilinguals' mind. In this way, elements like inhibitory control are taken as a mechanism to determine which language will be processed. Besides having Jackendoff's theory as a point of departure, MOGUL also echoes the Levelt's production model on the modular system structure in which lexical access/production follows sequential steps (Figure 4). Moreover, in the definition of bilingual competition, MOGUL resembles Hartsuiker et al.'s model in the sense that "when one lexical entry is

accessed there is a varying degree of activation for a host of competing candidates in all available language systems" (Truscott e Sharwood Smith, 2004, p. 13).

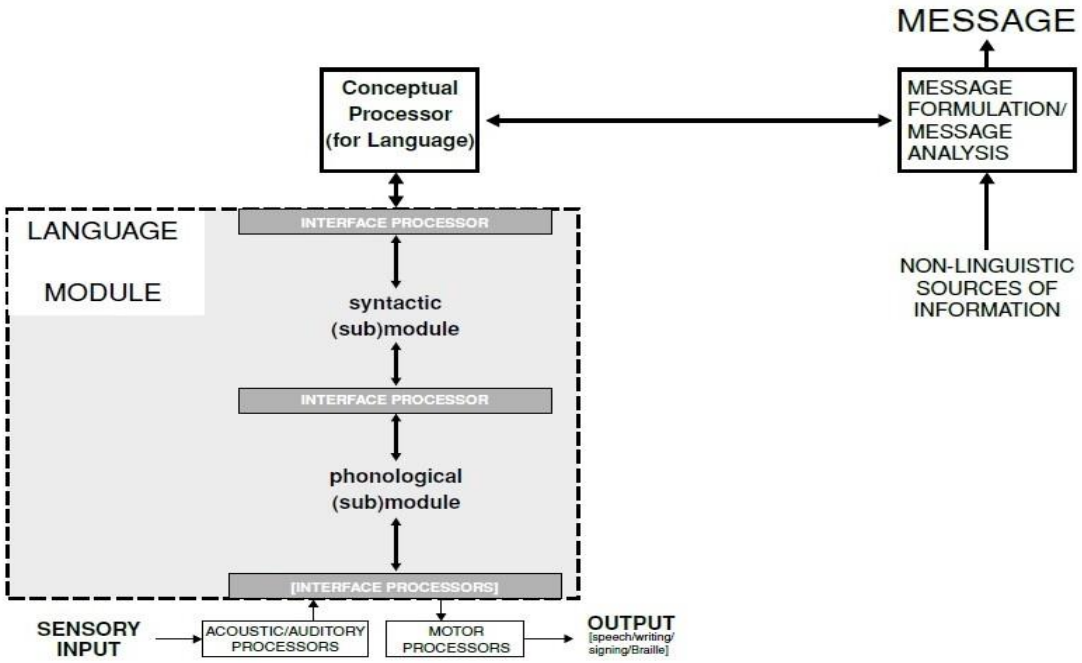


Figure 4 - The language module and adjacent cognitive systems: a first sketch.
(Truscott &Sharwood Smith, 2004, p. 2)

A preeminent endeavor of MOGUL is establishing the role consciousness in SLA, bringing to light Krashen's (1981) distinction between the unconscious process of acquiring and the conscious process of learning. Truscott (2015) affirms that the role of consciousness in SLA has to involve seven aspects: modularity, activation, contents of a short-term memory store, executive control, attention, value, and information, as displayed in Figure 5:

modularity	The mind is a made-of-units system, each unit has its specific function at the same time they serve an entity.
activation	Any component in the realm of consciousness is active and available to be accessed.
contents of a short-term memory store	The elements we are conscious of are those stored in short-term memory.
executive control	"Self is crucially involved in consciousness, as this involvement is a form of executive control and self can reasonably be seen as the ultimate executive" (Truscott, 2015, p. 417).
attention	There is a reciprocal relation between elements in our conscious and elements we address our attentional focus.

value	This biological notion assumes that mind would be selective according to what is important to it; it can be seen as matter of selection and evaluation. Truscott and Sharwood Smith (2011, p. 513) name this value as Activation Hypothesis, which means that "a representation is conscious if and only if its current activation level is above a given threshold value."
information	As a knowledge sample becomes conscious, it brings information that matters for the whole system, it they do not carry this load of information, they are erase from consciousness.

Figure 5- Aspects involved in the role of consciousness in SLA (adapted fromTruscott, 2015)

As we have seen so far, Levelt's model (1989), although alluding to lexical access/production of L1, offers us a common ground for the comprehension of L2 lexical access/production. Through the same conception of combinatoriality and assuming a shared processing engine, Hartsuiker et al.'s (2004) model concurs with the notion of in the conceptual level; lexical items are not language specific. Finally, by conceiving language as conscious, therefore influenced by the cognitive system as a whole, MOGUL incorporates the cognitive trait to language acquisition/processing construal. The three models presented are the basis for our definition of L2 proficiency is this work (which will be presented in the next subsection) as they encompass aspects of cognition which we accredit to be essential to a broader view of the proficiency construct.

2.3 L2 Proficiency as a construct swayed by models of memory and automaticity

Ultimately, L2 proficiency is a challenging construct to conceptualize and measure objectively. Notwithstanding, measurement of bilingual speakers' differential proficiency profiles is a matter of absolutely critical importance for the psycholinguistics of bilingualism. Because of the eminently experimental base of research in psycholinguistics, comparability and replication of results are fundamental for the advancement of the field.

Grosjean (1998) states that one of the difficulties that jeopardizes consensus in bilingualism studies is the lack of standardized procedures for describing and measuring differences in profiles across bilingual populations from which samples are drawn. Bilinguals' linguistic proficiency in both dominant and non-dominant languages is one of such relevant profile differences, according to the author. In addition, as pointed by Hulstijn (2012), some kind of measure of linguistic skill level is quite often taken as the

main—if not the single—independent variable of experimental studies in L2 acquisition and bilingual language processing. This fact alone should justify L2 scholars' careful theoretical consideration of which facets of L2 proficiency are selected for observation.

The relevance of the conceptualization of L2 proficiency is topped with the need for careful planning of how to measure it in efficient and practical manners in L2 acquisition and processing studies. Issues of the practicality of proficiency assessment administration haunt designers of language tests for educational and accreditation purposes. Of course, in experimental laboratory work such issues may be even more critical, as investigators can seldom afford the time required for administration of complex test batteries.

This scenario often leads researchers to employ sections or subtests within standard test batteries, or to use scales constructed to diagnose proficiency by measuring a single dimension, or but a few dimensions of the overall construct. It is our understanding that the problems with variability in psychometric instrumentation may be much worsened by lack of validation studies of the scales that researchers employ. Considering proficiency as a multidimensional construct, we understand that the validity of scales targeting specific dimensions to discriminate profiles in accordance to the variability in linguistic representation and processing is ultimately an empirical question.

There are several angles covering the definition of proficiency in L2, most of them converging with the idea of the capacity of using the language fluently in various communicative situations. Due to its multidimensional aspect, proficiency is often perceived as the product of manifold components rather than a singled-out definition (Bachman, 1990; Hulstijn, Anderson & Schoonen, 2010). By virtue of this multifaceted construct, different measures of proficiency are generally correlated in order to establish a more cohesive concept. In this way, tests covering vocabulary size, grammar knowledge, or reading skillsare compared and interrelated in order to establish a broader measure.

In order to define the construct of L2 proficiency, we must define the nature of knowledge involved: a metalinguistic knowledge about the language (explicit, declarative) or an inherent knowledge (implicit, procedural). Both implicit and explicit knowledge are different constructs and implicate in different conceptions of proficiency, as they encompass different loci in memory.

2.3.1 L2 implicit/explicit knowledge, automaticity and L2 proficiency

According to Ellis, R. (1994), there are two kinds of L2 knowledge to be accessed and assessed: explicit and implicit knowledge. The explicit knowledge is the one learners are able to talk about, to explain and describe what and why he is using such structures (Ellis, 2005, Bowles, 2011). Oppositely, implicit knowledge is intuitive, since it belongs to the realm of unconsciousness. Along these lines, learners are able to use the language but not explain it (Paradis, 1994, Ellis, R. 1994).

The explicit knowledge is also associated with controlled processes (Ellis, R. 1994). In such processes, learners are aware of the knowledge involved, as they are able to reflect upon their insights on the linguistic structure being used. In controlled processes, learners *know about* the language, in the sense that they are able to provide description and explanations for linguistic facts that appear in their production/perception. In order to access explicit knowledge, bilinguals' minds have to resort to cognitive strategies as mental appliances to build up their comprehension/production. Such operational mechanisms have a cost to the memory, which causes controlled processes to be slower than automatic ones.

Contrastively, the implicit knowledge is associated with automatic processes (Paradis, 1994). Automatic processes occur when linguistic rules, sounds, forms are embodied in the bilingual's mind. In this sense, learners *know* the language. Automatic processes are intuitive and unconscious by nature, as learners are able to use/understand pieces of language they would not be able to explain. Due to the fact that it is automatic, these processes are less costly to the memory, as less cognitive strategies are needed in the routine to the production/comprehension of the language. Ellis (1994) proposes a definition for types of knowledge in function of types of processing, as we can see in Figure 6:

Types of Knowledge	Types of Processing	
	Controlled	Automatic
Explicit	A new explicit rule is used consciously and with deliberate effort.	An old explicit rule is used consciously but with relative speed.

Implicit	A new implicit rule is used without awareness but is accessed slowly.	A thoroughly learnt implicit rule is used without awareness and without effort.

Figure 6– Types of Knowledge (Ellis, R. 1994, p. 86)

As presented in Figure 6, automatic language processing has been defined as opposed to controlled processing, with the latter being pricier to the memory than the former. Conforming to R Ellis (1994) and Segalowitz (1991), controlled processes are also based on strategic decisions, and they are more time consuming than automatic processing.

It is important to highlight that a high level of automaticity does not necessarily implicate in implicit knowledge. According to Segalowitz (1991) and Segalowitz (2006), a high reaction time on speeded tasks in the L2 could represent acceleration for controlled processes, thus not meaning automaticity. In these cases, automaticity means a quantitative process involving learners becoming faster in a task due to extensive practice (Jiang, 2007; Paradis, 2005). As a matter of becoming more accurate in a task, "automaticity entails just better, more efficient processing of the same kind as occurs when performance is not automatic" (Segalowitz, 2006, p. 386). Segalowitz (2006) presents a series of possibilities to operationalize the construct of automaticity regarding processing. According to him, automaticity can be construed based on fast, ballistic, load independent, effortless, unconscious, and shift-to-instance processing.

Although Segalowitz & Hulstijn (2005, p. 371) claimed that automaticity is "the prime psychological construct invoked for understanding frequency effects and how repetition leads to improvement in L2 skills," it must be clarified that we are conceiving automaticity as simply playing a role in language proficiency, and not paralleling with implicit learning of the L2. Moreover, like Ellis (2005), we assume that automaticity can be applied to access both implicitly and explicitly learned linguistic representations.

The automaticity as a facet of L2 proficiency is properly defined in Clahsen & Felser's (2006) exposure of findings from ERP studies. By discussing the role of L1 transfer, cognitive resource limitations, and maturational constraints, Clahsen & Felser (2006) assume that highly proficient L2 learners are able to reach a degree of automaticity close to that of native speakers in some subdomains of grammar. However, the difference between L1/L2 language processing remains distinctive in more intricate syntax.

In line with this multidimensional aspect of proficiency, Hulstijn (2011) proposes a model in which language proficiency (L1 or L2) can be allocated in *basic* and *higher* language cognition. Hulstijn's (2011) definition of basic language cognition (BLC) is consistent with the declarative/procedural model presented by Ulman (2001) (more of which later). This consistency has to do with the twofold notion of implicit computations of linguistic knowledge (procedural) and lexical representations stored in declarative memory. Contrarily, higher language cognition (HLC) embodies low frequent and more complex lexical items and morphosyntactic structures.

Hulstijn's (2011) model also assumes that, differently from L1 speakers who reach ceiling proficiency in BLC, high levels of HLC fulfillment among L1 speakers rely on the results of individual cognitive capacity and learning opportunities through life. In contrast, in the L2 dimension, the question concerning the full attainment of BLC (post-puberty) is still in discussion. The variability of bilinguals' L2 proficiency attainment is due to several factors including age of onset of L2 learning, aptitude, overall cognitive capacity, learning circumstances, and motivation, among others (Harley et al., 1990; Dörnyei, 2005).

According to Hulstijn (2011), BLC elements are added by high automaticity in processing. The author claims that automaticity would be related to the notion that BLC manifestation is usually high-frequency forms that are shared across different genres and registers. In the literature, automatic language processing is conceived as opposed to controlled processing, in the sense that the former depends upon a harder attentional effort than the latter (Segalowitz, 1991; Segalowitz & Hulstijn, 2005). Departing from this notion, we align ourselves with Segalowitz & Hulstijn (2005) in assuming that automaticity can be construed as a facet of high L2 proficiency, considering the understanding of language processing as relying on limited cognitive resources. If high L2 proficiency equates with the ability to master L2 in complex tasks, consequently such capacity requires aspects of the processing routines to be less demanding than those required for reasoning and reflection upon situation assessment.

Hence, according to Hulstijn (2011), at earlier stages in the ontogenesis of L2 BLC, L2 speakers would present less automatic processing than in later stages. As stated by Segalowitz & Hulstijn (2005, p. 371), automaticity is "the prime psychological construct invoked for understanding frequency effects and how repetition leads to improvement in L2 skill." We align ourselves with Hulstijn's proposal, as we understand automaticity in

processing to be an essential component of high L2 proficiency. Therefore, in the present study, we attempted the operationalization of some degree of automaticity in our tasks, a procedure we believe should be incorporated into any endeavor to analyze and measure L2 proficiency.

In empirical-based bilingual studies, the accountability of implicit/explicit knowledge distinction has been explored under several aspects in the sense of how to capture both forms of knowledge in behavioral tasks or on-ine processing procedures (Jiang, 2007). In some studies (Ellis, R. 2005, Hulstijn, 2011, Souza & Soares-Silva, 2015) implicit knowledge is seen as perceivable through measures of the learners' automaticity in psycholinguistic tasks. Since automaticity is a concept clearly related to time of processing, a reasonable way to measure learners' automaticity is by applying a time pressure in the task. According to Jiang (2007, p.7), "the underlying rationale is that the application of explicit knowledge takes time and, thus, will be discouraged or minimized when participants perform a task under time pressure." Therefore, one way to assume the degree of learners' L2 automaticity is by comparing their reaction time to native speakers.

A very frequently used technique in psycholinguistics studies to measure implicit L2 knowledge is the speeded Acceptability Judgment Task (AJT) (Ellis, N. 1993; Ellis, R. 2005; Jiang, 2007; Souza et al. 2015; Souza & Soares-Silva, 2015). AJT is an offline procedure in which participants are told to react to a series of stimuli presented on a computer screen. Offline means that data is collected after being processed by participants, and we only have access to what they respond as their interpretation—not to the exact moment that the processing is taking place (online tasks). The speeded version of the AJT attributes a time pressure to the task; thus, participants are given a limit to judge the sentences. Usually the time limit is statistically established *a priori* based on how much time native speakers need to judge the sentences.

Some authors (Ellis, R. 2005, Souza & Soares-Silva, 2015) consider the speeded AJT to capture learners' L2 implicit knowledge as the time limit deprives them from relying on metalinguistic knowledge or cognitive strategies to emit the judgment for the sentences. Therefore, if a participant does very well in judging sentences under a time pressure, we assume the knowledge being measured is implicit. Ellis, R. (2005) reported a study in which he tested both explicit and implicit L2 knowledge, by applying a non-

timed version of AJT to the former, and a timed version of AJT to the latter. Through the comparison with other tests and using statistic measures, results revealed that the speeded version of the task is able to apprehend L2 knowledge in the implicit level. Studies as reported by R Ellis are the theoretical and methodological basis for the administration of speeded AJT in this work.

As we can notice, both twofold concepts of controlled/automatic processes and explicit/implicit knowledge are directly related to models of memory that comprise instances of intuitive and self-explanatory knowledge. A model into this poles-apart knowledge conception is the Ulman's (2001) neurocognitive Declarative/Procedural Model of Lexicon and Grammar.

2.3.2 Ullman's neurocognitive Declarative/Procedural Model of Lexicon and Grammar.

Based on the twofold definition for language capacities, in which there is a mental lexicon (for lexical items) and a mental grammar (for rules that organize these items), Ullman (2001) relies on neural basis to propose a model for memory that covers this distinction: The Declarative/Procedural Model of Lexicon and Grammar. Ullman posits a distinction between a memory system, where "transformations' phonological and conceptual–semantic mappings are learned, stored, and computed" (p. 41); and a rules system, in which the rules "are represented as mental knowledge and implemented by mental operations" (p. 42). Ullman labeled the former as *declarative memory* and the latter *procedural memory.*

According to the model's proponent, the procedural memory is more confined in the sense that, differently from declarative memory, is not affected by other processes. Because the procedural memory regulates the learning and maintenance of cognitive skills, it holds the grammar rules while the declarative memory, which controls semantic knowledge (facts) and episodic knowledge (events), holds the semantic and common knowledge of words (p. 45).

One is a memory system implicated in the learning, representation, and use of knowledge about facts ("semantic knowledge") and events ("episodic knowledge"). The other underlies the learning and expression of motor and cognitive skills and habits. It is

argued that the first system's role in facts and events extends to words, whereas the second system's role in skills and habits extends to grammatical rules. In Ullman's model, the lexical item is seen as part of the declarative memory, since its meaning (semantic part) is likely to be held in a semantic locus. Moreover, Ullman (2001) presents neurological evidence that a patient with amnesia, which affects the brain area argued to store the words, presents problems in learning vocabulary. In a nutshell, the Declararative/Procedural model can be describe in Figure 7:

Declarative Memory	**Procedural Memory**
Rooted in temporal lobe structures;	Rooted in frontal/basal-ganglia structures
Appears to be specialized for associative binding;	May be specialized for sequences
Might not be informationally encapsulated	Appears to be largely informationally encapsulated
Underlies not only the learning and explicit (conscious) use of facts and events, but also of the sounds and meanings of morphologically simple and complex words—that is, the mental lexicon.	Underlies the implicit (unconscious) learning and expression not only of motor and cognitive skills and habits, but also of grammatical rules, in both syntax and morphology.

Figure 7: Summary of Declarative/Procedural model, adapted from Ullman (2001, p. 47)

As Ullman's model seems to claim, declarative memory holds a type of knowledge that is conscious, therefore explicit, while the procedure memory encompasses unconscious and therefore implicit knowledge. The question on how and if these two knowledges interact in a learner's mind have been a crucial empirical question to bilingualism studies. As R. Ellis (1994) argued, both controlled and automatic processes could occur in the two types of knowledge (Figure 6 above), however the notion of consciousness appears to be attached only to explicit knowledge. There are several divergences in literature on how both types of knowledge interact.

2.3.3 The interaction between implicit and explicit knowledge

There are well-defined positions claiming that there is no interface between both types of knowledge, and both knowledges are separated (Hulstijn, 2002, Paradis, 2004); A different kind of theorists argue that there is a strong interface in such a declarative knowledge that can be 'proceduralized' by practice (Anderson, 1982, Sharwood Smith 1981). A third view defends a weak interface, it means that both knowledges interact, but some rules must be settled (Ellis, R. 1994). We now continue to describe the three notions.

2.3.3.1 The strong interface hypothesis

Mainly represented in the eighties by Anderson (1982) and Sharwood Smith (1981), the strong interface hypothesis claims that there is a based-on-practice relationship between explicit and implicit knowledge. The authors claim that the declarative knowledge once subjected to extensive practice can be automatized at the point of becoming implicit. According to Sharwood Smith (1981), different from Krashen (1981), there is no distinction between learning and acquisition. Although the former entails formal instruction and in Ullman's term belong to the declarative system while the latter comprises unconscious and intuitive, both terms are interchangeable and they are not poles apart.

According to DeKeyser (1997), the knowledge of a language resembles any other cognitive ability in the sense that first it is acquired through instruction and it is practiced under the monitoring of the speaker. However, with practice, this knowledge is internalized turning into a habit. Agreeing with this idea, Gass & Selinker (2008, p. 245) argue that "regardless of what one is learning (e.g., language or tennis), learning progresses from knowledge that (declarative) relating to some skill or behavior to knowledge of how (procedural), and finally to automatization of procedural knowledge.

2.3.3.2 *The weak interface hypothesis*

The weak interface proposes a less antipodal position by arguing that explicit knowledge may become implicit through delimited conditions. Ellis, R. (1994) assumes that both a strong and a weak interface fall into important misconceptions. The author suggests a weak interface, in which declarative (explicit) knowledge can evolve to procedural (implicit) knowledge regarding bilinguals' L2 representational state. In this sense, learners' proficiency, as construed as involving both types of knowledge and reflecting the state of learners' interlanguage development, plays a crucial role in this shallow-to-deep knowledge routine:

> Explicit knowledge derived from formal instruction may convert into implicit knowledge, but only if the learner has reached a level of

> development that enables her to accommodate the new linguistic material. In such cases, the learner's existing knowledge constitutes a kind of filter that sifts explicit knowledge and lets through only that which the learner is ready to incorporate into the interlanguage system" (Ellis, R. 1994, p. 88).

In the weak interface, the formal instruction, which deals more directly with declarative knowledge, has the role of assisting learners to maintain their implicit knowledge and increase the automaticity of the explicit knowledge. In this sense, according do Ellis, R. (1994), explicit knowledge becomes implicit through three consequential moments. First, the learner notices a particular linguistic aspect during the input (*noticing)*, after that he analyzes it through the features he has already learned (*comparing*). Finally, he conceives new assumptions and inferences aiming at embodying the new features into his interlanguage system (*integrating*). Figure 8 below demonstrates the role of explicit knowledge.

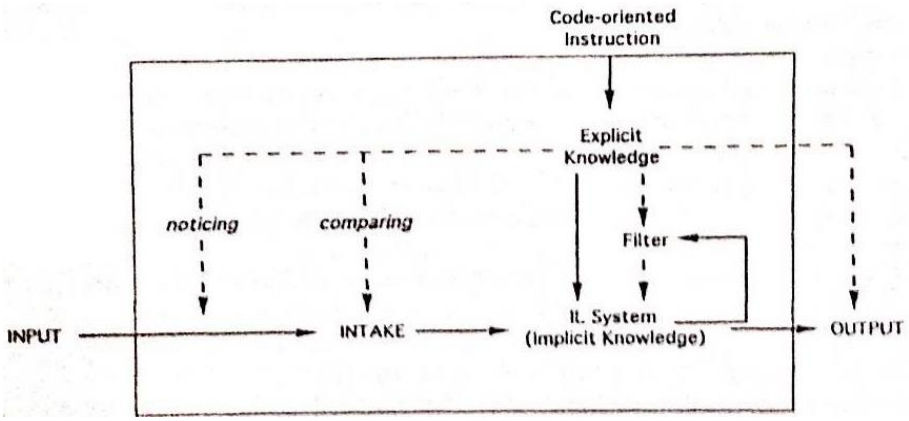


Figure 8 – The role of Explicit Knowledge in L2 Acquisition (Ellis, R. 1994, p. 97)

Another function of explicit knowledge is to 'facilitate' the learning of L2 since it helps learners to settle the association between form and meaning, and to serve as a problem-solving tool to learners (Ellis, N. 2005). Although the weak interface hypothesis assumes the possibility of the transformation of explicit knowledge into implicit, it will depend on the cognitive development of bilinguals. Thus, the path from implicit to explicit knowledge will occur only if the developmental rules in the learner's brain are appareled enough to bear a L2 level of interlanguage representation.

2.3.3.3 The non-interface hypothesis

According to Krashen's *Monitor Theory* "adults have two independent systems for developing ability in second languages, subconscious language *acquisition* and conscious language *learning"* (Krashen, 1981, p.1). These two processes entail different mechanisms and they are separated by nature. For instance, in a routine of language acquisition, individuals are not concerned with the rules and restriction of a given language; conversely, they are focused on the meaning the message conveyed. In this sense, "grammatical sentences 'sound' right, or 'feel' right, and errors feel wrong, even if we do not consciously know what rule was violated" (Krashen, 1982, p. 10).

In contrast, language learning is a conscious process in which learners can describe the language structure, which means "knowing the rules, being aware of them, and being able to talk about them" (op. cit., p. 10). In learning, the knowledge about the language serves as a tool for error correction and a conscious evaluation of language rules. According to Krashen (1981, 1982) both systems are not a continuum in which one can turn into another, contrarily, they remain separate regardless the level of bilinguals' linguistic knowledge of each language.

Agreeing with Krashen's distinction between conscious and subconscious knowledge, Paradis (1994, 2004, 2007) departs from neurolinguistics data from studies involving amnesia and aphasia to propose that the subconscious (implicit) and the conscious (explicit) knowledge involves two different types of memory (procedural and declarative) and are localized in different brain areas. According to Paradis (1994, p. 339), "the memory system that subserves the formal learning of a second language (declarative memory) is neurofunctionally and anatomically different from the one that subserves the first language or a foreign language acquired in conversational settings (procedural memory)." In this work, we align ourselves to Paradis' notion of non-interface hypothesis, since it offers both theoretical and methodological basis to our empirical endeavor.

Differently from the strong and the weak interface, Paradis (1994, 2004, 2007) assumes a non-interface hypothesis and claims that a declarative knowledge will never become procedural, even in the face of ostensive practice. Thus, the declarative knowledge can become more automatic as practice goes by and it provides learners with more agility in assessing the language rules. However, this knowledge will remain

declarative. Unlike Krashen's proposal, Paradis (2004) assumes the possibility of an L2 learner reaches the level of implicit knowledge according to his proficiency level. This way, as a learner becomes more proficient in the L2, the linguistic focal point goes from the declarative to the procedural knowledge.

Paradis's neurocognitive data gives us support to assess explicit/implicit knowledge, since they reveal what instances of language belong to what level of knowledge. For example, "all aspects of the acquisition and use of morphosyntax and phonology are compatible with the charateristics of procedural memory, while at least some aspects of lexicon would seem to be mainly within the purview of declarative memory"(Paradis, 1994, p. 398). Another example is the acquisition and use of vocabulary. To Paradis (1994, p. 398), vocabulary has two facets: "Their acquisition, which is conscious (one is made aware of pronunciation and meaning). And their use in context, which is automatic (one is not aware of the access mechanisms that select items during the microgenesis of an utterance)." By this vocabulary definition, we can infer that both acquisition and recognition of a lexical item belong to the declarative knowledge, since learners learn consciously *what is* and *how/where* to use such item. Thus, a test measuring lexical knowledge, through word association for example, would be a declarative knowledge matter.

These findings can lend support to empirical studies on explicit/implicit knowledge measurement. By assuming the discrepancy between morphosyntax and lexicon regarding the type of knowledge, we are able to administer psycholinguistics procedures (e.g. behavioral tasks) appareled with time restriction (in order to automatize the access) to measure learners' implicit L2 knowledge through morphosyntatic constructions/violations. These findings also contribute to assessment of L2 through processing tasks, since "the way in which L2 may be processed differently from L1 will depend on the extent of linguistic competence in L2. The weaker the linguistic competence is, the more the speakers will have to resort to metalinguistic knowledge" (Paradis, 1994, p. 414).

In this work, we are defining proficiency as a psychometric construct that can be validated as a methodological tool for delimiting bilinguals' profiles. This way, proficiency is a continuum that represents bilinguals' L2 explicit and implicit knowledge. In order to establish a measure of the explicit knowledge, we administered a vocabulary level test according to word frequency. In order to establish a measure that captures

implicit knowledge, we tested participants' ability to judge the grammaticality of L2 sentences in a given time. We are assuming that both results correlated can be an adequate measure of general L2 proficiency since it deals with the two types of knowledges.

As we mentioned, one of the dimensions of L2 proficiency that has stemmed reasonably practical tests (from the standpoint of test administration) is L2 lexical knowledge (Milton, 2009, Daller et al. 2007). De Groot (2011) asserts that a bilingual's high-proficiency level in the L2 is dependent not only on his or her accuracy in grammar, but also on a considerable level of vocabulary knowledge. Also, Hulstijn's (2011) model of split BLC and HLC does predict variability in lexical knowledge, as discussed above. Furthermore, there are considerable empirical studies suggesting a relation between differential vocabulary knowledge and measurable differences in bilinguals' language proficiency.

2.4 Vocabulary size as a measure of L2 proficiency

When it comes to measuring L2 vocabulary knowledge, more than one dimension can be taken into consideration. In the L2 mental lexicon literature, the dichotomy between 'vocabulary breadth' and 'vocabulary depth' is often cited as a descriptor of two broad dimensions in the organization and development of lexical competence (Meara, 1996; Read, 2000; Milton, 2010; Schmitt, 2014). Vocabulary breadth is understood as the quantity of words someone is capable of recognizing and connecting to a core meaning—in other words, an individual's vocabulary size. Vocabulary depth outreaches vocabulary recognition, as it entails at least access to information related to derivational morphology, collocation restrictions, subcategorization frames, membership to semantic fields and classes, and usage restrictions (MILTON, 2010; MEARA 2009).

According to Schmitt (2014), the research on measures of vocabulary breadth and depth shows that for learners with small L2 vocabularies and for high frequency words there is no distinction between the two measures; whereas for larger L2 vocabularies and low frequency words bilinguals tend to develop vocabulary depth more slowly than vocabulary size. As the dichotomy between size and depth is probably related to the distinction between receptive (recognition) and productive vocabulary (SCHMITT, 2014),

what the research suggests is that after a certain threshold of L2 vocabulary size bilinguals may be more able to recognize form-meaning links of L2 words than to accurately use L2 words.

Another dichotomy is used by Read (2000) to define the three dimensions of vocabulary knowledge. According to the author, the first dichotomy is related to the construct that grounds the test. In this way, a test will be *discrete* if the vocabulary is taken as an isolated object, apart from other language instantiation such as grammar. On the other hand, a test is considered *embedded* when the vocabulary measure is a component of a broad construct beyond the vocabulary knowledge itself. The second dichotomy has to do with the magnitude of vocabulary to be covered by the test. A test will be *selective* if it is composed of a set of previously selected isolated words to which test-takers must show some knowledge. On the contrary, a test is *comprehensive* if it considers the test-taker's knowledge of the totality of words in a given text. The third dichotomy has to do with the role of context in the vocabulary measurement. A test is *context-independent* if the word is presented without a circumstance of use. Contrastively, a test will be *context-dependent* if test-takers are given a situation involving the use of a given word. In Figure 9, we can see a summary of Read's (2000) definitions:

Discrete - A measure of vocabulary knowledge or use as an independent construct.	**Embedded** - A measure of vocabulary which forms part of the assessment of some other, larger construct.
Selective - A measure in which specific vocabulary items are the focus of the assessment.	**Comprehensive** - A measure which takes account of the whole vocabulary content of the input material (reading/listening tasks) or the test-taker's response (writing/speaking tasks).
Context-independent - A vocabulary measure in which the test taker can produce the expected response without referring to any context.	**Context-dependent** - A vocabulary measure which assesses the test taker's ability to take account of contextual information in order to produce the expected response.

Figure 9: Dimensions of Vocabulary Assessment (Read, 2000, p. 9)

Notwithstanding the distinction between vocabulary breadth and depth, there is ample evidence that vocabulary size measures are consistently correlated to estimates of performance level on the four language skills of speaking, writing, listening, and reading (ALDERSON, 2005, MILTON, 2013).

2.4.1 The relation between vocabulary size and receptive/productive abilities

Vocabulary size has proven to be highly associated with other English abilities. For example, Laufer (1992) indicates that there is empirical evidence suggesting that in order to reach a reasonable level of the L2 comprehension, learners should understand at least 95% of the words in a given text. Along the same lines, Nation (1990) points out that to reach such status, learners should have a vocabulary size of over 3,000 more frequent words (approximately 5,000 lexical items). There are several studies in the literature concerning the high correlation between vocabulary size and speaking, listening, writing, and reading abilities from Stæhr (2008), Zimmerman (2004), Milton (2010), and Koizumi & In'nami (2013).

For example, Stæhr (2008) investigated the relationship between vocabulary size within the four abilities, specifically with reading (.83) and writing (.73). Through a binary logistics analysis, Stæhr (2008) identifies that vocabulary size is responsible (in 72% of cases) for learners to reach the average score in the reading test. Similarly, Hu and Nation (2000) conclude that knowledge of almost every word in a text (98%) is needed to fully comprehend the text.

Zimmerman (2004) investigated the role of vocabulary size in speaking ability. Participants were native speakers of Spanish, Korean, Japanese, and Mandarin who were English L2 learners. The results of a vocabulary level test were compared to an institutional placement test that had a speaking component. The results suggest that the amount of L2 vocabulary highly correlates with the speaking ability in L2.

Concerning the relation of vocabulary and listening comprehension, Bonk (2000) developed an experiment by applying a listening task, which gradually increased the level of English words frequency, followed by a dictation test (N=59). Results showed that participants with a lexical coverage of 90% (or less) were less able to perform the task than the participants who had a lexical coverage of 95% (or more). Similarly, Stæhr (2009) investigated the relationship between vocabulary size and listening comprehension. Participants' profiles (N=115) were defined by the VLT, and a high correlation (.72*)* was found.

Milton (2010) conducted a study in order to analyze the impact of vocabulary as a dimension within the six L2 proficiency levels in the CEFR (Common European Framework of Reference for Languages). The study's focus was on the vocabulary threshold sizes necessary for bilinguals to perform according to the CEFR descriptors. The author compared the CEFR's levels with vocabulary size as measured by the XLex[5] test. The results show that L2 vocabulary range requirements steadily increase as the CEFR's levels move upwards. The correspondences found by Milton (2010) are summarized in Figure 10, where the estimates of vocabulary size refer to lemmatized items, i.e., to word families:

CEFR Level	**Vocabulary Descriptors**	**Vocabulary Size XLex (5000 max)**
C2	Broad lexical repertoire including idiomatic expressions and colloquialisms.	4,500 – 5,000
C1	Little obvious searching for expressions. Good command of idiomatic expressions and colloquialisms.	3,750 – 4,500
B2	Vocabulary for matters connected to his or her field and most general topics.	3,250 – 3,750
B1	Sufficient vocabulary to express him/herself with some circumlocutions.	2,500 – 3,250
A2	Sufficient vocabulary to conduct routine, everyday transactions involving familiar situations.	1,500 – 2,500
A1	Basic vocabulary repertoire of isolated words and phrases related to particular concrete situations.	< 1,500

Figure 10 - Vocabulary range criteria from Council of Europe (adapted from Milton, 2010)

2. 4.2 The Vocabulary Level Test (VLT)

The VLT (Nation, 1990; Schmitt, Schmitt & Clapham, 2001) is a five-level test elaborated to measure vocabulary size. It has 5 parts, each containing six items (totaling 30 items) in which test-takers must produce matching of three out of six words with three definitions. Therefore, each part of the VLT yields a ceiling of 18 correctly matched

[5] XLex (Meara & Milton, 2003) is a corpora frequency-based test in which participants have to affirm which words they know from a list. Then, it is calculated how many words for each list (frequency-based) each participant knows.

words. The VLT estimates vocabulary size levels by correspondence between level and word frequencies bands based on the Brown Corpus.

Successful completion of level 1 corresponds to knowledge of the 2,000 most frequent words; completion of level 2 corresponds to the 3,000 most frequent words. Level 3 corresponds to the 5,000 most frequent words, level 4 is a special section corresponding to academic and scientific vocabulary, and level 5 corresponds to knowledge of the 10,000 most frequent words. Depending on the L2 learners' L1, the fourth level presents some barriers. Because of the high frequency of words derived from Latin, Romanic L1 speakers can rely on the association by cognates.

According to Nation (1990), the cut-point for successful completion of a VLT level is 12 correct pairings out of the 18 that are possible per level. It should be noted that Souza, Duarte & Berg (2015) report no discriminatory effect for level 4 (academic and scientific vocabulary), which is interpreted as a byproduct of the fact that such lexical domain is heavily composed of Latin-originated words that form cognates with Portuguese words. Figure 11 is a demonstration of the layout of the VLT items.

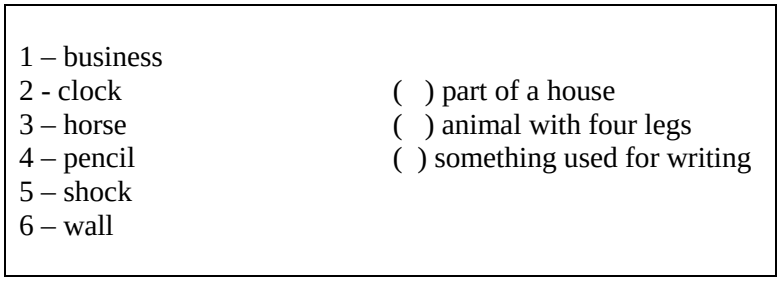


Figure 11 – Example of a question in the Vocabulary Level Test

According to Read (2000) and Read & Chapelle (2001) the VLT is a discrete vocabulary test, as the construct underlying the test relies exclusively on vocabulary knowledge (specifically, the meaning of words). The author also analyzes the VLT as being a selective test, as the words were chosen based on corpora frequency. Finally, Read describes the VLT as a context-independent vocabulary test, since it does not tap into knowledge of situations where words would be likely to occur. According to Milton (2009, p. 74), the VLT "allows rather more than passive recognition for word form to be tested and this form of test should allow an estimate of knowledge of words and their meanings to be formed."

In Nation's (1990) original proposal of the VLT, there is no specification of a time limit for test-takers. However, Laufer & Nation (2001) conducted a study of a computerized vocabulary test based on the VLT in which response latencies were analyzed. The researchers found a moderate and significant negative correlation between vocabulary size and response latency. In other words, raises in vocabulary breadth are related to a higher speed in linking word form and meaning.

In a study to explore the validity of the proposed 12-matching cut-point for Brazilian college level test-takers, Souza, Duarte & Berg (2015) implemented a temporal ceiling for completion of the VLT. This temporal ceiling suggests the administration of the VLT within a time window of 10 minutes, i.e.: 20 seconds per item or roughly 6.66 seconds per definition to be matched with a word. The authors' rationale was the integration of a component of automaticity—namely speed of task execution—to the VLT construct. Souza, Duarte & Berg (2015) suggest that 12, 13, and 14 correct matchings are equally discriminatory cut-points for the VLT.

Importantly, the integration of a speed requirement seems particularly useful in light of the authors' finding that the academic vocabulary section (level 4 of the VLT) does not discriminate English L2 vocabulary knowledge of the average Brazilian college student. By introducing the speed limit, the non-discriminatory section may be functioning as a modulator of how far the test-taker will be likely to reach in face of the temporal constraint for execution of the test's task. As described in the methods, in this study we employed it administration mode of the VLT described in Souza, Duarte & Berg (2015).

SLA and processing studies involving cognitive aspects of knowledge and proficiency has brought significant gains to the field. Due to the constant growth of languages being learned around the world, the majority of these studies rely on late bilinguals. According to the literature, those bilinguals who acquire L2 in post-puberty (majorly in formal settings) are named as *late bilinguals*. On the other hand, those who acquire both languages simultaneously since birth (with variation in the amount of exposure) are named as *early bilinguals.*

So far, we presented a series of studies involving L2 speakers as late bilinguals. The majority of bilingual studies deal with this type of bilingual, since it appears to be a more frequent population. The early bilinguals are less frequent phenomena, which

sometimes involve linguistic contact and immigration situations. The HL speakers fit in the early bilinguals' category. Due to some specificities of HL speakers, such as dominancy and age of acquisition, this type of bilingual poses several challenges to language proficiency studies.

2.5 Heritage Language speakers

It has been well over a decade since Montrul's (2002) first mention of what Polinsky and Kagan (2007) would name years later as a new branch of bilingualism studies: Heritage Language (HL) acquisition and the HL speakers. Valdés (2000, p. 1) has defined the adult HL speakers as those who have been "raised in a home where a non-English language [or other] is spoken, who speak or merely understand the HL , and those who are to some degree bilingual in English [or other language] and the heritage language."

In Montrul's (2012, p. 2) words, the HLs are "the languages spoken by immigrants and their children. Sociopolitically, the languages spoken by the wider speech community in the host country are majority languages with official status," and the HL speakers are "the children of immigrants born in the host country or immigrant children who arrived in the host country some time in childhood" (*op. cit*.).

Although the HL speakers are considered early bilinguals, Montrul (2012) defines two types of HL speakers according to the exposure to the dominant language: The *simultaneous* HL speakers are those who grew up speaking both the majority and the HL since birth. The *sequential* HL speakers are those who live in a monolingual context as children and gradually become bilingual as they go to school (majority language, at the age of 5 or 6). In both cases, the HL will be the non-dominant one as the speakers become adults. Montrul (2012) proposes a model (Figure 12) to differentiate the processes it takes for L2 and HL speakers develop the language:

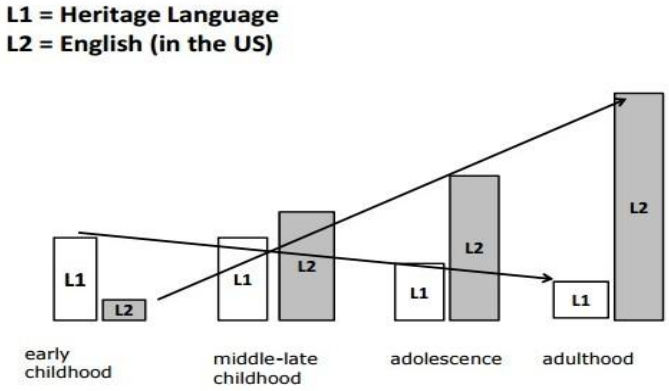


Figure 12 - Typical development of a heritage language (L1) in a majority language context. (MONTRUL, 2012, p. 4)

One aspect that differs L2 learners and HL speakers relies on the nature of the input, concerning timing, setting, and mode. While L2 learners receive late, instructed, aural/written input, HL speakers receive early, naturalistic, and only aural input. As similarities, both L2 learners and HL speakers receive input that is variable and restricted to environment (Montrul, 2012). In many cases as adults, the HL speakers do not take formal instruction on their HL, and that is why they present some lack of knowledge resembling the patterns investigated in SLA studies, such as case marking, gender agreement, etc. The HL speakers are also better in receptive abilities (for instance reading and listening) than in productive abilities (for instance speaking and writing) (MONTRUL, 2012).

Usually, HL speakers are fluent and functional in their dominant language, since they are exposed to it in informal and formal situations, including the contexts of educational settings. This means that, as a child, they spoke the HL with parents/caretakers, and later the dominance is inverted due to massive exposure to the other language. The HL speakers' proficiency in the HL may vary depending on how they kept it active throughout their lives. Limited exposure to the HL may result in confines such as a small amount of vocabulary, shallow grammar knowledge, and problems in pragmatic competence.

Montrul (2012) presents a series of empirical studies that investigated such differences as instantiated by phonology, lexicon, syntax, discourse-syntax, semantics, and morphology. Among these studies, there is ample empirical support for the claim that these differences in phonology and syntax privilege HL speakers in relation to L2

learners. For instance, Montrul (2012) investigated knowledge of gender agreement in Spanish among Spanish L2 learners (N=72), Spanish HL speakers (N=69), and a control group of native Spanish speakers (N=22). They were tested on their oral production, written comprehension, and written recognition of Spanish gender agreement. Results revealed a performance in a written task favorable to L2 learners, although it also revealed significant advantages of HL speakers in oral tasks regarding Spanish agreement recognition.

In another study, Montrul (2005) compared the linguistic knowledge of unaccusativity in the Spanish language of adult Spanish L2 learners (N=71) with adult HL speakers (N=36), and a native Spanish speaking control group (N=22). Despite presenting corresponding patterns, when proficiency was taken into consideration, the results displayed significant difference between the groups, among which high-proficiency HL speakers performed better than high-proficiency L2 learners.

Authors predict that HL speakers would be more accurate on gender agreement than L2 learners because their HL acquisition began early in childhood. According to Montrul (2012), these differences rely on: a) maturational aspects (such as age or moment of acquisition) since one of the language is learned before the other (Long, 2007, for instance); b) the nature of linguistic knowledge; c) degree of ultimate attainment, in which HL speakers perform more accurately than L2 learners.

Montrul (2010) also raised the issues regarding the lack of studies considering the relationship between the lexicon and the syntax in HL. In her words, "the relationship between grammar and the lexicon needs to be explored more closely in future psycholinguistic research, [since] this relationship has pedagogical and assessment potential for both L2 learners and heritage language learners" (op. cit. p. 6). Moreover, the studies on HL speakers' vocabulary knowledge seem to be important because usually their vocabulary relies on common words, related to simple words related to simple objects inside the context of their childhood. Hence, "HL speakers also have significant gaps in their vocabulary and find it difficult to retrieve words they do not use very frequently" (Montrul, 2010, p.6).

The data yielded by these studies provides convincing evidence that, in high-proficiency levels, HL speakers present a better (and in some cases, near-native)

performance than L2 learners in the non-dominant language. Montrul's findings lend support to the claim that L2 learners and HL speakers are quite similar in linguistic competence, however notable differences emerge when proficiency is taken into consideration. Montrul (2012) infers that these results are due to the exposure to primary input in early childhood. In this sense, although it has presented some sort of language loss, some parameters have been kept. According to Montrul (2005, p. 2), "L1 loss in a bilingual context is the flip side of the L2 acquisition.

In this chapter, we presented a theoretical overview of lexical-semantic interface, focusing the *Construction Grammar*. It was also discussed two models of lexical acess (Levelt's and Hartsuiker et al,' model), as well as a framework for the conception of language as a product of consciousness (MOGUL). Moreover, we presented a discussion on the construct of language proficiency in bilingualism studies, focusing on the role of vocabulary knowledge as a predictor of general proficiency. In next chapter, we will present the methods and procedures we adopted in this study. It will presented a detailed description of the experiments, participants and materials.

CHAPTER 3

METHODOLOGY

In this chapter, we will present an overview of Experiment 1 and Experiment 2, detailing the procedures for the study, a description of the participants, as well as the presentation of the materials employed in this empirical research. After that, we describe the AJ task, focusing on the types of the sentences, the form sentences were judged as well as the measures we counted on for this task. Following, we present the VLT in detail, its division by word frequency and the possible scores. Moreover, we describe the procedures we took to develop the Spanish version of the VLT. After that, we described the Bilingual History Questionnaire (BHQ), explaining each of its three sessions. Following, we describe the overall proficiency test administered in both experiments: Oxford Placement Test for Experiment 1 and Spanish Placement Test for Experiment 2. After that, we presented the participants in detail, showing biographic and linguistic data that concerns our assumptions in this study. Finally yet importantly, we presented the ROC curve methodology for diagnostic tests.

As stated above, the primary goal of the present study was to further validate the legitimacy of the VLT (English version) as a proficiency measure for Brazilian Portuguese-English bilinguals, and the VLT (Spanish version) for Spanish HL speakers. Specifically, we sought to (1) replicate previous research results suggesting correlations between vocabulary size and overall proficiency (Laufer, 1992; Nation, 1990; Stæhr, 2008; Zimmerman, 2004); (2) examine the behavior of VLT scores when distinguishing test-takers' performances in a timed version of the acceptability judgment task in both late bilinguals (L2 learners) and early bilinguals (HL speakers). An ancillary goal of the first experiment was the establishment of a minimum time window for the speeded judgment task with L2 speakers, and in order to achieve this goal we replicated the procedures described in Souza et al.'s (2015) study with monolinguals.

In order to meet the expectations of the present study, we designed two experiments. In the first experiment collected in Minas Gerais, participants (Brazilian Portuguese-English bilinguals) took an overall proficiency test in English (the Oxford Placement Test), the VLT (English version), and an acceptability judgment task with

sentences in English. In the second experiment collected in New York, participants (Spanish L2 learners and Spanish HL speakers) completed a language-dominance questionnaire, followed by an overall proficiency test in Spanish (the Oxford Placement Test), the VLT (Spanish version), and an acceptability judgment task with sentences in Spanish. In both judgment tasks, we covered morphosyntax and syntax-semantics interface phenomena. Each of these instruments is described in the following sub-sections.

3.1. Experiment 1: Overview

Experiment 1 was conducted with two main goals. First, we aimed to estimate the minimum time a bilingual (Portuguese/English) needs to conduct a judgment call of a sentence in English. This judgment is about the acceptability of the sentence concerning its grammatical accuracy according to the participants. To do so, we administered an Acceptability Judgment (AJ) (Appendix 1) task using a portable laptop. The AJ task is composed of randomly distributed grammatical and non-grammatical structures, according to the formal rules of the language (Bard, Robertson & Sorace, 1996). Through a Likert scale from 0 to 5, participants were supposed to judge each sentence, attributing a value to grammatical correctness.

In order to establish the minimum time, we applied an exploratory study of the judgment task's parameters. To do this we took sentence types of our experimental corpus and the level of proficiency as independent variables controlled in the experiment, and as dependent variables we considered the reaction time (RT) in each stimulus.

For the second goal, we aimed to verify: a) the correlation between an L2 proficiency test based on vocabulary knowledge and a placement test. Through statistical measures of correlation, we investigated wheather a vocabulary test and a placement test have similar measures to separate high levels of proficiency from non-high levels; b) the correlation among both proficiency tests and the judgment task in order to see whether the levels of proficiency converge with participants' judgment of the sentences.

All participants of Experiment 1 partook in the process in three consecutive stages. First, they took the paper-based placement test (maximum of 30 minutes); then they took the paper-based vocabulary test (10 minutes). Finally, they completed the judgment task

through specific software (DMDX[6]) on a laptop (with a time of 10 seconds allotted to each sentence). Following, we describe participants' profiles, detailed information about materials (tests and task), and the experiment's procedure.

3.2 Experiment 2: Overview

Similarly to Experiment 1, Experiment 2 first aimed to analyze the correlation between a Spanish VLT, an overall proficiency test, and a self-assessment test. Through statistical measures of correlation, we investigated where the three tests have similar measures to discern high levels of proficiency from low levels. Second, analyze the correlation among the three proficiency tests and the judgment task, in order to see whether the levels of proficiency we found in the tests are converged with participants' performances in the judgment task. Third, investigate the relation between language exposure types (home or school) and language exposure length with the level of proficiency found in the tests, in addition to performance in the judgment task.

In order to achieve those goals, we applied similar tasks from Experiment 1 with the addition of a questionnaire (BHQ). Aside from the questionnaire on participants' background, we administered a paper-based overall proficiency test (placement), a Spanish version of the VLT, and the acceptability judgment task with Spanish structures (following the procedures from Experiment 1).

The Spanish version of the VLT was created based on Nation's (1983, 1990) version, and we followed the score measures elaborated by Souza, Duarte & Berg, (2015). We based the word frequencies on the *Corpus de Referencia del Español Actual* (CREA[7]). In order to elaborate on word definitions, we had the help of three native Spanish-speaking (NS-S) assistants who were volunteers in the Neurolinguistics laboratory of the City University of New York (Graduate Center). The Spanish version of the Oxford Placement Test is a paper-version of an electronic test used by some universities for placement purposes. It is composed of 55 questions including grammar, reading, and context-based

[6] Availabe at < http://www.u.arizona.edu/~kforster/dmdx/dmdx.htm> accessed in December, 2013. For more information, see: FORSTER, K. I., & FORSTER, J. C. DMDX: A windows display program with millisecond accuracy. *Behavior Research Methods, Instruments, & Computers*, 35, 116-124, 2003.
[7] REAL ACADEMIA ESPAÑOLA: Banco de datos (CREA) [en línea]. Corpus de referencia del español actual. <http://www.rae.es>

material. The acceptability judgment task was composed of 16 target sentences. Of these sentences, 8 had grammatical violations concerning transitivity and morphosyntax, and 8 were grammatically correct.

All participants of Experiment 2 completed the process in four consecutive steps. First, they filled out the questionnaire, they then took the paper-based placement test (with a maximum of 30 minutes allowed); afterward they took the paper-based vocabulary test (with 10 minutes allowed). Finally, they completed the judgment task through specific software in the laboratory (each sentence was allotted up to 10 seconds). Following, we will describe the participants' profiles, detail information about materials (tests and task), and go over the experiment's procedure.

3.3 Materials

3.3.1 Acceptability Judgment (AJ) task

When it comes to administering a judgment task as part of a psycholinguistic investigation, sometimes the term 'grammaticaticality' appears, and other times 'acceptability' appears. Although both terms are often used interchangeably, the true meaning lie in different principles (Souza et al., 2015; Schültze, 1996). According to Bard, Robertson and Sorace (1996), grammaticality is a construct that belongs to the theoretical aspect of linguistics. It is beyond the possible variation in dialect, adequacy, and context. In other words, grammaticality refers to how accurate a sentence is under the formal rules of a given language. In contrast, the term acceptability is related to the way speakers perceive the language; it means how effective or ineffective a certain sentence sounds to a given speaker (Bard, Robertson & Sorace, 1996). Therefore, the term acceptability better fits our purpose in this research.

In language processing experiments, we can have offline and online procedures. Offline means that data is collected after being processed by participants, and we only have access to what they respond as their interpretation—not to the exact moment that the processing is taking place (online tasks). AJ is an offline procedure in which participants are told to react to a series of stimuli presented on a computer screen.

In this study, we employed the design reported in Souza et al.'s (2015) study of a timed acceptability judgment task with monolinguals of American English and Brazilian Portuguese. Accordingly, the present AJ task stimuli for both experiments were presented on a computer screen using DMDX software. During this procedure, participants were exposed to sentences (presented one-by-one) in the center of the screen. Then, they judged each sentence using a 5-point Likert scale (Figure 13). Responses were given using the numeric keys of a computer keyboard, and a time limit of 8 seconds was set for the judgment calls (this time limit of 8 seconds was reported in Souza et al.'s, 2015).

Numeric keypad	Judgment levels
1	Totally unacceptable
2	Not well-formed, almost unacceptable
3	Not well-formed, but maybe acceptable
4	Slightly ill-formed, almost perfect
5	Totally perfect

Figure 13 - Likert scale to the Acceptability Judgment task, adapted from Souza et al. (2015).

For both Experiments, our AJ task corpus was composed of 56 sentences, and 16 of them contained grammatical violations.. There were two types of sentence violations applied to 8 sentences each: argument structure realization violations involving unergative verbs in transitive syntax, and explicit morpho-syntactic violations involving long-distance dependencies (Wh-movement) and subject-verb agreement.

Argument structure realization violations were chosen because according to White (2003), L2 argument structure may pose a challenge to L2 learners, as "interlanguage lexical representations may not correspond to argument structures encoded in the lexicons of native speakers of the L2" (White, 2003, p. 206). Although unergative verbs do not transitivize in either Brazilian Portuguese or English, bilinguals with Brazilian Portuguese L1 and English L2 will need to learn which argument structures are licensed in their L2 and which ones are not, as there are several cases of argument structure patterns that are productive in English, but not in Portuguese. The induced-movement alternation (Souza, 2011; 2012), the dativized bi-transitive construction (Zara, Oliveira & Souza, 2013), and the resultative construction (Souza & Oliveira, 2014) are examples.

The sample of ungrammatical sentences due to explicit morphosyntactic violations encompassed 4 sentences with violations in subject-verb agreement (henceforth VAgr),

and 4 sentences with violations to WH-movement (henceforth WHm). WH-movement is a syntactic operation for questions or relative clauses in which the WH-word is moved to the position of the specifier of the CP. This movement will not be licensed when the specifier position is already filled and the WH-word goes out of the CP domain. For instance, the sentence: "[DP *who*]$_i$ *do they think* [DP]$_i$ *taught what?"* is licensed, however the sentence: "[DP *what*]$_i$ *do they think who taught* [DP]$_i$ *?"* has a violation on the syntactic movement of the WH-word.

Subject-verb agreement violations as well as WH-movement violations are cases of ungrammaticality in both the L1 and L2 of the bilingual population we observed. Nevertheless, it has been demonstrated that violations of this type are not necessarily perceived by L2 speakers (Clahsen & Felser, 2006; Jiang, 2007), an observation that has also been replicated with bilinguals of Brazilian Portuguese and English (Carneiro & Souza, 2012). Therefore, actual detection of such violations under the pressure of time may reflect more automaticity in access to grammatical knowledge, hence higher fluency and proficiency in L2 use.

As control sentences for Experiment 1, we employed a set of 8 sentences instantiating the induced-movement alternation of English (e.g.: The woman jumped her horse over the fence). As stated above, this is a case of argument structure realization that is not licensed in Brazilian Portuguese, but which is learnable by high-proficiency Brazilian Portuguese-English bilinguals (Souza, 2011; 2012). By employing this type of sentence, we wanted to check whether the sample of the bilingual population of interest to this study would be capable not only to perceive ungrammaticality in their L2, but also to inhibit a restriction of their L1 that is not applicable to their L2. Such inhibition should take place under the time constraint of our AJ task.

In addition, as control sentences, we applied grammatical change-of-state verbs, which can be externally or internally caused. The former is known as a *break*-type verb and the latter as a *bloom*-type verb (McKoon & Macfarland, 2000). When externally caused, the change-of-state verbs are able to alternate from causative form (e.g.: Mary broke the jar) to inchoative form (e.g.: The jar broke.). Conversely, when they are internally caused, change-of-state verbs would occur in inchoative form (e.g.: The roses bloomed.) but not causative form (e.g.: *The gardener bloomed the roses.) (Levin & Rappaport Hovav, 1995). As control sentences for Experiment 2, we also displayed 8

grammatical sentences with change-of-state verbs and 8 transitive sentences randomly selected from our stimuli. Sentences (1) through (5) below illustrate the type of stimuli employed in our AJ task for Experiment 1:

(1) *The man laughed the children during the party. *Transitivized unergative verb*
(2) *The girl give the cats milk twice a day. *Agreement violation:*
(3) * What did Steven read the book that Helen talked about? *WH movement violation:*
(4) The instructor ran the boys around the park. *Grammatical sentence (Induced movement alternation)*
(5) The girls melted the cheese in the bowl. *Grammatical sentence (change-of-state verbs)*

Sentences (1) through (5) below illustrate the type of stimuli employed in our AJ task for Experiment 2:

(1) * *El polvo denso tosió a los niños en el parque*. (The dense dust coughed the children in the park.) Transitivized unergative verb

(2) * *El criminal buscado por la policía fueron presos*. (The wanted criminal were arrested.) Agreement violation

(3) * *¿Qué leyó Steven el libro sobre el que Helen habló?* (What did Steven read the book that Helen talked about?) WH movement violation.

(4) *Las chicas derritieron el queso en el tazón* The girls melted the cheese in the bowl. Grammatical sentence (change-of-state verbs)

(5) *Los oficinistas limpiaron la oficina.* (The clerks cleaned the office). Grammatical sentence (transitives)

3.3.2 The Vocabulary Level Test (VLT)

Following Souza, Duarte & Berg (2015), in Experiment 1 participants were allowed up to 10 minutes to complete the VLT (Appendix 2) to the best of their abilities. This time limit is not applied in the original version of the test (Nation, 1983, 1990). In order to pass from one level to another, participants should score at least 14 (78%) of the 18 possible points per section. Therefore, we chose to employ the most conservative cut-point, despite the fact that Souza, Duarte & Berg did not observe differences between this cut-point and the 12-point cut-point originally proposed by Nation (1990). We considered participants measuring with high proficiency to be the ones who could achieve level 5 of

the VLT, based on previous studies with Brazilian Portuguese-English bilinguals that employed the same test as a screening method and that identified differential behavior related to high levels of proficiency thus measured (Souza, 2012; Souza & Oliveira, 2014).

The Spanish version of the VLT (Appendix 3) was elaborated following Nation's (1983, 1990) principles and Souza et al.'s (2014) scores redefinition. We developed the test in order to demonstrate its validity among the population we investigated: Spanish speakers (Spanish L2 leaners and Spanish HL speakers). As previously explained, the VLT is comprised of 5 parts with 6 questions in each part. The questions are composed of 6 words, and 3 definitions to those words. With this structure, participants must chose 3 (among the possible 6) definitions that fit the word. Each part of the VLT corresponds to a level of more frequent words: 2,000, 3,000, 5,000, the academic list, and 10,000.

In order to elaborate the Spanish version of the test, we collaborated with two researchers (PhD students) and 3 NS-S assistants. The test was elaborated in the context of the Neurolinguistic Laboratory in the City University of New York (Graduate Center). We based our elaboration of the first two levels, corresponding to the 2,000 and 3,000 more frequent words, on Casso's (2010) Master's dissertation in which he expanded a Spanish vocabulary test of the first VLT levels based on Nation's (1983, 1990) procedures. Casso's (2010) list of words is based on the *Diccionario de Frecuencias* (Almela et al., 2005), which is founded on the Spanish corpus Cumbre. We took the sample of Casso (2010) as a starting point in order to start the elaborations of the 5 bands of the Spanish version of the VLT.

The Corpus we used to elaborate the Spanish version of VLT is the *Corpus de Referencia del Español Actual* (CREA) of the Real Academia Española. CREA is constituted of texts of several natures, ranging from informal to formal texts. All the material is electronically stored, from which users are able to consult, research, and look for specific words/expressions and their contexts of appearance. CREA contains an unparalleled amount of over 60 million words in forms of written and oral texts from 1975 to 2004, comprehending all of the Spanish-language countries[8]. Consequently, 40 words (substantives and adjectives) were randomly selected from each level of frequency. As we were trying to avoid any cognate facilitation, in the 2,000, 3,000, 5,000, and 10,000

[8] Source: http://www.rae.es/recursos/banco-de-datos/crea

we avoided cognates as much as possible. In the academic lists, we used cognates, but we made a control (of which, more later).

Once we selected the words, each NS-S assistant had individual access to the words and created separate definitions. At this time, we had a meeting in which we all had access to every word definition, and together we decided which words would be better for a given frequency level. The NS-S assistants were also tasked with collectively deciding which the best definitions of the words were. An example of the process can be seen in the fragment of level 3 (5,000 words) below (Figure 14), with the words and definitions created by each NS-S assistant. The italicized definitions represent the ones the NS-S assistants chose as their favorites for test use.

Word	**English transl.**	**Cog**	**NS-S Assistant 1**	**NS-S Assistant 2**	**NS-S Assistant 3**
enferma	sick	n	que no goza de buena salud.	malo de salud	*sin salud*
pesadilla	nightmare	n	sueño no deseado que causa miedo.	*sueño que causa miedo*	un mal sueño
Fijado	attached	n	persona determinada a hacer algo.	Arraigado	*que no se mueve*
Mancha	stain	n	marca no deseadas en la ropa.	*marca de suciedad*	sucio
Trozo	piece	n	pedazo de un árbol.	fragmentos/ pedazos	*pedazo*
fantasma	ghost	n	algo no real o imaginario, que usualmente se relaciona con muertos.	imagen de una persona muerta que se aparece a los vivos	*espiritu de un muerto*
Casco	helmet	n	*objeto que protege la cabeza.*	pieza usualmente de plástico que cubre y protégé la cabeza	cubre la cabeza
Bruto	rough, dumb, brute	y	persona que no es inteligente o alguien que tiene mucha fuerza.	*poco inteligente*	sin entendimiento
caridad	charity	y	*ayudar a las personas necesitadas.*	Limosna	acto de ayudar al outro
bombero	fire fighter	n	persona que apaga incendios.	los que tienen oficio en extinguir incendios	*apaga incendios*
reposo	rest	n	descansar para recuperar fuerzas.	descanso de un trabajo o una actividad	*estado de descanso*

Figure 14 – Example of the words' definition elaboration for the VLT third level (3,000 words)

Immediately following, we chose the words/definitions that we would use as the targeted selections (the matches), and then filled the rest of the questions randomly with other definitions (the mismatches). After organizing the 5 parts of the test, we developed an informal pilot research. We sent the test to 5 Spanish native speakers, asking them to answer it and provide feedback on the definitions and meanings. After that, the NS-S

assistants read participants' feedback and considered some of the comments to make the appropriate changes. Finally we came up with the test to be applied. Below, Figure 15 is an example of a question from the fifth level:

1 - influjo 2 – reportaje 3 – huerta 4 – heredero 5 – tierno 6 – pastilla	(5) suave (3) pequeño terreno para sembrar (1) efecto de una cosa en outra

Figure 15– Example of a fifth-level question in the Spanish Vocabulary Level Test

3.3.3 The overall proficiency test (placement test)

For Experiment 1, we used the paper-version of the Oxford Placement Test (OPT) (the quick version), published by Oxford University Press (Appendix 4). The OPT is a 60-question multiple choice placement test containing questions on grammar, reading comprehension, and language usage. Participants are allowed up to 30 minutes to complete the entire test. In a typical OPT item, participants are required to indicate the best selection to create complete sentences. Below (Figure 16) there is an example of a grammar question from the OPT:

I don't remember ..... the front door when I left home this morning.			
( ) to lock	( ) locking	( ) locked	( ) to have locked

Figure 16– Example of a question of the Oxford Placement Test

The OPT scores roughly place test-takers in the CEFR levels. As previously mentioned, the CEFR levels (in ascending order) are A1, A2, B1, B2, C1, and C2. The corresponding scores with the CEFR are: from 0 to 17 points (A1); from 18 to 29 points (A2); from 30 to 39 points (B1); from 40 to 47 points (B2); from 48 to 54 points (C1), and from 55 to 60 points (C2).

Therefore, an OPT score tentatively allows an interpretation of overall proficiency gauged by the "can-do" list proposed by the CEFR. This "can-do" list specifies the communicative make-up of each CEFR level: A1-level learners can understand and use familiar everyday expressions and very basic phrases; A2-level learners can understand sentences and frequently used expressions. B1-level learners can understand the main points of clear, standard input on familiar matters; B2-level learners can understand the main ideas of complex text on both concrete and abstract topics. C1-level learners can understand a wide range of longer, demanding texts, and recognize implicit meaning; C2-level learners can easily understand virtually everything heard or read (Council of Europe, 2001, p. 24).

Similarly to Experiment 1, in Experiment 2 we applied an overall proficiency test containing grammar and reading questions. We made use of a paper-version of the Spanish Placement Test (SPT[9]). The SPT (Appendix 5) has six levels according to the CEFR: A1 (breakthrough or beginner) and A2 (waystage or elementary), where both are basic users; B1 (threshold or intermediary) and B2 (vantage or upper intermediate), where both are independent users; C1 (effective operational proficiency or advanced), andC2 (mastery or proficiency), where both are proficient users.

The paper-version of the SPT contains 48 objective questions, with each grouping of 8 corresponding to a CEFR level. Following the procedures of Experiment 1, in order to reach the proficient level, participants must reach a level of C2 (over 40 points). In Figure 17 below, there is an example of a question from the SPT:

Hace mucho frío en la calle. Por favor, ……(ustedes) la puerta cuando salgan.		
a-( ) cerrad	b-(X) cierren	c-( ) cierra

Figure 17 – Example of a grammatical question from the SPT

As we discussed in the first chapter, in this work we are assuming that proficiency tests such as VLT and overall grammar tests (placement tests, for instance) involve L2 explicit knowledge, since they allow participants to be aware of linguistic component being tested. Moreover, by associating words with their definition (as in VLT),

[9] http://www.lengalia.com/en/placement-test-spanish.html

participants consciously choose the right answer among a certain number of possibilities. Probably, after a test with this nature, a participant would be able to explain why he chose one answer instead the others. Differently, the AJ task, when applied under a certain time limit, captures L2 implicit knowledges, since the decision made in the judgment does not allow participants to rely in some cognitive strategies. Considering these aspects, in this work, as well as Souza & Soares-Silva (2015), we consider the AJ task as a proficiency test that is able to differentiate high-proficiency from low-proficiency participants.

3.3.4. The Bilingual History Questionnaire (BHQ)

The BHQ (Fernández, 2003) is a three-part questionnaire, specifically designed to delimitate the dominant language of participants, in the Spanish/English context of New York City[10]. The BHQ (Appendix 6) has been used in researches whose participants are assessed by their language dominance, especially the Spanish HL speakers (Fernández, 2002). The notion of dominance in a bilingual context involves a series of aspects that go from linguistic to political. (Birdsong, 2006; Gertken et al. 2014; Tocowicz, et al., 2004). To Gertken et al. (2014, p. 208), this process may involve "proficiency, fluency, ease of processing, 'thinking in a language,' cultural identification, frequency of use" among others.

According to Tocowicz et al. (2004), the bilingual dominance can be measured by self-assessment on abilities in both languages. In their studies, Tocowicz et al. (2004) classified Spanish/English speakers' dominance by using a self-report measure on their abilities in writing, speaking, listening, and reading. There are some interpretative measures to define dominance, such as Cutler et al. (1992), who asked participants if they were about to lose one language and if they had a choice, which one would they keep. The answer observed was considered to be the dominant language.

The first part of BHQ contains demographic questions of background information about age, place of birth, age of arrival in the United States, etc. In the first part, it already

[10] The paper-version of the BHQ is part of the project *Human Language Processing*, bound to the Department of Linguistics & Communication Disorders of the Queens College of the the City University of New York. Professor Eva Fernández coordinates the project.

has some overall questions on language exposure, such as the age of participants when they were first exposed to English/Spanish. Besides the objective questions, the first part also contains some open-ended questions, in which participants are able to write answers. Below (Figure 18), find an example of two questions of the BHQ's first part:

Age you were first exposed to Spanish: ……. Age you were first exposed to Spanish: ……
Briefly explain when you began learning each of your languages .....
……..

Figure 18 – Example of questions from BHQ's first part.

The second part of the BHQ contains more specific questions about a participant's linguistic background. Participants are asked about the amount of English or Spanish they generally use. To do so, participants have to use the following scale: 1-Spanish all the time (always); 2-Spanish usually more than English; 3-Spanish as much as English; 4-English usually more than Spanish; 5-English all the time (always); 6-does not apply.

In this part, participants are asked about language use in different situations (e.g.: home, school), with different people (e.g.: parents, siblings), and at different ages (e.g.: as a child, as a teenager). Differences of people that participants speak with are also questioned, in addition to inquiries about people who talk to participants (speaker or listener). Below (Figure 19), there is an example of a question from the second part of the BHQ:

When you were a child, how much Spanish/English did you speak:	always Spanish				always English	
at home, to your parents?.....	1	2	3	4	5	n/a
at home, to your brothers or sisters? ……..	1	2	3	4	5	n/a
at home, to your grandparents? ……..	1	2	3	4	5	n/a
at home, to other relatives? ……	1	2	3	4	5	n/a
to your friends? ……..	1	2	3	4	5	n/a
in other social contexts? ……	1	2	3	4	5	n/a

Figure 19 – Example a question from BHQ's second part

The third part of the BHQ is a self-assessment group of questions about a participant's proficiency in both Spanish and English in speaking, reading, writing, and comprehension. There is also a scale for the answers ranging from 'very good' to 'very poor.' Additionally, there are two questions asking if participants could pass as a monolingual speaking on the telephone who someone who does not know them (in both

English and Spanish). Below (Figure 20), there is an example of a question from the third part of BHQ:

Rate yourself according to the following categories (circle on each line):					
How would you rate your speaking ability in English/Spanish?					
ENGLISH	very good	somewhat good	so-so	somewhat poor	very poor
SPANISH	very good	somewhat good	so-so	somewhat poor	very poor

Figure 20 – Example of a question from BHQ's self-assessment part.

Participants took 20 minutes to complete the BHQ in the presence of the researcher. This way, any doubt about the questions could be solved. All participants filled out the whole BHQ. As we mentioned, participants of both experiments were divided into two groups: high-proficiency and low-proficiency, according to their scores in VLT (in correlation with the placement test and the acceptability judgment task).

3.4 Procedures of data analysis

In order to analyze the data we employed both descriptive and inferential statistics procedure. As our data are mostly based on measures of central tendency, the tests were essentially devoted to comparison and distribution of means. For the sake of using parametric tests, we submitted data from both experiments to Kolmogorov-Smirnov test for normality. For comparison between and within groups, we administered T-tests and Analysis of Variance (ANOVA) for both subject and items in the interest of assuring the effect value.

In this work, VLT, the placement test, and the acceptability judgment task were considered to be diagnostic tests, since we are looking for the presence or absence of a characteristic or condition of participants (in this case, the high level of proficiency). In the medical area, diagnostic tests are used to detect the presence of a clinical condition in the patients (Coelho-dos-Reis, 2008) and these tests must be as accurate as possible, since it is dealing with important diseases that need to be treated. In order to avoid misleading diagnosis, there are some statistics-based methods used to assess the probability of accuracy of diagnostic tests. One example widely used methodology of assessing diagnostic tests in the medical area is the Receiver Operating Characteristic (ROC) curve.

3.4.1 Receiver Operating Characteristic (ROC) curve

The ROC (Receiver Operating Characteristic) curve is a statistical methodology for estimating the accuracy of diagnostic tests usually in the medical area. The *ROC* curve is a graphic-displayed representation of the Sensitivity and the Specificity of a test. In a nutshell, the Sensitivity is the tool for accurately identifying individuals who have a certain disease, while Specificity accurately identifies individuals who do not have it (Metz et al., 1978 *apud* Coelho-dos-Reis, 2008). According to Fawcett (2006, p. 861), a ROC graph is "a technique for visualizing, organizing and selecting classifiers based on their performance."

The ROC curve, as displayed in a graph, is a product of two perpendicular axes. The upright axe represents the *Sensitivity* and the horizontal axe represents the *Specificity*. The curve encompasses several possible cut-off points for the test, therefore it is possible to delimit which is one accurately separates the two groups of individuals (Metz et al., 1978 *apud* Coelho-dos-Reis, 2008). Below (Figure 21), there is a graph that represents a two ROC curves.

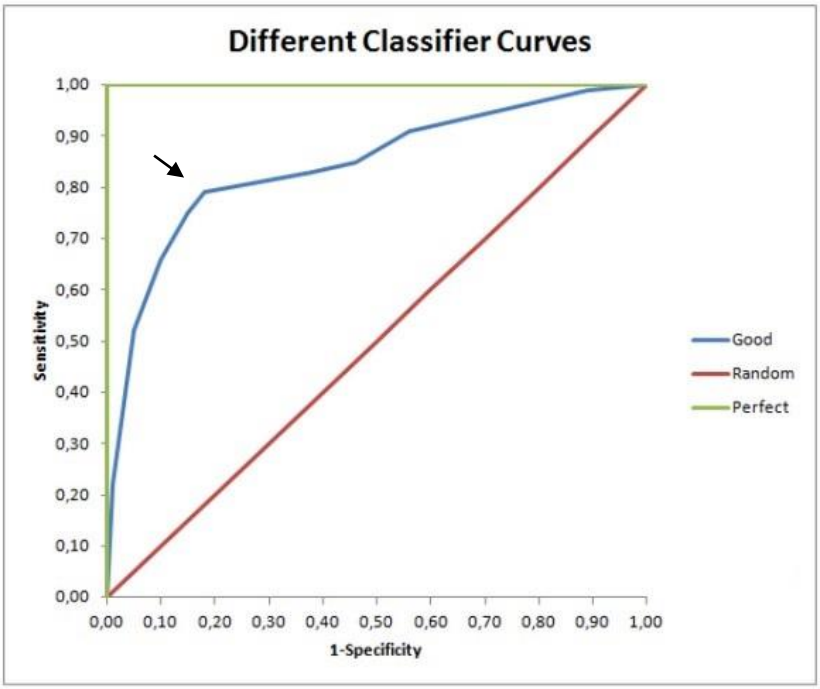


Figure 21: Example of a ROC curve[11]

[11] HOWE, B. *Statistics*: ROC Plot and Area under the Curve. Available at: < http://gerardnico.com>, access: 01/15/2016.

As displayed in Figure 21, the blue line is linking several possible cut-off points for a given test. In order to delimit a cut-off point that is accurate in distinguishing both groups of individuals, we are likely to choose the one closer to 1 in the true positive rate (Sensitivity) and to 0 in the false positive rate (Specificity). The estimated Area Under the Curve (AUC) is what determines how close the test is to the flawless accuracy (100,0). In the test represented by blue color, the narrow points the best cut-off point for the test. In the test represented by green color, the measures for true positives and false negatives generate a perfect curve, in which the AUC is 100%. In this case, the diagnosis is ideal to separate two distinct groups. A comparison between two tests is displayed in Figure 22:

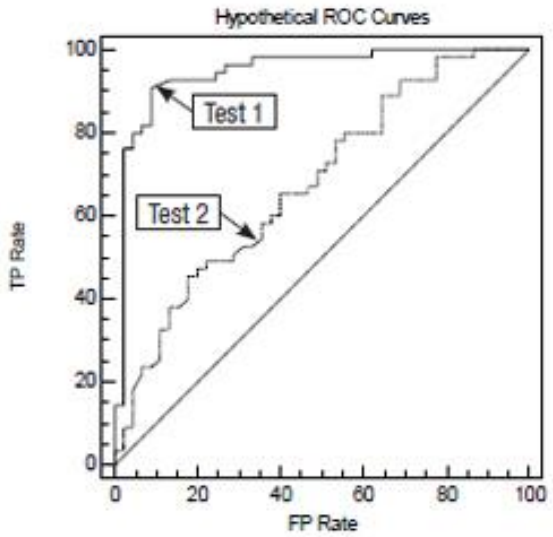


Figure 22: Hypothetical ROC curves for two diagnostic tests[12]

Considering that the ability of the test in separating two groups is proportional to the extent of the AUC, it is assumed that test 1 (Figure 22) is more adequate than test 2. Moreover, it is noticeable that for test one the curve is closer to an equal-to-1 true positive rate (Sensitivity) and an equal-to-zero false positive rate (Specificity).

In order to assess the ability of VLT in separate two groups of L2 proficiency, we submitted the scores to an ROC curve analysis. Moreover, since we are proposing a

[12] Savaloja, L. & Birdsong, G. Validating and verifying molecular tests in cytopathology, 2011. Available at: < http://www.captodayonline.com/>, access: 01/15/2016.

correlation among the VLT, the placement test and the acceptability judgment in both experiments, we also submitted those scores to ROC curve analysis. Our goal is to assess these tests' ability to differentiate accurately low-proficiency from high-proficiency individuals.

3.5 Participants

In the Experiment 1, 30 individuals took part in the research; 17 women and 13 men, with a mean age of 25.6 years (sd=6.03). Every participant completed or is in the process of completing at least a bachelor's degree. The participants were all right-handed, with good (or corrected) vision, living in Belo Horizonte/MG or Goiânia/GO. Each participant was bilingual (Portuguese/English), and was classified into two groups of English proficiency level based on the OPT and VLT test scores.

The proficiency classification employed both OPT and VLT scores. Following our criteria, in order to be considered as having high proficiency, participants needed to achieve 55 points (or more) on the OPT (91% is C2 level according to the CEFR descriptors), and also successfully complete level 5 of the VLT (10,000 frequent words). Level 4 of the VLT was not considered. As discussed above, this level was demonstrated not to differentiate the average college level Brazilian Portuguese-English bilingual with regards to L2 vocabulary knowledge (Souza, Duarte & Berg, 2015). Furthermore, as will be shown below, there were no significant differences among participants whose VLT scores placed them in levels 1-3 with respect to their performance in either the OPT or the AJ task. On the other hand, participants whose VLT scores placed them in level 5 showed performances that were significantly different from participants at VLT levels 1-3. Descriptive information about the proficiency-based stratification of our participant sample in Experiment 1 is displayed in Table 2.

Table 2: Participants' proficiency in Vocabulary Level Test (VLT) and Oxford Placement Test (OPT)

Tests			
VLT (Word Frequency level)		*OPT (CREF levels)*	
2000, 3000, 5000	10000	A2, B1, B2	C2
proficiency level		proficiency level	
Low	High	Low	High

participants (n=30)	18	12	18	12

Experiment 2 consisted of 2 groups. The first group contained 20 individuals, (age mean=21, sd=1.92), 9 of which were female. All participants were born in the United States of America and speak English as his or her first language. The parents of participants were also all born in the United States of America speaking English as their first language. All were exposed to Spanish at school with an average age mean of 13 years (sd=2.15). All of them had good (or corrected) vision; 18 were right-handed. They lived in New York City at the time of the research, and all had completed or are completing at least a bachelor's degree at the City University of New York (Queens College). The second group was composed of 40 individuals (age mean=23, sd=7.40), 32 of which were female. All of them had good (or corrected) vision, and 38 were right-handed. Each participant was born in a Spanish language context—some in a Spanish language country, and some in New York in a Spanish-speaking family. All participants had Spanish as their first language, and acquired it at home. Every participant was self-declared as being exposed to both languages since birth, and some of them were exposed to English a little later (mean age to English exposure: 3.9, sd=3.3).

Participants of both groups were gathered under the same processes. We displayed some flyers at the campus looking for participants, and we also used the Research Participant System (SONA) of the Department of Psychology of the City University of New York (Queens College). After the data was collected, we used the BHQ to classify them into Spanish L2 learners (group 1) or Spanish HL speakers (group 2). Descriptive information about the proficiency-based stratification of our participant groups in the Experiment 2 is displayed in Table 3.

Table 3: Participants' proficiency in Vocabulary Level Test (VLT) and Spanish Placement Test (SPT)

	Tests			
	VLT (Word Frequency level)		*SPT (CREF levels)*	
	2000, 3000, 5000	10000	A2, B1, B2	C2
	proficiency level		proficiency level	
	Low	High	Low	High
L2 learners (n=20)	20	-	20	-
HL speakers (41)	22	18	22	18

This chapter presented an overview of the methodology that guided our study. We described in details the main procedures, materials, participants and tools for analysis. In the next chapter, we will analyze the data and infer some aspects of our assumptions that emerge in the data. First, the Experiment 1 will be analyzed, concerning the study conducted in Brazil. After that, the second Experiment will be analyzed concerning the studied conducted in the New York City.

CHAPTER 4

DATA ANALYSIS

The data analysis for this study will be presented in two stages, concerning Experiment 1 and Experiment 2. Initially, we explore data from Experiment 1 aiming at observing the distribution of data through statistic test of normality. Following, we present the procedures for the establishment of the minimum time for the judgment. The next step was devoted to investigate the correlation between the VLT with the OPT and with the AJ task. Finally, we present the ROC curve analysis for the VLT, the OPT and the AJ task. Thenceforth, data from the second experiment will be presented. Following the same procedures from the Experiment 1, data were tested on their normality. Following, we will explore the data from the BHQ. After that, we will analyze the correlation between VLT, SPT, the self-assessment test and the AJ task. Finally, we will present the ROC curve analysis for SPT, VLT and also for the AJ task.

4.1 Experiment 1

4.1.1 Test for normality and estimative of time

First, an exploratory data analysis was conducted to determine if reaction time (RT) means for the 30 subjects in the Acceptability Judgment task were normally distributed for each target sentence. The goal for this analysis is to assert the normality of the data, allowing us to employ parametric tests. Results from a Kolmogorov-Smirnov test for normality indicated that the distribution of the RT means did not deviate significantly from a normal distribution in all four cases and in the overall case. Normality test results, means, and standard deviation are displayed in Table 4:

Table 4: Means, standard deviation and normality test of reaction time means to sentence type (n=30)

Sentence type	RT (msec)	Sd (msec)	Kolmogorov-Smirnov test
Verbal transitivity violation	4821	.746	.121*
Morphosyntactic violation	5488	.912	.104*
Induced-movement alternation	4809	.807	.143*
Grammatical	5104	.796	.130*
Overall results	5056	1018	.152*

* p > .05

Following the confirmation of the normality of data, we must estimate the maximum time one would take to make a judgment call for the sentences. In order to establish the estimate, we considered the RTs by level of proficiency, based on the procedures developed by Souza et al. (2015). The difference of means for all sentences combined (grammatical/ungrammatical) between low-proficiency (M=5569, SD=.426) and high-proficiency (M=4509, SD=.655) was significant, t(4) 8, 268, p< .005. We suggest that a value of one RT mean of lower proficiency added with a standard deviation is the maximum time necessary in an Acceptability Judgment under this configuration. Thus, we suggest that 6000 milliseconds is the maximum time that a bilingual (Portuguese/English) takes to judge a sentence written in English with approximately 40 characters (spaces excluded).

4.1.2 Correlational analysis

We then proceeded to the confirmatory investigation of the correlation of VLT scores and a general proficiency measure (the OPT in Experiment 1). A Pearson product-moment correlation coefficient (*r*) was computed to assess the relationship between the two tests in order to verify the degree of correlation between the scores each one produces as diagnosis of L2 proficiency. To achieve this, we computed the total number of scores reached by low and high-proficiency participants in OPT and VLT. Our hypothesis was that there would be a positive correlation between tests for each group of proficiency. The data displayed in Table 5 confirms this hypothesis:

Table 5: Correlation between OPT and VLT proficiency levels

	Low proficient (OPT)		**low proficient (VLT)**		
N	Mean	SD	Mean	SD	**Pearson's *r***
18	35.9	10.3	43.4	5.6	.559*
	high proficient (OPT)		**high proficient (VLT)**		
	Mean	SD	Mean	SD	
12	55.7	.86	86.9	2.6	.586*

*$p < .05$

These results reveal that the VLT scores indicating high proficiency (VLT level 5 in our definition) correlate significantly with differences in OPT scores. The lower correlation among low-proficiency subjects in both tests (r=.559) can be attributed to the higher variation among the subjects' performances, as attested by the clearly higher standard deviations observed with the low-proficiency group. We interpret this result as a satisfactory confirmation for the claim that a measure of vocabulary size correlates to gauges of overall proficiency for the population of interest to Experiment 1, namely college-level Brazilian Portuguese-English bilinguals. These results agree with Milton (2013) and Stæhr (2008) on the notion that vocabulary size measures correlate with overall proficiency test, including grammar and reading abilities.

In the interest of verifying the correlation among the proficiency tests (the OPT and VLT) and our timed AJ task, we first analyzed the judgments elicited by the collapsed grammatical and ungrammatical sentences in the two proficiency level groups. As demonstrated below in Figure 23, differences were found in the behavior of the two groups. Such differences are related to both the grammaticality status of the stimuli and the proficiency level of the participants.

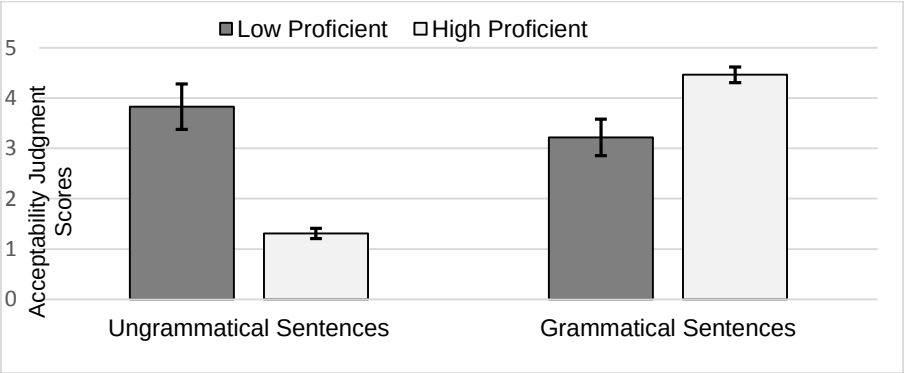


Figure 23: Means of sentence judgments by proficiency level

There were significant differences between the judgment calls for the ungrammatical sentences made by the low-proficiency group (M=3.8, SD=0.45) and the high-proficiency group (M=1.3, SD=0.10), t(28)=18,97, $p< .01$. This clearly indicates that the low-proficiency participants were generally unable to detect the violations of the ungrammatical sentences within the 6-second time window of our AJ task. This observation supports Jiang's (2007) proposal that some bilinguals may experience difficulty integrating morphosyntactic information when processing L2 stimuli. We interpret this situation as an indicator of the lower level of automaticity of the low-proficiency sample when compared to the higher proficiency sample of the present study. It should be noted that the high-proficiency participants (M=4.46, SD=0.15) actually fared better at indicating the grammatical sentences than the low-proficiency group (M=3.21, SD=0.36), t(28)=11,14, $p< .01$.

As discussed below, we interpret this observation as a probable effect of the inclusion of the induced movement alternation sentences among our grammatical sentences. It were only the high-proficiency participants' responses that yielded a significant difference between the ungrammatical (M=1.30, SD=0.10) and the grammatical sentences (M=4.46, SD=0.15), t(11)=54,10, $p< .01$. Among the low-proficiency participants sample that we observed, the pattern actually indicates a tendency for misjudgment, with ungrammatical sentences (M=3.83, SD=0.45) yielding higher mean judgments than grammatical sentences (M=3.21, SD=0.36), t(17)=4.25, $p< .05$. All in all, we interpret the pattern of our results as showing that only the participants whose VLT and OPT scores classify as high-proficiency had sufficiently automatic access to their L2 grammatical representations to perform satisfactorily within the average 6-second ceiling of our timed AJ task.

Consequently, we analyzed the specific role of each target sentence type in the timed AJ task for the two proficiency groups. A repeated-measures analysis of variance of AJ scores means of the low-proficiency group across the four groups of sentences indicated a main effect of sentence type considering subjects as a random factor, $F1(3,51)=9.45$, $p< .001$, $\eta_p^2 = .357$, and items as a random factor $F2(3,21)=4.20$, $p< .05$, $\eta_p^2 = .375$. The repeated-measures analysis of variance of AJ scores means for the high-proficiency group also revealed a main effect of sentence type when subjects were taken

as random factor: F1(3,33)= 1159, p< .001, η_p^2 = .991, as well as item as random factor: F2(3,21)=685, p< .001, η_p^2 = .990. Results are displayed in Figure 24.

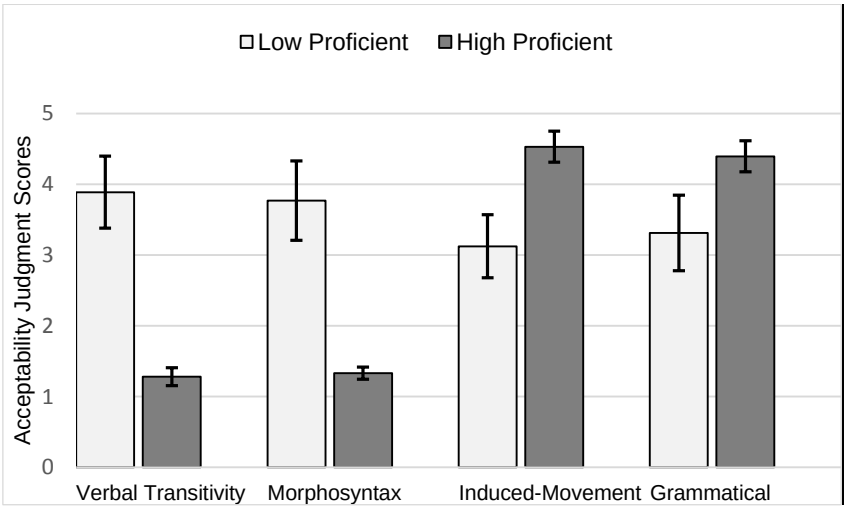


Figure 24: Means of sentence type judgments by proficiency level

Observable in Figure 24, it is only with the high-proficiency group that consistent detection of ungrammaticality as opposed to grammaticality takes place. Therefore, it is clear that the sentence effect yielded among the low-proficiency participants in the 6-second ceiling of our task is not driven by access to grammatical representations (some of which are shared by their L1), such as agreement and WH-movement violations. This supports our interpretation that only the participants whose VLT/OPT scores indicate high proficiency demonstrate sufficient automaticity to access L2 grammatical representations under strict time constraints.

4.1.3 ROC curve analysis

As we previously mentioned, VLT, OPT, and the AJ task were considered diagnostic tests for proficiency levels in our study. The vocabulary level test (VLT), the overall grammar knowledge (OPT), and the perception of (un)grammaticality (AJ task) were administered with the purpose of distinguishing two proficiency profiles among the participants: L2 English high-proficiency and L2 English low-proficiency individuals. Conducive to ascertaining the accuracy of the three tests, we applied the ROC curve methodology. First, we submitted VLT and OPT scores to ROC curve analysis for the

purpose of analyzing their ability to detect high levels of proficiency accurately, through measures of Specificity and Sensitivity. Results are displayed in Figure 25:

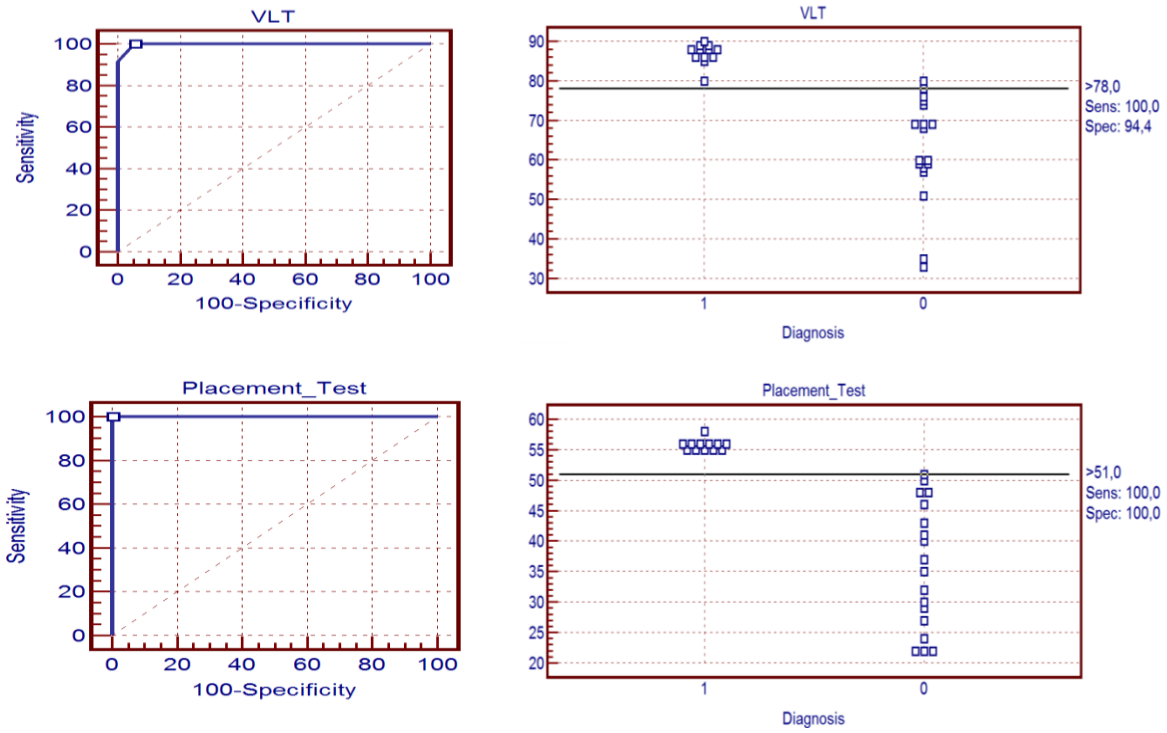


Figure 25: ROC curve graphs and Interactive Dot Diagrams for VLT and OPT

As evidenced in Figure 18, VLT scores with the cut-off point we proposed generated an almost perfect curve, since the Sensitivity is up to the total (100,0) and the Specificity is close to the total (94,4). Considering that Sensitivity represents the group of high-proficiency individuals that the test accurately identifies as positive, and Specificity represents the group of low-proficient individuals that the test accurately identifies as negatives, the point of convergence represented in the graph shows a perfect level of Sensitivity to detect the high proficiency. Differently, it shows no total efficacy in distinguishing the low-proficiency group as true negatives.

This result can be seen in more detail in the Interactive Dot Diagram (Figure 18), with the first column (number 1) representing our group of high-proficiency individuals, the second column (number 2) representing the low-proficiency individuals, and the line as the cut-off point we proposed (78,0). In this Figure, it is evidenced that the high-proficiency group is entirely above the line, indicating that the cut-off point we proposed for the test is accurate in detecting those who are highly proficient in English (as true

positives). On the other hand, if we look at the second column, there are some dots above the line. It means that some low-proficiency individuals scored above the cut-off point. This result shows that the test is not perfect in detecting low-proficiency individuals, since the Specificity presented in the ROC curve is higher than zero. Based on this result, we still consider VLT to be an adequate diagnostic tool for English L2 proficiency measures. The results for the cut-off points we suggested for both tests (78) are summarized in Table 6 below:

Table 6: Sample of Cut-off point values, Sensitivity and Specificity of VLT and OPT scores for Experiment 1

	Area under the curve (standard error)	Criterion (cut-off point)	Sensitivity (C.I.)	Specificity (C.I.)	+LR	-LR
VLT	0,998 (0,010)	78	100,0 (73,4 – 100)	77,8 (52,4 – 93,5)	4,50	0,00
OPT	1,000 (0,000)	51	100,0 (73,4 – 100)	100,0 (81,3 – 100,0)	-	0,00

As we can see in table 5, the ROC curve analysis for the VLT scores evidenced an AUC close to 1 (0,998), which gives VLT a considerable level of reliability as a diagnostic test, since the perfect result is 1,000. The results also revealed a perfect accuracy in detecting the high-proficiency individuals (Sensitivity) (100,0, C.I.=73,4 – 100), although its power to accurately identify low-proficiency individuals as truly negatives is not one hundred percent accurate (Specificity) (94,4, C.I.=72,6 – 99,1).

Following, we submitted the AJ task to the ROC curve analysis in order to examine if the task used as a proficiency test is accurate in distinguishing those individuals who have a high level of proficiency from those who do not. In order to achieve this, we organized the scores of ungrammatical and grammatical sentences to be measured separately. The reason for that relies in the fact that both structures generate opposite scores among the participants in the Likert scale (from 1 to 5). For instance, a low-proficiency participant who has a mean of 3,3 in the grammatical structures and 3,0 in the ungrammatical structures would have 3,1 as overall mean. On the other hand, a high-proficiency participant who scores a mean of 4,4 in the grammatical sentences and 1,8 in the ungrammatical sentences would also have an overall mean of 3,1. In this way, the overall result does not represent participants' performances on the test. For that reason, we divided the AJ task scores in

ungrammatical and grammatical. This operation allowed us to see properly how both groups of proficiency performed in the AJ task, as we can see in Figure 26 below:

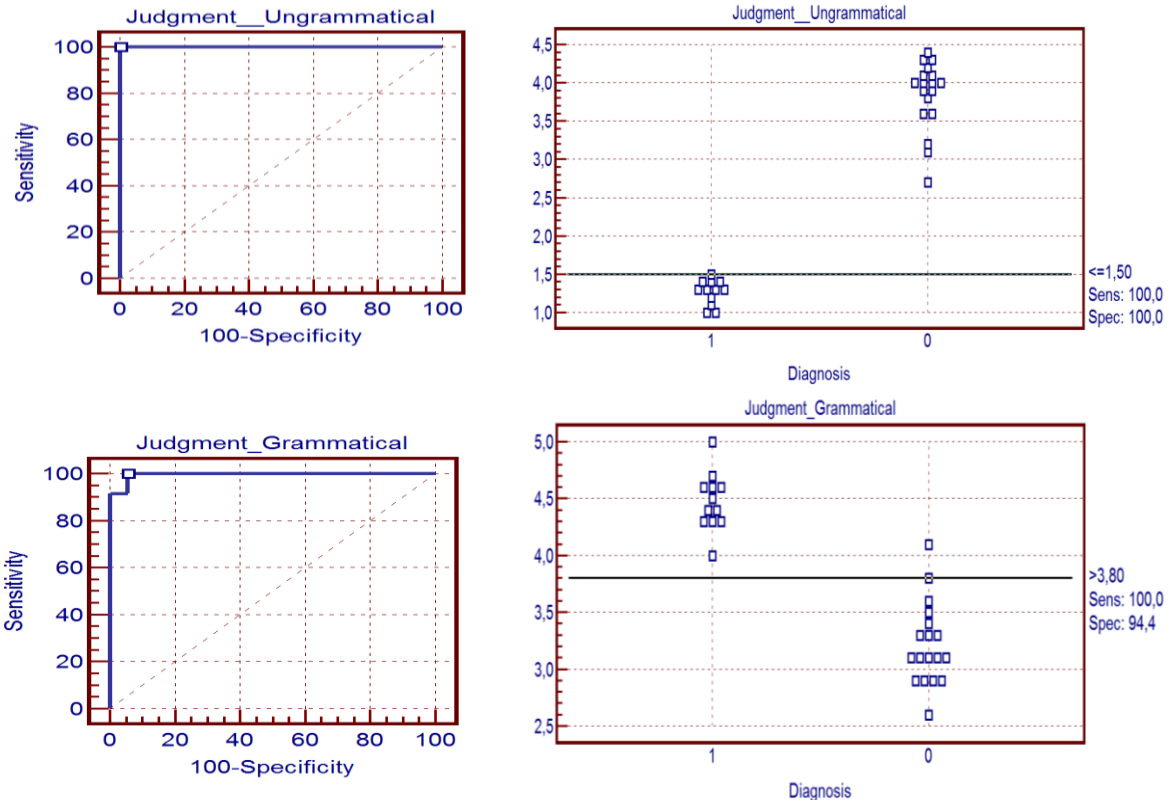


Figure 26: ROC curve graphs and Interactive Dot Diagrams for AJ task (Ungrammatical and Grammatical sentences) for Experiment 1

As demonstrated above, the AJ task is an adequate diagnostic test for proficiency (Figure 19), especially with the grammatically unlicensed structures. The ROC curve for ungrammatical structures is perfect since the cut-off point divides both groups of proficiency with precise Sensitivity (true positives), and Specificity (true negatives). In the Dot Diagram, it is visually represented that all participants who were classified as high-proficiency are under the line (cut-off point), while all participants who were classified as the low-proficiency group are above the line. A slightly different situation is seen with the grammatical structures, because the ability of detecting the low-proficiency group as true negatives is not totally perfect. The Dot Diagram for the grammatical structures reveals a high Sensitivity (100,0), and the high-proficiency group remains above the line, while the Specificity is not totally accurate (94,4). As we see, some participants who were classified as low-proficiency are above the cut-off point line. Below these results are summarized in table 7:

Table 7: Sample of Cut-off point values, Sensitivity and Specificity of AJ task (Ungrammatical and Grammatical sentences) for Experiment 1

	Area under the curve (standard error)	Criterion (cut-off point)	Sensitivity (C.I.)	Specificity (C.I.)	+LR	-LR
AJ (Ungr)	1,000 (0,000)	1,5	100,0 (73,4 – 100)	100,0 (81,3 – 100,0)	-	0,00
AJ (Gram)	0,995 (0,014)	3,8	100,0 (73,4 – 100)	94,4 (72,6 – 99,1)	18,00	0,00

The data reveals that the twofold analysis uncovers that the AJ task is an adequate instrument for English proficiency diagnosis among the population we studied, mainly concerning the ungrammatical structures. Both types of structures present an accurate Sensitivity; this means the ability of detecting high-proficiency participants as true positives, while the Specificity in the ungrammatical sentence analysis is not perfect (94,4). It is important to mention that in a diagnostic test, we expect ROC curve analysis to show an AUC of 1,000 (total). This expectation is confirmed with the ungrammatical sentences. Differently, the ROC curve analysis for the grammatical sentences shows an AUC of 0,995 due to Specificity data.

4.2 Experiment 2

4.2.1 Normality test and reaction time analysis

Following the procedures of Experiment 1, an exploratory data analysis was conducted to determine if reaction time (RT) means for the 30 subjects in the Acceptability Judgment task were normally distributed for each target sentence. Results from a Kolmogorov-Smirnov test for normality indicated that the distribution of the RT means did not deviate significantly from a normal distribution in all four cases. Normality test results, means, and standard deviations are displayed in Table 7:

Table 7: Normality test of Reaction Time means by sentence type of L2 learners and HL speakers

	Spanish L2 learners (n=20)			Spanish HL Speakers (n=40)		
Sentence type	**RT (msec)**	**Sd (msec)**	**Kolmogorov Smirnov**	**RT (msec)**	**Sd (msec)**	**Kolmogorov Smirnov**
Transitivity violation	3593	.914	181*	3732	.640	.118*
Morphosynt. Violation	3843	.832	.134*	3884	.548	.124*
Change of state verbs	3484	.814	.123*	3506	.660	.098*
Transitive verbs	3371	.733	.095*	3284	.617	.077*
Overall results	3633	1.139	.104*	3616	.622	.101*

* $p > .05$

Next, we investigated whether there is an effect in the RT means of group type (L2 learners versus HL speakers), proficiency level (low versus high), or in the sentence type (ungrammatical versus ungrammatical). First, we determined if there is a difference in the RT between HL speakers (n=41, m=3592, sd=.504) and L2 learners (n=20, m=3566, sd= 751). An independent sample t-test showed no significant difference between both groups in their RT to judge the target sentences of the Experiment 2, t(59)=0,159, p> .05.

Subsequently, we investigated if the RT means for the judgment task would be affected by proficiency. In order to do this, we administered an independent-sample one-way ANOVA for three groups of proficiency, taking subjects as a random factor: L2 leaners (n=21, m=3566, sd=.751), low-proficiency HL speakers (n=23, m=3772, sd=.412), and high-proficiency HL speakers (n=18, m=3362, sd=.528). The ANOVA test showed no effect of proficiency across the means, F(2,58)=2,58, p > .05. Post-hoc tests with Bonferroni corrections demonstrated no significance among the three groups in the pairwise comparison.

With all of this in mind, we set out to determine whether there would be a sentence type effect on participants' RT in the judgment task. To do so, first we compare the HL speakers' RT means for both ungrammatical (m=3800, sd=.539) and grammatical (m=3391, sd=.586) sentences. A paired-sample t-test revealed a significant difference between both sentence type RTs, t(40)=5,23, p< .01. Similar results were found for the L2 learners regarding RT for ungrammatical (m=3718, sd=.816) and grammatical (3425, sd=.729) sentences: t(19)=3,94, p< .01.

These results suggest that, according to their level of proficiency, participants are not different in relation to time they take to judge the grammaticality of a sentence in the task we presented. These results also suggest that both L2 learners and HL speakers take more time to judge ungrammatical sentences than those that are grammatical. This may be due to the fact that ungrammatical sentences are harder to process and comprehend.

4.2.2 Analysis of the Bilingual History Questionnaire

Our next step was to explore the participants' self-reported data through the BHQ in consideration of better understanding their relationship with both languages throughout their lives, as well as their perception about their proficiency in both languages from birth to present day. To commence, we analyzed the exposure of both languages to L2 learners and HL speakers. Utilizing a scale from 1 to 5 in which 1 is Spanish and 5 is English, participants declared their exposure to both languages in the different moments of their lives. Data from both groups is displayed in Figure 27 below:

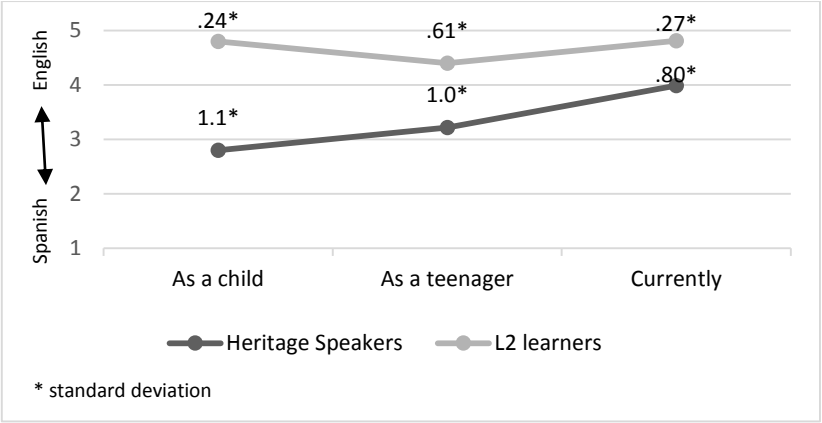


Figure 27 – Language exposure through life periods by learner type

As expected, Figure 27 exhibits that L2 learners were massively exposed to English from birth through present day. English is their first language, as well as the first language of their parents. The only slight difference we can observe is in their teenage years, where they had a higher exposure to Spanish. The variation among participants' mean regarding the exposure to both languages in the L2 learners groups is relatively low,

and it has a higher variation as a teenager (.61). This higher variation could be explained by the fact that, among L2 learners, some of them were exposed to Spanish in middle school (approximately 11-14 years), and some of them in high school (approximately 14-18 years).

The HL speakers group has a different behavior. As shown, participants reported that as a child they were exposed to both languages almost equally, which would differentiate from the hypothesis that they are massively exposed only to Spanish at home. However, there are three facts that can explain this datum. First, the variation between the answers is relatively high (1.1) which demonstrates that some of them answered that the Spanish exposure was almost the only language exposure they had, and some of them that they were exposed even more to English. Second, is the type of interactions that participants had (as speaker or as listener).

The BHQ inquires about participants' exposure to languages in two forms: actively (participants speaking with someone else) or passively (someone else talking to participants). This way, that variation may be explained by the difference in their interactions. Third, this variation may be due to *where* they were/are exposed to both languages. The BHQ addresses home and outside language use, which can bring great differences in answers. For instance, a person who was born in the USA in a Spanish language environment (parents who are Spanish native speakers) may speak only Spanish at home with family, and only English outside (social interactions). As displayed in Figure 18, as teenagers when their exposure to English increases, there is no significant difference among the means and also the variations continue to be high (1.0). Currently, HL speakers reported to be more exposed to English than Spanish, which goes together with the idea of inversion of dominance.

In order to solve the problem with the exposure type (passive/active) and the locus of exposure (home/outside), we scrutinized their report about language exposure, separating both exposure type and local. Below, Figure 28 and Figure 29 represent this information respectively.

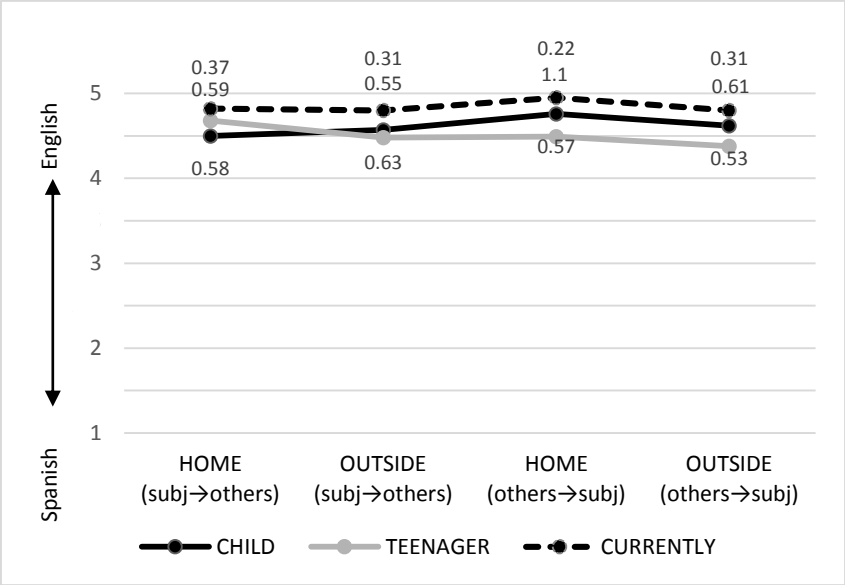


Figure 28 – Language exposure through life periods by L2 learners

As evidenced in Figure 28, the exposure type or locus seem not to interfere in L2 learners' report on their language history. We recognize that as teenagers they end up being exposed to Spanish more than as a child, and more than currently. However, there is no significant difference in any means regarding their exposure to language. Another point worth noting is the fact that their exposure to Spanish (passively or actively) is a little higher outside than at home, which is expected since in the context of New York the Spanish language is present in several social situations. A quite different situation is presented in Figure 29:

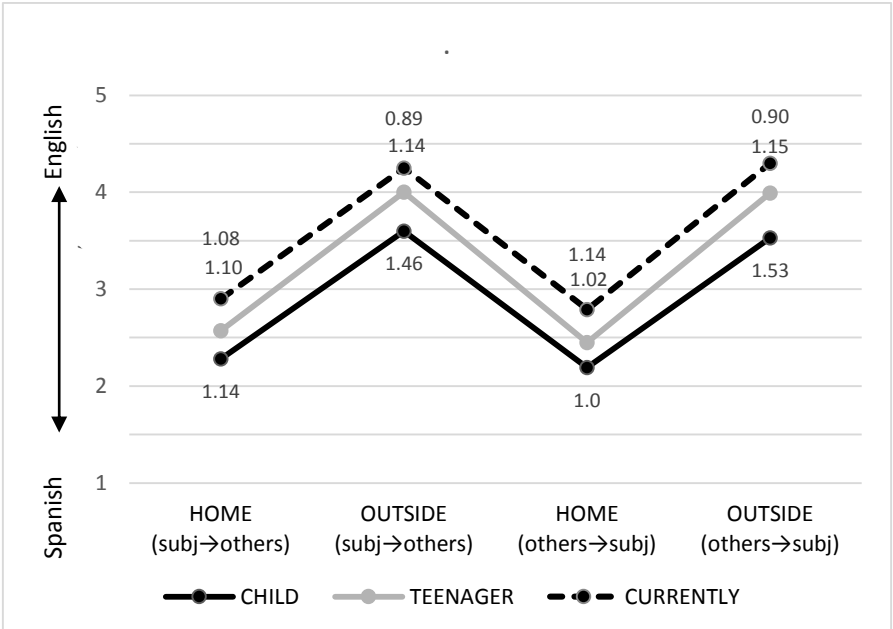


Figure 29 – Language exposure through life periods by HL speakers

As demonstrated in Figure 29, differently from the L2 learners', HL speakers' data have a different behavior about their exposure to both languages. As a child, teenager, and currently, participants seem to be more exposed to Spanish at home (speaking the language or listening to it) than outside. Even with a considerable variation, there is a notable difference. This confirms what Singleton (1999) pointed out that early bilinguals as infants in a naturalist context appear to receive ostensive input from their interlocutors. Moreover, data confirm the hypothesis that, even as adults, HL speakers keep using their non-dominant language among relatives at home. Another significant observation is that gradually HL speakers use more English as they grow, independently of the context of use. That is another suggestion for the inversion of dominance.

4.2.3 Correlational analysis

We then proceeded to the confirmatory investigation of the correlation of VLT scores, the general proficiency diagnostic measure (the SPT in the present study), and the self-assessment of language abilities. The first step was to analyze how both L2 learners and HL speakers self-assessed their abilities in speaking, writing, listening, and reading in both English and Spanish. Figure 30 and Figure 31 below display these assessments:

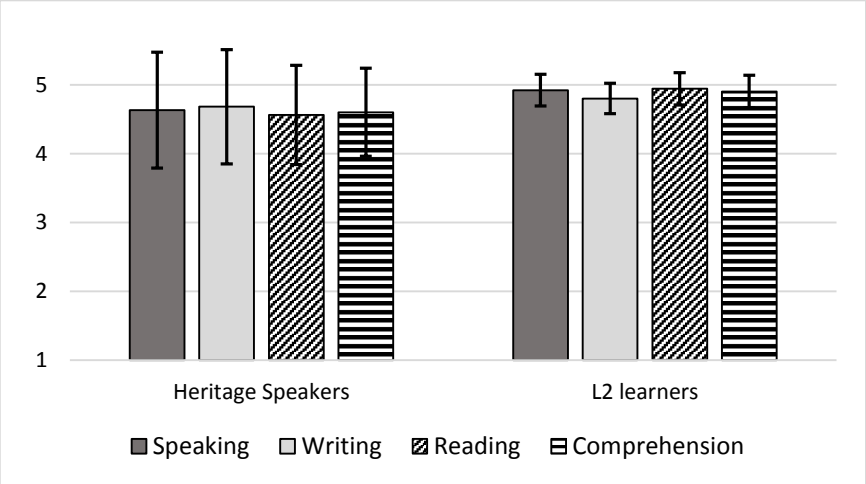


Figure 30 – HL speakers' and L2 learners' Self-assessment in English skills

As evidenced in Figure 30 above, both L2 learners and HL speakers have a high self-assessment in the four abilities in English that we are observing. This result confirms

the idea that as adults, HL speakers have already acquired a high level of their dominant language and also suggests the process of inversion of dominance. As we expected, Spanish L2 learners had a very high self-assessment of their abilities, since it is their first/dominant language. The only lower evaluation was on their writing ability. Below we display the self-assessment from both L2 learners and HL speakers on Spanish:

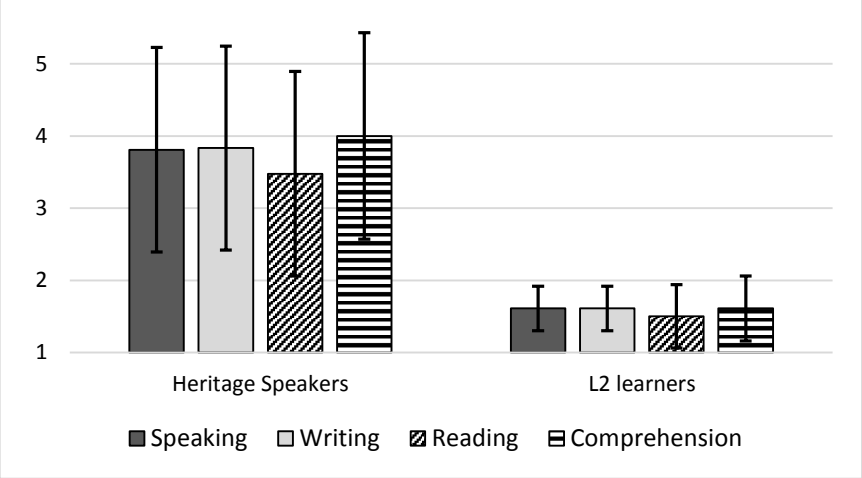


Figure 31 – HL speakers' and L2 learners' Self-assessment in Spanish skills

Figure 31 above is informative about participants' perception of their abilities in Spanish. The L2 learners groups have a very low evaluation of themselves on the four abilities. This result confirms the proficiency profiling we created based on VLT and SPT scores, in which all of the L2 learners were classified as having low proficiency in Spanish.

The HL speakers' self-assessment in Spanish presents a strong variation on this graph. This variation may be due to fact that, among the groups of HL speakers, there are low and high-proficiency participants. This way, in order to better see the HL speakers' self-assessment data, we have to separate them in two groups according to their proficiency level. In Figure 32 and Figure 33 below, we display the self-assessment data of HL speakers in both English and Spanish.

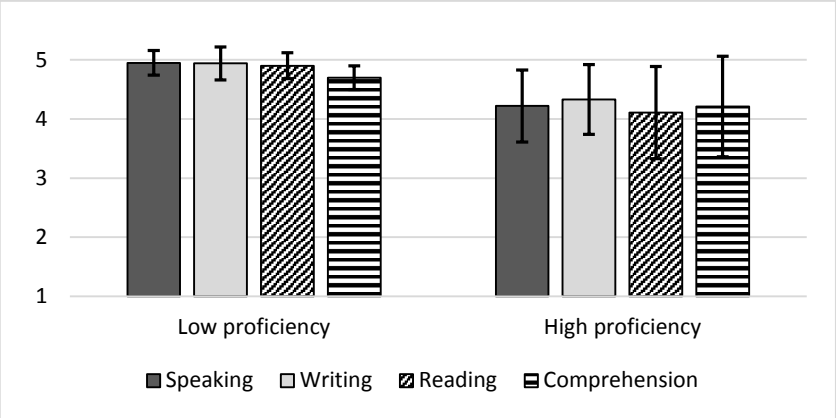


Figure 32 – Self-assessment in English skills by proficiency (HL speakers)

As displayed above (Figure 32), the Spanish low-proficiency HL speakers evaluate themselves as having high-proficiency in English, which is an interesting data confirming that their language may have inverted through their experiences throughout the years. The same logic applies to the Spanish high-proficiency group, which assessed themselves lower than the first groups. It may suggest that they must somehow have their non-dominant language stronger than the first group. Figure 32 also shows that, differently from what Montrul (2012) affirmed that speakers, regardless of their proficiency level are not better in receptive abilities (reading and comprehension) than in productive abilities (speaking and writing). There were no significant differences to support this claim. In Figure 33 below, we show results from the self-assessment test of Spanish from both low and high-proficiency HL speakers.

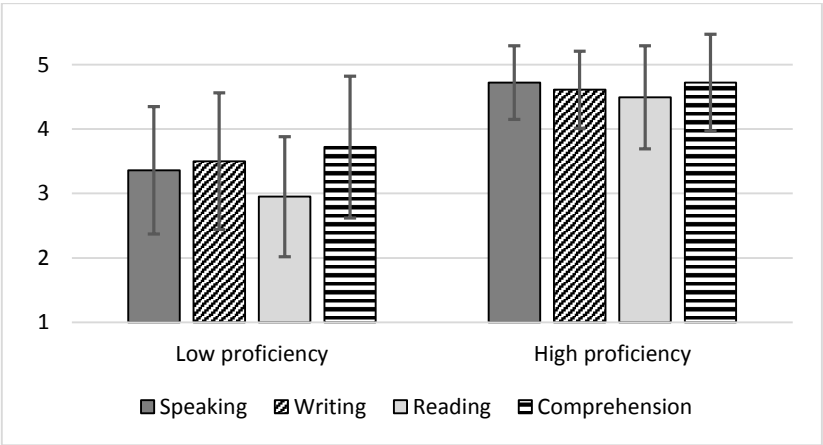


Figure 33 – Self-assessment in Spanish skills by proficiency (HL speakers)

The data displayed in Figure 33 above confirms the results of both proficiency measures we adopted in this study (VLT and SPT). The high-proficiency groups (according to our tests) also highly assessed themselves in Spanish, while the low-proficiency group evaluated themselves lowly and with more variation. These results so far agree with Meara & Malcoy (2010) about VLT being a valuable instrument to measure overall proficiency, especially due to the correlation it presents with other forms of assessing. These findings also come along with Singleton (1999) about the idea that the consciousness of word level plays a crucial role in the L2 learning. From now on, we describe participants' performances in the AJ task.

To begin, we analyzed the overall performance of both L2 learners and HL speakers on ungrammatical and grammatical sentences. HL speakers in general demonstrated more accuracy in differentiating a sentence's grammaticality than L2 learners. L2 learners were no able to perceive the status of the sentence and their judgments are higher for sentences that are ungrammatical, and lower for grammatical sentences. Data are displayed in Figure 34 below:

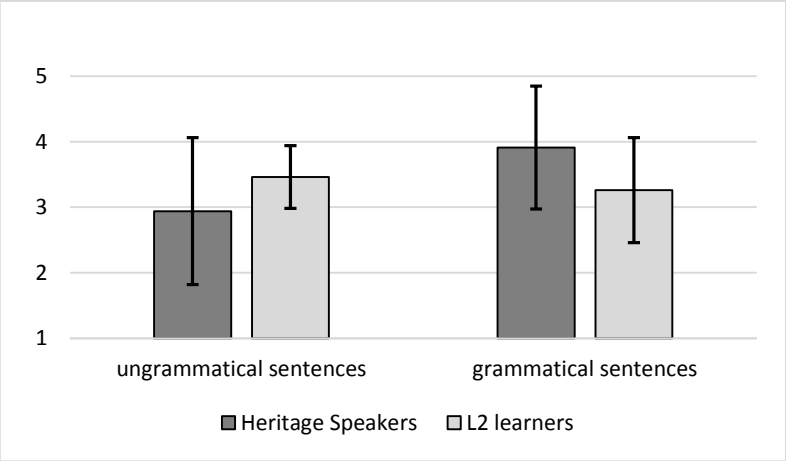


Figure 34 – Judgment scores of groups for sentence status

Subsequently, we studied the difference between and within both groups in relation to their performance on the AJ task. An independent-sample t-test was applied considering the judgment for ungrammatical sentences by both L2 learners (n=20, m=3.46, sd=.48) and HL speakers (n=41, m=2.95, sd=1.13), t(59)=-1,92, p> .05. As we can see there is no difference between both groups judging ungrammatical sentences. This may due to the fact that HL speakers are contained by low and high-proficiency, which

can explain the high variation (sd=1.13). Then we apply the same test statistic procedure to see participants' performances in the grammatical sentences. Below (Figure 35) there is is a descriptive demonstration of results:

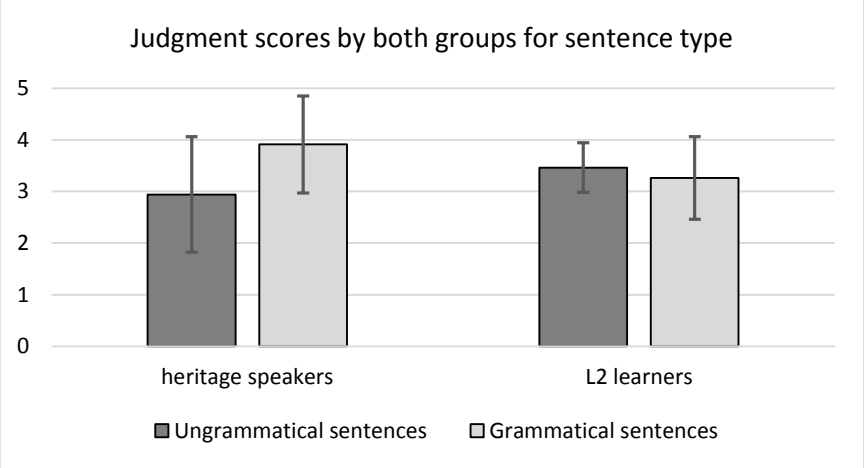


Figure 35 – Judgment scores within groups for sentence status

As displayed, the same pattern occurs for L2 participants (n=20, m=3.91, sd=.95) since they are not able to distinguish the grammaticality status. The same pattern also happens to the HL speakers group (n=41, m=3.91, sd=.95), as they differentiate both sentence statuses. Different from the ungrammatical sentences judgment, to grammatical sentences, both groups revealed no statistical difference between them: t(59)=2,62, p< .05. Next, we analyzed the mean difference within groups in order to see if they significantly differ regarding the grammaticality status of the sentences.

First, we analyzed L2 learners' judgments on the sentences that were ungrammatical (n=20, m=3.46, sd=.48) and grammatical (n=20, n=3.26, sd=.80), t(19)=1,02, p> .05. As can be noted, there is no significant difference in both sentence status judgments. This result confirms that, in our experiment, L2 learners with low proficiency were not able to differentiate grammatical from ungrammatical sentences in their L2. Consecutively, we checked HL speakers' performances on the sentences that were ungrammatical (n=40, m=2.94, sd=1.1) and grammatical (n=41, m=3.91, sd=.94), t(41)=-3,82, p< .01. As we can see, there is a difference between means. It may suggest that (no matter the proficiency level) the heritage-speakers group was able to separate the sentences according to their grammaticality.

As we have observed from the results so far, HL speakers are divided in two groups of proficiency (low and high). Based on that assumption, and on the high variation we found when HL speakers were counted as one general group, we created three sub-groups according to the proficiency level: L2 learners (low-proficiency), HL speakers (low-proficiency), and HL speakers (high-proficiency). Descriptive data is displayed in Figure 36 below:

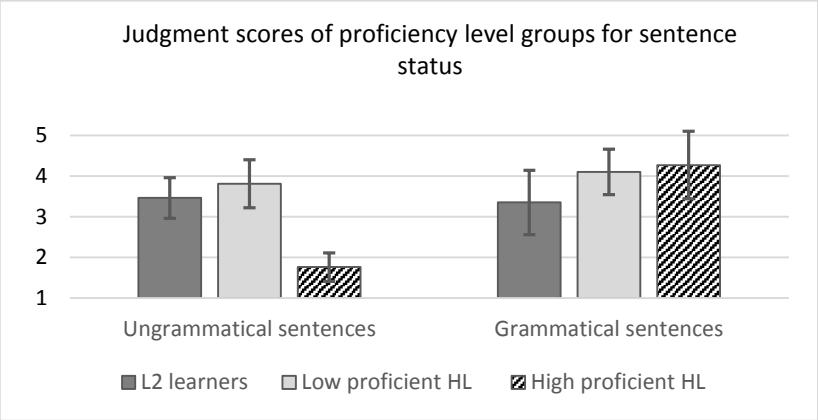


Figure 36 – Judgment scores of three proficiency groups for sentence status

In order to see if there was an effect of proficiency among the three groups on their performances in the judgment task, we applied a one-way ANOVA for independent samples for both ungrammatical and grammatical sentences. First, we analyzed the effect of proficiency of L2 learners with low proficiency ((n=20, m=3.46, sd= .48), low-proficiency HL speakers (n=23, m=3.83, sd= .53), and high-proficiency HL speakers (n=18, m=1.76, sd= .35). We found an effect for both subject as a random factor, F1(2.58)=106,6, p< .01, and item as a random factor, F2(2,30)=199,6, p< .01. A post-hoc test with Bonferroni corrections showed significance in all of multiple comparisons: between low-proficiency L2 learners and low HL speakers (p=.04), which is unexpected, since both groups have low proficiency. Significance was also found between low-proficiency L2 learners and high HL speakers (p<.01), and also between low and high-proficiency HL speakers (p<.01), which we expected since they differ in levels of proficiency.

The next step was to see if an effect of proficiency existed in the three groups' performances for the grammatical sentences. The same statistical procedure was applied for low-proficiency L2 learners (n=20, m=3.26, sd=.80, low-proficiency HL speakers

(n=23, m=3.64, sd=.96), and high-proficiency HL speakers (n=18, m=4.27, sd=.83). An effect of proficiency was revealed when subjects were taken as a random factor, $F(2,58)=6,36$, $p<.05$, as well as item as a random factor, $F2(2,30)=47,58$, $p<.01$. In multiple comparisons, Bonferroni corrections revealed a significant difference only between low-proficiency L2 learners and high-proficiency HL speakers ($p<.05$). There was no significance between low and high-proficiency HL speakers ($p=.081$), or between low-proficiency L2 learners and low-proficiency HL speakers ($p>.05$).

At this moment, we conducted a descriptive analysis with the three groups according to the proficiency level and their performance in all four of the target sentence types: transitivity violation, morphosyntactic violation (ungrammatical structures), change-of-state verbs, and transitives (grammatical structures). The data that we found are displayed in Figure 37 below:

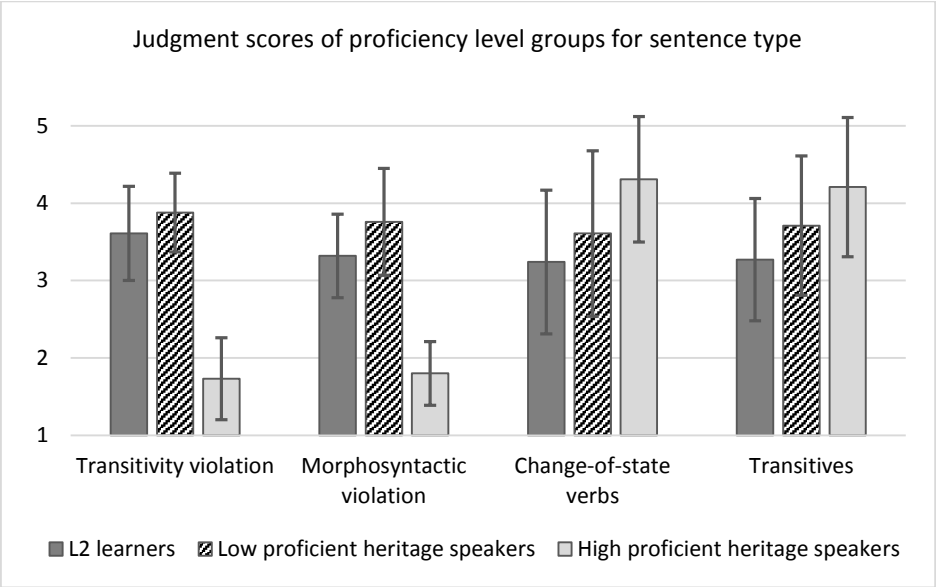


Figure 37 – Judgment scores of three proficiency groups for sentence type.

As Figure 37 shows, high-proficiency HL speakers was the only group that was able the capture the grammaticality status of all four target sentence types. Low-proficiency HL speakers seem to keep an average close to 4 in all for structures. It suggests that no matter the grammatical status of sentences, they evaluate them higher, close to the ceiling (5 points). Similarly, low-proficiency L2 learners are not able to identify the grammaticality of all sentence types. These results confirm that the

proficiency cut we made through our test was accurate, and attest them as adequate to measure the level of proficiency.

In order to confirm the ability of our proficiency tests in profiling participants' proficiency, we tested the correlation of the three proficiency tests: Spanish Placement Test, Vocabulary Level Test, and Self-Assessment Test. A Pearson product-moment correlation coefficient (*r*) was computed to assess the relationship between the three tests in order to verify the degree of correlation between the scores each one produces as diagnosis of L2 proficiency. To do so, we computed the total number of scores reached by low and high-proficiency participants in the OPT and VLT, together with the means from low and high-proficiency from the Self-Assessment Test. Our hypothesis was that there would be a positive correlation among tests for each group of proficiency. The data displayed in Table 3 confirms this hypothesis:

Table 8 – Correlation of SPT, VLT and Self-Assessment tests

	SPT	VLT
Spanish Placement Test		
Vocabulary Level Test	.922**	
Self-Assessment	.782**	.791**

**Correlation is significant at the 0.01 level (2-tailed)

As evidenced above, there is a positive correlation among the tests. This result suggests that the proficiency level of the participants in Experiment 2 was sensitive to the tests, and also their proficiency levels are correspondent in each test. It confirms our prediction that the VLT is an adequate predictor of proficiency. After that, we tested the correlation among proficiency tests with each of our target sentences types. Table 9 below displays these results:

Table 9– Correlation of SPT, VLT, Self-Assessment tests and sentence types

	PT	VLT	Spea.	Read.	Comp.	Writ.	Trans.	Morph.	Chang.
Placement test									
VLT	.922**								
Speaking	.827**	.834**							
Reading	.701**	.695**	.875**						
Comprehension	.673**	.678**	.837**	.861**					
Writing	.771**	.800**	.901**	.868**	.855**				
Transitivity	-.638**	-.690**	-.535**	-.485**	-.414**	-.598**			
Morphosyntax	-.593**	-.655*	-.527**	-.473**	-.419**	-.541**	.818**		
Change of state	.556*	.499*	.556**	.510**	.480**	.514**	-.251	-.294*	
Transitives	.459**	.419**	.449**	.427**	.445**	.479**	-.230	-.244	.834**

**Correlation is significant at the 0.01 level (2-tailed)

*Correlation is significant at the 0.05 level (2-tailed)

It is clear that there is a significant overall correlation between both the Placement test and the VLT, with the learners assessing each ability, and also with each sentence type from our AJ task. Correlation was not significant in two cases: when means for sentences with transitivity violation was correlated with change of state verbs ($r=-.251$, $p>.05$) and with grammatical transitive sentences ($r=-.230$, $p>.05$). These results confirm Ellis' (2005) notion that grammaticality judgment tasks are able to differentiate the state of L2 (and/or non-dominant language) representations. A similar conclusion was made by Souza & Oliveira (2014) attesting that the acceptability judgment as a behavioral task is able to detect high and low levels of language proficiency.

Results from Table 9 confirm Stæhr's (2008) findings about the relationship between vocabulary size and reading and writing abilities. As we can see, there was a significantly high correlation between reading ability with both SPT ($r=.701$, $p<.05$) and VLT ($r=.695$, $p<.05$). Also, there was significance when writing ability was correlated with SPT ($r=.771$, $p<.05$) and VLT ($r=.800$, $p<.05$). Similarly, the relationship between speaking ability and vocabulary knowledge found by Zimmerman (2004) is also confirmed in our study.

There was a significant correlation between speaking with SPT ($r=.827$, $p<.05$) and VLT ($r=.834$, $p<.05$). Moreover, correlation between listening ability with vocabulary knowledge found by Bonk (2000) was also found in our experiment. This shows that listening ability correlates with both SPT ($r=.673$, $p<.05$) and VLT ($r=.678$, $p<.05$).

4.2.4 ROC curve analysis

Following the procedures of Experiment 1, we submitted the Spanish version of VLT, the Spanish Placement Test (SPT), and the Spanish version of the AJ task to the ROC curve analysis in order to assess their ability in producing accurate diagnosis of proficiency in Spanish among this specific population. We decided to compute the scores of the HL speakers only, because the group formed with L2 Spanish speakers did not produce two groups of proficiency. Considering the ROC curve as a methodology for diagnosis, we considered the low-proficiency and the high-proficiency HL speakers. First, we generated the ROC curve for the VLT and SPT scores. Data are displayed in Figure 38 below:

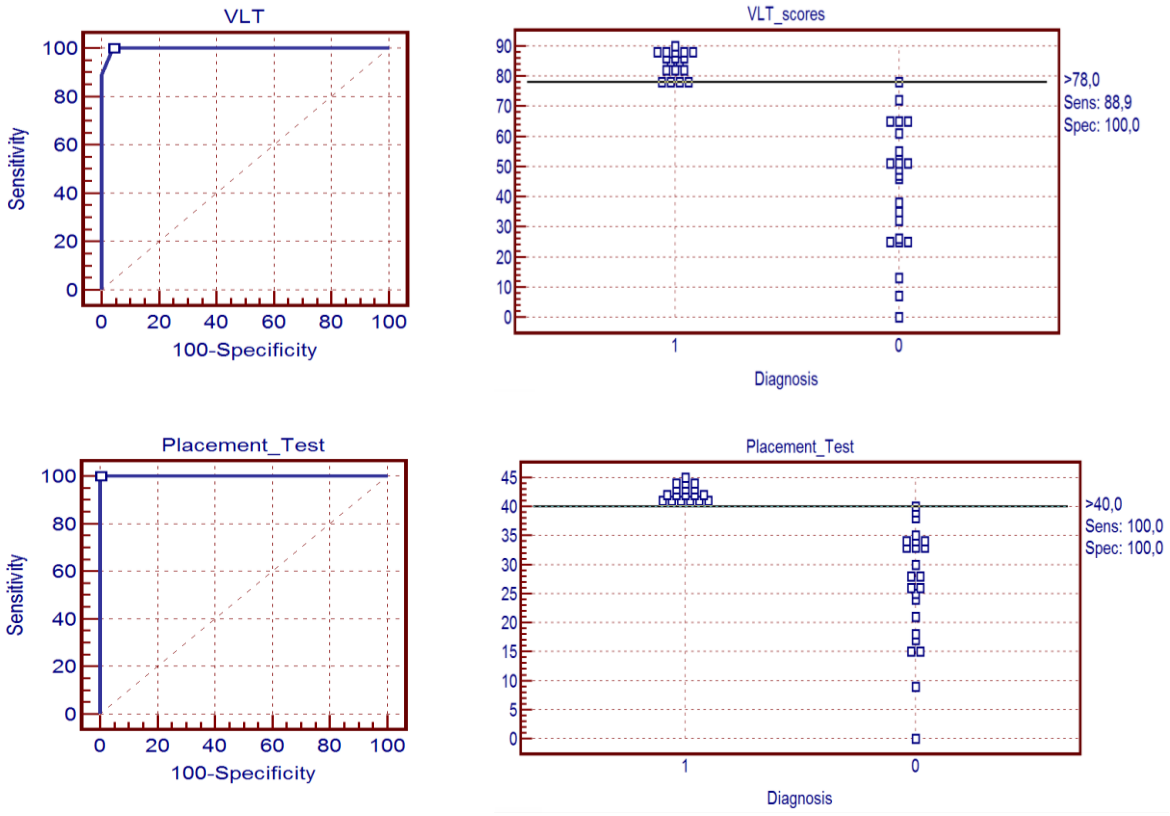


Figure 38: ROC curve graphs and Interactive Dot Diagrams for VLT and SPT for Experiment 2

As seen in Figure 38, VLT scores with the cut-off point we proposed generated an almost perfect curve, since both Sensitivity and Specificity are close to the ideal value. The point of convergence represented in the graph shows a perfect level of Specificity to detect the low proficiency. In contrast, it shows no total efficacy in distinguish the high-

proficiency group as true positives (Sensitivity). Such result can be seen in more detail in VLT's Interactive Dot Diagram (Figure 38). The high-proficiency group is supposed to be located above the line, indicating that every high-proficiency participant is above the cut-off point. This demonstrates that there is a portion of individuals who are under the line, which indicates that the cut-off point we suggest is not completely accurate in delimiting high-proficient individuals as true positives.

On the other hand, if we consider the second column, we can see that the low-proficiency individuals scored under the cut-off point. This result shows that the test was perfect in detecting low-proficiency individuals, since the Specificity presented in the ROC curve is equal to zero. Based on these results, we consider the VLT to be an adequate diagnostic tool for English L2 proficiency measure. The results for the cut-off point we suggested (78) are summarized in table 10 below:

Table 10: Sample of Cut-off point values, Sensitivity and Specificity of VLT and SPT scores for Experiment 2

	Area under the curve (standard error)	Criterion (cut-off point)	Sensitivity (C.I.)	Specificity (C.I.)	+LR	-LR
VLT	0,998 (0,008)	78	88,9 (65,2 – 98,3)	100,0 (85,6 – 100,0)	-	0,11
SPT	1,000 (0,000)	40	100,0 (81,3 – 100)	100,0 (85,6) – 100,0)	-	0,00

In Table 10 above, the ROC curve analysis for the VLT scores revealed an AUC close to 1 (0,998), suggesting that VLT has a considerable level of reliability as a diagnostic test. The results revealed a perfect accuracy in detecting the low-proficiency individuals (Specificity) (100,0, C.I.=85,6 – 100), although its power to accurately identify high-proficiency individuals as truly negatives is not one hundred percent accurate (Sensitivity) (88,9, C.I.=65,2 – 98,3).

Following the same procedures of Experiment 1, we then applied the ROC curve analysis for the AJ task to assess its ability to precisely distinguish individuals who have a high level of proficiency from those who do not. For the very same reasons as Experiment 1, we analyze ungrammatical and grammatical sentences separately. The ROC curve graph and the Interactive Dot Diagram are displayed in Figure 39:

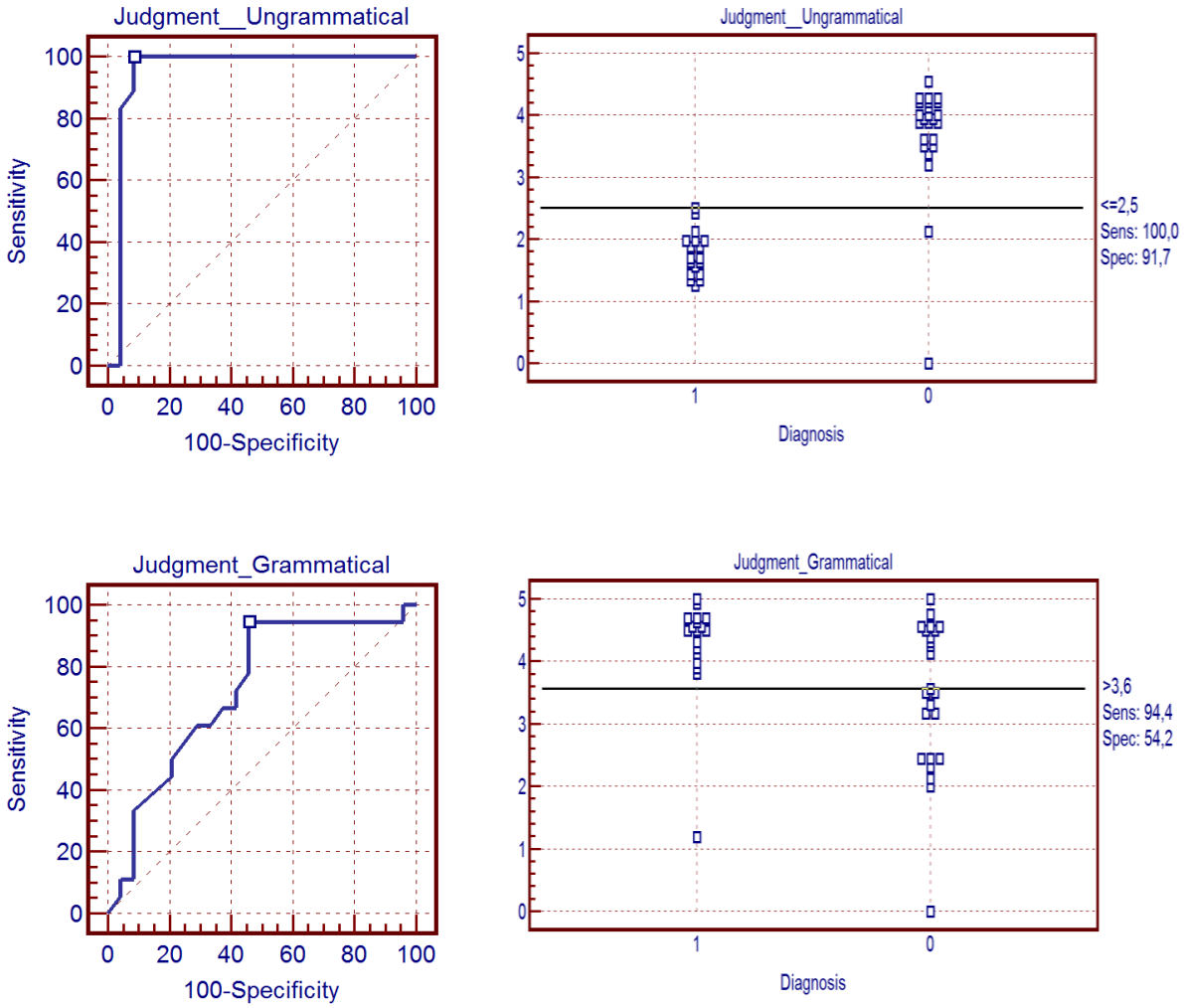


Figure 39: ROC curve graphs and Interactive Dot Diagrams for AJ task (ungrammatical and grammatical sentences) for Experiment 2

The AJ task seems to be an adequate diagnostic test for proficiency (Figure 39) when the ungrammatical structures are at stake. The ROC curve for ungrammatical structures reveals that the cut-off point (2,5) divides both groups of proficiency with precise Sensitivity (true positives), and almost fully precise Specificity (true negatives). In the Interactive Dot Diagram, it is shown that all participants who were classified as having high-proficiency are under the line (Sensitivity: 100,0), while the majority of participants who were classified as the low-proficiency group are above the line (Specificity: 91,7).

A quite different situation is presented with the grammatical structures, because the ability of detecting the low-proficient group as true negatives is not accurate. The Dot Diagram for the grammatical structures reveals a considerably high Sensitivity (94,4), and the majority of high-proficiency individuals remain above the line, while the Specificity is considerably less accurate (54,2), as we see that participants who were classified as having low proficiency are randomly distributed in the graph without a specific criteria in relation to the cut-off point (3,6). Below, these results are summarized in Table 11:

Table 11: Sample of Cut-off point values, Sensitivity and Specificity of AJ task (Ungrammatical and Grammatical sentences) for Experiment 2

	Area under the curve (standard error)	Criterion (cut-off point)	Sensitivity (C.I.)	Specificity (C.I.)	+LR	-LR
AJ (Ungr)	0,953 (0,033)	2,5	100,0 (81,3 – 100,0)	91,7 (73,0 – 98,7)	12,00	0,00
AJ (Gram)	0,726 (0,081)	3,5	94,4 (72,6 – 99,1)	54,2 (32,8 – 74,4)	2,06	0,10

Table 11 reveals that the AJ task is an adequate instrument for Spanish proficiency diagnosis among HL speakers, especially considering the ungrammatical structures that present an accurate Sensitivity (100,0), it means the ability of detecting high-proficient participants as true positives. The Specificity in the ungrammatical sentence analysis is not wholly accurate, but it is close to the total (94,4). The area under the ROC curve for the ungrammatical sentences is considerably ample (0,953). The ROC curve analysis for the grammatical sentences is largely different, since the AUC is smaller (0,726). Moreover, although the Sensitivity is high (94,4), the Specificity is only 54,2, which makes the test with grammatical structures unable to accurately distinguish the low-proficiency participants are true negatives.

In this chapter, we analyzed the data of both experiments. In the first experiment we stipulated the minimum temporal window a bilingual need to do a judgment call on a sentence in the L2. Moreover, we submitted the VLT, the OPT and the AJ task to a correlational analysis and to the analysis their Sensitivity and Specificity as diagnostic tests by applying the ROC curve methodology. In Experiment 2, we applied similar procedures of correlation, adding the self-assessment test into the equation. We

also explored the data from the BHQ as demographic and language exposure tool. Equivalently, we submitted the VLT, the SPT and the AJ task to the ROC curve analysis.

CHAPTER 5

CONCLUSION

Succinctly, the three aims of the present study were (i) to explore the capacity of the Vocabulary Level Test, in English and Spanish version, of producing language proficiency profiles whose scores correlate with an overall grammar test (OPT in English and SPT in Spanish), and with a AJ task (English and Spanish version). Moreover, a complimentary goal of estimating the minimal time a bilingual needs to make a sentence judgment; (ii) to investigate whether VLT (both versions), the placement tests (both versions), and the AJ task (both versions) are adequate in producing diagnosis about language proficiency level according to the ROC curve analysis; and (iii) to investigate whether highly proficient L2 learners and HL speakers perform better than low ones in the speeded acceptability judgment task, assuming that the former group has a higher level of automaticity in the L2 processing.

In order to accomplish such objectives, we proposed two questions for this study: Are the VLT (as a diagnostic test for explicit knowledge of L2 vocabulary size) and the Acceptability Judgment Task (as a diagnostic measure for implicit knowledge of L2 representation) able to discriminate L2 proficiency profiles that correlate among L2 learners and HL? Assuming that the increase in the proficiency level reflect on the type of memory learners rely on: Do highly proficient L2 learners and HL speakers perform better than low ones in the speeded acceptability judgment task? The answer is yes for both questions followed by some important observations.

5.1 Discussion

In Experiment 1, we developed a comparison between VLT scores and another measure of proficiency in English L2, the Oxford Placement Test (OPT). We verified moderate and significant correlations between the scores of test-takers who achieved the highest level in the VLT (whom we refer to as “high-proficiency), and the last level of the OPT. Moreover, we found moderate and significant correlations between VLT scores up

to level 3, and lower bands in the OPT, which are tentatively associated with CERF labels describing levels of ability in an L2 below full communicative mastery.

In addition, we investigated the performance in a timed acceptability judgment task with English sentences of a sample of bilinguals whose VLT scores indicate "high-proficiency" in L2 English as compared to a sample diagnosed through the same test as having "low-proficiency". This timed task was assigned with a temporal ceiling of 8 seconds, the average performance having taken place within 6 seconds. The results showed a striking contrast between the high-proficiency and the low-proficiency groups, with only the latter being able to make judgment calls that converge with the L2 grammar.

Three quarters of the violations instantiated in our stimuli for the acceptability judgment task were actually grammatical restrictions that do apply to the bilinguals' L1. Therefore, we interpret the overall failure of the low-proficiency participants to accurately detect such violations as a failure to fully access grammatical knowledge when using the L2 under strict temporal restrictions. Difficulty to integrate certain types of grammatical information when processing the L2 has been previously suggested to be a factor modulating bilingual language processing (Clahsen & Felser, 2006; Jiang, 2007).

It is crucial to mention that the task employed in the present study does not elicit samples of online language processing. Nevertheless, in light of the requirement for speeded performance of our task, we interpret our results as suggestive that the measurement of large L2 vocabulary size does not only indicate higher fluency in lexical access, as suggested by Laufer & Nation (2001), but also fluency in access to grammatical representation repositories. As fluency comes along unplanned and subliminal performance as dimensions of automaticity (Segalowitz & Hulstijn, 2005), we believe that the measurement of L2 vocabulary can be also indirectly informative of differential profiles in L2 automaticity.

As previously mentioned, the procedures adopted for stablishing the window time for the AJ task in the Experiment 1 was based on the procedures of Souza et al. (2015). We agree with the authors on the fact that the judgments performed under time pressures is more reliable about participant's implicit knowledge than no timed judgments.

Following Jiang's (2012) concepts and Souza et al's. (2015) findings, we assume that the speeded AJ task may inhibit participants from relying on a set of cognitive strategies in their judgment on the licitness of the sentences, and, consequently, these judgments are exemplary of participants' mental representation. This result supports Jiangs's (2007) assumption that some bilinguals are more able than others in the perception pf morphosyntatic information during L2 processing depending on their level pf L2 proficiency.

Similarly to the first experiment, in the second experiment there was a significant correlation among VLT, SPT, AJ task and, in addition, to the self-assessment test. This result also confirms our hypothesis that VLT, which have demonstrated high correlation with the four abilities skills (Milton, 2010, Stæhr (2008) is highly correlated with an overall grammar test and with a psycholinguistic test. These results from both Experiment 1 and 2 confirm the Hypothesis 1.

In the second experiment, the BHQ with linguistic and biographic information was applied with the objective of screening participant's linguistic trajectory. As we previously mentioned, the literature define the HL speakers as those who are are first exposed to their family's language (Spanish in this study), and as they grow up, the other language becomes predominant. The data reported in the BHQ confirmed this definition, since participants who were classified as HL speakers reported an exposition to Spanish majorly bigger than English as a child (in some cases, only Spanish), while as an adults, the process was the opposite, when English became the language of massive exposition.

On the contrary, Spanish L2 learners reported almost none contact with Spanish in childhood, then some contact as teenagers (mostly at school), and then less contact in adulthood. Since this exposition report comprehended both receptive and active use of language, we can assume that the HL speakers inverted their linguistic dominance from Spanish to English. This result agrees with Montrul (2005, 2010). An interesting aspect to be noticed in the HL speakers' report is their predominance of Spanish use at home even as adults. Although, there is an inversion of dominance in any other aspects of their lives (school, work), at home Spanish remains as the main language.

We submitted the VLT, the OPT and the AJ task to the ROC curve analysis in order to test their ability to diagnosing proficiency by separating accurately individuals who are high proficient from those who are low proficient in English as L2 (in both experiments). Results from Experiment 1 revealed that VLT scores produced a Specificity value of 77,8, the Sensitivity was total (100,0) and the AUC was considerably extent (0,998). These findings allow us to consider VLT to be an adequate instrument for proficiency diagnosis. Differently, OPT results revealed complete accuracy (Sensitivity and Specificity) as a diagnostic test. The ROC curve analysis for the AJ task revealed an the AUC of 0,995 for grammatical structures with a Sensitivity of 100,0 e a Specificity of 94,4. The results for the ungrammatical sentences were more accurate, with an AUC, Sensitivity and Specificity of 100,0. This slight difference among sentence grammaticality does not interfere significantly in our results, since the main focus of the AJ task is on the ungrammatical structures, being the grammatical ones a counterbalance part.

For the Experiment 2, *the ROC curve* analysis revealed that the Spanish VLT is an adequate instrument for L2 proficiency diagnosis, since it has a moderate ability of detecting individuals who are high proficient (Sensitivity=88,9) and a perfect ability to detect true negatives (Specificity=100,0). The ROC curve analysis in the second experiment revealed an AUC of 0,998 for the VLT, with a Specificity of 100,0, which is a considerably high number for a diagnostic test, although the Sensitivity was 88,9. A similar result was found in the ROC curve analysis for the AJ task, in which the AUC for the grammatical structures (0,726) was not as high as in the ungrammatical structures (0,953). These findings confirm the Hypothesis 2

In both experiments, only the high-proficient groups were able to detect the ungrammaticality of the sentences. Both Spanish L2 learners and low-proficient HL speakers were not able to perceive the ilictness of the structures. These results agrees with Souza (2011) on the finding that the perception the grammaticality in specific argument structure (such as induced movement alternation) in the L2 is only perceived by bilinguals with a high level of language proficiency. Moreover, as postulated by Ellis, R. (1995) and Ellis, R. (2005), the implicit knowledge has a higher level of unconsciousness and intuitiveness. However, it is important to evoke Segalowitz's (1991, 2006) observation that a high level of automaticity in a given task does not

entangle implicit knowledge necessarily, since it could mean only an acceleration of controlled processes. In this study, the major part of sentence type we chose to compose the AJ task usually are not explicitly taught in L2 formal instructions, for instance the induced-movement alternation, the change of state verbs, the WH-movement. For that reason, and based on previous studies (Souza, 2011, 2012,), we assume that a higher performance in a speeded AJ task for these structures demands implicit L2 knowledge and a considerable level of automaticity. These findings confirm the Hypothesis 3.

Results from both experiments in the AJ task reveal a pattern with the forged sentences with verbal transitivity. In the first Experiment, such violations imposed in English L2 (for instance: **The man laughed the child during the party*) also occurs in Portuguese L1 (**O homem sorriu a criança durante a festa*). Similarly, in the second experiment, the transitivity violation displayed in Spanish L2 (**El polvo denso tosió a los niños en el parque)* is also ilicit in English L1 (**The dense dust coughed the children in the park*). Although we imposed a time limit to the judgment that implies implicit knowledge, we are not able to assume for sure that the difference in the judgment scores for this structure regarding proficiency level relies on the fact that high proficiency bilinguals learned the structure; therefore, they could judge it as ungrammatical. The reason for that is that in their first language, such construction is also unlicensed; it means that their judgment could have been supported by their implicit knowledge of the L1. The same idea can be applied to the judgment scores of the morphosyntactic violation. Both WH-movement and agreement violations occur in participants' L1 and L2, blurring any inference concerning the learnability of the structure in the L2.

Differently, if we look at the grammatical structure induced-movement alternation in the Experiment 1 (*The instructor ran the boys around the park)*, we will see a significant difference ($p < .05$) between high and low proficiency groups (Figure 24). Considering that the induced-movement alternation is not licensed in Brazilian Portuguese (**O instrutor correu os garotos pelo parque)*, we can assume that this structure agrees with Goldberg's (1995) notion of the learnability of the construction involved in this alternation. Moreover, it confirms Souza's (2012) findings about the difference between low and high proficiency individuals in perceiving such argument structure pattern in English as L2.

In both experiments, data about the AJ task concerning the processing of transitivity violations (despite of being ungrammatical in participants' both L1 and L2) and the grammatical induced-movement alternation in experiment 1 offer us subsides to consider the clear interface between lexical-semantic and syntax (Jackendoff, 1997, 2002). This is due to the restriction imposed by the verbal semantics on the syntactic configuration (Cullicover & Jackendoff, 2005) especially in the second case.

As we discussed in the theoretical chapter, both in English and Portuguese motion verbs like *walk, jump* and *run* have an intransitive interpretation (*John jumped*), requiring only an agent in the subject position. However, in English it is also possible a transitivization of such verbs by adding a direct object underlying the meaning: *x made y to z* (*John jumped the horse over the fence).* Results revealed that L2 high proficiency participants were able to detect the licitness of this construction under the time pressure we imposed in the AJ task. It may indicate that, when they access the item *jump,* the possible sintatic configuration with an direct object (plus a PP) is also activated and available in participants' mind to be filled. Differently, it may indicate that low proficiency participants'mind when access the item *jump* only offers the intransitive configuration, therefore they are not able to detect the grammaticality of the sentences.

This result also agrees with speech production model (Levelt, 1989) in the sense of the levels and the path a word takes to be accessed. According to Souza (2011) the linguistic realization of arguments consists in the route from the conceptual representation in the mind to its manifestation in morphological and syntactic structures and White (2003) affirmed that L2 speakers (regarding their level of L2 knowledge) reach a representational level for lexical items and map from the argument structure to syntax.

The processing of the induced-movement alternation by English L2 speakers suggests that the processing of the motion verbs relies on the two stages proposed in Level's (1989) and Hartsuiker et al's (2004) models concerning the concept of lemma. When access the lemma for *jump* for example in the formulator, there is also an activation for all possible meaning and constraint for articulation of it (grammar encoder). These results confirms Goldberg's (1995) and Souza's (2012) assumption that induced-movement alternation is a learnable construction and aligns with Oliveira's (2013) finding on the learnability of resultatives.

5.2 Limitations of the study

During the participants screening and the development of the experiments, we faced some limitations. In Experiment 1, we did not apply a questionnaire for screening participants in relation to their language experience, type of exposure, language use. By the time of the first experiment, we could not find an instrument that we considered to be adequate to the type of bilingual we were dealing with.

Another limitation in Experiment 1was the lack of a self-assessment test in order to correlate its scores with the other proficiency tests. Such implementation would give us more information about the population of our study such as it has in the Experiment 2. In the Experiment 2, there was a limitation in relation the balance for the groups of proficiency. We were able to gather HL speakers with low and high proficiency; however, only low-proficiency participants composed the group of L2 Spanish learners. Such limitation deprived us from a undertaking a broader comparison among the groups regarding their proficiency level.

5.3 Contribution of the study

In this study, we advanced the work reported by Souza, Duarte & Berg (2015) aiming at validating a measure of vocabulary size—the Vocabulary Size Test, or VLT (Nation, 1990)- as a diagnostic tool to assess bilinguals. In Souza, Duarte & Berg (2015) the bilingual group was composed by Brazilian Portuguese-English bilinguals at college level, similarly to the Experiment 1 of this study. Moreover, we advanced in the VLT validation as a diagnostic tool, by applying the same procedures with English-Spanish bilinguals as well as Spanish HL speakers. We consider to be important studies that assess tests of vocabulary size, since most of them are practical and easy to administer tests that have been used as screening procedure in many studies of bilingualism and second language acquisition (Hulstjin, 2012). The VLT has also been previously employed with such a purpose in studies with Brazilian Portuguese-English bilinguals (Souza & Oliveira, 2011; Souza, 2012; Oliveira & Souza, 2014).

The correlation between the VLT and AJ task in both experiments represents an advance in the use of VLT as a proficiency measure. As we previously mentioned, the literature has shown that in several studies (Stæhr, 2008; Zimmerman, 2004, Milton, 2010; Koizumi & In'nami , 2013), the VLT revealed significant correlation with the four abilities (speaking, listening, writing, reading). The results of the correlational data in this study expands the reach of VLT by correlating it with an psycholinguistic task in which the knowledge type involved is from a different nature. In other words, we assumed VLT requires the explicit knowledge, since after taking the test, a participant would be able to verbalize and explain his choices, to talk about the knowledge being measures. In addition, the questions allow the test-takers to rely on cognitive strategies such as association, elimination. On the other hand, the AJ task with those specific structures, and with a time pressure, requires a more implicit knowledge from the individuals.

It is important to mention that there are several proficiency tests, including test of vocabulary size, that produce more than two levels of proficiency, including intermediate levels. Depending on the way scores are computed, even VLT is able to produce different levels of proficiency. In this study, following the same procedures of Souza, Soares-Silva & Silva, 2016), our purpose was to administer the VLT as a diagnostic test, in other words, we were looking for the detection of high-proficient individuals from non-high proficient individuals (low). For this purpose, VLT demonstrated to be adequate in generating two groups of proficiency, being the high-proficiency group significantly different from the low-proficiency group. The VLT's adequacy as a diagnostic test was confirmed in the ROC curve analysis, considering the AUC of VLT's scores in both experiments were close to the totality.

These findings are aligned with distinction of implicit versus explicit knowledge (Ellis, 1994, Paradis, 1994), specifically with the distinction between automatic/controlled processes. Data revealed that the level of automaticity (as a type of processing) and the level of implicitness (as a type of knowledge) (cf. Figure 5) relies on the level of proficiency. In order words, high proficient participants are able to process the languages with more automaticity with an implicit knowledge, while low proficient participants process language in a more controlled way and explicit knowledge (Segalowitz & Hulstijn (2005)). As assumed in Segalowitz (1991), controlled processes depend on strategic decision and are more time consuming in

relation to automatic processes, that is why low proficient participants were not able to judge properly the sentences under the time pressure we imposed in the AJ task for both experiments. Simmilarly, these results agree with the Hulstijn (2011) idea that in the primary stages of BLC (Basic Language Cognition), the L2 speaker perform less automatically than in advanced stages.

These results of automaticity as a facet of L2 proficiency found in this study also agrees with Clahsen & Felser's, (2006) findings from ERP studies. Similarly to their studies, in which highly proficient L2 participants showed a level of automaticity close to the native speakers' performance, our results revealed that our diagnostic tests classified as high proficient participants those who were able to detect the grammaticality of the structures in the time window we stipulated.

As presented in Alderson (2015), Milton (2013), Stæhr (2008), and Zimmerman (2004), the measure of vocabulary size has been significantly correlated with the four language abilities (speaking, writing, listening, reading). The results of the self-assessment test we applied in the Experiment 2 confirmed this assumption, since participants self-reported levels of proficiency in the four skills that correlated significantly with their scores in in the VLT.

Methodologically, this study advances in the validation of the VLT as a diagnostic test. Through the measures of Sensitivity and Specificity, the ROC curve analysis revealed that the VLT produces scores that differentiate two profiles of language proficiency in a considerable level of adequacy. Moreover, the ROC curve analysis for the AJ tasks confirmed our assumption that this type of psycholinguistic task with the sentence types we included can be used as a measure for language proficiency diagnostic, especially the ungrammatical sentences.

REFERENCES

ALDERSON, J. C. *Diagnosing Foreign Language Proficiency*: The Interface between Learning and Assessment. London/New York: Continuum, 2005.

ALMELA, R.; CANTOS, P.; SÁNCHEZ, A.; SARMIENTO, R. & ALMELA, M. *Frecuencias del Español*. Diccionario y estudios léxicos y morfológicos. Madrid: Editorial Universitas S.A., 2005.

BACHMAN, L. *Fundamental considerations in language testing*. Oxford: Oxford University Press, 1990.

BARD, E. G.; ROBERTSON, D. & SORACE, A. Magnitude estimation for linguistic acceptability. *Language,* 72(1), 1996, p. 32-68.

BEDORE, L. M.; PEÑA, E. D.; SUMMERS, C. L.; BOERGER, K. M.; RESENDIZ, M. D.; GREENE, K.; BOHMAN, T. & GILLMAN, R. B. *The measure matters:* Language dominance profiles across measures in Spanish-English bilingual children. Bilingualism: Language and Cognition, 15 (3), 2012, p. 616-629.

BERNOLET, S. *Lexical-syntactic representations in bilingual sentence production*. Unpublished PhD Dissertation, Ghent University, 2008.

BERNOLET, S., HARTSUIKER, R. J. & PICKERING, M. J. From language-specific to shared syntactic representations: the influence of second language proficiency on syntactic sharing in bilinguals. *Cognition,* 127(3), 2013, p. 287-306.

BIRDSONG, D. Dominance, proficiency, and second language grammatical processing. *Applied Psycholinguistics*, 27, 2006, p. 46- 49.

BONK, W. J. Second language lexical knowledge and listening comprehension. *International Journal of Listening*, 14(1), 2000, p. 14-31.

BOWLES, M. Measuring implicit and explicit linguistic knowledge: What can heritage language learners contribute? *Studies in Second Language Acquisition*, 33, 2011, p, 247-271.

CARNEIRO, M. M. & SOUZA, R. A. Observação do processamento online: uma direção necessária para o estudo experimental da sintaxe bilíngue. *ReVEL,* 10 (18), 2012, p. 107-127.

CASSO, J. *Análisis y revisión crítica de los materiales de evaluación de la competencia léxica*. Elaboración de un test de vocabulario de nivel umbral. Master oficial en lingüística

aplicada a la enseñanza de español como lengua extranjera, Universidad Nebrija, Madrid, 2010.

CLAHSEN, H. & FELSER, C. Grammatical processing in language learners. *Applied Psycholinguistics,* 27 (1), 2006, p. 3-42.

COOK, V. *Chomsky's Universal Grammar*: an introduction. Oxford: Blackwell, 1988.

COUNCIL OF EUROPE. *Common European Framework of Reference for Languages*. Cambridge: Cambridge University Press, 2001.

CULICOVER, P & JACKENDOFF, R. S*impler Syntax*. Oxford University Press: Oxford, 2005.

CUTLER, A.; MEHLER, J.; NORRIS, D. & SEGUI, J. The monolingual nature of speech segmentation by bilinguals. *Cognitive Psychology*, 24(3), 1992, p. 381-410.

DE BOT, K. A bilingual production model: Levelt's Speaking model adapted. *Applied Linguistics,* 13 (1), 1992, p. 1–24.

DE GROOT, A. M. B. Late foreign learning and lexical representation. In: DE GROOT, A. M. B. Language and cognition in bilinguals and multilinguals. New York: Psychology Press, 2011

DeKEYSER, R. *Implicit and Explicit learning.* In: Doughty, C. & Long, M. (eds.). The Handbook of Second Language Acquisition. Malden, MA: Blacwell Publishing, 2003, p. 313-348.

DÖRNYEI, Z. *The Psychology of the Second Language Learner*: Individual Differences in Second Language Acquisition. Mahwah, NJ: Lawrence Erlbaum Associates, 2005.

DUNN, A. L. & FOX TREE, J. E. A quick, gradient Bilingual Dominance Scale. *Bilingualism: Language and Cognition,* 12 (3), 2009, p. 273-289.

EDWARDS, J. *Foundations of Bilingualism.* In: BHATIA, T.; RITCHIE, W. (Eds.) The Handbook of Bilingualism – 2nd Edition. Malden, MA: Wiley-Blackwell, 2006.

_____. *Bilingualism and Multilingualism*: Some Central Concepts. In: In: BHATIA, T.; RITCHIE, W. (Eds.) The Handbook of Bilingualism and Multilingualism, Chichester: John Wiley & Sons, 2012.

ELLIS, N. C. *The Weak-Interface, Consciousness, and Form-focussed instruction*: Mind the Doors. In S. Fotos & H. Nassaji (Eds.), Form Focused Instruction and Teacher

Education: Studies in Honour of Rod Ellis, pp. 17-33,Oxford: Oxford University Press, 2007.

_____. (ed) *Implicit and Explicit learning of Languages*, Sandiego: Academic Press, 1994.

_____. At the interface: Dynamic interactions of explicit and implicit knowledge. *Studies in Second Language Acquisition*, 27, 2005, p. 305-352.

ELLIS, R. Measuring implicit and explicit knowledge of a second language: A psychometric study. *Studies in second language acquisition*. 27 (2), 2005. p. 141-172.

_____. *A Theory of instructed second language acquisition*. In: Ellis, N. C (ed) Implicit and Explicit learning of Languages, Sandiego: Academic Press, 1994, p. 79-112.

_____. *The study of second language acquisition*. Oxford, UK: Oxford University Press, 1994.

FAWCETT, T. An introduction to ROC analysis. *Pattern Recognition Letters* 27, 2006, p. 861–874

FERNÁNDEZ, E. *Bilingual Sentence Processing*- Relative Clause Attachment in English and Spanish. Amsterdam/Philadelphia: John Benjamins, 2003.

FRANCHI, C. & CANÇADO, M. Reexame da noção de hierarquia temática. *Revista Estudos Linguísticos*, 11(2), 2003, p. 125-153.

GASS, S. M.; SELINKER, L. *Second Language Acquisition*: an introductory course. New Jersey: 3rd edition Lawrence Erlbaum Associates, 2008.

GERTKEN, L. M.; AMENGUAL, M. & BIRDSONG, D. *Assessing language dominance with the Bilingual Language Profile.* In: LECLERCQ, P.; EDMONDS, A & HILTON, H. (Eds.) Measuring L2 proficiency: Perspectives from SLA. Bristol: Multilingual Matters, 2014, p. 208-255.

GROSJEAN, F. Studying bilinguals: methodological and conceptual issues. *Bilingualism: Language and Cognition*, 1 (2), 1998, p. 131-149.

_____. *Studying Bilinguals*. Oxford: Oxford University Press, 2008.

_____. *Bilingualism:* A short introduction. In: GROSJEAN, F. & Li, P. (eds.) The Psycholinguistics of Bilingualism. Oxford: Wiley-Blackwell, 2013.

.

HARLEY, B.; CUMMIS, J.; SWAIN, M. & ALLEN, P. *The nature of language proficiency*. In HARLEY, B.; ALLEN, P.; CUMMIS, J. & SWAIN, M. (orgs.). The development of second language proficiency. Cambridge: Cambridge University Press, 1990.

HARTSUIKER, R. J., PICKERING, M. J., & VELTKAMP, E. Is syntax separate or shared between languages? Cross-linguistic syntactic priming in Spanish–English bilinguals. *Psychological Science*, 15(1), 2004, p. 409–414.

HU, M. & NATION, P. Unknown vocabulary density and reading comprehension. *Reading in a Foreign Language*, 13(1), 2000, p. 403-430.

HULSTIJN, J. Towards a unified account of the representation, processing and acquisition of second language knowledge. *Second Language Research*, 18, 2002, 193–223.

_____. *Linking L2 proficiency to L2 acquisition*: Opportunities and challenges of profiling research. In: BARTNING, I.; MARTIN, M. & VEDDER. I. Communicative proficiency and linguistic development: intersections between SLA and language testing research: EuroSLA, 2010, p. 233-238.

HULSTIJN, J.; ANDERSON, C.; SCHOONEN, R. *Developmental stages in second language acquisition and second language proficiency:* Are there links between them? In: BARTNING, I.; MARTIN, M. & VEDDER. I. Communicative proficiency and linguistic development: intersections between SLA and language testing research: EuroSLA, 2010.

HULSTIJN, J. Language proficiency in native and nonnative speakers: An agenda for research and suggestions for second language assessment. *Language Assessment Quarterly*, 8 (3), 2011. p. 229-249.

_____. The construction of language proficiency in the study of bilingualism from a cognitive perspective. *Bilingualism: Language and Cognition*, 15(1), 2012, p. 422-433.

_____. *Language Proficiency in Native and Non-native Speakers*. Amsterdam/Philadelphia: John Benjamins Publishing Company, 2015.

JACKENDOFF, R., 1997. The Architecture of the Language Faculty. MIT Press: Cambridge, MA, 1997.

_____. *Foundations of Language*: Brain, Meaning, Grammar, Evolution. New York: Oxford, 2002.

_____. A Parallel Architecture perspective on language processing. *Brain Research*, 1146. 2007, p. 2– 22.

JIANG, N. Selective integration of linguistic knowledge in adult second language learning. *Language Learning*, 57 (1), 2007, p. 1-33.

JUFFS, A. *An overview of the second language acquisition of links between verb semantics and morpho-syntax*. In: ARCHIBALD, J. (Ed). Second Language Acquisition and Linguistic Theory. Malden, MA: Blackwell Publishing, 2000, p. 197-227.

KRASHEN, S. *Second Language Acquisition and Second Language Learning*. Oxford: Pergamon Press, 1981.

_____. *Principles and Practice in Second Language Acquisition.* Oxford: Pergamon Press, 1982.

LAUFER, B. *How much lexis is necessary for reading comprehension?* In: BEJOINT, H. & McMILLAN, A. (Eds). Vocabulary and applied linguistics, 1992, p. 126-132.

LAUFER, B. & NATION, P. A vocabulary-size test of controlled productive ability. *Language Testing*,16 (1), 1999, p. 33-51.

LAUFER, B. & NATION, P. Passive vocabulary size and speed of meaning recognition: Are they related? *EUROSLA Yearbook*, 2001, p. 7-28.

LEVELT, W. J. M. *Speaking:* from intention to articulation. Cambridge: Cambridge MIT Press, 1989.

LEVELT, J. M. W., ROELOFS, A., & MEYER, A. S. A theory of lexical access in speech production. *Behavioural and Brain Sciences*, 22, 1999, p. 1-75.

LEVIN, B. & RAPPAPORT HOVAV, M. *Unaccusativity:* At the Syntax-Lexical Semantic Interface. MIT Press, Cambridge, MA, 1995.

LEWIS, S. & PHILLIPS, C. Aligning Grammatical Theories and Language Processing Models. *Journal of Psycholinguistics Research*, 44(1), 2015, p. 27–46.

LONG, M. H. *Problems in SLA*. Mahwah, NJ: Lawrence Erlbaum Associates, 2007.

MARANTZ. A. Generative linguistics within the cognitive neuroscience of language. *The Linguistic Review*, 22, 2005, p. 429–445.

MCKOON, G. & MACFARLAND, T.. Externally and internally caused change of state verbs. *Language*, 76(4), 2000, p. 174-196.

MEARA, P. *Connected Words:* Word Associations and Second Language Vocabulary Acquisition. Amsterdam/Philadelphia: John Benjamins, 2009.

_____. *The dimensions of lexical competence*. In: BROWN, G.; MALMKJAER, K. & WILLIAMS, J. (Eds). Performance and Competence in Second Language Acquisition. Cambridge: Cambridge University Press, 1996.

MEARA, P. & ALCOY, J. Words as Species: An Alternative Approach to Estimating Productive Vocabulary Size. *Reading in a Foreign Language*, 22 (1), 2010, p. 222-236.

MEARA, P. & MILTON, J. *X_Lex*, the Swansea Levels Test. Newbury: Express, 2003.

MILTON, J. *Measuring the contribution of vocabulary knowledge to proficiency in the four skills.* In: BARDEL, C.; LINDQVIST, C. & LAUFER, B. (Eds.) L2 Vocabulary Acquisition, Knowledge and Use: New Perspectives on Assessment and Corpus Analysis. European Second Language Association, 2013.

_____. *The development of vocabulary breadth across the CEFR levels*. In: BARTNING, I.; MARTIN, M. & VEDDER, I. (Editors). Communicative proficiency and linguistic development: intersections between SLA and language testing research, EUROSLA, 2010. p. 211-232.

MONTRUL, S. 'Se' o no 'se': A look at transitive and intransitive verbs in Spanish interlanguage. *Spanish Applied Linguistics*, 3(1), 1999, p.145-194.

_____. Incomplete acquisition and attrition of Spanish tense/ aspect distinctions in adult bilinguals Bilingualism: *Language and Cognition*, 5 (1), 2002.

_____. Second language acquisition and first language loss in adult early bilinguals: exploring some differences and similarities. *Second Language Research*, 21 (3), 2005, p. 199-249.

_____. Current Issues in Heritage Language Acquisition. *Annual Review of Applied Linguistics*, 30(1), 2010, p. 3–23.

_____. *Is the heritage language like a second language?* In: LEAH, R.; LINDGYST, C.; BARDEL, C. & ABRAHAMSSOM, N. (eds.) EUROSLA Yearbook, 12, p. 1-29, 2012.

NATION, P. Testing and teaching vocabulary. *Guidelines*, 5(1), 1983, p. 12–25.

_____. *Teaching and Learning Vocabulary*. Boston, MA: Heinle & Heinle, 1990.

PARADIS, M. *Neurolinguistics aspects of implicit and explicit memory*: implications for bilingualism and SLA. In: Ellis, N. C (ed) Implicit and Explicit learning of Languages, Sandiego: Academic Press, 1994., p. 393-418.

POLINSKY, M. & KAGAN, O. Heritage languages in the 'wild' and in the classroom. *Language and Linguistic Compass*, 1, 2007, p. 368-395.

READ, J. *Assessing vocabulary*. Cambridge: Cambridge University Press, 2000.

COELHO-DOS-REIS, J. C. *Desempenho da pesquisa de imunoglobulinas G por citometria de fluxo aplicada ao diagnóstico e prognóstico da infecção pelo HTLV-1*. Master Thesis, Universidade Federal de Minas Gerais, 2008.

SAUSSURE, F. de. *Curso de Linguistica Geral*. 18a ed. Trad. de Antonio Chelini, Jose Paulo Paes e Izidoro Blikstein. Sao Paulo, Cultrix, 1995 [1971].

SCHMITT, N. Size and depth of vocabulary knowledge: What the research shows. *Language Learning*, 64 (4), 2014. p. 913–951.

SCHMITT, N.; SCHMITT, D. & CLAPHAM, C. Developing and exploring the behaviour of two new versions of the Vocabulary Levels Test. *Language Testing*, 18 (1), 2001, p. 55-88.

SCHÜLTZE, C. *The empirical base of linguistics*: Grammaticality judgments and linguistic methodology. Chicago, IL: University of Chicago Press, 1996.

SEGALOWITZ, N. Does advanced skill in a second language reduce automaticity in the first language? *Language Learning*, 41 (1), 1991. p. 59-83.

SEGALOWITZ, N. & HULSTIJN, J. *Automaticity in bilingualism and second language learning.* In: KROLL, J. & DE GROOT, A. (EDS.). Handbook of Bilingualism – Psycholinguistic Approaches. Oxford/New York: Oxford University Press, 2005.

SEGALOWITZ, N. Does advanced skill in a second language reduce automaticity in the first language *Language Learning:* A Journal of Applied Linguistics, (41) , 1991, p. 59-83.

SEGALOWITZ, N. Automaticity and Second Languages. In: DOUGHTY, C. & LONG, M., (eds.). The handbook of second language acquisition, Malden, Ma: Blackwell Publishing Ltd., 2006, p. 382-408.

SHARWOOD SMITH, M. Consciousness-raising and the second language learner. *Applied Linguistics*, 2, 1981, p.159–169.

SKEHAN, P. Modelling Second Language Performance: Integrating Complexity, Accuracy, Fluency, and Lexis. *Applied Linguistics*, 30(4), 2009, p. 510–532.

SOUZA. R. Argument structure in L2 acquisition: language transfer re-visited in a semantics and syntax perspective. *Ilha do Desterro,* 60(1), 2011. p. 153-187.

SOUZA. R. & OLIVEIRA, F. L. P. Is knowledge of a non-dominant L2 activated by bilinguals using their dominant L1? Insights from an on-line psycholinguistic study. *Organon,* Porto Alegre, 51, 2011, p. 103-128.

SOUZA. R. A.Two languages in one mind and the online processing of causatives with manner-of-motion verbs. *ReVEL, special issue*, 6(1), 2012. p. 220-239.

SOUZA. R. & OLIVEIRA, C. The learnability of the resultative construction in English: a comparative study of two forms of the acceptability judgment task. *Revista da ABRALIN,* 13 (2), 2014, p. 375-410.

SOUZA. R. & MELLO, H. Realização argumental na língua do aprendiz de línguas estrangeiras – possibilidades de exploração da interface entre semântica e sintaxe. *Revista Virtual de Estudos da Linguagem – ReVEL*, 5 (8), 2007.

SOUZA. R.; DUARTE, D. & BERG, I. *A avaliação de habilidades linguísticas em L2*: uma questão metodológica em estudos de influências translinguísticas. In: BUCHWEITZ, A. & MOTA, M. B. (eds.) Linguagem e Cognição: Processamento, Aquisição e Cérebro. Porto Alegre: EdiPUCRS, 2015.

SOUZA. R.; OLIVEIRA, C.; SOARES-SILVA, J.; PENZIN, A. & SANTOS, A. Estudo sobre um Parâmetro de Tarefa e um Parâmetro Amostral para Experimentos com Julgamentos de Aceitabilidade Temporalizados. *Revista Estudos da Linguagem*, 23(1), 2015, p. 211-244.

SOUZA. R. & SOARES-SILVA, J. Exploring the measurement of vocabulary size to differentiate Brazilian Portuguese-English bilinguals' access to grammatical knowledge in the L2. *LinguíStica*, 11 (1), 2015, p. 187-204.

SOUZA. R.; SOARES-SILVA, J. & SILVA, G. Brazilian Portuguese-English bilingualism does not afect metalinguistic awareness of L1 constraints in two argument structure constructions. *Ilha do Desterro*, 69(1), 2016, p. 17-31.

SINGLETON, D. *Exploring the Second Language Mental Lexicon*. Cambridge: Cambridge University Press, 1999.

STÆHR, L. Vocabulary size and the skills of listening, reading and writing. Language *Learning Journal*, 36(2), 2008, p. 139-152.

_____. Vocabulary knowledge and advanced listening comprehension in English as a foreign language. *Studies in Second Language Acquisition*, 31(4), 2009, p. 577-607.

TOKOWICZ, N., MICHAEL, E. & KROLL, J. (2004). The roles of study-abroad experience and working-memory capacity in the types of errors made during translation. Bilingualism: *Language and Cognition*, 7(1), 2004, p. 255-272.

TRUSCOTT, J. & SHARWOOD SMITH, M. Acquisition by processing: A modular perspective on language development. Bilingualism: *Language and Cognition*, 7, 2004, p. 1-20.

TRUSCOTT, J. Consciousness in SLA: A modular perspective. *Second Language Research*, 31, 2015, p. 413-434.

ULLMAN, M. The neural basis of lexicon and grammar in first and second language: the declarative/procedural model. *Bilingualism: Language and Cognition*, 4 (2), 2001. p. 105-122.

VALDÉS, G. *Spanish for native speakers:* AATSP professional development series handbook for teachers K-16 (vol.1): New York: Harcourt College Publishers, 2000.

WHITE, L. *Second language acquisition and Universal Grammar*. Cambridge: Cambridge University Press, 2003.

ZARA, J.; OLIVEIRA, F. & SOUZA, R. A. Selective transfer in the acquisition of the double object construction by Brazilian learners. *Alfa – Revista de Linguística*, 57 (2), 2013. p. 519-544.

ZIMMERMAN, K. *The role of vocabulary size in assessing second language proficiency*. MA Thesis, Brigham Young University, 2004.

APPENDIX

APPENDIX 1 – THE ACCEPTABILITY JUDGMENT TASK FOR EXPERIMENT 1

The man laughed the children during the party.
The scientist appeared her paper after years.
The director spoke the actor about the play.
The woman played the children until bed time.
The farmer fell the apple from the tree.
The teenager arrived his friend for his appointment.
The staff flowered the plants before spring.
The president resigned the minister after the incident.
The girl give the cats milk twice a day.
Does the professor talk often about his research?
Where are the articles that contains the information?
Don have always written nice poems.
Who did Jane call her friend after she saw?
What did Steven read the book that Helen talked about?
Moses imagined to whom what he said.
Which case did the detective say was worried?
The instructor ran the boys around the park.
The lady walked her child down the street.
The man swam his dog to his boat.
The rider jumped his horse over the fence.
The scientist flew her balloon above the clouds.
The captain marched his troop into the town.
The child floated her toys on the pond.
The woman danced the man across the room.
The coach told the players to jog around the field.
The officer warned the tourists not to drive in the park.
The instructor asked the girls to skate across the lake.
The veterans ordered the recruits to leap over the hole.
The guards forced the prisoners to march through town.
The teacher told the students to wait for the bell.
The pilots made the planes loop in the sky.
The policeman told the driver to fix his brake lights.
The girls melted the cheese in the bowl.
The fighter broke the chair in the corner.
The servant dried the sheets in the garden.
The woman baked her bread at her place.
The cook froze his dinner at the restaurant.
The general burned the bridges to the village.
The students cooled the cakes on the sink.
The man warmed his soup on the stove.
The ball rolled down the street for a few minutes.
The horse leaped over the fence without warning.
The athletes paraded along the avenue after the victory.
The kids played in the park all morning long.
The boys walked by the stadium at night.
The teacher graded the students' essays at home.
The president spoke about the crisis to the press.
The workers returned to the factory after two days.
The actress answered the questions of the interviewer.
The doctor neglected the patient's complaints about his headache.
The students read the novels over the semester.
The clerks cleaned the office after work.
The driver took the visitors to the wrong place.
The nurse fed the baby every three hours.
The hunter killed the deer with a gun.
The criminals burned the papers from the safe.

APPENDIX 2 – THE ENGLISH VERSION OF THE VOCABULARY LEVEL TEST (VLT)

01

1. original			1. apply		
2. private	☐	complete	2. elect	☐	choose by voting
3. royal	☐	first	3. jump	☐	become like water
4. slow	☐	not public	4. manufacture	☐	make
5. sorry			5. melt		
6. total			6. threaten		

1. blame			1. accident		
2. hide	☐	keep away from sight	2. choice	☐	having a high opinion of yours
3. hit	☐	have a bad effect on something	3. debt	☐	something you must pay
4. invite	☐	ask	4. fortune	☐	loud, deep sound
5. pour			5. pride		
6. spoil			6. roar		

1. basket			1. birth		
2. crop	☐	money paid regularly for doing a job	2. dust	☐	being born
3. flesh	☐	heat	3. operation	☐	game
4. salary	☐	meat	4. row	☐	winning
5. temperature			5. sport		
6. thread			6. victory		

02

1. administration			1. bench		
2. angel	☐	managing business and affairs	2. charity	☐	part of a country
3. front	☐	spirit who serves God	3. fort	☐	help to the poor
4. herd	☐	group of animals	4. jar	☐	long seat
5. mate			5. mirror		
6. pond			6. province		

1. coach			1. marble		
2. darling	☐	a thin, flat piece cut from something	2. palm	☐	inner surface of your hand
3. echo	☐	person who is loved very much	3. ridge	☐	excited feeling
4. interior	☐	sound reflected back to you	4. scheme	☐	plan
5. opera			5. statue		
6. slice			6. thrill		

1. discharge			1. annual		
2. encounter	☐	use pictures or examples to show the meaning	2. blank	☐	happening once a year
3. illustrate	☐	meet	3. brilliant	☐	certain
4. knit	☐	throw up into the air	4. concealed	☐	wild
5. prevail			5. definite		
6. toss			6. savage		

03

1. alcohol
2. apron
3. lure
4. mess
5. phase
6. plank

( ) cloth worn in front to protect your clothes
( ) stage of development
( ) state of untidiness or dirtiness

1. circus
2. jungle
3. nomination
4. sermon
5. stool
6. trumpet

( ) speech given by a priest in a church
( ) seat without a back or arms
( ) musical instrument

1. apparatus
2. compliment
3. revenue
4. scrap
5. tile
6. ward

( ) set of instruments or machinery
( ) money received by the government
( ) expression of admiration

1. bruise
2. exile
3. ledge
4. mortgage
5. shovel
6. switch

( ) agreement using property as security for a debt
( ) narrow shelf
( ) dark place on your body caused by hitting

1. blend
2. devise
3. embroider
4. hug
5. imply
6. paste

( ) hold tightly in your arms
( ) plan or invent
( ) mix

1. desolate
2. fragrant
3. gloomy
4. profound
5. radical
6. wholesome

( ) good for your health
( ) sweet-smelling
( ) dark or sad

04

1. affluence
2. axis
3. episode
4. innovation
5. precision
6. tissue

( ) introduction of a new thing
( ) one event in a series
( ) wealth

1. deficiency
2. magnitude
3. oscillation
4. prestige
5. sanction
6. specifition

( ) swinging from side to side
( ) respect
( ) lack

1. configuration
2. discourse
3. hypothesis
4. intersection
5. partisan
6. propensity

( ) shape
( ) speech
( ) theory

1. anonymous
2. indigenous
3. maternal
4. minimum
5. nutrient
6. modification

( ) without the writer's name
( ) least possible amount
( ) native

1. elementary
2. negative
3. static
4. random
5. reluctant
6. ultimate

( ) of the beginning stage
( ) not moving or changing
( ) final, furthest

1. coincide
2. coordinate
3. expel
4. frustrate
5. supplement
6. transfer

( ) prevent people from doing something they want to do
( ) add to
( ) send out by force

05

1. acquiesce			1. blaspheme		
2. contaminate	☐	work at something without serious intentions	2. endorse	☐	give care and food to
3. crease	☐	accept without protest	3. nurture	☐	speak badly about God
4. dabble	☐	make a fold on cloth or paper	4. overhaul	☐	slip or slide
5. rape			5. skid		
6. squint			6. straggle		

1. auxiliary			1. anterior		
2. candid	☐	full of self importance	2. concave	☐	small and weak
3. dubious	☐	helping, adding support	3. interminable	☐	easily changing
4. morose	☐	bad-tempered	4. puny	☐	endless
5. pompous			5. volatile		
6. temporal			6. wicker		

1. dregs			1. auspices		
2. flurry	☐	worst and most useless parts of anything	2. casualty	☐	being away from other people
3. hostage	☐	natural liquid present in the mouth	3. froth	☐	someone killed or injured
4. jumble	☐	confused mixture	4. haunch	☐	noisy and happy celebration
5. saliva			5. revelry		
6. truce			6. seclusion		

APPENDIX 3– THE SPANISH VERSION OF THE VOCABULARY LEVEL TEST (VLT)

VOCABULARY LEVEL TEST – Spanish Version

You should match each definition with one of the six words in each cluster. So, three words will match the definitions and three words will not be used.

Here is an example:

1- business	
2- clock	( ) part of a house
3- horse	( ) animal with four legs
4- pencil	( ) something used for writing
5- shoe	
6-wall	

You answer it the following way:

1- business	
2- clock	(**6**) part of a house
3- horse	(**3**) animal with four legs
4- pencil	(**4**) something used for writing
5- shoe	
6-wall	

Level 1 (2000)

1- reforma	
2- empleado	( ) persona que trabaja para otra
3- objeto	( ) parte de un territorio
4- político	( ) charla o discurso ante el público
5- región	
6- sentencia	

1- boca	
2- cadáver	( ) esquina
3- dueño	( ) actor principal
4- protagonista	( ) muerto
5- clásico	
6- sitio	

1- rincón	
2- especial	( ) fuerte y estable
3- ético	( ) sin compañía
4- firme	( ) conservador
5- tradicional	
6- solitario	

1-administración	
2- esposo	( ) aparece en el cine
3- hijo	( ) gestión de lo público y privado
4- imagen	( ) marido
5- materia	
6- conferencia	

1- bajo	
2- claro	( ) muy grande
3- desnudo	( ) mucha luz
4- importante	( ) sin ropa
5- inmenso	
6- mujer	

1- anuncio	
2- audiencia	( ) produce sangre en el cuerpo
3- calor	( ) principio
4- dedo	( ) se usa en publicidad
5- herida	
6- inicio	

Level 2 (3000)

1- actriz	
2- cura	( ) persona con mucho poder
3- emperador	( ) objeto envuelto
4- filósofo	( ) hace películas
5- paquete	
6- reloj	

1- agujero	
2- curiosidad	( ) preguntas para detectar la opinión pública
3- encuesta	( ) deseo de saber una cosa
4- humor	( ) cigarrillos
5- mentira	
6- tabaco	

1- asiento	
2- circo	( ) persona que asiste a una presentación
3- despido	( ) acto de expulsar del trabajo
4- espectador	( ) parecido a una silla
5- perfil	
6- subsidio	

1- cementerio	
2- disco	( ) narración oral o escrita
3- clínica	( ) donde atienden a los enfermos
4- fábrica	( ) es típico de los cumpleaños
5- regalo	
6- relato	

1- amargo	
2- crudo	( ) contrario de dulce
3- elegante	( ) muy bonito
4- juvenil	( ) comida no cocinada
5- precioso	
6- publicación	

1- comunicado	
2- delincuente	( ) característica de una persona o cosa
3- editorial	( ) donde están los que cometen un delito
4- prisión	( ) empresa que se dedica a la edición
5- remoto	
6- rasgo	

Level 3 (5000)

1- fijado			1- caridad	( )	estado de descanso
2- mancha	( )	apaga incendios	2- reposo	( )	recolección de frutos de cultivo
3- bombero	( )	descubrimiento	3- enorme	( )	insignia de una nación
4- placa	( )	que no se mueve	4- bandera		
5- roble			5- cosecha		
6- hallazgo			6- nula		

1- placer			1- trozo		
2- bruto	( )	iglesia pequeña	2- pañuelo	( )	pedazo
3- llanto	( )	marca de los pies por donde se pisa	3- prohibición	( )	tela usada para limpiarse
4- aparente	( )	lagrimas y lamentos	4- inmóvil	( )	fijo en un lugar
5- huella			5- arroyo		
6- capilla			6- pálida		

1- norma			1- laguna		
2- cáscara	( )	ágil	2- plazo	( )	buena salud
3- casco	( )	objeto que protege la cabeza	3- árbitro	( )	tiempo para completar algo
4- ligero	( )	poca cantidad	4- sano	( )	autoridad en el terreno de juego
5- pesadilla			5- magistrado		
6- escasez			6- manga		

Level 4 (academic list)

1- adecuado			1- tambor		
2- lectura	( )	apropiado	2- tesoro	( )	compuesto de varios elementos
3- enferma	( )	posición entre dos puntos en una escala	3- devoto	( )	comportamiento
4- grado	( )	leer y comprender un texto	4- prestación	( )	persona religiosa
5- grueso			5- complejo		
6- consistencia			6- conducta		

1- espesor			1- cheque		
2- primo	( )	nivel profesional o académico	2- ambiguo	( )	serie de cosas en línea
3- sugerencia	( )	persona mayor de edad	3- cañón	( )	más de una interpretación
4- rango	( )	numero no divisible	4- jornal	( )	periódico
5- terraza			5- fila		
6- jerarquía			6- colega		

1- escenario			1- aldea		
2- breve	( )	donde se realizan presentaciones	2- categoría	( )	documento oficial
3- década	( )	corta duración	3- anual	( )	sucede cada año
4- tópico	( )	periodo de diez años	4- carta	( )	mitad de algo
5- flecha			5- débil		
6- fantasma			6- media		

Level 5 (10000)

1- escritorio	
2- pasajero	( ) esencia de algo
3- entraña	( ) escrito provisional
4- trasera	( ) grupo de personas con un mismo propósito.
5- concejo	
6- borrador	

1- instancia	
2- acoso	( ) situado
3- anhelo	( ) propiedad privada
4- ubicado	( ) deseo profundo
5- banquero	
6- inmueble	

1- aduana	
2- suelto	( ) saturado
3- cohesión	( ) regula el paso de mercancias entre países
4- cargado	( ) libre
5- embajada	
6- vergüenza	

1- influjo	
2- reportaje	( ) suave
3- huerta	( ) pequeño terreno para sembrar
4- heredero	( ) efecto de una cosa en otra
5- tierno	
6- pastilla	

1- átono	
2- hueco	( ) orificio
3- tarjeta	( ) barrera que separa dos lados
4- sigla	( ) iniciales
5- muralla	
6- gota	

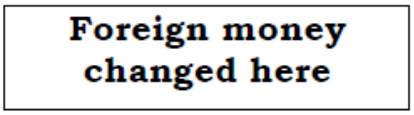
1- acceso	
2- llanura	( ) elegante
3- garrido	( ) un terreno plano
4- cortejo	( ) personas que acompañan en una ceremonia
5- academia	
6- eventual	

APPENDIX 4– THE OXFORD PLACEMENT TEST (OPT)

1

A in a shop
B in a hotel
C in a taxi

2

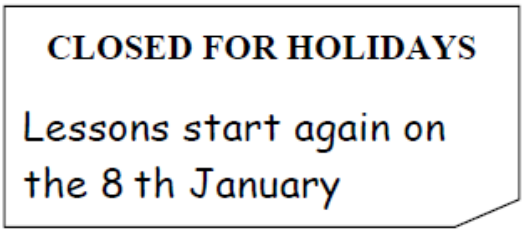


A in a library
B in a bank
C in a police station

3

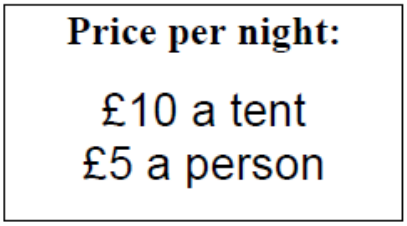
A outside a theatre
B outside a supermarket
C outside a restaurant

4



A at a travel agent's
B at a music school
C at a restaurant

5



A at a cinema
B in a hotel
C on a camp-site

Scotland

Scotland is the north part of the island of Great Britain. The Atlantic Ocean is on the west and the North Sea on the east. Some people **(6)** ..... Scotland speak a different language called Gaelic. There are **(7)** ..... five million people in Scotland, and Edinburgh is **(8)** ..... most famous city.

Scotland has many mountains; the highest one is called 'Ben Nevis'. In the south of Scotland, there are a lot of sheep. A long time ago, there **(9)** ..... many forests, but now there are only a **(10)** ..... .

Scotland is only a small country, but it is quite beautiful.

6	**A**	on	**B**	in	**C**	at
7	**A**	about	**B**	between	**C**	among
8	**A**	his	**B**	your	**C**	its
9	**A**	is	**B**	were	**C**	was
10	**A**	few	**B**	little	**C**	lot

Alice Guy Blaché

Alice Guy Blaché was the first female film director. She first became involved in cinema whilst working for the Gaumont Film Company in the late 1890s. This was a period of great change in the cinema and Alice was the first to use many new inventions, **(11)** ..... sound and colour.

In 1907 Alice **(12)** ..... to New York where she started her own film company. She was **(13)** ..... successful, but, when Hollywood became the centre of the film world, the best days of the independent New York film companies were **(14)** ..... . When Alice died in 1968, hardly anybody **(15)** ..... her name.

11	**A** bringing	**B** including	**C** containing	**D** supporting
12	**A** moved	**B** ran	**C** entered	**D** transported
13	**A** next	**B** once	**C** immediately	**D** recently
14	**A** after	**B** down	**C** behind	**D** over
15	**A** remembered	**B** realised	**C** reminded	**D** repeated

UFOs – do they exist?

UFO is short for 'unidentified flying object'. UFOs are popularly known as flying saucers, **(16)** ..... that is often the **(17)** ..... they are reported to be. The **(18)** ..... "flying saucers" were seen in 1947 by an American pilot, but experts who studied his claim decided it had been a trick of the light.

Even people experienced at watching the sky, **(19)** ..... as pilots, report seeing UFOs. In 1978 a pilot reported a collection of UFOs off the coast of New Zealand. A television **(20)** ..... went up with the pilot and filmed the UFOs. Scientists studying this phenomenon later discovered that in this case they were simply lights on boats out fishing.

16	**A**	because	**B**	therefore	**C**	although	**D** so
17	**A**	look	**B**	shape	**C**	size	**D** type
18	**A**	last	**B**	next	**C**	first	**D** oldest
19	**A**	like	**B**	that	**C**	so	**D** such
20	**A**	cameraman	**B**	director	**C**	actor	**D** announcer

21 The teacher encouraged her students ..... to an English pen-friend.

A should write **B** write **C** wrote **D** to write

22 They spent a lot of time ..... at the pictures in the museum.

A looking **B** for looking **C** to look **D** to looking

23 Shirley enjoys science lessons, but all her experiments seem to ..... wrong.

A turn **B** come **C** end **D** go

24 ..... from Michael, all the group arrived on time.

A Except **B** Other **C** Besides **D** Apart

25 She ..... her neighbour's children for the broken window.

A accused **B** complained **C** blamed **D** denied

26 As I had missed the history lesson, my friend went ..... the homework with

A by **B** after **C** over **D** on

27 Whether she's a good actress or not is a ..... of opinion.

A matter **B** subject **C** point **D** case

28 The decorated roof of the ancient palace was ..... up by four thin columns.

A built **B** carried **C** held **D** supported

29 Would it ..... you if we came on Thursday?

A agree **B** suit **C** like **D** fit

30 This form ..... be handed in until the end of the week.

A doesn't need **B** doesn't have **C** needn't **D** hasn't got

31 If you make a mistake when you are writing, just ..... it out with your pen.

A cross **B** clear **C** do **D** wipe

32 Although our opinions on many things ..... , we're good friends.

A differ **B** oppose **C** disagree **D** divide

33 This product must be eaten ..... two days of purchase.

A by **B** before **C** within **D** under

34 The newspaper report contained ..... important information.

A many **B** another **C** an **D** a lot of

35 Have you considered ..... to London?

A move **B** to move **C** to be moving **D** moving

36 It can be a good idea for people who lead an active life to increase their ..... of vitamins.

A upturn **B** input **C** upkeep **D** intake

37 I thought there was a ..... of jealousy in his reaction to my good fortune.

A piece **B** part **C** shadow **D** touch

38 Why didn't you ..... that you were feeling ill?

A advise **B** mention **C** remark **D** tell

39 James was not sure exactly where his best interests ..... .

A stood **B** rested **C** lay **D** centred

40 He's still getting ..... the shock of losing his job.

A across **B** by **C** over **D** through

The tallest buildings - SKYSCRAPERS

Nowadays, skyscrapers can be found in most major cities of the world. A building which was many **(41)** …….. high was first called a skyscraper in the United States at the end of the 19th century, and New York has perhaps the **(42)** …….. skyscraper of them all, the Empire State Building. The **(43)** …….. beneath the streets of New York is rock, **(44)** …….. enough to take the heaviest load without sinking, and is therefore well-suited to bearing the **(45)** …….. of tall buildings.

41	**A** stages	**B** steps	**C** storeys	**D** levels
42	**A** first-rate	**B** top-class	**C** well-built	**D** best-known
43	**A** dirt	**B** field	**C** ground	**D** soil
44	**A** hard	**B** stiff	**C** forceful	**D** powerful
45	**A** weight	**B** height	**C** size	**D** scale

SCRABBLE

Scrabble is the world's most popular word game. For its origins, we have to go back to the 1930s in the USA, when Alfred Butts, an architect, found himself out of **(46)** …….. . He decided that there was a **(47)** ……. for a board game based on words and **(48)** ……. to design one. Eventually he made a **(49)** ……. from it, in spite of the fact that his original **(50)** ……. was only three cents a game.

46	**A**	earning	**B**	work	**C**	income	**D** job
47	**A**	market	**B**	purchase	**C**	commerce	**D** sale
48	**A**	took up	**B**	set out	**C**	made for	**D** got round
49	**A**	wealth	**B**	fund	**C**	cash	**D** fortune
50	**A**	receipt	**B**	benefit	**C**	profit	**D** allowance

51 Roger's manager ..... to make him stay late if he hadn't finished the work.

A insisted **B** warned **C** threatened **D** announced

52 By the time he has finished his week's work, John has hardly ..... energy left for the weekend.

A any **B** much **C** no **D** same

53 As the game ..... to a close, disappointed spectators started to leave.

A led **B** neared **C** approached **D** drew

54 I don't remember ..... the front door when I left home this morning.

A to lock **B** locking **C** locked **D** to have locked

55 I ..... to other people borrowing my books: they always forget to return them.

A disagree **B** avoid **C** dislike **D** object

56 Andrew's attempts to get into the swimming team have not ..... with much success.

A associated **B** concluded **C** joined **D** met

57 Although Harry had obviously read the newspaper article carefully, he didn't seem to have ..... the main point.

A grasped **B** clutched **C** clasped **D** gripped

58 A lot of the views put forward in the documentary were open to ..... .

A enquiry **B** query **C** question **D** wonder

59 The new college ..... for the needs of students with a variety of learning backgrounds.

A deals **B** supplies **C** furnishes **D** caters

60 I find the times of English meals very strange – I'm not used ..... dinner at 6pm.

A to have **B** to having **C** having **D** have

APPENDIX 5– THE SPANISH PLACEMENT (SPT)

PLACEMENT TEST – Full name: _____ Age: _____ Code: _____

1. ¿Cómo se llama la letra "V"?
.a- ( ) be
b- ( ) uve
c- ( ) u

2. ¿Qué número es: "96"?
a- ( ) sesenta y nueve
b- ( ) nueventa y seis
c- ( ) noventa y seis

3. Mi amiga Ana tiene los ojos…
a- ( ) negro.
b- ( ) azulos.
c- ( ) verdes.

4. ¿Qué hora es? "9:40"
a- ( ) Son las diez menos veinte.
b- ( ) Son las nueve menos veinte.
c- ( ) Es la diez menos veinte.

5. - ¿..... vas a la fiesta?
- Voy a las diez.

a- ( ) Por qué
b- ( ) Cómo
c- ( ) Cuándo

6. ¿Dónde..... encontrar una farmacia en es pueblo?

a- ( ) tengo
b- ( ) debo
c- ( ) puedo

7. COMPRENSIÓN DE LECTURA. Leer el texto y elegir la respuesta correcta.

Un día en la vida de Lena

Lena tiene veintidós años, es alemana y muy guapa. Es de Hamburgo y vive en una casa muy grande con sus amigas. L estudia Económicas en la Universidad de Hamburgo. ¡Es muy inteligente! Por la mañana, Lena se despierta a las siete, temprano! Se ducha, se viste, se lava los dientes y desayuna: hoy leche con cereales y una manzana. ¡Lena es muy sana!

Lena desayuna y después se viste.
a- ( ) Verdadero
b- ( ) Falso

8. COMPRENSIÓN DE LECTURA. Leer el texto y elegir la respuesta correcta.

En una boutique del barrio de Salamanca

DEPENDIENTA.— Buenas tardes. ¿Qué desea?
CARRIE.— Hola, me gustan esos zapatos de ahí. Son muy bonitos.
DEPENDIENTA.— ¿Cuáles? ¿Esos de ahí que están en la estantería?
CARRIE.— Sí, sí, esos. ¿Cuánto cuestan?
DEPENDIENTA.— Cuestan 350 euros.
CARRIE.— Son muy caros, pero son muy bonitos y elegantes. Me encantan.
DEPENDIENTA.— Sí, son de Manolo Blahnik.

Los zapatos que le gustan a Carrie son...
a- ( ) bonitos y baratos.
b- ( ) elegantes y caros.
c- ( ) bonitos y económicos.

9. Cuando entra, todos miran. Lleva un sombrero negro y un vestido rojo de fiesta. Hoy..... la presentación de su última película.
a- ( ) está
b- ( ) es
c- ( ) hay

10. -¿De quién son esas bicicletas que hay ahí?
-Son..... (mis hermanos y yo).

a- ( ) mías
b- ( ) nuestras
c- ( ) suyas

11. Hoy Cristina..... muy temprano y desayuna sola en la cocina.
a- ( ) se despierta
b- ( ) se levantan
c- ( ) te despiertas

12. Desde ..... monumento del Alcázar se puede ver la Catedral y la Giralda.
a- ( ) este
b- ( ) esto
c- ( ) aquello

13. Muchas gracias, de momento no quiero....., pero si necesito tu ordenador, sé que puedo ir a tu casa.
a- ( ) algunos
b- ( ) nada
c- ( ) alguno

14. Sí, esta mañana..... la cocina de la casa nueva que han comprado.
a- ( ) han pitado
b- ( ) habemos pintado
c- ( ) han pintado

15. Alguien pregunta: "¿Puedo abrir la ventana?". ¿Cómo respondes formalmente que no?
a- ( ) Disculpe, pero está resuelta.
b- ( ) Sí, sí, ábrela.
c- ( ) Perdone, pero estoy resfriado.

16. COMPRENSIÓN DE LECTURA. Leer el texto y elegir la respuesta correcta.

¡Menudo día!

Esta mañana me he levantado de la cama muy contento porque hoy es mi cumpleaños. Ya tengo 18 años y ya puedo hace examen de conducir porque mi padre ha dicho que, de regalo de cumpleaños, ¡me compra un coche! Después he ido a cla Todos mis amigos saben cuándo es mi cumpleaños, pero la primera persona que me ha felicitado de todos ha sido Sonia, chica más guapa de mi clase.

¿Quién me ha sido la primera persona en felicitarme?
a- ( ) Sonia
b- ( ) Mis amigos en el colegio
c- ( ) Mi padre, después de levantarme

17. Cuando era joven..... de una chica de su clase y tuvo una aventura con ella, aunque rompieron al llegar a la universidad.
a- ( ) me enamoré
b- ( ) se enamoró
c- ( ) me relacioné

18. - Al final, ¿qué hicisteis ayer?
- Pues..... una vuelta por el casco viejo de la ciudad.

a- ( ) dimos
b- ( ) habíamos dado
c- ( ) hemos dado

19. Cada vez que ellos..... tiempo libre, iban de copas a un bar de marcha que estaba cerca de la playa, junto a zona comercial de Matalascañas.
a- ( ) salían
b- ( ) tenían
c- ( ) habían tenido

20. Supo que ya..... de puente cuando vio que su maleta no estaba en el armario; seguramente ahora estaba haciendo turismo.
a- ( ) se había marchado
b- ( ) se marchó
c- ( ) se marchaba

21. -¿Y no sabes nada más de él?
- Pues no. No me escribe ..... hace tres meses, por lo menos.

a- ( ) desde
b- ( ) sobre
c- ( ) de

22. Son muchos los ecologistas que aconsejan usar la bici y no el coche. Es una cuestión medioambiental: la bici contamina..... que el coche.
a- ( ) menos
b- ( ) tanto
c- ( ) tan

23. Cuando nos dicen: "No está, ¿quiere que le diga algo?", ¿qué nos preguntan indirectamente?
a- ( ) Si queremos hablar con ella.
b- ( ) Si queremos llamar otra vez.
c- ( ) Si queremos dejar un mensaje.

24. COMPRENSIÓN DE LECTURA. Leer el texto y elegir la respuesta correcta.

Hermanos gemelos

Cuando estábamos en el instituto yo salía con Carmen, una chica morena muy guapa y Juan salía con Elena, una rubia muy guapa también. Aunque Carmen era más guapa que Elena, Juan siempre lo discutía. Pero yo tenía razón. En la Universidad los dos estudiamos Medicina porque los dos queríamos ser médicos. Los dos estudiábamos mucho para conseguir mejores notas que el otro. Acabamos la Universidad con unas notas muy parecidas, pero las notas de Juan fueron un poco peores que las mías, evidentemente, yo siempre he sido más listo que él. Ahora los dos somos excelentes cardiólogos. Nuestra reputación es conocida y respetada en todo el mundo.

¿Quién sacaba mejores notas en la Universidad?

a- ( ) El narrador del texto
b- ( ) Mi gemelo
c- ( ) Los gemelos sacan notas iguales.

25. He oído que..... a tu padre de la empresa. En fin, que le vamos a hacer. Esto ya no tiene arreglo.
a- ( ) despedirán
b- ( ) decidirá
c- ( ) habrán elegido

26. Y pensar que dentro de muy poco los coches que conocemos hoy en día..... o serán solo objetos de museo.
a- ( ) habrá aparecido
b- ( ) habrán desaparecido
c- ( ) habrán estado

27. Si todavía no has ido, ... hacerlo lo antes posible. Sevilla es una de las ciudades más bonitas de España.
a- ( ) tendrías
b- ( ) deberías
c- ( ) tenerías

28. Hace mucho frío en la calle. Por favor, ... (ustedes) la puerta cuando salgan.
a- ( ) cerrad
b- ( ) cierren
c- ( ) cierra

29. Mejor dicho, cuando ..... a Marta, decidle que me llame.
a- ( ) veáis
b- ( ) veis
c- ( ) veréis

30. Espero que el tren..... ya. No me gusta tener que esperar demasiado en esta estación.
a- ( ) ha llegado
b- ( ) haya llegado
c- ( ) está llegando

31. ¿Cómo propondrías a tu grupo de amigos quedar con Raúl, a quien ya hace mucho que no veis?
a- ( ) Bueno, al final, ¿quedamos con Raúl?
b- ( ) Supongo que no querréis quedar con Raúl, ¿verdad?
c- ( ) Estaría bien salir algún día con Raúl. Llevamos meses sin verlo.

32. COMPRENSIÓN DE LECTURA. Lee el texto y elige la respuesta correcta.

La Tomatina

Era agosto de 1945. En Buñuel, como en otros puntos de la península, se celebraba un desfile de Gigantes y Cabezudos, que es una cabalgata en la que los lugareños se disfrazan y pasean por el pueblo. Ese año, todo el mundo quería disfrazarse. Cuentan, entonces, que un hombre, enfadado porque no le habían permitido participar en el desfile, golpeó a otro que estaba disfrazado. La pequeña pelea, desencadenó otra más grande, en la que alguien, usó las verduras de un puesto de hortalizas para lanzarlas, cogiendo y tirando luego los tomates que allí estaban. De repente, una lluvia de tomates tomó las calles, y la gente empezó arrojar tomates sin parar: la Tomatina había nacido. Se tomaron medidas contra los culpables de esa revuelta, pero al año siguiente, la gente, que recordaba en su memoria tal situación, repitió el episodio, reuniéndose en la plaza con miles de cajas de tomates. El intento del gobierno por paliar lo sucedido fue en vano. La Tomatina había llegado para quedarse.

El origen de la tomatina coincidió con el día del espectáculo de los Gigantes y Cabezudos.
a- ( ) Verdadero
b- ( ) Falso
c- ( ) No se sabe a ciencia cierta

33. ¿Sería mucho pedir que..... la mirada al hablar? No es que quiera ligar contigo, simplemente quiero estar seguro de que me estás escuchando.
a- ( ) mantuvieras
b- ( ) mantendrías
c- ( ) sujetarías

34. Los sindicatos quisieron que los trabajadores..... a cobrar el subsidio de desempleo antes de haber firmado el despido.
a- ( ) hubieran empezado
b- ( ) hayan empezado
c- ( ) empezarían

35. Indica a qué tiempo verbal corresponden el verbo en negrita en la siguiente frase.
Sus datos se conservarán en la notaría.
a- ( ) condicional compuesto
b- ( ) condicional simple
c- ( ) futuro simple

36. -En resumidas cuentas, si tenemos que captar clientes, siempre ..... nuevos productos que se pueden pagar a plazos.
- Veo que tu empresa está ya en auge.

a- ( ) ofertamos
b- ( ) endeudamos
c- ( ) adjudicados

37. No teníamos nada y sin embargo teníamos todo; una mujer barbuda llamada Oscar y un payaso tan patoso que era torpe hasta cuando no.....
a- ( ) actuaba
b- ( ) ofertaba
c- ( ) sabía

38. ¿Me permites darte un consejo? Si no..... tu interés, pon fin a la relación. Sólo sientes por él afecto, no tienes un sentimiento profundo.
a- ( ) despierta
c- ( ) despertará
d- ( ) despertaría

39. ¿Qué significa esta expresión: "Me importa un pimiento"?
a- ( ) Te gustan los pimientos.
b- ( ) Te es indiferente.
c- ( ) Te importa mucho.

40. COMPRENSIÓN DE LECTURA. Lee el texto y elige la respuesta correcta.

Nuevo signo zodiacal

Un revuelo está ocurriendo entre los astrónomos, astrólogos y gentes aficionadas al horóscopo, ya que acaban de publicarse recientes declaraciones de parte del emisario de la Minnesota Planetarium Society, Park Kunkle, quien llega a asegurar la designación de Ofiuco como parte del Zodíaco. Esto es debido a los cambios en la posición del planeta respecto de la ubicación de las constelaciones en los últimos dos mil años. "Cuando se desarrolló la astrología, observábamos una posición parecida a ésta. Y el Sol podría entrar en una constelación en particular. Pero después de que pasaran miles de años, esa posición ha llegado a cambiar y ahora hay que admitir que el Sol está en una constelación diferente en la misma fecha", declaró Kunkle. "Cuando los astrólogos dicen que el sol está en Piscis, en realidad está en Acuario", agregó.

Kunkle es:
a- ( ) Astrónomo.
b- ( ) Astrólogo.
c- ( ) Posee ambas especialidades.

41. Cuando se enteró de todas las mentiras que le había dicho, me dijo: "..... te enteres, estoy hasta el gorro de ti".
a- ( ) Para
b- ( ) Para que
c- ( ) A que

42. La basura constituye un problema en muchas sociedades. ..... solución, cada vez resulta más habitual ver en muchos países europeos contenedores que, al reciclar, te abonan una pequeña cantidad económica.
a- ( ) A fin de darle
b- ( ) A fin de cogerle
c- ( ) A fin de haberle

43. ..... que la meteoróloga advirtió de las inclemencias del tiempo, no estábamos dispuestos a suspender nuestro mágico encuentro después de tantos años.
a- ( ) Porque
b- ( ) Si mal
c- ( ) Por más

44. Esa lo que quiere es desplumarlo, ¿no ves que es una fresca? En su familia todas son iguales,..... que la madre ha sido buena maestra.
a- ( ) sin embargo
b- ( ) habida cuenta de
c- ( ) no es

45. Los republicanos tendieron una emboscada al bando nacional, ... para conseguir detenerlos a la entrada de Barcelona y finalmente entraron.
a- ( ) aunque fue suficiente
b- ( ) aunque no fue suficiente
c- ( ) por lo suficiente

46. Completa el hueco con la palabra correcta: "No te lo tomes tan…"
a- ( ) a la cabeza
b- ( ) a pecho
c- ( ) al corazón

47. ¿Cómo preguntamos si alguien recuerda un acontecimiento?
a- ( ) Mañana tienes médico a las cinco, ¿no?
b- ( ) Mañana tienes médico a las cinco. ¿Lo tienes presente?
c- ( ) Mañana tienes médico a las cinco, ¿verdad?

48. COMPRENSIÓN DE LECTURA. Lee el texto y elige la respuesta correcta.

Una Hypatia del S.XXI

Rita Levi-Montalcini es una desconocida para la mayoría, al igual que lo fue la bella Hypatia, primera astrónoma de la etapa neoplatónica de Alejandría o al igual que Marie Curie, investigadora de la radiactividad. Todas ellas tienen algo en común: ser mujeres científicas en un campo dominado por los hombres. Esta mujer que fue Premio Nobel de Medicina.

De origen judío, Rita nació en la Italia de principios del S.XX, en el seno de una familia culta pero de una moral victoriana. Sus ojos, brillantes y verdes, suavemente ajados por el paso del tiempo, son el espejo de su alma, un corazón libre, el reflejo de una vida tejida en hilos de esperanza, vida que ella misma cosió con mucho esfuerzo, y a pesar de ello, logró doctorarse en Neurocirugía. Tan pronto como habla del fascismo que le tocó vivir, su expresión se torna amarga, y relata apesadumbrada todo lo que luchó para conseguir sus objetivos.

Por lo tanto, al conocer su biografía, no puedes evitar acordarte de la enigmática Hypatia. Reflexionas sobre el paralelismo que comparte con Hypatia, y ves a dos mujeres que separadas por miles de siglos, lucharon por sus hipótesis científicas cuando el campo de la ciencia estaba y está dominado por el sometimiento masculino. Las dos, víctimas de un trasfondo social peligroso: Hypatia, laica, rodeada en el siglo IV, del caos de una guerra de religión y Rita, en el escenario fascista y sangriento de la II Guerra Mundial. Ninguna de las dos se casó – "Yo soy mi propio marido" – decía siempre Rita entre risas. En un contexto así, es maravilloso ver que hay mujeres que, cuando menos te lo esperas, dejan la orilla atrás y se lanzan a "la mar".

¿Qué semejanzas comparten Hypatia y Rita Levi-Montalcini?
a- ( ) Las dos son feministas y pioneras de grandes descubrimientos en el campo científico.
b- ( ) Lucharon por aportar nuevas teorías científicas en un dominio donde imperan los hombres.
c- ( ) Son célebres y reconocidas científicas en un terreno dominado por los hombres.

APPENDIX 6 - THE BILINGUAL HISTORY QUESTIONNAIRE (BHQ)

SAMPLE LANGUAGE HISTORY QUESTIONNAIRE: BILINGUAL

BACKGROUND INFORMATION

All personal information you provide will maintained in strict confidentiality. Feel free to use the back of the sheet if you need more room.

Subject No.: .....

Age:..... Sex: ☐ male ☐ female

Age you were first exposed to Spanish: ..... Age you were first exposed to English: .....

Briefly explain when you began learning each of your languages:.....
.....
.....

Where did you learn each language? Spanish: .....
English: .....

What language(s) does your mother speak? ..... your father? .....

If you've lived outside of the US, describe briefly where, when, and for how long.....
.....
.....

Age you arrived to US: .....

Do you speak any languages other than Spanish and English? .....
(Indicate "fluent" or "only a little".) .....

City/Country of origin: ..... Occupation: .....

Are you right handed or left handed? ☐ right handed ☐ left handed
Do you have any left-handed blood relations? (please list them).....

Education background (check all that apply):

elementary school	in Spanish	☐ in English
high-school	in Spanish	☐ in English
college	in Spanish	☐ in English
graduate school	in Spanish	☐ in English
other: please explain	in Spanish	☐ in English

.....

Where did you attend elementary school?	in US	☐ elsewhere (indicate where).....
Where did you attend high-school?	in US	☐ elsewhere.....
Where did you go to college?	in US	☐ elsewhere.....

For the following questions, circle the number that corresponds with the amount of English or Spanish that you generally use. Follow the scale below:

1	2	3	4	5	n/a
Spanish all the time (always)	Spanish usually more than English	Spanish as much as English	English usually more than Spanish	English all the time (always)	does not

When you were a child, how much Spanish/English did you speak:	always Spanish				always English	
at home, to your parents?	1	2	3	4	5	n
at home, to your brothers or sisters?	1	2	3	4	5	n
at home, to your grandparents?	1	2	3	4	5	n
at home, to other relatives?	1	2	3	4	5	n
to your friends?	1	2	3	4	5	n
in other social contexts (to neighbors, people at the supermarket, etc.)?.	1	2	3	4	5	n

When you were a child, how much Spanish/English did the following people speak to you?	always Spanish				always English	
your parents	1	2	3	4	5	n
your brothers or sisters	1	2	3	4	5	n
your grandparents	1	2	3	4	5	n
other relatives	1	2	3	4	5	n
your friends	1	2	3	4	5	n
other people	1	2	3	4	5	n

When you were a teenager, how much Spanish/English did you speak:	always Spanish				always English	
at home, to your parents?	1	2	3	4	5	n/a
at home, to your brothers or sisters?	1	2	3	4	5	n/a
at home, to your grandparents?	1	2	3	4	5	n/a
at home, to other relatives?	1	2	3	4	5	n/a
to your friends?	1	2	3	4	5	n/a
to your teachers at school?	1	2	3	4	5	n/a
in other social contexts?	1	2	3	4	5	n/a

When you were a teenager, how much Spanish/English did the following people speak to you?	always Spanish				always English	
your parents	1	2	3	4	5	n/a
your brothers or sisters	1	2	3	4	5	n/a
your grandparents	1	2	3	4	5	n/a
other relatives	1	2	3	4	5	n/a
your friends	1	2	3	4	5	n/a
your teachers at school	1	2	3	4	5	n/a
other people	1	2	3	4	5	n/a

(The scale is repeated here for your reference.)

1	2	3	4	5	n/a
Spanish all the time (always)	Spanish usually more than English	Spanish as much as English	English usually more than Spanish	English all the time (always)	does not

How much Spanish/English do you speak now:	always Spanish				always English	
at home, to your spouse, living companion, roommate?	1	2	3	4	5	n/
at home, to your children?	1	2	3	4	5	n/
at home, to your relatives?	1	2	3	4	5	n/
to your friends?	1	2	3	4	5	n/
to your colleagues at work?	1	2	3	4	5	n/
to people at school?	1	2	3	4	5	n/
in other social contexts?	1	2	3	4	5	n/

How much Spanish/English do the following people speak to you now?	always Spanish				always English	
your spouse, living companion, roommate	1	2	3	4	5	n/a
your children	1	2	3	4	5	n/a
your relatives	1	2	3	4	5	n/a
your friends	1	2	3	4	5	n/a
your colleagues at work	1	2	3	4	5	n/a
people at school	1	2	3	4	5	n/a
other people	1	2	3	4	5	n/a

Rate yourself according to the following categories (circle one on each line):

How would you rate your speaking ability in English/Spanish?

ENGLISH:	very good	somewhat good	so-so	somewhat poor	very poor
SPANISH:	very good	somewhat good	so-so	somewhat poor	very poor

How would you rate your reading ability in English/Spanish?

ENGLISH:	very good	somewhat good	so-so	somewhat poor	very poor
SPANISH:	very good	somewhat good	so-so	somewhat poor	very poor

How would you rate your writing ability in English/Spanish?

ENGLISH:	very good	somewhat good	so-so	somewhat poor	very poor
SPANISH:	very good	somewhat good	so-so	somewhat poor	very poor

How would you rate your comprehension in English/Spanish?

ENGLISH:	very good	somewhat good	so-so	somewhat poor	very poor
SPANISH:	very good	somewhat good	so-so	somewhat poor	very poor

Could you pass as a monolingual speaking on the telephone with someone who doesn't know you?

IN ENGLISH:	always	almost always	sometimes	almost never	never
IN SPANISH:	always	almost always	sometimes	almost never	never

Could you pass as a monolingual speaker in a face-to-face conversation with a stranger?

IN ENGLISH:	always	almost always	sometimes	almost never	never
IN SPANISH:	always	almost always	sometimes	almost never	never

THANKS FOR YOUR COOPERATION!

Please take a moment now to ensure that you have filled in all the blanks.