Prifysgol Abertawe
Swansea University

Developing a unified feature-based model for L2 lexical and syntactic processing

Gabriele Luoni

Submitted to Swansea University in fulfilment of the requirements for the

Degree of Doctor of Philosophy

Swansea University

2021

Abstract

Research on lexical processing shows that lexical representations of L2 speakers are less developed, so frequency and vocabulary size affect the way they use lexical information. Specifically, reduced access to lexical features hinders the processing system of L2 speakers from working efficiently, having an impact on their ability to build syntactic structures in a native-like manner. The present research project aims to construct and test a unified model that explains how lexical and sentence processing interact. First, it develops and validates a productive vocabulary task for L2 Italian to measure vocabulary size. The task, called I-Lex, is based on the existing LEX30 for English, and uses frequency to determine lexical knowledge. Then, adopting the formalism of Head-Driven Phrase Structure Grammar, a framework that associates all the information relevant to the grammar with the lexicon, the research project develops a model that explains the effects of lexical access on syntactic processing. The model is tested in two empirical studies on L2 speakers of Italian. The first study, using an Oral Elicited Imitation task, and the I-Lex productive vocabulary task investigates the effects of frequency and vocabulary size on cleft sentences. The second study, using the same productive vocabulary task and a Self-paced Reading task, investigates frequency and vocabulary effects on relative clauses. The results reveal that frequency and vocabulary size interact with the ability of L2 speakers to process both cleft and relative clauses, providing evidence that accessing lexical features is a crucial stage for processing syntactic structures. Based on the results, a feature-based lexical network model is constructed. The model describes how lexical access and the activation of structural links between words can be described using the same set of lexical features. In the last chapter, the model is applied to the results of the two studies.

DECLARATION AND STATEMENTS

DECLARATION

This work has not previously been accepted in substance for any degree and is not being concurrently submitted in candidature for any degree.

Signed [redacted]

Date 30/07/2021

STATEMENT 1

This thesis is the result of my own investigations, except where otherwise stated. Other sources are acknowledged by footnotes giving explicit references. A bibliography is appended.

Signed [redacted]

Date 30/07/2021

STATEMENT 2

I hereby give consent for my thesis, if accepted, to be available for photocopying and for inter-library loan, and for the title and summary to be made available to outside organisations.

Signed [redacted]

Date 30/07/2021

Table of Contents

Acknowledgements

I would like to express my gratitude to my supervisor, Dr. Vivienne Rogers, for all she has done throughout this long and challenging path. Her energy and passion have always helped me take on the most difficult moments. I am sure that a great part of what I have written on these pages would not have been possible without her support. I would like to thank Dr. Cornelia Tschichold, whose feedback has been truly helpful and has made me improve.

I would like to thank Tess Fitzpatrick and the Applied Linguistics team at Swansea University. They really have the ability to make every student feel at home and to provide the most positive and stimulating environment.

A special thanks goes to all the staff at Swansea University and the Library team for being always kind and welcoming. I would like to thank all the colleagues with whom I shared my happy moments along with the sad ones. They have always been there with the right words.

I would like to thank my parents for being always present and supportive, with long and uplifting Skype meetings from Italy. My deepest gratitude goes to Antonio, Tami, Tori, Octavia, and Nic for being there when I needed them. A special and big thanks to Dusica: her love has been the brightest light during the most difficult moments.

List of Tables

List of Figures

List of Examples

Chapter 1 Introduction

In a typical classroom where Italian, the fourth most studied language in the world, is taught every day, a typical lesson would have a vocabulary section. The section would be kept firmly separate from the other sections, and learners will hold on to the idea that the vocabulary is the list of words that is needed to speak about many things. However, vocabulary does more than that. The aim of the present research project it to show that it does all the "nitty-gritty" work needed to have meaningful strings of language.

In recent years, a wealth of research has shown how the strength of lexical representations, frequency and vocabulary size of L2 speakers affects lexical processing (Bialystok et al., 2008; Brysbaert et al., 2017, 2018; Cop et al., 2015; De Deyne et al., 2013, 2019; de Groot & van Hell, 1998a; Diependaele et al., 2013; Dijkstra et al., 2019; Dijkstra & van Heuven, 2002; Duyck et al., 2008; Fitzpatrick & Izura, 2011; Lemhöfer et al., 2008; Lemhöfer & Broersma, 2012; Mainz et al., 2017; Meara, 2006; Whitford & Titone, 2012; Wilks & Meara, 2002; Yap et al., 2009, 2012). At the same time, several syntactic frameworks have emerged that have opted for encoding more grammatical information into lexical items, in the form of specific structural – but also semantic - features (e.g., Boas & Sag, 2010; Bresnan, 2016; Chomsky, 1995; Pollard & Sag, 1994). Some of these frameworks have converged on constraint-based models that consider every type of information relevant to language processing as lexically encoded (MacDonald et al., 1994; McRae & Matsuki, 2013; Pollard & Sag, 1994; Sag & Wasow, 2011, 2014; Wasow, 2019).

Overall, research coming from different perspectives agrees on the necessity of assigning a prominent role to lexical information. However, there are still theoretical differences between frameworks and differences in scope between processing models and syntactic theories. Most current models of L2 sentence processing are based on so called syntax-first models (Clifton & Frazier, 1989; Frazier, 2013; Juffs & Rodrigues, 2015) which posit a staged use of lexical information. A first automatic stage where grammatical information is automatically accessed is followed by a second stage where meaning information is added to complete the interpretation of the sentence (Clahsen & Felser, 2006, 2018; Frazier, 2013). Such differences have not helped to create a common framework to set out by which mechanism lexical information is encoded and accessed in L2 speakers. Hence, research on lexical processing and on sentence processing have proceeded separately, with one camp not exploiting fully the results of the other camp.

However, some notable exceptions have found experimental evidence that there are, indeed, links between lexical knowledge sentence processing, and the ability to produce specific syntactic structures (David et al., 2009; Hopp, 2016; Kawaguchi, 2016; Rogers et al., 2018; Tily et al., 2010). The present research project follows up on these studies by addressing the gap between the two strands of lexical and sentence processing research. It incorporates the notions that have proven to be relevant for L2 lexical processing and accommodates them into an existing theoretical framework that is entirely lexically driven, the Head-driven Phrase Structure Grammar (henceforth HPSG, see Pollard & Sag, 1994). In doing so the project addresses three main points. The first is to have a reliable measurement task that could offer an estimate of vocabulary size. Measuring the breadth of lexical knowledge has the potential of determining the amount of lexical information that L2 speakers rely on when processing language. Having a reliable measurement tool of lexical knowledge is pivotal in comparing vocabulary size to the ability of processing language. The second point concerns how to use an existing syntactic framework, namely, HPSG, to create a formal machinery that could explain how lexical information is encoded and then accessed to be used in language processing. Such a model needs to have the ability to describe lexical representations in a principled manner and the flexibility to be adjusted to a lexically driven model of grammar. The third point is to extend the model to L2 speakers of Italian, a language that, thus far, has not received the same amount of research as other languages. Italian is the fourth most studied language in the world, with a number of students estimated to be around 150 million (Balboni, 2019). However, only in the last two decades there has been a significant increase of research on language teaching and learning (Rastelli, 2013, 2018). Studies on Italian as a second language are still in the early stages and any contribution is important.

The points addressed in the previous section underlie the two main aims of the project that can be summarised as follows:

1. Is there a relationship between vocabulary knowledge, lexical access, and the ability to process language in on-line tasks?
2. Is it possible to integrate lexical and syntactic knowledge in a unified model, exploiting the properties of an existing framework?

Turning to the first research aim, the present project builds on the few studies that have used frequency as the main tool to explore the links between lexical and sentence processing (Hopp, 2016; Tily et al., 2010). The choice of frequency is based on two assumptions. In Chapter 2, I will show that frequency effects are crucial in lexical processing research and they interact with vocabulary size (Bialystok et al., 2008; Brysbaert et al., 2017, 2018; Cop et al., 2015; Diependaele et al., 2013; Lemhöfer et al., 2008). Moreover frequency underlies the mechanism by which words are stored and integrated into lexical representations (e.g., De Deyne & Storms, 2008, 2015; Meara, 2009; Meara & Fitzpatrick, 2000; Steyvers & Tenenbaum, 2005). Chapter 2 also outlines the notion of feature availability, which claims that for L2 speakers it is more difficult to access lexical features because their lexical representations are weaker and less organised (e.g., Brysbaert et al., 2017; Diependaele et al., 2013; Meara, 2009). Overall, L2 speakers need more resources to make lexical features available to the same degree as L1 speakers and this hinders their processing system from working efficiently.

To investigate the relationship between lexical and syntactic processing, it is necessary to measure vocabulary size. To date, there are no tasks to measure productive vocabulary size for Italian. A receptive vocabulary test, specifically developed to measure L2 vocabulary size, was recently published but was not available at the outset of this PhD (Amenta et al., 2021). I-Lex, the productive vocabulary task that I develop in Chapter 3, is adapted to Italian from an existing test, namely Lex30 (Fitzpatrick & Clenton, 2010, 2017; Fitzpatrick & Meara, 2004; Meara & Fitzpatrick, 2000) and it estimates vocabulary size based on the number of infrequent words provided by the test takers. I will examine the findings of the validation study using three distinct groups. A group of 30 pre-intermediate L2 speakers, a group of 82 advanced L2 speakers and a control group of 36 L1 speakers of Italian. The study, besides validating the productive measurement task, aims to unveil whether frequency is a reliable tool to measure the different degree of complexity in L1 and L2 lexical representations. To further establish the construct validity of the test, I will compare the I-Lex results with results from other studies that have used Lex30.

Achieving the second research aim encompasses a theoretical stage and an experimental stage. In the theoretical stage I will lay the groundwork to develop and implement a model that is able to define how lexical information is encoded and used in language processing. The process, described in Chapter 4 and Chapter 5, will encompass two phases. A theoretical phase where the relevant factors related to features-based representations are described based on the formal machinery developed by HPSG. In Chapter 4, I will show that, as a framework, HPSG can be regarded as constraint-based and lexically driven, in that lexical entries encode phonological, semantic and syntactic information.

In Chapter 5 I will focus on the applied phase. I will evaluate the theoretical assumptions drawn from the previous chapters, in terms of lexical processing models and I will compare them against current models of sentence processing. I will show that a feature-based model of lexical representations is best-suited to explain the interaction between vocabulary size and frequency. Adopting the view that all the information that is relevant to language processing is lexically encoded means that it can be simultaneously used to access words and process syntactic structures. I will then show that this claim ties in with the type of formal representations of HPSG. Furthermore, based on research on the properties and setup of lexical networks (Cancho, 2005; Cancho & Solé, 2001; De Deyne & Storms, 2008a, 2015; Meara, 2006; Steyvers & Tenenbaum, 2005) I will show that the HPSG feature system can be incorporated into the network-based model of lexical representations outlined in Chapter 2. The result is a unified model called the feature-based lexical network model that describes the process of lexical access and the process of activating structural links between words using the same set of features.

To empirically test the second research aim, Chapter 6 and Chapter 7 will describe the experimental stage. A productive on-line task, namely, an Oral Elicited Imitation Test and a Self-paced Reading Task will be used in two different studies. Based on previous research that used frequency to measure lexical effects on syntactic processing, the syntactic structures examined in the experiments are two different types of filler-gap dependencies (Hopp, 2016; Reali & Christiansen, 2007; Tily et al., 2010). The Oral Elicited Imitation Test investigates frequency effects on cleft sentences and the Self-Paced Reading Task investigates the frequency effects on relative clauses. To evaluate the effects of vocabulary size and lexical knowledge, the results of two tasks will be compared to the results of the I-Lex task and frequency will be also used as a predictor in the experimental design. In the first study, I will compare a group of 21 high intermediate L2 speakers to a group of 12 L1 speakers. In the second study, I will again compare a second group of 20 high intermediate L2 speakers to the same group of 12 L1 speakers.

Subsequently, the results of the two experiments will be couched in the formal apparatus developed by HPSG, showing how, in effect, the feature system of HPSG can easily explain the interactions between lexical and syntactic features. This leads to the use of one level of formal representation that accounts for semantic and syntactic effects in sentence processing, with a high degree of precision. The concluding sections of Chapter 6 and Chapter 7 will show how a feature-based lexical network can offer a simple framework to represent the interaction between lexical knowledge and sentence processing. It provides a precise system of lexical features that underlies lexical access and determines how a link between two words can be activated.

Finally, Chapter 8 will offer a formal description of the two key properties of the feature-based lexical network. It will describe in detail how lexical features are accessed and how links between words are activated to form or interpret grammatical strings of language. The formal description of the link activation process will draw on examples from the Self-Paced Reading task, while the formal description of lexical access will draw on examples from the Elicited Oral Imitation task. I will then discuss the limitations of the model and I will propose future directions to implement it.

The present research project develops the first unified model to investigate lexical and syntactic processing in L2 Italian. This is a novel approach that has the potential to improve the knowledge of how Italian is processed by learners and to extend it to language learning and teaching. In addition to the feature-based lexical network model, the study also provides a productive vocabulary task that can be put into use for future studies.

Chapter 2 Literature Review

2.1 Introduction

Most of the research carried out in second language acquisition has drawn a clear-cut line between vocabulary and sentence processing (see Juffs & Rodrigues, 2015, pp. 1-15 for a review). Researchers have considered sentence processing in terms of syntactic processing, assigning vocabulary knowledge an ancillary role: when syntactic processing needs to be revised, word meaning knowledge can be used (e.g., Frazier, 2013). While this has been the predominant view, there are notable exceptions. Usage-based models and constraint-based models have put lexical information at the forefront, but they have not delved much into the details of the various components of lexical information (e.g., Ellis, 2011; Ellis, 2002; MacDonald et al., 1994; McRae et al., 1998; McRae & Matsuki, 2013). In addition, most of the research in constraint-based models has focused on L1 speakers, whereas usage-based models have focused on L2 speakers to a greater extent (e.g., Ellis et al., 2014; Ellis & Ferreira-Junior, 2009). On the contrary, vocabulary research has reached a high level of precision. There are models for lexical processing that explain, in great detail, how individuals can access their L2 lexicon, showing how the meaning and the form of a word interact (de Groot & van Hell, 1998b; Dijkstra et al., 2019; Dijkstra & van Heuven, 2002; Grosjean & Byers-Heinlein, 2018). Research on vocabulary has also shown what kind of information is encoded in lexical items and how it affects lexical processing, focusing on frequency and various aspects of vocabulary knowledge (e.g., Brysbaert et al., 2017, 2018; Cop et al., 2015; Diependaele et al., 2013; Duyck et al., 2008; Gollan et al., 2008, 2011). Extensive research on finding the best approach to measure vocabulary has developed and implemented several tasks that can offer reliable estimates of vocabulary knowledge (e.g., Fitzpatrick & Clenton, 2017; Fitzpatrick & Meara, 2004; Laufer & Nation, 1995, 1999).

This research project aims at exploiting the results of vocabulary research and adding them to models of syntactic processing that incorporate, in their design, lexical information. In doing so, it makes use of a grammar framework that has put lexically encoded information as the key mechanism to model all levels of linguistic representations (e.g., Abeillé & Borsley, 2019; Pollard & Sag, 1994; I. Sag, 2012; I. A. Sag et al., 2003). However, claiming that lexical information needs to be integrated into lexical processing needs to take into account all aspects of vocabulary knowledge, lest it becomes a general claim not adequately supported by the available knowledge (McRae & Matsuki, 2013).

This chapter focuses on vocabulary knowledge and lexical representations in L2 speakers. The first part of the chapter lays out briefly the central features of vocabulary knowledge by looking at the complex interweaving of word features. Evidence comes both from theoretical models (e.g., Fitzpatrick, 2006; Nation, 2013) and experimental research into lexical processing (e.g., Brysbaert et al., 2017; Diependaele et al., 2013). Theoretical models and lexical processing show that fine-grained representations can be associated to words, in terms of their meaning, their form, and the way they combine with other words. The chapter then discusses how vocabulary knowledge might be represented. The model more common in the field is that of the network (Aitchison, 2003) where words are the nodes and the links represent the type of connections that exists between words (Cong & Liu, 2014). The network model is very flexible, and recent research based on corpora and word association tasks showed the useful implications that arise once it is fully adopted (e.g., De Deyne et al., 2019; Meara, 2009; Solé et al., 2010; Wilks & Meara, 2002). There is rising evidence that similar mathematically grounded properties emerge in lexical networks built out of distinct types of connections based on meaning (De Deyne et al., 2013, 2019; De Deyne & Storms, 2008a; Steyvers & Tenenbaum, 2005), phonological form (Vitevitch, 2008), and combinatorial links (Cancho et al., 2004; Cancho & Solé, 2001).

The chapter then looks briefly at another aspect of lexical representations, namely the role played by the L1. Next the chapter focuses on frequency and how it is extremely effective in capturing many factors related to lexical representations and lexical processing. The section shows two other corpus-based measures, namely contextual diversity and semantic diversity, and their relationship with frequency. Frequency is examined considering all various aspects of vocabulary knowledge, appearing to be best suited to account for the reduced availability of lexical features in L2 speakers.

2.2 Complexity of Vocabulary Knowledge

Vocabulary knowledge is a rich and multi-layered construct. One useful way to capture the key properties of a word is to start by examining its form and meaning. An effective way to frame these two properties is to regard form as the orthographic or phonological representation of a word, while meaning can be linked to the set of semantic properties. The form of a word has a degree of complexity that arises from different sources (Schmitt, 2010). First, the way a word is processed needs a different amount of cognitive resources: Processing written words is easier than processing spoken words (Lemhöfer et al., 2008; Schmitt, 2014). Speech is indeed fast and unpredictable by its very nature, so that when a word has been uttered, there is no means to go back to its form. Speech tends also to generate more repetitions and uses a body of formulaic language mainly drawn from high-frequency words (Milton, 2009). In contrast, written language is not meant to be conveyed quickly, and written words can be read more than once if interpretation stalls. Repetitions are also normally avoided; more infrequent words and more specific lexical items are used more than in spoken language (Milton, 2009). In addition to the difference between written and spoken language, the form of a word may undergo morphological processes such as verb inflection and derivational rules, which adds further grammatical and meaning information (Nation, 2013). Morphological features can be especially difficult to process for L2 speakers (e.g. Clahsen & Felser, 2006), as it seems that L2 speakers store and process affixes differently from L1 speakers (e.g. Ullman & Lovelett, 2018; Vanpatten, 2015).

The meaning of words is an even more complex concept to examine. In fact, meaning describes two different properties that words have. First, meaning refers to the way a word is linked to concepts or entities in the real world. Second, the meaning of a word also depends on the associations that arise between the word itself and the context (Nation, 2013). For instance, the words *substance* and *branch* in Example **2.1** and Example **2.2** take quite different meanings in different contexts.

2.1 substance

a) the most important or essential part of something

b) an intoxicating, stimulating, or narcotic chemical, especially an illegal one.

2.2 branch

a) a part of a tree which grows out from the trunk or from a bough.

b) a division or office of a large business or organisation, having a particular function.

The word *substance* has at least two different meanings: the first may be found in philosophy texts, whereas the second may be found in newspapers or police reports. The word *branch*, in addition to its core meaning indicating a real-world object, also has a meaning related to a completely different context of business and marketing. Although the meaning of a word varies along with the context it occurs in, sentence context can facilitate or hinder the interpretation of words in both L1 and L2 speakers (e.g., Kaan, 2014) and reduce or eliminate the effects caused by ambiguous words (e.g., Starreveld et al., 2014; Van Assche et al., 2011). The relation between word meaning and context can even facilitate the interpretation of ambiguous syntactic structures (e.g., Pickering & Gompel, 2006; Spivey-Knowlton & Tanenhaus, 1998; Van Dyke & Johns, 2012).

Both form and meaning are difficult notions to pin down. They work as useful labels, but to reflect the complexity of the information encoded into words, a finer-grained approach is needed. Nation (2013) adds the notion of use to those of meaning and form; use refers to the patterns into which a word occurs, the type of words it occurs with, and the constraints on how it is used. Two factors that Nation puts under the label of use, namely the patterns and type of words with which a word can be used, have been actively studied in word association tasks (see Fitzpatrick & Thwaites, 2020 for review). In word association tasks, participants must provide one or more words to a set of stimulus words. Once the task is completed, the associates produced can be examined to gain information about the type of links between them and the stimulus words. Investigating the type of links that a word forms with other words improves the knowledge of the relations that underlie the mental lexicon as well (De Deyne & Storms, 2015; Fitzpatrick & Thwaites, 2020). The study of the mental lexicon and the relationship that hold amongst words has led to the idea that words interact as nodes within a network. In lexical networks, nodes stand for words, and the links are determined by the multiple types of connections that occur between words (Meara, 2006, 2009).

2.3 Lexicon as a Network

The idea of representing the lexicon as a network has been a longstanding metaphor in SLA studies (e.g. Aitchison, 2003). It is a useful metaphor, for it shows that the mental lexicon is a tightly interconnected system made up of different layers. However, Meara (2006) noted that researchers should be wary of using the network analogy as a straightforward comparison without examining the full set of implications. The properties of a lexical network can be singled out by examining the number of nodes it contains, the number of links, and the distance between nodes, measured as the number of links needed to travel between one node to another (Cancho & Solé, 2001; Cong & Liu, 2014; Meara, 2006; Solé et al., 2010). An L1 English speaker knows thousands of words, between 15000 and 20000 (Schmitt, 2010). When it comes to L2 speakers, the figures are much lower but, as proficiency increases, so does the number of words they know (Schmitt, 2010). Milton (2010) has shown, for instance, that L2 speakers at the intermediate level know more than 3000 words. Given the remarkably high number of words and connections, investigating a lexical network might certainly seem like a complex task. However, a series of studies demonstrated that all lexical networks have certain well-known properties that make them highly organised objects (Cancho, 2005; Cancho et al., 2004; Cancho & Solé, 2001; Cong & Liu, 2014; De Deyne et al., 2013, 2019; De Deyne & Storms, 2008b; Fitzpatrick & Thwaites, 2020; Meara, 2006, 2009; Steyvers & Tenenbaum, 2005; Vitevitch, 2008; Wilks & Meara, 2002).

The first property concerns how many links must be travelled through to get from one node to the other. As a lexical network of a language speaker is made up of many thousands of words, intuitively it seems that many nodes will be at a great distance from one another. However, research has shown that the number of links that connect two randomly picked words in the network is usually extremely small. A series of studies on lexical networks of different languages has consistently shown an average path length around 2.5 and never higher than 10. These figures apply to networks containing syntactic links (Cancho et al., 2004; Cancho & Solé, 2001), meaning links (De Deyne et al., 2013, 2019; De Deyne &

Storms, 2008a; Steyvers & Tenenbaum, 2005) and phonological links (Vitevitch, 2008). Having to use a few links to move across the entire network enhances its efficiency, because short paths afford a quick and highly structured information transfer (Solé et al., 2010)

Another property that arises from language networks concerns their connectivity. All language networks that researchers have studied thus far are extremely interconnected, i.e., highly clustered (Cong & Liu, 2014). Clustering reflects the possibility that two nodes connected to one other node are also mutually connected (De Deyne & Storms, 2008a). Clustering is expressed as a value between 0 and 1. A clustering value of, for instance, 0.25 shows that, approximately 25% of the time, two nodes connected to the same node will also be connected. Clustering may be thought of as the proportion of isolated nodes out of the total number of nodes within the network (Meara, 2009). A high clustering value entails having very few isolated nodes so that the network is highly interconnected. Clustering also captures important meaning relations in lexical networks, in that words whose neighbours are strongly interconnected may be regarded as semantically coherent (De Deyne & Storms, 2015). All networks that display the two properties of being highly clustered and having few links to travel form one node to any of its neighbours are called small-world networks (Watts & Strogatz, 1998).

There is another property that applies to language networks, which is related to the first two. All the links of lexical networks are heterogeneously distributed, with very few nodes that have a much larger number of links and many nodes that have few links. In mathematical models of networks, nodes that have a considerable number of links are called hubs (Cong & Liu, 2014; Solé et al., 2010). Hubs appear because new nodes are attached to exiting nodes in a probabilistic fashion. The higher the number of links of a node, the higher the probability that a word is attached to one of those already existing links: this process is called preferential attachment (Barabási, 2002; Solé et al., 2010). Hubs serve the purpose of speeding up navigating throughout the links, working as shortcuts between nodes, thus playing a key role in enhancing connectivity across the network (Solé et al., 2010).

In syntactic networks, that is, networks formed by syntactic relations amongst words, hubs correspond to functional words such as prepositions, articles, and conjunctions (Solé et al., 2010). In semantic networks (that is networks made up of meaning relations) the characteristics of hubs are not as easy to single out. For example, De Deyne & Storms (2015) reported the ten most frequent hubs of two large-scale word association norms, for English and Dutch. Hubs for the English word association network in De Deyne & Storms (2015) are *money, food, water, love, word, car, music, time, happy* and *green*. Those for the Dutch network are *water, food, money, tasty, music, car, pain, sea, beautiful,* and *warm* (De Deyne & Storms, 2015, p. 476). The two lists of hubs show a degree of similarity, sharing half of their words, suggesting that hubs "might indicate that certain properties are universally more central in the human semantic system" (De Deyne & Storms, 2015, p. 477)

A key property related to language networks, is the mathematical relation between the number of links of a word and its frequency. De Deyne & Storms (2008a) found that the most frequent words have the highest number of links to other words in the network and are acquired earlier. This property is not entirely surprising, as it is well-known that the most frequent words tend to appear extensively over many types of different contexts. Several studies have shown a relationship between word frequency and number of links, focusing on lexical networks based on syntactic, semantic, and phonological connections (e.g., Cancho, 2005; Cancho et al., 2004; De Deyne et al., 2013, 2019; Steyvers & Tenenbaum, 2005; Vitevitch, 2008).

The meaning of words may also be explained as an emergent property of the network. In fact, words can be viewed as lexical entries in a thesaurus whose meaning is determined by the set of other words associated with it (Steyvers & Tenenbaum, 2005). Adding a new word to the network differentiates the existing meaning representations, since it introduces a meaning variation to the meaning of the existing pattern of nodes and its related concept. This process is called semantic differentiation (Steyvers & Tenenbaum, 2005).

When examined through precise mathematical modelling, lexical networks show some precise properties about their organisation based on the number of nodes and the number of links that connect them. In a series of studies, Paul Meara (Meara, 2006, 2009; Wilks & Meara, 2002) has moved along this line of research, examining the overall structure of lexical networks in terms of their degree of density. Density can be defined as the proportion between the actual links in the network and the number of links if all nodes in the network were connected (Meara, 2009). Examining the mental lexicon of language users through the concept of density offers a new insight into the differences between L1 and L2 speakers (Meara, 2006, 2009). A well-established fact about the vocabulary of L2 speakers is that it is smaller (e.g. Milton, 2010; Nation, 2013; Schmitt, 2010). The notion applies to the lexical networks of L2 speakers, which have fewer links and fewer nodes than those of L1 speakers (Meara, 2009; Wilks & Meara, 2002). Because they know fewer words, lexical networks of L2 speakers are expected to exhibit fewer links and fewer nodes, namely, to be less connected than those of the L1 speakers (Meara 2009).

Density has also been employed to frame the gap between productive and receptive vocabulary knowledge that has consistently been noted with regard to L2 speakers (Meara, 2009). Although the extent of the gap is not quite clear, research shows that the receptive vocabulary of L2 speakers is larger than their productive one (e.g. Schmitt, 2019). This difference between productive and receptive vocabulary knowledge may be reshaped in terms of lexical density. A word which has fewer links is likely to be difficult to access and be known only receptively, whereas a word with numerous links is likely to be easier to access and to be used productively (Meara, 2006). Network density has also been used to analyse the difference between vocabulary breath and vocabulary depth. While breadth can be easily defined as the number of words known to a learner, the concept of vocabulary depth is more elusive. Depth consists of what a learner knows about words, though it is not straightforward to model what the components are that conspire to increase it (Schmitt, 2014). Using density as a foothold might be a better way to explain depth: the more connections a single word has in the network, the deeper is the knowledge of the word associated to it (Meara, 2009, pp. 65-85). If, as shown by Steyvers & Tenenbaum (2005), the meaning of a concept may be represented as patterns of connection between words, numerous links between a word as its neighbours correspond to deeper meaning representation.

2.4 The Role of the First Language

Having to cope with linguistic representations of two (or more) languages, L2 speakers need to be able to avoid mixing them. Having to navigate through a lexicon that encompasses words from both languages, L2 speakers must be particularly deft at keeping the language not in use under control (Bialystok et al., 2008). However, despite the high level of control, the role of the first language remains crucial in lexical representations of L2 speakers. Lexical errors can often be easily explained through the influence of the L1, and translation from L1 greatly improves the ability of L2 speakers to recall L2 words (e.g. Schmitt, 2019). The similarity of the L1 and the L2 at the phonological, morphological, and semantic level can facilitate the learning of L2 words (e.g. de Groot & Brink, 2010; Spinelli et al., 2020).

The most compelling evidence about the importance of the L1 comes from studies that have shown that L1 and L2 lexical representations are subserved by a common system of lexical features. A considerable body of research has demonstrated that bilingual lexical representations are not accessed independently (see Tokowicz, 2014, for a review). The lexicon of L2 speakers is not a compound which contains two separate sets of lexical items, whereby a set contains L2 words only and a set contains L1 words only. As a matter of fact, research has shown that the lexical features of the L1 and the L2 are not stored separately in the mental lexicon (e.g., Dijkstra, 2003; Grosjean & Li, 2013). The effects of cross-linguistic overlap have been labelled with the term non-selectivity. Non-selectivity means that L2 speakers cannot choose to access lexical features of one language only, in fact features of the language not in use receive a certain amount of activation as well. As a result, despite the overall ability in preventing that one language affects the use of the other, L2 speakers cannot fully switch off one of them (e.g., de Groot & van Hell, 1998; Dijkstra et al., 2019; Dijkstra & van Heuven, 2002).

The extent to which one of the two languages can be selected depends on the purposes, the domains and the requirements under which L2 speakers communicate. An L2 speaker, for instance might be in an entirely bilingual mode, they might need to switch from one language to the other or use their L2 in specific circumstances, e.g. at work only (Grosjean & Li, 2013, p 5-25). The notion has been formalised in the Complementary Principle, according to which there is a continuum along which selecting a word may shift from effortless to difficult according to communicative purposes (Grosjean, 2008).

Although L2 speakers can be in various modes when they use either language, the cross-linguistic overlap characterises the way L2 lexical representations are organised. It affects lexical and syntactic processing in an automatic way, ranging from the level of lexical features to the level of words (Bernolet et al., 2007, 2013; Chambers & Cooke, 2009; Duyck et al., 2007; Elston-Gúttler & Friederici, 2005; Hartsuiker et al., 2004; Hartsuiker & Bernolet, 2017; Libben & Titone, 2009; Poort & Rodd, 2019; Schoonbaert et al., 2007; Schwartz & Kroll, 2006a; Spivey & Marian, 1999; Starreveld et al., 2014; Van Assche et al., 2011, 2013). Interlexical neighbours, i.e. words which share a wide proportion of phonemes, can slow down lexical processing compared to words that did not have any phonological overlap (e.g., Chambers & Cooke, 2009; Spivey & Marian, 1999). Studies that have examined homographs, i.e. words that share their form but not their meaning, have also found similar slowdowns in lexical processing, in terms of reading times and gaze fixations (Dijkstra & van Heuven, 2002; Lemhöfer & Dijkstra, 2004; Libben & Titone, 2009; Schwartz & Kroll, 2006a). Some studies have shown facilitating effects of cognates, i.e. words that share their form and their meaning. Cognates are read faster than control words, not only when the task requires to process single words, but also when cognates are embedded in sentence context (Lemhöfer & Dijkstra, 2004; Libben & Titone, 2009; Poort & Rodd, 2019; Starreveld et al., 2014; Van Assche et al., 2011). Cognates can speed up lexical access, thus facilitating processing in complex syntactic dependencies (Hopp, 2017; Miller, 2014a). Apart from cognates, non-selectivity has been shown to have a significant impact on syntactic representations as well. Hartsuiker et al. (2004), for example, found a cross-linguistic priming effect for passive sentences, while Schoonbaert et al. (2007) showed the same effects for the prepositional dative and double object construction (*give something to someone* and *give someone something*). Both studies showed that the priming effects increased when translation equivalents were used. Effects of syntactic shared representations have also been found for several types of constructions and seem to be lexically driven; syntactic patterns are shared based on the word around which they constructed (e.g., Bernolet et al., 2013; Branigan & Pickering, 2017; Cai et al., 2011; Hartsuiker & Bernolet, 2017).

In short, the fact that the lexical representations of L2 speakers are shared between two languages means that when lexical features overlap, they are likely to be stored together. As a consequence, the role of the L1, whose lexical representations are stronger or more "entrenched" (Diependaele et al., 2013), becomes more relevant. Thus, the effects of the L1 are mediated by the fact that L2 speakers organise their mental lexicon by storing lexical features in a non-selective way. It is crucial to view the role of the L1 as caused by the specific quality of lexical representations in L2 speakers rather than by mere interference. That the peculiar character of the lexical representations of the L2 speakers is what determines the effects of the L1 is supported by the presence of a reverse, albeit weaker effect from the L2 to the L1 (e.g. Bernolet et al., 2013; Dijkstra & van Heuven, 2002; Whitford & Titone, 2012). If the L1 interfered, then it would be hard to explain why the L2 can also affect the lexical processing of the L1.

2.5 Frequency and the Lexicon

The factor which has been mostly used to describe vocabulary knowledge is frequency (e.g., Ellis, 1997; Laufer & Nation, 1999). As noted by Schmitt (2010), frequency "permeates all aspects of vocabulary behaviour. It is arguably the single most important characteristic of lexis that researchers must address." (p. 63).

There are many reasons that underlie the claim. First, frequency bears on the organisation of the lexicon: In normal language usage, some types of word occur far more often than others. Function words such as prepositions, articles, and pronouns have extremely high frequency compared to content words. Function words are also hubs within lexical networks (see 2.3). Furthermore, function words mostly serve the purpose of providing scaffolding for the sentence, whereas content words with more specific meanings tend to convey the main message (Milton, 2009). The link between frequency and word meaning extends beyond the difference between content and function words. Words with a general meaning tend to be more frequent, while words with a specific meaning tend to be less frequent and used in narrower contexts. Schmitt, for instance, compared the adjective *odd* to the less frequent *eccentric*. Although the two words have very similar meaning, *odd* is used to describe anything that is out of the usual or normal, whereas the less frequent *eccentric* is commonly used only to describe a person or their behaviour (Schmitt, 2010, p. 63).

A similar relationship between frequency and meaning applies to Italian, as shown in Table 2.1. Comparing verbs like *avere* (*to have*) or *fare* (*to do*) and adjectives like *buono* (*good*) and *cattivo* (*bad*) with the more specific synonyms like *possedere* (*to possess*) or *realizzare* (*to realise)* and *amabile* (*lovely*) or *malvagio* (*evil*) shows a striking difference in terms of frequency. Frequency was checked in the Corpus Coris, which is a corpus of written Italian, containing 150 million words drawn from various types of written texts (Rossini Favretti, 2000).

Table 2.1

Frequency Difference Between Low-content Words (General Meaning) and more Specific Synonyms (Specific Meaning)

General Meaning				Specific Meaning			
Avere (have)	fare (do)	buono (good)	cattivo (bad)	possedere (possess)	realizzare (realise)	amabile (lovely)	malvagio (evil)
1,525,926	548,148	55,857	8,722	10,097	23,312	482	872

Note. Frequency figures were drawn from the Corpus Coris.

The figures show the same interaction between meaning and frequency as found by Schmitt (2010). Italian words with a specific meaning are less frequent and tend to be used in narrower contexts. For example, the adjective *amabile* (*lovely*), when used to describe a person, means that they inspire fondness and affection, while in the same context, *buono* (*good*) refers, in Italian, to a general inclination to do good deeds.

A key factor related to frequency is the notion of coverage, which is the proportion of words that are likely to be known in a text (Milton, 2009; Nation, 2013). Coverage is extremely wide for the most frequent words (Milton, 2009). The relation between frequency and coverage depends on the way the message is conveyed. Spoken language tends to rely more on high-frequency words and formulaic language, whereas written language tends to incorporate more infrequent and specific words (Milton, 2009). This suggests that coverage may require slightly fewer words in spoken language. For example, Schmitt (2010) showed that, in English, the most frequent 3000 words are needed to attain the 95% coverage. It means that if a person is engaging in a conversation and knows those most frequent 3000 words, they will be likely to know 95% of the words (Milton, 2009; Nation, 2013). Figures concerning written language are, expectedly, fairly different: estimates for English, based on the British National Corpus (BNC), suggest that the knowledge of around 8,000 to 9,000 words ensures a degree of coverage of about 98% (Nation, 2006). Data for frequency and coverage in Italian are given in Table 2.2 (De Mauro, 2016). Looking at written language with the most frequent 1000 words, the coverage is around 80%, while the first 5000 words guarantee a 95% of coverage: these figures are lower but in line with the English ones.

Table 2.2

Frequency and Coverage for the 5000 most frequent Italian words

Frequency Band	Written	Spoken
1000	80%	86%
1-2000	84%	91%
1-5000	93%	95%

Note. Adapted from the data available in De Mauro (2016)

The figures for the spoken language present a pattern that differs from that of English as well. To reach the same level of 95% of coverage, Italian needs some 5000 words, a much higher level than the English threshold. Despite some degree of difference, data from English and Italian show that, in terms of language learning, gaining around 95% of coverage, requires to learn at least 5000 words.

2.6 Frequency Effects on L2 Lexical Processing

Frequency is also a main factor in marking the difference between L1 and L2 speakers lexical processing. Although all language users are faster at processing high-frequency words than they are with low-frequency words across a number of different tasks, frequency effects are systematically larger for L2 speakers than for L1 speakers (e.g., Brysbaert et al., 2018; Duyck et al., 2008; Gollan et al., 2011). An explanation for the strongest frequency effects in L2 speakers is that they are caused by a lower rate of exposure to the L2 (Brysbaert et al., 2017; Diependaele et al., 2013). Due to lower exposure, L2 speakers have weaker lexical representations, and this is what leads them to show more pronounced frequency effects. The Lexical Entrenchment Hypothesis (Diependaele et al., 2013) argues that vocabulary size tends to be a significant predictor for frequency effects, given the strong relationship between lexical representations and frequency. For example, Diependaele et al. (2013) reanalysed reaction times from a large-scale study on lexical recognition of high- and low-frequency words (Lemhöfer et al., 2008). They found that the interaction between frequency and vocabulary size was significant in both the L1 and L2 groups. Importantly, when the scores in the vocabulary knowledge task were controlled, the effects of frequency on the L2 speakers were not significantly greater than those of the L1 speakers. Brysbaert et al. (2017) analysed reaction times in lexical processing of high- and low-frequency words. They found that the reaction times of the L2 speakers showed a high correlation with those of the L1 group, although the L2 speakers were slower and less accurate in lexical decision rates. Like in Diependaele et al. (2013), when the vocabulary size was added to the statistical analysis, there was no interaction between frequency and language in both the L1 and the L2 groups.

Additionally, a complex interaction occurred between vocabulary knowledge and frequency. With high-frequency words or low-frequency words, no significant effect of vocabulary size was found. The effect was indeed significant only for the stimulus words that fell in the middle of the frequency spectrum.

The studies described above show that there is a relationship between frequency and vocabulary size. When vocabulary size increases, the effects of frequency decrease, although the relationship is not simply linear. Brysbaert et al. (2018) discuss a model based on reaction times in lexical decision tasks. The model predicts that low-frequency words show the most prominent effects on individuals with a low vocabulary size. In individuals with a high vocabulary size, the effects of low-frequency words are less marked and tend to reduce quickly as soon as frequency starts rising. The model entails that L2 speakers, who normally have a smaller vocabulary size, will be more sensitive to the effects of low-frequency words.

An explanation as to why frequency impacts L2 speakers to a greater extent has been put forward by Gollan and colleagues and is based on language exposure (Gollan et al., 2008, 2011). Gollan et al. (2008) compared L1 and L2 speakers of English. Reaction times were measured on a picture naming task. The results showed that the L2 speakers had, in general, slower reaction times and this interacted with frequency, since the difference between the L1 and L2 speakers was more pronounced for low frequency words. Gollan et al. (2008) interpreted the results as caused by the fact that L2 speakers tend to use their second language less often than L1 speakers. According to them L2 speakers must split their time of language use between the L1 and the L2. As they tend to speak one language at a time, at least in many of the communicative events, this reduces the amount of exposure to the lexical items of both languages. Hence, L2 knowledge of frequency patterns ends up lagging behind that of L1 speakers: this assumption is called Frequency Lag Hypothesis (Gollan et al., 2008, 2011). The hypothesis ties in with the age effects reported in the study. When age was used as a predictor in the study, frequency effects on L2 speakers were reduced in the eldest group to the point that they produced equivalent frequency effects both in the L1 and the L2.

Gollan et al. (2011) examined the effects of the Frequency Lag Hypothesis on production and comprehension in two groups of English L2 speakers along with a control group of English L1 speakers. The study included a first experiment on production and a second on comprehension. To gauge the role played by context, high- and low-frequency words had to be processed, respectively, in isolation or in embedded in high- and low-constraint sentences. The difference between the two types of sentences is that, in high-constraint sentences, the context strongly primes the reading of the target word; whereas, in low-constraint sentences, the context is not skewed towards any specific interpretation. In the production experiment, results showed a significant effect of context and frequency: both groups were faster at producing high frequency words in high constraint sentences, but the effects were larger for the L2 speakers. Furthermore, frequency and context interacted significantly as L2 speakers were slower than L1 speakers in producing low-frequency words in low-constraint sentences. The results of the comprehension experiment were measured in an eye-tracking design. The results showed shorter gaze durations on high-frequency words and on words embedded in high constraint sentences. The effects were greater in the L2 speakers group compared to the L1 speakers, but unlike the production experiment, no context-frequency interaction appeared. The results suggest that lexical access is driven by frequency in comprehension, while in production it is affected by the interaction of frequency effects and semantic constraints.

The studies considered so far show that frequency interacts with other several factors, such as exposure, the type of communicative process (i.e. comprehension and production) and the quality of lexical representations. As I will explain in the next section, it is not surprising that frequency normally correlates with several other word properties such as word length and meaning similarity. It is also a reliable predictor of reaction times in various lexical processing tasks, such as word naming or lexical decision (Brysbaert et al., 2018; De Deyne et al., 2019; Diependaele et al., 2013). Frequency is also a predictor of syntactic processing. According to the Lexical Bottleneck Hypothesis (Hopp, 2014, 2016, 2017) the difficulties that L2 speakers show in lexical access affect their ability to build correct syntactic structures. I will discuss this hypothesis in Chapter 6, where I will investigate the effects of frequency on L2 Italian syntactic processing.

2.6.1 *Factors Related to Frequency*

Despite having a key role in accounting for lexical processing, there are other factors that interact with frequency. Brysbaert et al. (2018) notes that, although it is true that frequency is linked to vocabulary learning, it is not always the case that all high-frequency words are learned first. Second there are many low-frequency words that are well known and generate reaction times as fast as those of high-frequency words. For example, from the very start of language courses, L2 speakers of Italian learn words like *lavagna* (*board*) or *pennarello* (*felt-tip pen*). They are words that refers to objects relevant to teaching but are not frequent. According to the frequency list published by De Mauro (2016), for example, the two words are not in the first 5000 most frequent Italian lemmas (see Chapter 3, Section 3.8.4) for a description of the list.

Furthermore, frequency effects build on the assumption that low-frequency words are met more rarely, and so are more difficult to learn and to process. Although this is true in many cases, not every encounter with a word has the same impact on learning processes (Brysbaert et al., 2018). Some words may simply be easier to learn than others despite their frequency because their referents can be available to speakers for cultural reasons. This is captured by word prevalence, which refers to the number of people who know a word (Brysbaert et al., 2016, 2019; Keuleers et al., 2015). As a measure of lexical knowledge, word prevalence reflects the fact that some words not very frequent in corpora, like *toolbar* or *screenshot*, are known better than other words in the same frequency band (examples from Brysbaert et al., 2019). Word prevalence has been shown to explain some 7% of effects on response times in lexical decision studies (Brysbaert et al., 2016).

In addition, two other factors have showed to be reliable predictors in explaining variance in lexical processing yielding equal or even higher results than frequency. One factor is contextual diversity, that is the number of distinct texts in a corpus in which a word occurs. If a word occurs in a variety of texts it is likely to be widely known by language users. In word naming and lexical decision tasks, when compared to frequency, contextual diversity yields a significant improvement in explaining the variance in the results of L1 and L2 speakers' performances (Adelman et al., 2006; Hamrick & Pandža, 2019).

Another factor is semantic diversity, which measures how many types of context a word appears in. If a word has an elevated level of semantic diversity, it appears in a variety of texts and in different contexts. For example, the word *bank* can occur in contexts and texts linked to finance and economics or in other contexts where river banks are involved (Johns et al., 2016, p. 2). A word with a high degree of semantic diversity is more likely to be needed in many different communicative events, thus increasing the possibility of becoming well known in many of its meanings. On the contrary, if a word appears in a variety of texts that are remarkably similar in terms semantic content, that word is less likely to become needed and known, for it is tightly linked to the same communicative events (Johns et al., 2016).

Semantic diversity has shown to be a better predictor then frequency and contextual diversity in lexical access for both L1and L2 speakers (Hamrick & Pandža, 2019; Johns et al., 2016). For L2 speakers, semantic diversity can explain up to 30% of the variance in addition to that accounted for by frequency and is significantly correlated with age. Specifically, in a lexical decision task, older bilinguals benefited more if the stimulus words have a large semantic diversity, indicating that semantic diversity depends on how familiarity with the language develops over time (Jones et al., 2012).

Although they may outweigh frequency as a predictor, when semantic diversity and contextual diversity are compared to frequency, they are always highly correlated (Adelman et al., 2006; Hamrick & Pandža, 2019). If a word has a high corpus frequency, it is likely to appear in many diverse contexts. The two properties are related, although their statistical effects can be separated. Therefore, the high correlation between the three predictors may well show the simple fact that a word needs to be frequent to appear in a variety of texts. In this respect McDonald & Shillcock (2001) noted that:

> [f]requency of occurrence provides an uninformative model of the contextual distribution of any word; given no other information (treating the corpus as a randomly ordered "bag" of words), the expected probability of a particular context word occurring in the immediate lexical environment of any target word is simply the relative frequency of that context word. (McDonald & Shillcock, 2001. p. 306)

Furthermore, frequency is tightly linked to the other factors that underpin vocabulary knowledge in L2 speakers. In Section 2.3, the analysis of the properties of lexical networks showed that the more frequent words are the ones that have more links. Steyvers & Tenenbaum (2005) argue that the meaning of a word depends on its number of connections with other words in the lexical network. Therefore, a high-frequency word appears in numerous and diverse contexts, and it is likely to have a complex meaning and a large number of links. This explains, in terms of network density, the relationship between semantic and contextual diversity and frequency.

The relationship between frequency and non-selectivity might seem more blurred. At first glance, the effects caused by non-selectivity might seem unrelated to a lack of knowledge about L2 frequency patterns. However, it turns out that, when added as a predictor, frequency also affects the effects of shared representations. Homographs that have a higher frequency in one language yield larger inhibitory effects in the other language (Dijkstra & van Heuven, 2002). The assumption that underlies these results is that high-frequency words would compete for selection with more strength than low-frequency words. Furthermore, L2 speakers have less precise subjective frequency patterns than L1 speakers. Given that they have less exposure to the L2, their knowledge of word frequency is fragmentary so that their lexical representations tend to be activated more slowly (Dijkstra et al., 2019; Dijkstra & van Heuven, 2002; Gollan et al., 2011).

Looking more thoroughly at the effects of contextual and semantic diversity, there is one important aspect to consider. Both factors were evaluated in terms of the improvement they were able to add to the fit of statistical models. However, studies on contextual and semantic diversity investigated the processing of single words. They are silent about the role played by sentence context. When studies combine the effects of frequency with the effects generated by the sentence context, a significant interaction appears (Gollan et al., 2011). Importantly, these effects were found in studies that did not investigate lexical frequency *per se*, but its role in limiting the processing of certain type of syntactic contexts in L1 speakers (e.g. Reali & Christiansen, 2007; Tily et al., 2010) and L2 speakers (Ellis et al., 2014; Hopp, 2016). When measurement on lexical processing is integrated into sentence context, frequency seems to be well-suited to bridge the gap between measurements at the level of lexical processing and of sentence processing.

In summary, frequency has consistently shown a close relationship with other more specific measures. This is because, overall, frequency reduces the number of lexical features to which the processing system can have access. However, some lexical features tend to be more informative than others and affect the experimental task to a different extent. In this respect, frequency is not informative about which lexical features will be harder to access, as this depends on how a lexical item is going to be exploited in a specific task. Hence, frequency cannot allow one to make specific predictions about which features are going to be available and their effect in the experiment. If a task aims at measuring the effects caused by specific features, then there are other factors that can be more informative than frequency. If the task aims to estimate how a reduced pool of lexical features affects processing overall, then frequency is still the most reliable predictor. This claim will be discussed in Chapter 5 and in the experiments described in Chapter 6 andChapter 7.

2.7 Conclusions

The chapter has shown the main factors that underpin lexical knowledge and how they interact. As shown in Sections 2.5 and 2.6, frequency is linked to vocabulary knowledge, and it underlies the process through which lexical networks are built. Hence, the next chapter is going to examine the effectiveness of frequency as proxy for vocabulary knowledge (e.g., Laufer & Nation, 1999). The chapter will show how, based on these assumptions, a brand-new measurement tool for Italian has been developed and validated. As noted at the beginning of the chapter, this research project aims to bring together the results of vocabulary research and incorporate them into models of syntactic processing. Therefore, measuring lexical knowledge is crucial for investigating the mechanisms that drive words to combine into larger constructions.

Chapter 3 Development of I-lex

3.1 Introduction

In the previous chapter I described the multi-layered nature of vocabulary knowledge and how frequency is interwoven with all the various components that are encoded into words. To compare vocabulary knowledge to other language phenomena, it is important to have a tool which is able to measure its various components. Thus, the present chapter aims to develop and validate a new productive vocabulary measurement task for L2 Italian and to subsequently use it as a tool to weigh lexical effects on language processing. This is going to be the focus of Chapter 6 and Chapter 7 where vocabulary knowledge and frequency are going to be used as variables to investigate syntactic processing.

As I will show in the chapter, productive vocabulary knowledge taps into the two processes of recalling the meaning and encoding it into a specific lexical item (e.g., Fitzpatrick & Clenton, 2017; Levelt & Meyer, 1999; Nation, 2013; Schmitt, 2019). Therefore, productive vocabulary tasks are more informative than receptive tasks, which are based only on the process of recognising a word (Schmitt, 2019). For this reason, to be able to capture many more aspects related to vocabulary knowledge, I opted for the development and implementation of a new productive vocabulary measurement task for L2 Italian named I-Lex.

The task is, to my knowledge, the first attempt to construct a productive vocabulary task for L2 Italian. At the moment, there are two receptive vocabulary measurements that have been validated and have been shown to be reliable. Stella et al. (2000) developed a task adapted from the Peabody Picture Vocabulary Test-Revised (Dunn et al., 1965). The test has been consistently used in L2 Italian (e.g., Spinelli et al., 2020). A more recent receptive vocabulary task has been developed and validated by Amenta et al. (2021) based on

LexTALE (Lemhöfer & Broersma, 2012). In a validation study, the task was shown to be reliable in distinguishing between L1 and L2 speakers and in showing a clear relationship between the score and L2 proficiency (Amenta et al., 2021).

First, the chapter explains how, based on the analysis carried out in Chapter 2, frequency reliably reflects various aspects of vocabulary knowledge. Then the chapter deals with the nature of vocabulary knowledge and its reliability as a predictor of several linguistic abilities (e.g., Andringa et al., 2012; Lemhöfer & Broersma, 2012; Milton, 2009). Then the discussion focuses on the differences between productive or receptive knowledge, describing the problems that need to be addressed to develop a reliable measurement tool of productive vocabulary knowledge. The research conducted so far has shown that for a productive measurement to be reliable, it should generate many words with a large degree of lexical diversity. The focus shifts to an existing English productive vocabulary task, Lex30 which, in a series of studies, was shown to meet both requirements (e.g., Fitzpatrick & Clenton, 2010, 2017). The chapter discusses the design of Lex30, focusing on the key role that frequency plays in the creation and the validation process of Lex 30.

The main part of the chapter shows how the novel productive measurement for Italian, I-Lex, was designed, developed, and validated with L1 and L2 speakers. The task is based on the word association format and the frequency-based measurement criterion used in Lex30. The results of I-Lex will be discussed in terms of productive vocabulary knowledge and the effects of proficiency. Some simple measurements related to the lexical networks generated by participants' responses will also be analysed and explained.

3.2 Frequency and Vocabulary Measurement

As I explained in Chapter 2, frequency is interwoven with all other main factors of lexical representations. It shows also a high level of correlation with contextual and semantic diversity and it appears to be related to non-selectivity (e.g., Dijkstra & van Heuven, 2002). Furthermore, frequency and vocabulary knowledge interact, in that frequency effects depend on the lexical knowledge of L2 speakers (Brysbaert et al., 2017; Diependaele et al., 2013). Therefore, frequency can be regarded as a reliable measure of which lexical features are more readily available to L2 speakers. With high-frequency words, many lexical features become easily available, so that lexical processing is facilitated overall. On the contrary, with low-frequency words, lexical features are more difficult to access, and lexical processing occurs more slowly (Brysbaert et al., 2017; Diependaele et al., 2013; Gollan et al., 2011).

One aspect of frequency that I have described in Chapter 2 is particularly important here, namely, the relationship between frequency and the structure of lexical networks. Lexical networks contain frequent words that have numerous links and are densely connected, and infrequent words that are more sparsely connected but not distant form them (De Deyne & Storms, 2008a; Meara, 2009). In addition, the growth of vocabulary knowledge can be viewed as the process of attaching novel words to the most frequent and connected words within the lexical network (De Deyne & Storms, 2008a; Steyvers & Tenenbaum, 2005). Since infrequent words are added to the network to differentiate meaning representations, they can be regarded as "advanced" words that reflect lexical sophistication. As Laufer & Nation (1999) pointed out, "the true lexical quality of a piece of writing is determined by the proportion of all the other words at the more advanced frequency levels" (p. 318). Therefore, having a reliable estimate of the ratio between infrequent and frequent words shows whether the lexical knowledge and the lexical networks that shape it are becoming more complex. For this reason, the main aim that many vocabulary measurements pursue is to generate a sufficient number of infrequent words (e.g., Laufer & Nation, 1995, 1999; Meara, 2009; Meara & Fitzpatrick, 2000).

The importance of frequency and its tight relationship with vocabulary knowledge raises the question of what kind of vocabulary task is better suited to capture the various properties encoded in lexical features. A productive vocabulary task seems to be the most suitable format. Producing a word relies on wider pool of features as the same concept can be represented through several distinct lexical items, each with its own set of word-form, meaning, and combinatorial features (e.g., Levelt & Meyer, 1999). In a receptive task, the feature activation path is determined by word-form features to a great extent, and only one specific lexical item is strongly activated (Schmitt, 2019). The key notion is that the effects of feature availability are more prominent in productive vocabulary tasks. As Schmitt (2014) notes:

> Overall, as the frequency level decreases, the recognition–recall gap increases. This means that learners are more likely to have both form recognition and form recall mastery at the higher frequencies (i.e., smaller gap) and less likely to have form recall mastery at the lower frequency levels (i.e., form recognition mastery only and a wider gap). (p. 924)

Since frequency is related to all lexical characteristics, it is also expected to relate to other aspects of language proficiency in the same way as vocabulary knowledge. Therefore, frequency reflects that L2 speakers who have a large vocabulary are more proficient in a wide range of language skills than learners who have a smaller vocabulary knowledge (Schmitt, 2010). Therefore, before examining how to build a reliable productive vocabulary measurement task based on frequency, it is important to review the importance of vocabulary knowledge as a major component of almost all aspects of L2 proficiency.

3.3 Vocabulary Knowledge and Language Proficiency

When using language, it is necessary to access words that are stored in the mental lexicon and use the information encoded in them to link form features to meaning features (see Evans & Levinson, 2012, for review). At this point, the word is ready to be interpreted or used, for instance, in a written text or during an ordinary conversation. Knowing the form of a word, what a word means, how it can be used, and the words it is usually associated with, is essential to communicate efficiently (Milton, 2009; Schmitt, 2010). Since vocabulary knowledge enables learners to achieve a good level of language ability in various areas of language use, it is a useful tool in predicting other proficiency measures (e.g., Laufer & Nation, 1999; Milton, 2010, 2013)

The skill that has traditionally shown the highest degree of correlation with vocabulary knowledge is reading. Correlations found between vocabulary knowledge and reading skills are generally high, ranging from 0.50 to 0.85 (Milton, 2013; Stæhr, 2008). Some studies have looked at the relationship between reading and vocabulary knowledge considering two different parameters. One is vocabulary breadth that indicates the size of vocabulary knowledge; the other is vocabulary depth, a measure of how well a word is known, for instance, the knowledge of its collocations, its grammatical patterns and its register (Schmitt, 2014). Although both measures showed significant correlations with reading comprehension, vocabulary depth can provide an explanation for additional variance (Qian & Schedl, 2004). An important link between reading and vocabulary depth is that vocabulary depth is crucial in how successfully L2 speakers can infer the meanings of unknown words in a reading task. Indeed, strong lexical representations seem able to back better lexical inferencing skills and improve reading comprehension (e.g., Prior et al., 2014; Schmitt, 2010). There is also a strong link between writing and vocabulary knowledge, although it is usually less strong in comparison to reading. Examining EFL learners, Stæhr (2008) found that vocabulary knowledge accounted for around 50% of the variance in writing scores. However, while some studies have shown an equally strong correlation (Shi & Qian, 2012) others studies have found smaller figures, showing that vocabulary only predicts some 25% of writing scores (Qian & Lin, 2019).

Reading and writing are the skills most often reflected in vocabulary measurement. The fact that speaking and listening unfold in real time can make them more difficult to deal with for L2 speakers. This does not rule out the possibility that aural skills can also be captured. Some studies have indeed found that vocabulary knowledge is able to explain variance in listening comprehension tasks (e.g. Andringa et al., 2012; Stæhr, 2008, 2009; Wang & Treffers-Daller, 2017). Some results showed a significant correlation between listening and vocabulary size (e.g. Andringa et al., 2012; Qian & Lin, 2019; Wang & Treffers-Daller, 2017), while others showed a correlation between listening and vocabulary depth (Stæhr, 2009).

The relation between vocabulary knowledge and speaking has also shown that vocabulary can be a reliable predictor of fluency and accuracy. For example, word association tasks have shown a significant relationship with oral fluency (Clenton et al., 2020; Uchihara et al., 2020; Uchihara & Saito, 2019). Vocabulary tasks presented in an aural format, whereby individuals recognize words after hearing them, have also proven to be a reliable predictor of speaking skills (Milton et al., 2010). Focusing on lexical depth, Crossley et al. (2014) found that collocation accuracy was able to explain more than 80% of variance in the scores of spoken (and written) texts produced by learners with different levels of proficiency. Comparable results concerning oral production, were found by Saito et al. (2016). In their study, the level of comprehension by L2 speakers was correlated with several lexical variables such as the use of concrete and complex polysemous words, measures that are indicative of vocabulary depth.

The studies examined so far investigated the relationship between vocabulary knowledge and tasks designed to measure the four language skills. Furthermore, vocabulary knowledge is a reliable predictor for on-line tasks that tap into specific aspects of lexical processing. Vocabulary knowledge correlates with faster reaction times in lexical decision tasks and word naming (Brysbaert et al., 2017; Diependaele et al., 2013; Lemhöfer & Dijkstra, 2004; Mainz et al., 2017; Yap et al., 2012). Overall, vocabulary knowledge can give reliable estimates of other types of language skills and can capture complex aspects of linguistic knowledge.

3.4 Measuring Productive Vocabulary Knowledge

Measurement of productive vocabulary knowledge raises several problems, especially in comparison to the measurement of receptive vocabulary knowledge (Milton, 2009). Some of the problems arise from the qualitative difference that exists between productive and receptive vocabulary knowledge. Productive knowledge seems to depend on different processes and is harder to achieve than receptive knowledge (Nation, 2013).

Receptive and productive vocabulary knowledge needs two distinct types of cognitive processes. Receptive knowledge builds on the recognition process that occurs when a language user is presented with the form of a word and must recognise the meaning of it. Productive vocabulary knowledge builds on the process of providing both the meaning and the form of a target word based on a stimulus designed to elicit it (Nation, 2013; Schmitt, 2010). As formalised in the Levelt model (Levelt & Meyer, 1999), producing a word involves two cognitive operations, carried out by two distinct components. The first component, the conceptualiser, underpins the process of recalling a word from the mental lexicon according to the meaning that the speaker needs to convey. The second component, the formulator, underlies the process of attaching the structural properties to the recalled word to put it into the appropriate segment of the linguistic structure already built (Levelt & Meyer, 1999). While producing a word requires both processes to occur, in the case of receptive vocabulary knowledge, the information which is handled in the formulator is already available and has only to be matched with the relative meaning (Schmitt, 2019).

In addition to cognitive processes, there are other factors that make productive vocabulary knowledge more difficult to measure. Productive vocabulary requires L2 speakers to master many patterns of spoken and written features to be effective, while receptive vocabulary can achieve a comparable level of efficiency with fewer features (Nation, 2013; Schmitt, 2014). In production, many aspects of word knowledge need to be activated in the mental lexicon to fit into the context that the processing system is creating (Schmitt, 2014). This requires the processing system to access lexical features that encode various bits of information, such as the syntactic structure a word is a part of, meaning relations with other words, potential collocations, and so forth (e.g., Levelt & Meyer, 1999; Schmitt, 2014). In comprehension, only lexical features that link the form to the meaning of a word need to be recognised (Schmitt, 2019). Although competition among features may arise (e.g. Evans & Levinson, 2012), once the activation settles on the most likely candidate, the form meaning link is enough to extract the communicative value form a lexical item (Schmitt, 2019). Other lexical features that encode information like combinatorial potential, meaning relationship, and collocates are provided by the context (Schmitt, 2014, 2019). This reduces the number of features that the processing system needs to handle when recognising a word in receptive tasks.

The design of the task also reflects the complexity of productive vocabulary knowledge, since constructing a reliable measure for it has been proven challenging for researchers. With receptive vocabulary measurement tasks, the features of the words to be measured can be selected in advance. The researcher can easily choose words from different frequency bands (e.g., Laufer & Nation, 1999), determine their register, and control for any effect caused by the context such as, for instance, the frequency or the meaning of preceding words (Milton, 2009; Schmitt, 2019). These adjustments are not possible when productive vocabulary is measured. Since L2 speakers must provide the language that will be measured, they are expected to vary their vocabulary and retrieve words from a wide range of frequency bands, depending on the communicative goals of the task and on the vocabulary that is most readily available to them (Read, 2019; Schmitt, 2010). Furthermore, regardless of the task, L2 speakers are going to generate a sizable proportion of functions words, like articles or prepositions, that serve the only purpose to pursue textual and grammatical coherence. Function words do not really offer a fruitful insight into the breadth of their vocabulary knowledge, but, at most, into their linguistic competence (Meara & Fitzpatrick, 2000). It is also true that some factors that are normally available in the design of receptive vocabulary tasks may also be used in productive vocabulary tasks. For instance, L2 users may be recommended to use a specific register or to produce a text about a subject to narrow down the subset of words to be provided (Milton, 2009). However, due to the very nature of a productive vocabulary task, L2 users will tend to rely mainly on the lexicon they can handle more easily (Fitzpatrick & Meara, 2004; Meara & Fitzpatrick, 2000).

For these reasons, it is clear that it is difficult to collect enough words to have a reliable estimate of the productive vocabulary knowledge. Researchers have tirelessly worked to construct tasks that could prompt test-takers to generate enough words without being limited by the design of the task itself. The aim has been to build a task that could generate significant numbers of words that can be considered an actual estimate of the productive vocabulary of the learner. As noted, this is not easy, as producing a text brings in many unwelcome lexical items, such as function words and words related to the specific context and type of the task. These words are not indicative of the number of words that the test-taker can correctly retrieve from their mental lexicon. Since lexical knowledge is organised into a network with specific properties, it is important to tap into the sections of the network that are most informative of its density. As noted by Wilks & Meara (2002) some sections of lexical networks, even in L1 speakers, can have higher levels of links, whereas other sections can be sparsely connected or contain isolated words. A reliable vocabulary task must be able to capture both sections of the network. In the following sections, I will examine some existing productive vocabulary tasks to find out to what extent they capture the multifaceted nature of vocabulary knowledge.

3.5 The Design of Lexical Frequency Profile and Vocabulary Level Test

Laufer & Nation (1995) made a first attempt to capture the complexity of productive vocabulary knowledge in L2 English with the Lexical Frequency Profile. It requires the test-taker to produce a text instead of single words. In the Vocabulary Level Test, test-takers must write a 300 word composition on a very general topic. The topics are given in the form of a statement followed by a prompt, e.g., whether a government should limit the number of children a family can have. Written production is analysed through a software called RANGE that measures the frequency of words provided in the written text. Frequency is profiled into the first 1000 band, the second 1000 band, the Academic Word List (AWL), and any words absent from the lists. The rationale behind the task is that as proficiency increases, the L2 speakers should know more infrequent words.

In a follow up study, Laufer & Nation (1999) developed the Productive Vocabulary Levels Test, which has a different format from Vocabulary Level Test. It is a test of controlled productive ability, where responses are primed by the sentence context. The task consists of five sets of 18 gap-filling sentences, and both the frequency and the meaning of the words that must be inserted into the gaps are established by the examiner beforehand. The

18 items are drawn from five frequency bands: the first 2000, 3000, and 5000 most frequent words, the University Word List (UWL, Guoyi & Nation 1984) and the last 10000 words. Every set of selected words captures the potential knowledge that a test-taker is supposed to have at each frequency band. To prevent a test-taker from writing a word that might be correct but drawn from a different frequency band, the first letters of the target words are given.

Both tests have been successfully validated through a series of studies and are widely used in the field of second language acquisition. Laufer & Nation (1999) found that the Productive Vocabulary Levels Test was reliable in capturing vocabulary growth. The Vocabulary Level Test could reliably discriminate among L2 speakers of different proficiency levels, showing a significant correlation with other vocabulary measurement tasks (Laufer & Nation, 1995) and being able to capture changes in L2 speaker's writing (Laufer, 1994). Both tests have been extensively used in a plethora of studies, proving to be reliable in many different situations, even when compared to other tasks (e.g., Chui, 2006; Fitzpatrick & Clenton, 2017; Fitzpatrick & Meara, 2004; Laufer & Ravenhorst-Kalovski, 2010; Walters, 2012; Webb, 2005). The key point to highlight here is that both tests rely on frequency as the discriminant factor in vocabulary growth. As Laufer & Nation (1999) noted, "All things being equal, words should be learned roughly in order of their frequency of occurrence, with high-frequency words being learned first." (p.35)

While the tests clearly show how frequency can be a reliable factor in the design of a productive vocabulary task, both measurements are not particularly well-suited to generate a large number of infrequent words. Discussing the Productive Vocabulary Level Test, Meara and Fitzpatrick (2002) noted that this kind of controlled productive vocabulary test is likely to work reliably with low-level L2 users who have limited vocabulary knowledge. Drawing 18 words out of a productive knowledge of about few hundred words can be a representative estimate. As proficiency increases, a sample of 18 words out of a lexical knowledge of two or three thousand words is too small to gauge vocabulary breadth. The Vocabulary Level Test has similar problems. In fact, a wide proportion of words that appear in a text is usually made up of function words and a large body of high-frequency words so that the number of infrequent words tends to be small (Fitzpatrick & Clenton, 2017; Meara & Fitzpatrick, 2000). Both tests raise the question of what kind of task can use frequency as the main measurement tool while generating a large number of words. The next section seeks an answer to this question.

3.6 Word Association Tasks and Vocabulary Knowledge

As noted above, the ratio between frequent and infrequent words appears to be a reliable criterion to measure productive vocabulary knowledge (Laufer & Nation, 1999). The tasks examined in the previous section suggest that an advisable criterion to exploit the importance of frequency should be that of having a high number of infrequent words. One type of task that can generate many infrequent words in a brief period of time is word association (Fitzpatrick & Thwaites, 2020). Word association tasks can easily avoid function and high-frequency words that are normally brought about when writing short texts (Meara & Fitzpatrick, 2000). In word association tasks, participants are asked to provide one or more associates to a list of stimulus words. This format makes it easy for each participant to provide more than one associate with a long list of stimulus words in a very short time. According to De Deyne et al. (2013), for instance, their participants were able to produce around 50 words - three associates to 18 stimulus words - in less than 10 minutes. As a task, word associations can be easily scaled up without putting much strain on the participants. Furthermore, word association tasks can provide information on the type of links between stimulus words and associates. Examining the type of associated produced reveal how words are stored in the mental lexicon (De Deyne & Storms, 2015; Fitzpatrick & Thwaites, 2020). Traditionally, lexical relations have been framed into either the paradigmatic or syntagmatic category, although, over the years, the categories have been expanded and refined (Fitzpatrick, 2007, 2009).

As outlined in Chapter 2, categories depend on the type of links between the stimulus words and the associates produced by the test-takers. They can be based, for example, on broad meaning relations like conceptual links or more specific meaning relations like synonymy; other types of relations can be position-based collocational and phrasal relations or be based on the form of the associate and the stimulus word (Fitzpatrick, 2006, 2007, 2009). Since the choice of an associate is related to specific aspects of lexical representations, it is not surprising that there are several categories to describe the types of associations (De Deyne & Storms, 2015; Fitzpatrick, 2006, 2007, 2009; Fitzpatrick & Thwaites, 2020; Meara & Fitzpatrick, 2000).

Since the type of links in a lexical network is based on the properties encoded into words, the properties of lexical networks based on word association tasks have been shown to correlate with word level phenomena. Measures such as clustering or the average number of links between words exhibit high correlations with several effects in various lexical processing tasks, at the level of semantic and phonological processing (De Deyne et al., 2013, 2019; De Deyne & Storms, 2008a; Vitevitch, 2008). For example, De Deyne et al. (2013) measured various density measures against a large database containing reaction times in a lexical decision task (Brysbaert et al., 2018). They found significant correlations between reaction times and measures of the overall structure of the network, such as the average number of links between words. De Deyne et al. (2019) found that the most frequent responses in the word association network showed a high correlation with a lexical decision and a picture naming task and a lower, yet still significant, correlation with a semantic decision task.

Although links between network properties and lexical processing tasks appeared in large-scale lexical networks, they suggest that the word association format can equally tap into general and specific properties of lexical representations. The word association format seems well suited to capture the most tightly connected and the most sparsely connected sections of lexical networks. Since connections in lexical networks depend on word frequency, the use of a word association format that can be measured against frequency seems to be a valid tool to capture several aspects of lexical representations.

3.7 The Design and Validation of Lex30

Meara and Fitzpatrick (2000) developed a task, for English, called Lex30 which takes the format of a word association task. The word association format was chosen to generate many infrequent words without being limited by using short texts. They set up three criteria that the stimulus words had to meet in order to be considered reliable in providing the suitable type and number of responses. First, stimuli were drawn from the first 1,000 most frequent English words, to avoid the risk that test-takers would encounter words they have no knowledge of. Second, the number of infrequent responses generated by learners should not exceed 50% of the overall number of words provided by participants. Third, every stimulus word should not generate a single associate more frequent than 25% of the total given responses. To establish whether the stimuli would meet this last criterion, the Edinburgh Associative Thesaurus (Kiss et al., 1973) was used as a baseline.

To complete the Lex30 task, a test-taker is presented with a list of 30 stimulus words, and they must provide up to 4 responses for each of them. Test-takers are given complete freedom in choosing their responses. The instructions of the task clearly prompt them to write down, next to each of the stimulus words, any other word that it makes them to think of. For example, the first word stimulus word in Lex30 is *attack*, and test-taker must write 4 associates to the word based on their own criteria.

Once all responses are collected, they are measured against a frequency list, and every word that is not present in the list receives one point. The frequency lists have changed overtime: initially responses were scored against Nation's first 1,000-word list (Fitzpatrick & Meara, 2004; Meara & Fitzpatrick, 2000). In its latest version of Lex30, uses the frequency lists JACET 8000 (Fitzpatrick & Clenton, 2010). In all versions of the task, the frequency list contains the most frequent 1000 words of English. Hence, infrequent words in Lex30 are all those words that are outside of the first 1000 band. Lex30 comes in a pencil and paper version or in an online version (Meara & Fitzpatrick, 2015). In the printed version, responses are first checked for spelling, then tallied and scored. The online version, after asking the test-takers to confirm their choices, assigns scores automatically and directly provides them to the test-taker. Because of its word association format Lex30 does not generate many function words that normally occur throughout every text, and so it seems to be well-suited to produce a fairly large number of infrequent words. Furthermore, by using 30 stimulus words, Lex30 activates multiple semantic fields that generate a broad and representative sample of lexical representations (Fitzpatrick & Clenton, 2017). The 46 L2 speakers examined in Meara and Fitzpatrick (2000) generated, for instance, a mean of 32% infrequent responses. Fitzpatrick and Meara (2004) obtained a mean of infrequent words of 30%, Fitzpatrick (2007) reported a mean of 27% of infrequent words, while Fitzpatrick and Clenton (2017) 43%.

Lex30 has been validated through an extensive series of experimental studies. Regarding its reliability, Lex30 was able to produce correlated scores in a take-and-retake procedure in both a small-scale study (Fitzpatrick & Meara, 2004) and a large-scale one

(Fitzpatrick & Clenton, 2010). In both studies, the correlation was significant and high, suggesting that the Lex30 scores can be regarded as generally representative of the individual lexicon (Fitzpatrick & Clenton, 2010). To exclude that changes in performance are not caused by practice, Fitzpatrick & Clenton (2010) tested two parallel forms of Lex30, based on two different lists of stimulus words. Like the results with the take and retake procedure, the scores correlated but did not differ significantly. Fitzpatrick & Clenton (2010) also found an acceptable level of internal consistency (Cronbach's alpha = 0.866). Fitzpatrick & Clenton (2010) also compared the commonly used written version of Lex30 with its spoken counterpart, finding a significant but weak correlation. They argued that the weak correlation could show that the two tests do not work precisely in the same way or that the oral ability of the L2 speaker might be behind their written one. They also cautioned that the sample of written vocabulary produced by a test-taker should not be expected to mirror their sample of spoken vocabulary (Fitzpatrick & Clenton, 2010).

Lex30 was also tested in terms of criterion-validity. Criterion validity can be achieved by comparing one set of language user scores with the scores of other language users that are at a different level of language proficiency (Bachman 1990). First, Fitzpatrick and Meara (2004) used a group of L1 speakers as a criterion, finding a significant difference between L1 and L2 speakers, although there was some overlap in the scores of the two groups. Fitzpatrick & Clenton (2010), using the same group of L2 speakers, compared the Lex30 scores of the two groups after a teaching intervention period. The difference between the two means was indeed significant and the two scores were correlated. The criterion validity procedure shows that Lex30 is sensitive to the improvements of L2 speakers, to the difference between L2 and L1 speakers, and to the proficiency of L2 speakers.

Three other studies looked at the ability of Lex30 to capture distinct levels of proficiency. (González & Píriz, 2016) used Lex30 to measure the effects of CLIL (Content and Language Integrated Learning) on productive vocabulary knowledge. Lex30 scores after one year of CLIL lectures increased significantly, although the growth was greater in low proficiency students than in high proficiency students. Similarly, Moreno Espinosa (2009) showed that participants in CLIL modules had higher Lex30 scores than participants not enrolled in CLIL modules. Finally, in a study carried out by Walters (2012), Lex30 was able to differentiate among three groups of different levels of proficiency.

Lex30 has shown relationships with some other lexical measurement tasks with mixed results. Fitzpatrick & Clenton (2017) compared Lex30 with the Lexical Frequency Profile (Laufer & Nation, 1995) without finding a significant statistical correlation between the two tasks. On the contrary, a significant but weak correlation was found between Lex30 and the Productive Level Test (Fitzpatrick, 2010; Fitzpatrick & Meara, 2004; Walters, 2012), while Catala & Moreno Espinosa (2005) found a significant but small correlation between Lex30's scores of young learners and the scores of the Vocabulary Level Test. The correlation with yes/no test of receptive vocabulary knowledge gave mixed result, showing both weak (González & Píriz, 2016) and high correlations (Meara & Fitzpatrick, 2000). Medium-size effects with translation tests have appeared to be steady across the studies (Fitzpatrick & Clenton, 2010; Fitzpatrick & Meara, 2004).

Lex30 has proven to be an effective measurement task in a series of studies that have used it along with other types of proficiency measures. A series of studies have found that Lex30 correlates with fluency. Uchihara & Saito (2019) found a significant medium correlation (r = .34) correlation between fluency and Lex30. Comparable results were found comparing speech fluency with Lex30 and other vocabulary measurements that looked at sophistication and diversity (Clenton et al., 2020; Uchihara et al., 2020). Walters (2012) considered the ability of Lex30 to reflect the difference between the capacity to simply recall a word or to use it meaningfully. The study could be regarded as an attempt to evaluate the depth of knowledge that Lex30 is able to capture. Participants in the study were required to produce sentences using all the infrequent words they were able to provide in Lex30. Subsequent analysis compared Lex30 scores with the scores generated by the number of corrected sentences produced using infrequent responses. Both scores could significantly differentiate among the groups of participants almost to same extent. Based on these results, Walters (2012) argued that Lex30 not only measures the ability to recall words, but it also reflects the ability to use them in a meaningful and appropriate way.

The series of studies described so far show that Lex30 is a reliable task in measuring productive vocabulary knowledge. The word association format ensures that the task generates a large number of infrequent words. The fact that it shows significant relationships with a variety of other tasks suggests that the task can tap into several dimensions of vocabulary knowledge. Fitzpatrick & Clenton (2017) have highlighted that Lex30 captures four levels of vocabulary knowledge at the same time. The first level is the ability to produce the form of a word, the second is the ability to link a word to an appropriate referent or an equivalent L1 word, the third is the ability to use words with semantic appropriateness, and the fourth the ability to use words in a suitable semantic context with grammatical accuracy (Fitzpatrick & Clenton, 2017). An important feature of Lex30 is that it uses word frequency as a reference to assign scores. The fact that several studies found the task shows a strong relationship between Lex30 and language proficiency resonates with the claim that L2 speakers tend to produce first the most frequent words adding in infrequent words as their proficiency increases (Milton, 2009).

3.8 The Development and Validation of I-Lex

Overall, Lex30 is related to some of the key factors that underlie lexical representations in L2 speakers as presented in Chapter 2. It gives a glimpse into lexical networks and exploits frequency, which is a key component of vocabulary knowledge (Brysbaert et al., 2017, 2018; Diependaele et al., 2013). Hence, I have decided to design and develop a productive vocabulary test for L2 Italian called I-Lex, based on the same word association and on the same frequency-based scoring system. The following sections show the stages and the criteria used to construct the I-Lex task, the validation process and the first results from a study conducted on three diverse groups: L1 speakers, L2 advanced speakers, and L2 pre-intermediate speakers.

3.8.1 *Research Questions*

Research question (1). To what extent do the I-Lex's stimulus words meet the two criteria set up to generate a high number of infrequent words?

Meara and Fitzpatrick (2000) established a threshold to consider the stimulus words sufficiently reliable in producing lexical diversity and enough infrequent words. Concerning lexical diversity, the 30 stimulus words should not generate an individual response recurring more than the 25% of all given responses. In terms of frequency, responses should typically produce a fairly large set - around 50% - of infrequent words, i.e. words that are not included within the first 1000 most frequent words. I-Lex should produce the same degree of lexical variety and lexical diversity as Lex30. In line with Fitzpatrick & Clenton (2010), the internal consistency of the items will also be investigated. This norming stage of the analysis will be carried out on the L1 group of native speakers only, as they are going to be used as the baseline group.

Research question (2). To what extent does I-Lex produce the same results as found for Lex30, thus achieving concurrent validity?

The study will investigate whether the pattern of results generated by I-Lex is the same as that generated by Lex30. Specifically, the analysis will focus on the difference between L1 speakers and L2 speakers and within the L2 speaker group, expecting the scores to be significantly related to proficiency. A significant difference between the level of proficiency will reproduce and extend the effects found in many studies carried out using Lex30 (e.g., Fitzpatrick & Clenton, 2010; González & Píriz, 2016; Walters, 2012). Finally, the study will measure whether the same group of L2 speakers, performing the same task twice, will reach a similar score and a similar variance (Fitzpatrick & Clenton, 2010; Fitzpatrick & Meara, 2004).

Research question (3). To what extent do the lexical networks generated by each group show the same patterns found in the subject analysis?

The third research question focuses on the overall number of responses given by each group of participants. This type of analysis has as source the relationship found between general properties of lexical networks and word-level properties using large-scale word association networks (De Deyne et al., 2013, 2019). In fact, grouping together word association tasks of many participants produces, at all effects, small-scale lexical networks whose properties can be fully analysed, offering a glimpse into the lexical network of each group (Fitzpatrick & Thwaites, 2020). The analysis will examine network-level properties, namely, the number of infrequent words and the lexical diversity, generated by the three groups. The lexical networks of the three groups are expected to reflect the difference in proficiency, so that L1 speakers are likely to express a higher level of lexical diversity and more infrequent words compared to the two L2 groups.

3.8.2 *Methodology*

The methodology section contains two parts. The first part describes which frequency lists were used as a reference for the scores and how the stimulus list for I-Lex was constructed. The second part describes the validation study carried out by three groups of participants.

3.8.3 *The Design of I-Lex*

The design of the I-Lex task encompassed three steps. The first step was to prepare the frequency list for Italian based on the most recent available frequency list and corpus-driven data (Chiari & Mauro, 2010, 2015). This step provides the reference list that could be used to measure the number of infrequent words produced by the participants. In the second step, I set up the 30 stimulus words and tested them with a control group of Italian speakers to ensure that the I-Lex stimulus word would meet the same criteria as those established for Lex30 by Meara & Fitzpatrick (2000). This step is crucial to ensure that the I-Lex task can produce results comparable to those of Lex30 using a similar design of stimulus words. The third step consists of the actual validation process in which L1 and L2 speakers at two different proficiency levels were compared. The criterion for establishing the validity of I-Lex was the comparison between its results and those of Lex30. Obtaining equivalent and comparable results can be regarded as concurrent validity for the I-Lex, as both tests measure the same kind of knowledge using the same criteria.

3.8.4 *The Frequency Lists*

The list used as the benchmark for I-Lex is the New Basic Vocabulary of Italian (NVDB) recently published by De Mauro and colleagues (Chiari & De Mauro 2014; De Mauro 2016) and currently available in an online version (De Mauro, 2016). The list is the most recent update of that originally published, as an appendix, in De Mauro (1980). The updated version has been drawn from a corpus containing 18 million words formed by combining six different sub-corpora of three million words each. All sub-corpora contain texts whose chronological span ranges from 2001 to 2011. The corpus was purposely assembled to gather the most diverse number of lexical items and to offer a comprehensive insight into the usage of Italian. The list of the sub-corpora is given below:

- Press (newspapers and periodicals)
- Literature (novels, short stories, and poetry)
- Non-fiction (textbooks, essays, and encyclopedia)
- Entertainment (theatre, cinema, songs, and television shows)
- Computer-mediated communication (forum, newsgroup, blog, chat, and social networks)
- Spoken language.

(from Chiari & De Mauro 2014, p. 29)

The resulting new NVDB list was divided into three bands of around 2000 words each. The first band with the most frequent 2000 lemmas provides 90% coverage. The second band with the second most frequent 2700 lemmas provides a further 6% of coverage, and the third band of around 2300 lemmas contains high availability vocabulary, i.e. lemmas that are not highly frequent but are widely known by native speakers (Chiari & Mauro, 2010, 2015).

The NVDB list contains more words than the list used in the original Lex30 validation study (Meara and Fitzpatrick 2000). The number of infrequent words provided by Lex30 participants was scored against the first 1000 most frequent words. Consequently, to have a comparable result from I-Lex, I decided to further divide the NVDB list to get a list with the first 1,000 most frequent words. Unfortunately, the raw data used in De Mauro (2016) were not available: in order to circumvent the issue, I relied on another corpus, namely the CoLFIS Corpus developed by Bertinetto et al. (2005).

The composition of CoLFIS is different from the NVDB list: it is actually a lexical database that builds on a balanced corpus of written Italian only of 3 million words. It has been assembled based on the reading habits of the Italian population as surveyed by ISTAT (The Italian National Institute of Statistics). The list of the entire corpus ranked by absolute frequency was used to set up the 1000-band list in the following manner. I compared the

CoLFIS frequency list with the NVDB first 2000-word band: as soon as a word that was within the first 1000 most frequent words of the CoLFIS list also appeared on the NVDB list it was used to form a new list. Words that appeared on the CoLFIS list but not on the De Mauro list were not taken into account. The procedure was manually repeated until a new list containing 1000 words was completed.

After the process, I ended up with two lists. The first one, with the 1000 most frequent words, was named I-Lex1000, and the second one, with the 2000 most frequent words entirely based on the NVDB list, was named I-Lex2000. I am aware that the method has some limitations. While the NVDB list has a part of spoken texts, the CoLFIS one is based on written texts so that I-Lex1000 list is skewed towards the written section of the NVDB. For this reason, both lists were used in the validation process. The I-Lex1000 was used to directly compare my results with those obtained with Lex30. In contrast, since the I-Lex2000 word list is entirely based on the NVDB, it has the high degree of reliability provided by the variety of the corpus on which the NVDB is based. Analysis of the data with the I-Lex2000 provided a further reliable criterion to evaluate the reliability of the I-Lex.

3.8.5 *The Stimulus Words*

Most of I-Lex stimulus words were translated to Italian from the Lex30 stimulus words. A few I-Lex stimulus words that could not be translated had a close meaning-based relationship with the Lex30 stimulus words (see Section 3.8.6 for a detailed explanation). The assumption behind the use of translation equivalents is to provide further ground for the comparison between I-Lex and Lex30. Since translation equivalents share several abstract and concrete meaning features, the results from the Lex 30 and the I-Lex tasks could be partially ascribed to those shared features. Although it is extremely difficult to determine the extent to which such a process would occur, nonetheless it would still provide more common ground for comparison. For example, the Lex30 stimulus word *hold* and the Italian translation *tenere* indicate the action of grasping something carried out by an entity with the thematic role of agent on an object with the thematic role of theme. The two stimulus words also have several metaphorical usage-based expressions that overlap in the two languages like, for example, *hold a conference* and *tenere una conferenza* or *hold the stage* and *tenere la scena*. Considering this overlap of meaning features, it is likely that the some of the shared meaning features that prompted the English responses to the stimulus word *hold* also prompted the Italian responses to the stimulus word *tenere*.

There is evidence to support this reasoning. For example, two studies that have translated stimulus words from a pre-existing English list to create an Italian database of semantic norms were able to replicate most of the English patterns (Montefinese et al., 2013; Fairfield et al., 2017). Rosenzweig (1961), using a word association task, found a remarkably high cross-linguistic overlap among English, French, German. Similarly, Kremer & Baroni (2011), comparing two sets of semantic norms for Italian and German to a pre-existing set of semantic norms for English (McRae et al. 2005), found that the three languages also shared a sizeable proportion of meaning features. More importantly, there is also evidence that word associations tasks specifically tap into shared conceptual representations, at least to some extent. In Chapter 1 (see Section 2.3) I reviewed a study which found that half of the ten most connected words in two large-scale English and Dutch word association networks overlapped (De Deyne & Storms, 2015). As the authors of the studies noted, the overlap shows that there are some meaning properties that have a more central place in meaning relations that do not seem to be language-related (De Deyne & Storms, 2015, p. 477).

3.8.6 *The Construction of the Stimulus Word List*

Since the I-Lex stimulus words were given as lemmas, some grammatical differences arose. Unlike in English where the part of speech a word belongs to might be ambiguous, in Italian words are marked by specific suffixes, and ambiguity is much lower than in English.

For example, the first Italian stimulus word *attaccare* is a verb form bearing the infinitive suffix -*are*, while the English equivalent *attack*, being morphologically unmarked, can be a verb and a noun. The words *attack, hold,* and *rest* were translated as verbs, whereas *experience*, *kick*, and *trade* were translated as nouns. In English, the adjective *dirty* can also be used as a verb. The Italian translation was the adjective form *sporco,* which is a homophone to the first person of the present indicative of the equivalent verb *sporcare*[1]. The English stimulus words *board*, *close,* and *seat* have different meanings when used as verbs, nouns, or adjectives. In the original papers in which Lex30 was validated there were no indications about the grammatical category of the stimulus words. In fact, as previously explained, stimulus words were chosen for their frequency and their ability to produce both unique and infrequent words (Meara & Fitzpatrick, 2000; see also Section 3.7). Since the Italian translations do not have the same polysemic meanings, I looked at the examples given in the appendixes of Meara & Fitzpatrick (2000) and Fitzpatrick & Clenton (2010) to see whether the predominant English response would suggest the grammatical category of the stimulus word. The stimulus word *board* had predominant responses suggesting it was interpreted as a noun (i.e., *plane*, *wood*, *airport*, *boarding pass*, *snow*, *games*, *sports*, *surf*); the stimulus word *close* had predominant responses suggesting it was interpreted as a verb (i.e., *lock*, *avenue, finish*, *end*, *up*, *shave*, *shop*, *comfort*). Finally, the stimulus word *seat* had predominant responses suggesting it was interpreted as a noun (i.e., *bench, sit*, *sofa*, *belt, free, car, chair*). Based on that, the Italian stimulus words were translated using equivalents that align with the predominant interpretation of the Lex30 stimulus words. Thus, the word *board* was translated as the noun *tavolo*, the word *close* as the verb *chiudere* and the word *seat* as

[1] When I examined the results, it was not possible to determine if the responses were also elicited by the verb interpretation, as the collocational patterns are extremely similar.

the noun *posto*. Some Italian stimulus words had homophones in Italian but not in English: they appear in bold in Table 3.1 and the two meanings are given.

Table 3.1

List of I-Lex Stimulus Words in Lex 30 and I-Lex

Lex30	I-Lex	Meaning in Italian
attack	**attaccare**	stick; attack
board	**tavola**	board; table
close	chiudere	close/lock/put to an end
cloth	**abito**	suit/dress; abode/live
dig	TOGLIERE	remove/take off
dirty	sporco	dirty/sleazy/corrupt
disease	malattia	disease/mental disease
experience	esperienza	experience
fruit	frutto	fruit/result/outcome
furniture	**mobile**	forniture; mobile
habit	abitudine	habit/routine
hold	tenere	hold
hope	speranza	hope
kick	**calcio**	kick; calcium
map	CARTOLINA	postcard
obey	OBBLIGO	obligation/duty
pot	CUOCERE	cook
potato	patata	potato
real	**reale**	real; royal
rest	riposare	rest
rice	**riso**	rice; laughed
science	scienza	science
seat	posto	place/seat/spot
spell	SCRIVERE	write
substance	sostanza	substance
stupid	stupido	stupid
television	televisione	television
tooth	dente	tooth
trade	commercio	trade
window	finestra	window

Note. The Italian items in bold have homographs. The translation equivalent is followed by the English translation of the Italian homophone

For the five stimulus words, *dig*, *map*, *obey*, *pot*, and *spell*, a translation that also would fall within the NVDB list was not available. Translation equivalents that were not in the NVDB list were swapped with other items that were as close as possible in terms of meaning relatedness. As a metric for relatedness, I used the lexical lists given in the website FrameNet (Baker et al., 2003), an online collection of lexical frames. Frames are broad-scope cognitive structures that encompass the set of knowledge, beliefs, and schemes of practice that underpins the way experiences are shaped. As Fillmore & Baker (2009) put it, “Frame Semantics is the study of how linguistic forms *evoke* or activate frame knowledge and how the frames thus activated can be integrated into an understanding of the passages that contain these forms.” (p.317). An excerpt of the kind of information available for each frame is given in Figure 3.1. It is taken from the website FrameNet (http://framenet.icsi.berkeley.edu/) which has been built on the basis of the developments and implementations of Frame Semantics carried out in the last four decades (e.g. Baker et al., 1998, 2003; Fillmore, 1982).

Figure 3.1

An Excerpt of the Cooking_creation Frame

Frame Element	Number Annotated	Realization(s)
Cook	(1)	NP.Ext (1)
Ingredients	(1)	NP.Obj (1)
Produced_food	(5)	NP.Obj (4) NP.Ext (1)
Purpose	(2)	NP.Dep (1) PP[for].Dep (1)
Time	(1)	PP[in].Dep (1)

Frame Element	Number Annotated	Realization(s)
Container	(53)	DEN.-- (53)
Contents	(16)	A.Dep (1) AJP.Dep (2) N.Dep (3) PP[of].Dep (8) Poss.Gen (2)
Material	(16)	N.Dep (14) AJP.Dep (2)
Owner	(3)	Poss.Gen (3)
Relative_location	(2)	PP[in].Dep (1) PP[on].Dep (1)
Type	(7)	N.Dep (7)
Use	(3)	PP[for].Dep (1) AJP.Dep (1) N.Dep (1)

The right-hand side of Figure 3.1 shows the frame Cooking_creation and, as an example, the lexical entry for the verb *to cook*. The varied colours illustrate the so-called frame elements: in the case of a verb, they represent semantic roles and adjuncts. Under the column *Realisation* there is a list of the syntactic structures associated with each element (the column *Number Annotated*, not important in the present discussion, contains the number of exemplar sentences available on FrameNet). Importantly, frames and frame elements are associated to lists of lexical units, that is, lexical items whose meaning is determined by and linked to some aspect of the frame (Fillmore, 1982; Fillmore & Baker, 2009). After associating each of the five stimulus words to be replaced with the frame to which they belonged, I checked words that were semantically related and were within the list of lexical units of the frame. The noun *map* was changed into *postcard* (*cartolina*) due to an underlying Travel Frame; the verb *obey* (*obbedire*) was changed into *obligation* (*obbligo*) due to an underlying Compliance Frame; the noun *pot* was changed into the verb *to cook* (*cuocere*) due to an underlying Cooking_creation. For the word *spell* I considered the meaning of "writing a word in the correct form" and I resorted to the hypernym *to write* (*scrivere*) due to the underlying Spelling and Pronouncing Frame. The verb *to dig* has both a concrete (i.e. extracting something from the soil) and an abstract meaning (i.e. bringing out something after an investigation). Drawing on the concrete meaning of *extract* and on the Removing Frame, I opted for the Italian verb *remove* (*togliere*). Although in Italian the word can have an abstract meaning of *setting free* and *abolishing,* relative to the English verb to dig, the concrete meaning of *taking away* something is prevalent. Interestingly, six stimulus words, when translated into Italian, had a homophone not available for their English equivalents (see Table 3.1). I decided to keep them on the grounds that they could generate a more diverse set of responses.

3.8.7 *Establishing the Validity of I-Lex*

The first type of validation of the I-Lex draws on concurrent validity (e.g. Schmitt, 2010). A classic procedure to pursue concurrent validity is to administer a new test to a group of test-takers and then give to the same group a reliable and previously validated test, which works as a criterion measure. If the correlation between the two is high and significant, then the new test is likely to capture what is supposed to capture (Schmitt, 2010). Therefore, obtaining the same results using a similar version of the same task adds confidence that the same construct is currently captured by the two tasks in two different languages. In the present case the comparison is drawn between a well-grounded productive vocabulary measurement task and a new one that has the same stimulus words and the same format. If a similar pattern appears, this will ensure that I-Lex measures the same sort of knowledge as Lex30. For this reason, the first phase of the validation of I-Lex closely reproduces the procedures that have been applied to validate Lex30. The validation process of I-Lex is based on the following analyses.

- Examining the difference in scores between taking and retaking the same task. If the results correlate showing the same variance, then I-Lex can be regarded as broadly representative of the lexical knowledge of L2 speakers. The correlation would suggest that the first score can be a predictor of the second score, while the same variance indicates that the scores have not changed across the group. This same pattern was found for Lex30 (Fitzpatrick & Clenton, 2010; Fitzpatrick & Meara, 2004).

- To have a more direct comparison with the analysis adopted by (Fitzpatrick & Clenton, 2010) the internal consistency evaluation was carried out using the Cronbach's alpha test. Recall that in Fitzpatrick & Clenton (2010) Lex30 has shown an α value of .86, normally considered an acceptable level of internal consistency (Field et al., 2012).

- A way to assess the reliability of Lex30, is through the criterion validity, which is obtained by comparing "one set of learner data with data from a group of individuals known to be at a different level of language ability from those learners" (Fitzpatrick & Clenton, 2010, p. 543). Following Fitzpatrick & Meara (2004), L1 speakers scores are used as a criterion expecting a significant difference from those of the L2 speakers. Furthermore, Fitzpatrick & Clenton (2010) found a significant difference within the same L2 group after a teaching intervention. Rather than revaluating the same group in a longitudinal study, the score of the most proficient L2 group will be used as a criterion, expecting it to be significantly higher than the less proficient L2 group and significantly lower than the L1 group.

- As noted above, the overall responses of each group of participants can be regarded as a lexical network representative of the vocabulary knowledge of the group itself. Since L1 speakers know more infrequent words than L2 speakers and the difference tends to diminish as proficiency increases (Laufer & Nation, 1995, 1999; Meara & Fitzpatrick, 2000) lexical networks based on homogenous groups should reflect the same tendency.

3.8.8 *Participants*

The validation study of the I-Lex employed three groups: a control group of 36 native speakers of Italian, 82 advanced L2 learners of Italian with different L1s and 30 pre-intermediate L2 learners of Italian with Chinese as their first language (a summary is given in table 3.2 below).

The L1 speakers group (N = 36) encompasses two subsets: 12 participants were part of the control group for study two and three and took the productive vocabulary task along with the other tasks, namely an Elicited Oral Imitation Test and a Self-paced Reading (see Chapter 6 and Chapter 7). The other 24 participants were recruited by email: the email contained a brief summary of the research project and the link to an online version of I-Lex created using Google Forms. The emails were sent to a pool of potential participants ($N \approx 50$) from which 24 individuals accepted to take part in the study. The resulting L1 group matched the L2 speakers in terms of education and age.

L2 speakers are divided into two groups based on their language proficiency. The first group is formed by 82 advanced students who were enrolled into a three-week intensive Italian language course. To be enrolled in the programme, students need to provide a language certificate that is at least B2. The level of the course is C1: students are taught Italian for 4 hours every morning, from Monday to Friday, and, in the afternoons, they must attend a series of conferences on subjects related to Italian culture. A subset of this group also took part into the studies described in Chapter 6 and Chapter 7, taking the productive vocabulary task along with either the Elicited Oral Imitation Test (Chapter 6) or the Self-Paced Reading (Chapter 6).

The second group of L2 speakers is made up of 30 L1 Chinese students. They were all enrolled into The Marco Polo project. Students who take part in the project must have an A2 level certificate to be admitted to a language course of Italian before enrolling in Italian universities. The language course lasts 8 months and, by the end of it, students are expected to have reached the B2 level. The I-Lex task was delivered during the second week of the course so that the proficiency level was likely not higher than pre-intermediate (A2). Table 3.2 summarises all relevant information about all groups of participants who took part into the study.

Table 3.2

Data of Participants as collected in the Questionnaires

	Age	Education	Gender
L1 (N=36)	*M* = 28 *SD* = 4.02	BA = 24 MA = 7 PhD = 5	F = 21 M = 15
L2 Advanced (N=82)	*M* = 28.13 *SD* = 7.69	BA = 57 MA = 18 PhD = 7	F = 63 M = 19
L2 Pre-Intermediate (N=30)	*M* = 22.21 *SD* = 1.23	BA = 30	F = 12 M = 18

3.8.9 *Procedure*

One third of the L1 speakers (N = 12) completed the productive vocabulary task on a Mac laptop with a 13-inch Retina screen and the responses were typed directly into a spreadsheet using Numbers (Version 3.6.2.). The other L1 participants (N = 24) completed an online version of I-Lex that was assembled using the format available with Google Forms. All L1 speakers also completed a consent form, a brief about the purpose and the methodology of the study and a questionnaire containing questions about their age, their education and their L2s (see Appendix 2). The data from all participants was then downloaded as a csv file containing the I-Lex responses and the questionnaire. The advanced group of L2 speakers (N = 82), completed the I-Lex at separate times. Data from the first subset of participants (N = 40) participants were collected in August 2016, those of the second subset (N = 22) in August 2017 and those of the last group (N = 20) in August 2018. The three groups performed the task in pencil and paper form in one session that lasted approximately 15 minutes. Before completing the I-Lex, all L2 speakers received the same consent form, brief, and questionnaire as L1 speakers (see Appendix 2). The pre-intermediate group of L2 speakers (N = 30) also performed the I-Lex in pencil and paper format. Unlike the advanced L2 group, this second pre-intermediate group took the test twice: I-Lex was given again to the same students four days later. On both occasions, the task lasted around 20 minutes. Before the task, they were given the same consent form, brief, and questionnaire as L1 and L2 advanced speakers (see Appendix 2). The Ethics Assessment Status for the study was approved (Su-Ethics-student-051020/3106 see Appendix 1).

The task responses of the advanced and pre-intermediate L2 speakers were manually transcribed into a spreadsheet using the software Numbers (Version 3.6.2.), one file for each participant. Unlike the L2 speakers, all the L1 speakers completed the I-Lex in a spreadsheet, so their data did not need to be transcribed. The author of the study manually transcribed all the data.

3.8.10 *Scoring of the Responses*

Two different scores were measured. The first score was based on the I-Lex1000 list created by comparing the Colfis and the NVDB list. One point was assigned to every word provided by participants that fell outside the list of 1,000 most frequent Italian words. The second score was based on the I-Lex2000 list based on the NVDB. One point was assigned to every word that fell outside the list of the 2000 most frequent Italian words.

Following Fitzpatrick & Meara (2004), before scoring the responses, the spelling was checked and corrected when necessary: the criterion used was that, despite the wrong spelling, the word would still be clearly recognisable. Following Meara & Fitzpatrick (2000) the responses were also normalised: I removed proper names, English words commonly used in Italian such as *software* or *joystick*, names of famous brands like *Windows* or *Apple*, and foreign words related to food like *croissant* or *cheesecake.* All these types of responses, despite being well known to Italian speakers, are not Italian words, and consequently they do not appear in neither of the two lists used for the scoring procedure. Thus, keeping them would have warranted a higher score for participants not necessarily based on real Italian vocabulary knowledge. These words were extremely scarce and removing them led to a loss of 71 responses out of 14,098 responses.

To automatically score I-Lex, following Paul Meara's advice, the scoring was performed using a custom programme created together with Brian Rogers using the software Perl (version 18, subversion 2 (v5.18.2)). To be readable by Perl, the csv files had to be exported into a txt format file, one file containing the list of responses generated by every participant. Once the list of responses produced by each participant is entered into the programme, it automatically uploads the I-Lex1000 and I-Lex2000 lists and compares them to the list of each participant. The output given by the programme consists of the number of words that simultaneously appears in either the I-Lex1000 or the I-Lex2000 list and in the list of words provided by each test-taker. This number is then subtracted from the total amount of words provided by the test-taker. The result amounts to the number of words that are not in either the I-Lex1000 or the I-Lex2000 lists, and it forms the actual score, that is, the number of infrequent words given by every participant. This procedure was performed for each participant.

To answer the third research question, the responses given to each stimulus word were also analysed for each of the three groups of participants. All responses from each group were collected and gathered in txt files, one for each stimulus word. This resulted in 1 file for the L1 speakers, 1 file for the advanced L2 speakers, and 1 file for pre-intermediate group. The same programme written in Perl was used to analyse the responses in terms of frequency. To evaluate lexical diversity of the responses, AntConc software was used (Anthony, 2020). Every file containing a set of responses of each group was treated, to all effects, as a small corpus and loaded into AntConc. It was then analysed using the "word list" function that counts all the words in the corpus, presenting them in a list ordered based on the number of occurrences within the corpus. This stage was used to count the number of unique and predominant words in relation to the third research question. As the corpus was previously normalised in the scoring stage, there was no need to carry out any tagging procedure.

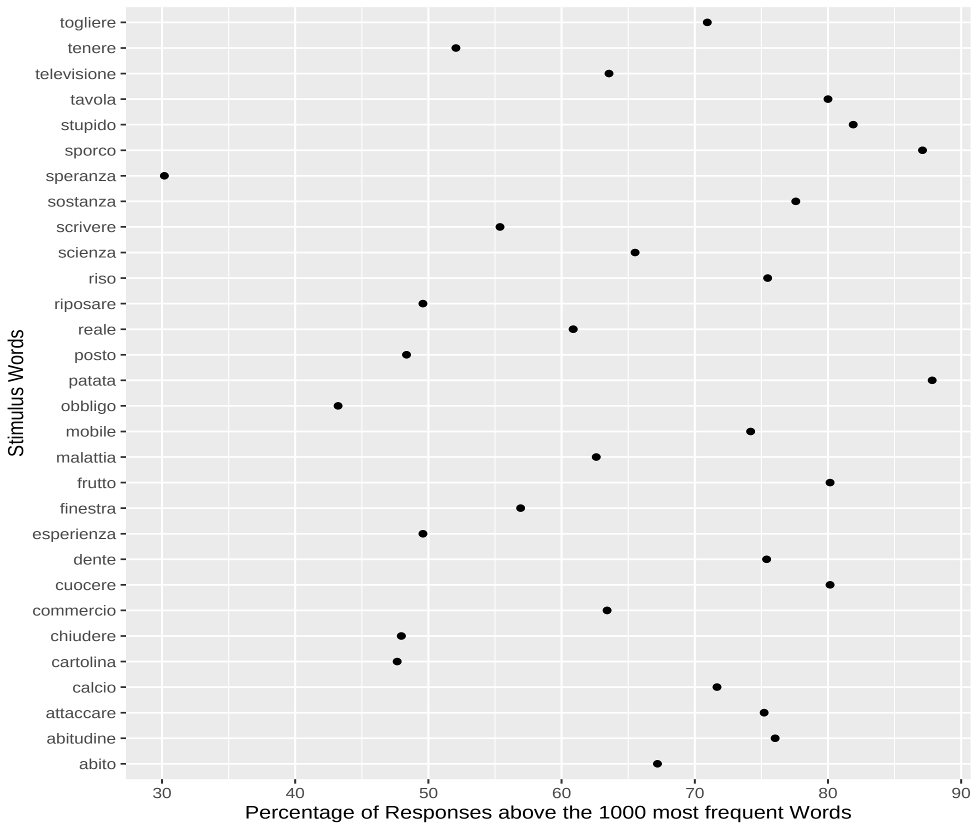
3.8.11 *Results*

Research question (1). To what extent the I-Lex's stimulus words meet the two criteria set up to generate a high number of infrequent words?

The I-Lex stimulus words did not rely on a word association list but were translated from Meara & Fitzpatrick (2000). Therefore, the results of the L1 speakers were used as a reference to evaluate whether the 30 stimulus words of I-Lex would generate enough infrequent and diverse responses. Following Meara & Fitzpatrick (2000) I first tallied the number of infrequent words produced to make sure that every stimulus word generated at least 50% infrequent responses. The results are in Figure 3.2.

Figure 3.2

Plot with the Percentage of Infrequent Responses Generated by Stimulus Words

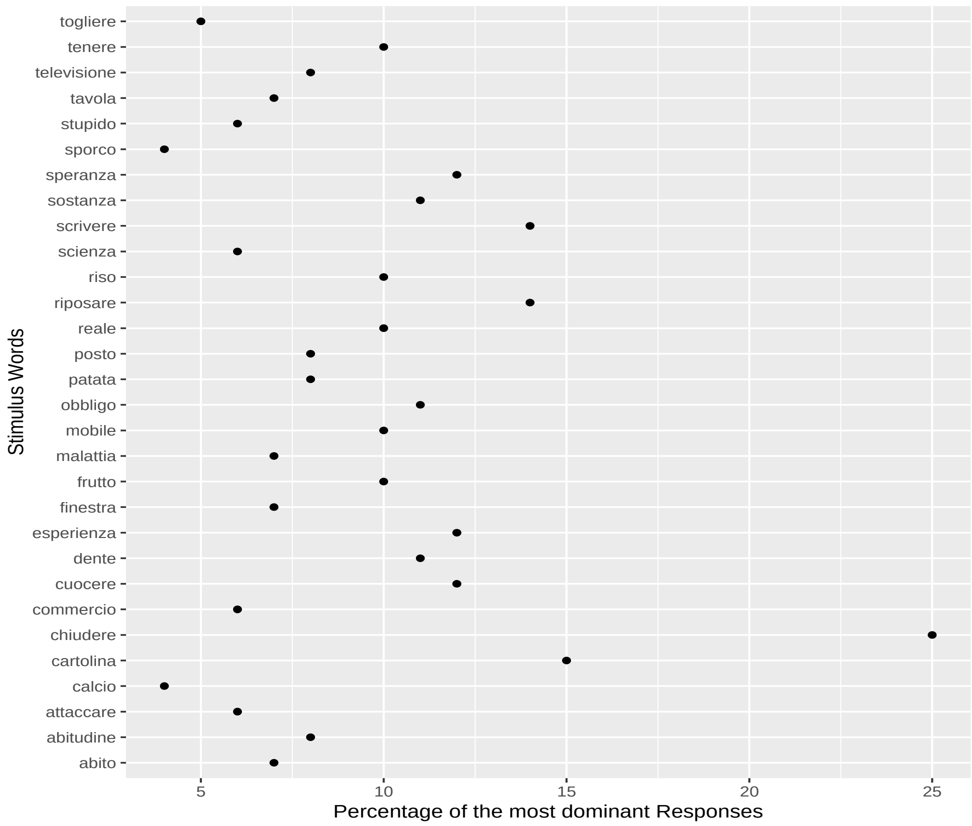


Looking at the list of stimulus words, five of them did not elicit the required number of infrequent responses. They were *cartolina* (*postcard*): 47%; *chiudere* (*close*): 48%; *obbligo* (*obligation*/*duty*): 43%; *posto* (*seat*): 48%; *speranza* (hope): 30%. However, not all of the five stimulus words were equally problematic. Three ranked close to the expected level *(chiudere*, *cartolina, posto*) while one was fairly below it, yet not extremely far (*obbligo*). There remains the case of the item *speranza* (*hope*) that ranked far below the 50% threshold. Two of them, *cartolina* (postcard) and *obbligo* (obligation/duty), are among the items that were adjusted from the original Lex30 stimulus words based on semantic relatedness.

Based on the second criterion set up in Meara and Fitzpatrick (2000) responses of the L1 speakers' group were analysed to make sure that they would not generate a single associate more frequent than the 25% of the total given responses. The results are in Figure 3.3.

Figure 3.3

Plot with the Percentage of most Dominant Responses Generated by Stimulus Words



In this case the L1 speakers met the expected criteria, with only the item *chiudere* (close) that displayed a low degree of diversity although it did not exceed the recommended threshold.

Overall, the response to the first research question is that, using the L1 speakers' group as the norm, the two criteria recommended by Meara & Fitzpatrick (2000) for the stimulus word were only partially met. The L1 group generated enough diverse responses, with only one item reaching the recommended 25% threshold without exceeding it. In terms of infrequent words cued by stimulus words, five items were problematic: less than 50% of the responses generated by them were infrequent words. Not all of them were equally problematic. Four were slightly short of the target (*cartolina* (*postcard*): 47%; *chiudere* (*close*): 48%; *obbligo* (*obbligation*/*duty*): 43%; *posto* (*seat*): 48%;) while one returned a limited number of infrequent responses (*speranza* (*hope*): 30%.). The stimulus word *chiudere* (*close*) was slightly problematic in both recommended thresholds.

Research question (2). To what extent will I-Lex produce the same results found for Lex30, thus achieving concurrent validity?

The first step in answering the research questions is to analyse the internal reliability. Following Fitzpatrick & Clenton (2010), the internal reliability was assessed by calculating Cronbach's alpha. Statistical analysis, here and in the following two research questions, was carried out using the software R, version 3.5.1 (R Core Team, 2018). The result of Cronbach's alpha (30 items; $\alpha = .82$) shows that I-Lex has an acceptable level of internal consistency. The results for all items are shown in Figure 3.4. Standard alphas are the Cronbach values, while raw alphas are the values of the overall alpha if that item is not included in the calculation. Comparing the two values helps to evaluate whether removing an item is likely to improve the overall Cronbach's α values. If the output of the analysis shows that the overall alpha value does not improve by dropping any of the items, then the set of stimulus words as a whole appears to have internal consistency (Field et al., 2012). Although some stimulus words showed values slightly below .80, results of the analysis did not return problematic stimulus words.

Figure 3.4

The Raw and Standard Alpha Values for I-Lex's Items

```
 Reliability if an item is dropped:
            raw_alpha std.alpha
abito            0.80      0.81
abitudine        0.81      0.81
attaccare        0.79      0.79
calcio           0.81      0.82
cartolina        0.80      0.81
chiudere         0.79      0.80
commercio        0.80      0.81
cuocere          0.81      0.82
dente            0.80      0.80
esperienza       0.77      0.79
finestra         0.79      0.79
frutto           0.80      0.80
malattia         0.80      0.81
mobile           0.80      0.81
obbligo          0.80      0.81
patata           0.81      0.82
posto            0.77      0.79
reale            0.81      0.81
riposare         0.79      0.79
riso             0.79      0.79
scienza          0.81      0.81
scrivere         0.78      0.79
sostanza         0.82      0.82
speranza         0.78      0.79
sporco           0.81      0.82
stupido          0.80      0.81
tavola           0.81      0.82
televisione      0.80      0.80
tenere           0.80      0.80
togliere         0.80      0.81
```

Then the I-Lex score analysis was carried out to complete the answer to the second question. The aim is to find out if a similar pattern will appear, thus providing support to the concurrent validity. I am going to first examine the data provided by the three groups using the I-lex1000 frequency list. Because all the studies on Lex30 have been examined using the 1000 band frequency list, the pursuit of concurrent validity of the I-Lex task needs to build on the same frequency band for I-Lex to be reliably compared to Lex30. Data for the L1 and L2 groups are given in Table 3.3.

Table 3.3

Absolute Scores (abl) and Percentages (%) Scores of I-Lex1000

	I-Lex1000 (abl) *M* (*SD*)	I-Lex1000 (%) *M* (*SD*)
L1 Speakers	71 (12)	66 (7)
L2 Advanced	55 (14)	56 (9)
L2 Pre-intermediate	37 (13)	46 (10)

Note. Percentage of L1 out of 108. Percentages of L2 Advanced out of 97. Percentage of L2 Pre-Intermediate out of 80.

The first analysis was carried out on the absolute scores attained by the three groups. A preliminary analysis of normality using the Shapiro-Wilk normality test showed that the scores in all three groups were normally distributed (L1 speakers: $W = 0.91$, $p = 0.12$; L2 advanced; $W = 0.9$, $p = 0.59$; L2 pre-intermediate: $W = 0.95$, $p = 0.22$). A one-way ANOVA that showed a significant effect of group ($F(2) = 48.23$, $p < 0.001$). A post hoc Tukey test (see Table 3.4) revealed that the L1 group ($M = 71$, $SD = 12$) achieved a significantly higher score than the advanced group ($M = 55$, $SD = 14$) and the difference with the pre-intermediate group ($M = 37$, $SD = 13$) showed a significantly higher score (L1s *vs* advanced L2: ß = 16.12, $SD = 2.75$, $t = 5.85$, $p < 0.001$; L1s *vs* pre-intermediate L2: ß = 33.42, $SD = 3.4$, $t = 9.81$, $p < 0.001$). The L2 pre-intermediate group ($M = 37$, $SD = 13$) showed lower scores than the advanced L2 group ($M = 37$, $SD = 13$) and the difference was significant (ß = 17.3, $SD = 2.9$, $t = 5.88$, $p < 0.001$).

Table 3.4

Post hoc Results: Absolute Scores of I-Lex 1000

	L1 Speakers	Advanced L2	Pre-Intermediate L2
L1 Speakers	-		
Advanced L2	< .0001	-	
Pre-Intermediate L2	< .0001	< .0001	-

Note. Only p values are shown

Because not all participants provided the same number of responses, I wanted to rule out the possibility that differences in the scores might be affected by the number of provided responses. Thus, scores were transformed into percentages by calculating the proportion of infrequent responses out of the total number of responses given by every participant (Table 3.3, first right-hand column). A preliminary analysis of normality using the Shapiro-Wilk normality test showed that the scores in all three groups were normally distributed (L1 speakers: $W = 0.97$, $p = 0.71$; L2 advanced; $W = 0.98$, $p = 0.77$; L2 pre-intermediate: $W = 0.98$, $p = 0.83$). A one-way ANOVA showed a significant effect of the groups (F (2) = 35.43, $p < 0.001$). A post hoc analysis with the Tukey contrast method (see Table 3.5) again showed a significantly larger number of infrequent responses given by L1 speakers ($M = 66$, $SD = 7$) than the advanced group (M = 56, $SD = 9$). The difference with the pre-intermediate group ($M = 46$, $SD = 10$) was also significant (L1s *vs* advanced L2s: ß = 9.96, $SD = 1.72$, $t = 5.06$, $p < 0.001$; L1s *vs* pre-intermediate L2s: ß = 17.78, $SD = 2.13$, $t = 8.34$, $p < 0.001$). The L2 pre-intermediate speakers ($M = 46$, $SD = 10$) gave fewer responses than the L2 advanced group ($M = 56$, $SD = 9$) and the difference was significant (ß = 8.11, $SD = 1.83$, $t = 4.41$, $p < 0.001$).

Table 3.5

Post hoc Results: Percentage Scores of I-Lex1000

	L1 Speakers	Advanced L2	Pre-Intermediate L2
L1 Speakers	-		
Advanced L2	< .001	-	
Pre-Intermediate L2	< .001	< .001	-

Note. Only p values are shown

Results of I-Lex are in line with those of Lex30, confirming that L1 speakers score consistently higher the L2 speakers (Fitzpatrick & Meara, 2004) and the L2 scores increase with proficiency level (Fitzpatrick & Clenton, 2010; Walters, 2012), The same pattern of results appeared in both absolute and percentage scores.

The next step in the analysis for the second research question was to examine the effects of the Take and Retake procedure. As a part of the Lex30 validation process, the same group of participants was tested twice to find out whether it would produce comparable results after taking the test on two separate occasions (Fitzpatrick and Meara, 2004; Fitzpatrick & Clenton, 2010). The same procedure was extended to the analysis of I-Lex1000. The Take and Retake of the I-Lex was carried out only by the pre-intermediate group. The group took the test twice, the second time after four days: the choice of four days was chosen as a compromise between the schedule of the Italian course and the three-day interval used in Fitzpatrick and Meara (2004). The take and retake procedure should reflect the ability of the I-Lex to consistently capture the same type of competence over time. If the amount of variance is the same in the two tasks there is confidence that this is the case. Table 3.6 gives the scores.

Table 3.6

I-Lex: Absolute (abl) Scores and Percentages (%) Scores of the Take and Retake

	Take (abl) *M* (*SD*)	Retake (abl) *M* (*SD*)	Take (%) *M* (*SD*)	Retake (%) *M* (*SD*)
L2 Pre-intermediate	38 (13)	42 (7)	47 (12)	46 (8)

Note. Percentage of Take: out of 80. Percentages of the Retake: out of 92

Both absolute and percentage scores were very close in both tasks, suggesting that the variation between the two tasks was very small. A preliminary analysis of normality using the Shapiro-Wilk normality test showed that the scores were normally distributed (Take: $W = 0.98$, $p = 0.83$; Retake; $W = 0.98$, $p = 0.93$). A Welch's paired *t*-test showed that the difference between the scores of the first and the second I-Lex1000 was not significant (t (29) $= 1.71$, $p = 0.09$). The fact that there was no significant difference between the two scores indicates that the number of infrequent responses given by the same students did not change and the score was consistent across both the take and retake tasks. Furthermore, a Pearson's correlation test showed a high and significant positive correlation between the two sets of scores ($r = .53$, $p = 0.002$). This means that the amount of variance in the scores of the second I-Lex1000 task can be significantly predicted from the score of the first I-Lex1000 task. Our analysis was carried out on percentages to match the results of Fitzpatrick and Meara (2004). In addition, I also analysed the data using absolute scores, that is, the absolute number of responses given by each subject (see Table 3.6). The difference between the take and the retake was again not significant (t (29) $= 1.93$, p $= 0.07$) and there was a greater significant positive correlation ($r = .65$, $p < 0.005$). To sum up, if the I-Lex task is carried out twice the number of infrequent responses remains constant and the scores correlate, thus indicating that it seems to capture the same broad type of knowledge.

There is always the risk that the similarity in the scores of the take and retake is the result of producing a large number of responses twice in both tasks (Fitzpatrick & Meara, 2004). Therefore, the responses provided exclusively during the take, those provided exclusively during the retake, and those that appeared in both the take and the retake were tallied. The results are shown in Table 3.7 and Figure 3.5. Table 3.7 also reports the results found by Fitzpatrick & Meara (2004): the patterns are quite similar, apart from a steeper increase in the present study for the retake scores.

Table 3.7

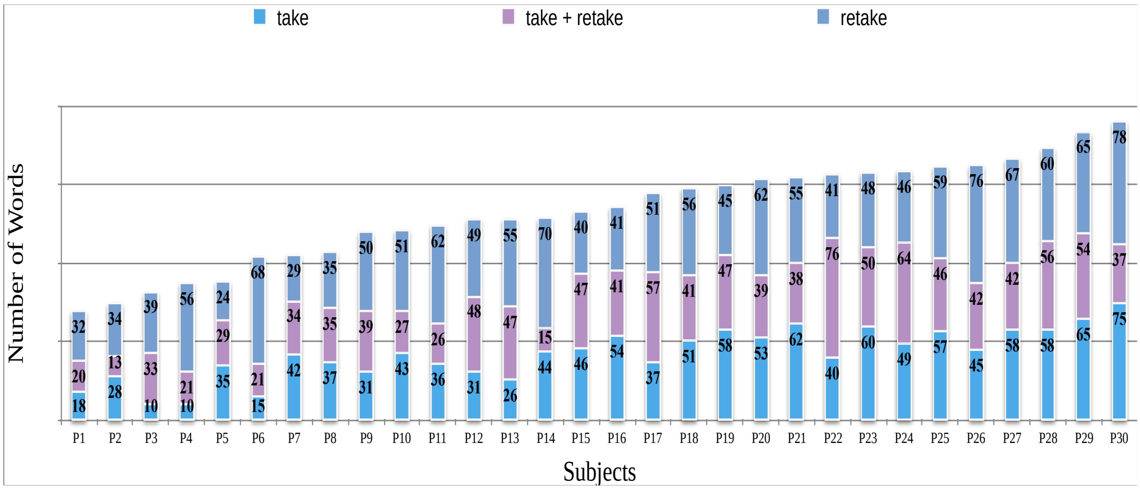
Number of Words in the Take, in Take and Retake and in Retake

	Take *M* (*SD*)	Take and Retake *M* (*SD*)	Retake *M* (*SD*)
L2 pre-Intermediate	42 (16)	39 (14)	52 (13)
Fitzpatrick and Meara (2004)	38 (18)	42 (14)	39 (15)

Overall, roughly 30% of responses are repeated across the two times of the task and approximately a third of new words are provided during the retake. Looking at the plot in Figure 3.5 this proportion holds for most participants. Hence, although a proportion of responses changes across time, nonetheless, the frequency pattern tends to remain consistent. Providing different words as responses while maintaining the same score indicates that the task is consistently tapping into the same sections of lexical networks (Fitzpatrick & Clenton, 2010; Fitzpatrick & Meara, 2004).

Figure 3.5

Number of Words in the Take (blue), the Take and Retake (purple) and the Retake (dark blue)



I also analysed the results of the scores calculated against the I-Lex2000 which corresponds to NVDB list of the 2000 most frequent Italian words. The scores of the I-Lex2000 are shown in Table 3.8.

Table 3.8

Absolute Scores (abl) and Percentages (%) Scores of I-Lex2000

	I-Lex2000 (abl) *M* (*SD*)	I-Lex2000 (%) *M* (*SD*)
L1 Speakers	43 (12)	40 (8)
L2 Advanced	29 (10)	30 (9)
L2 Pre-intermediate	18 (7)	22 (7)

Note. Percentage of L1 out of 108. Percentages of L2 Advanced out of 97. Percentage of L2 Pre-Intermediate out of 80.

I first examined the absolute scores. A preliminary analysis of normality using the Shapiro-Wilk normality test showed that the scores in the three groups were normally distributed (L1 speakers: $W = 0.98$, $p = 0.9$; L2 advanced; $W = 0.98$, $p = 0.37$; L2 pre-intermediate: $W = 0.98$, $p = 0.95$). A one-way ANOVA showed a significant effect of group (F (2) = 51.33, p < 0.001). A post hoc Tukey test (see Table 3.9) revealed the same pattern as that produced by I-Lex1000. The L1 group ($M = 43$, $SD = 12$) achieved a significantly higher score compared to the advanced group ($M = 29$, $SD = 10$). The difference with the pre-intermediate group ($M = 17$, $SD = 7$) was also significant (L1s *vs* advanced L2s: ß = 13.89, $SD = 2.09$, $t = 6.36$, $p < 0.001$; L1s *vs* pre-intermediate L2s: ß = 26.04, $SD = 2.58$, $t = 10.06$, $p < 0.001$). The difference between the pre-intermediate group ($M = 17$, $SD = 7$) and the advanced group ($M = 29$, $SD = 10$) was also significant (ß = 12.15, $SD = 2.23$, $t = 5.44$, $p < 0.001$).

Table 3.9

Post hoc Results: Absolute Scores of I-Lex2000

	L1 Speakers	Advanced L2	Pre-Intermediate L2
L1 Speakers	-		
Advanced L2	< .0001	-	
Pre-Intermediate L2	< .0001	< .0001	-

Note. Only p values are shown

As in the analysis based on the I-Lex1000, to rule out the possibility that the differences in the scores might be affected by the total number of responses provided. I also analysed the scores transformed into percentages (Table 3.8, second right-hand column). A preliminary analysis of normality using the Shapiro-Wilk normality test showed that the scores in all three groups were normal distributed (L1 speakers: $W = 0.97$, $p = 0.62$; L2 advanced; $W = 0.98$, $p = 0.62$; pre-intermediate L2: $W = 0.98$, $p = 0.93$). A one-way ANOVA showed a significant effect of group ($F(2) = 31.01$, $p < 0.001$). A post hoc analysis with the Tukey test (see Table 3.10) showed that the L1 speakers ($M = 40$, $SD = 8$) achieved higher scores than the advanced group ($M = 30$, $SD = 9$). The difference with the pre-intermediate group ($M = 18$, $SD = 7$) was also significant (L1s *vs* advanced L2s: ß = 10.04, $SD = 1.77$, $t = 5.66$, $p < 0.001$; L1s *vs* pre-intermediate L2s: ß = 16.90, $SD = 2.19$, $t = 7.71$, $p < 0.001$). The L2 pre-intermediate speakers ($M = 18$, $SD = 7$) had a lower score than the L2 advanced group ($M = 30$, $SD = 9$) that produced a significant effect (ß = 6.86, $SD = 1.89$, $t = 3.62$, $p = 0.001$).

Table 3.10

Post hoc Results: Percentage Scores of I-Lex2000

	L1 Speakers	Advanced L2	Pre-Intermediate L2
L1 Speakers	-		
Advanced L2	< .0001	-	
Pre-Intermediate L2	< .0001	.001	-

Note. Only p values are shown

The second research question asked whether I-Lex was able to replicate the same patterns found in Lex30. The answer is that it did. Like Lex30, I-Lex had an acceptable level of internal consistency, the L1 speakers score consistently higher than L2 speakers (Fitzpatrick & Meara, 2004), and the L2 scores increase with proficiency level (Walters, 2012). The scores of the L2 groups could discriminate between L2 speakers with different levels of proficiency. The pattern of results obtained with the I-Lex1000 and I-Lex200 lists overlapped, proving that both lists have the same reliability. This result is important because it demonstrates the potential skewness of the I-Lex1000 towards written texts was not problematic. Finally, in a Take and Retake task, I-Lex could again replicate the result of Lex30. Participants' scores were significantly correlated but were not significantly different. This suggests that I-Lex consistently measures the same number of infrequent words, and the score's variance in the Take predicts the variance in the Retake.

Research Question (3). Do the lexical networks generated by each group show the same patterns found in the subject analysis?

As stated earlier, in word association tasks, the number of associates produced by a group of participants can be regarded as a homogeneous lexical network (De Deyne & Storms, 2015; Fitzpatrick & Thwaites, 2020). Lexical networks have shown to be highly predictive in lexical processing tasks, arguing for a direct link between large-scale properties and word-level properties. These assumptions underpin the third research question, in which I will analyse the lexical networks produced by each of the three groups. For each group, all responses were collapsed across all participants, obtaining three lexical networks. For each network, the number of infrequent words, the number of unique responses, and the number of dominant responses were measured. To investigate each of the three networks, three main statistical models were used. The three models can be represented in the form of a regression equation, where the term on the left-hand side indicates what variable is being measured, while the term on the right-hand side indicates the mean score provided by each group in relation to the measured variable. A simplified form is given below.

- Model 1. Infrequent responses = L1 + advanced L2 + pre-intermediate L2
- Model 2. Unique responses = L1 + advanced L2 + pre-intermediate L2
- Model 3. Dominant responses = L1 + advanced L2 + pre-intermediate L2

As said, the predictors in the models were obtained by collapsing the scores of all participants within each of the three groups. To minimise the effects of the different number of responses provided by the three groups of participants, the analysis was carried out on the percentages. (For completeness, absolute scores are also given for each model). Diversity was determined by means of predominant and unique responses in each of the three networks. The absence of predominant responses and the presence of a fairly high number of unique associates indicate that the lexical networks generated by each group show a low number of repeated words, suggesting lexical diversity (De Deyne et al., 2013).

Table 3.11 gives the data of the infrequent responses generated by each group. Data suggest that the number of infrequent responses varies together with the proficiency level of participants.

Table 3.11

Number of Infrequent Responses, Absolute (abs) and Percentages (%)

	Infrequent (abs) *M* (*SD*)	Infrequent (%) *M* (*SD*)
L1 Speakers	79 (17)	65 (14)
L2 advanced	153 (46)	54 (15)
L2 pre-intermediate	40 (21)	48 (17)

Note. Percentage of L1 out of 121. Percentages of L2 Advanced out of 281. Percentage of L2 Pre-Intermediate out of 82.

A preliminary analysis of normality using the Shapiro-Wilk normality test showed that the scores in all three groups were normally distributed (L1 speakers: $W = 0.97$, $p = 0.70$; L2 advanced; $W = 0.95$, $p = 0.26$; L2 pre-intermediate: $W = 0.95$, $p = 0.23$). A one-way ANOVA showed that there was a significant effect of groups (*F* (29) = 9.84, $p < 0.001$). A post hoc analysis with the Tukey contrasts method (see Table 3.12) revealed that L1 speakers produced more infrequent responses (*M* = 65, *SD* = 14) compared to the advanced L2 group (*M* = 54, *SD* = 15) and the pre-intermediate L2 group (*M* = 48, *SD* = 17) and the differences were significant (L1s *vs* advanced L2s: ß = 10.7, *SD* = 4.15, *t* = 4.41, $p < 0.001$; L1s *vs* pre-intermediate L2s: ß = 18.3, *SD* = 4.15, *t* = 4.41, $p < 0.001$). The difference between the L2 pre-intermediate group (*M* = 48, *SD* = 17) and the L2 advanced group (*M* = 54, *SD* = 15) was not significant (ß = 7.63, *SD* = 4.15, *t* = 1.8, $p = 0.16$).

Table 3.12

Post hoc Analysis of Infrequent Responses

	L1 Speakers	Advanced L2	Pre-Intermediate L2
L1 Speakers	-		
Advanced L2	< .001	-	
Pre-Intermediate L2	< .001	.16	-

Note. Only p values are given

It seems that when proficiency is at the ceiling, as in the case of L1 speakers, the stimulus words work quite well in eliciting infrequent responses, whereas, as the level of proficiency decreases, stimulus words tend to generate fewer infrequent words. When the measurement shifts from single participants to the lexical networks of the three groups, the analysis can capture the difference between L1 and L2 speakers. However, the lack of a significant difference between the two L2 groups also suggests that the network-based measurement is not sensitive to smaller variations in proficiency, like those between the two L2 groups.

To investigate Model 2, the number of unique responses across the three groups was analysed. Unique responses are words provided only once for every stimulus word. The notion behind this analysis is that finding a fairly large number of unique words might give some more evidence of the degree of lexical diversity within each of the three groups. More unique words provided as responses can be, indeed, a good index of a more varied lexical network. Data are given in Table 3.13.

Table 3.13

Number of Unique Responses, Absolute (abl) and Percentages (%)

	Unique (abl) *M* (*SD*)	Unique (%) *M* (*SD*)
L1 Speakers	52 (8)	42 (7)
L2 advanced	82 (19)	30 (9)
L2 pre-intermediate	35 (10)	43 (17)

Note. Percentage of L1 out of 121. Percentages of L2 Advanced out of 281. Percentage of L2 Pre-Intermediate out of 82.

A preliminary analysis of normality using the Shapiro-Wilk normality test showed that the scores in all three groups were normally distributed (L1 speakers: $W = 0.96$, $p = 0.42$; L2 advanced; $W = 0.96$, $p = 0.38$; pre-intermediate L2: $W = 0.95$, $p = 0.29$). A one-way ANOVA showed that there was a significant main effect of groups (F (29) = 16.69; $p < 0.001$) and a post hoc analysis with the Tukey contrast method (see Table 3.14) showed that the L2 advanced group ($M = 30$, $SD = 9$) gave fewer unique responses than the pre-intermediate group ($M = 43$, $SD = 17$) and the L1 group ($M = 42$, $SD = 7$) and the differences were significant (advanced L2s *vs* pre-intermediate L2s: ß = - 0.17, $SD =$ 0.31, $t = -5.53$, $p <$

0.001; advanced L2s *vs* L1s: ß = 0.13, *SD* = 0.31, $t = 4.19$, $p < 0.001$). The L1 group (*M* = 42, *SD* = 7) provided almost the same number of unique responses as the Pre-Intermediate group (*M* = 43, *SD* = 17) and the difference was not significant (ß = -0.04, *SD* = 0.031, $t = -1.34$, $p = 0.37$).

Table 3.14

Post hoc Analysis of Unique Responses

	L1 Speakers	Advanced L2	Pre-Intermediate L2
L1 Speakers	-		
Advanced L2	< .001	-	
Pre-Intermediate L2	0.37	< .001	-

Note. Only p values are given

Overall, degree of lexical diversity, measured as the number of unique responses, seems to be affected by proficiency. However, the post hoc analysis gives a more complex pattern. On that the one hand, the fact the L1 speakers provided a significant higher number of unique responses suggests that proficiency could be a factor. On the other hand, the fact that L2 pre-intermediate group produced a higher percentage of unique words than the advanced group suggests that proficiency might not be the only explanation. Perhaps unique responses might not be informative enough about the structure of the network or L1 speakers tend to have more homogeneous lexical networks than L2 speakers (De Deyne et al., 2013).

Lexical diversity was also measured in terms of the absence of predominant responses. The threshold is the same applied to L1 speakers, that is, dominant responses should not exceed 25% of the total. Table 3.15 shows the data relative to the three groups.

Unlike the figures for the infrequent responses, proficiency means, and standard deviation are almost the same across the three groups.

Table 3.15

Number of Dominant Responses, Absolute (abl) and Percentages (%)

	Dominant (abl) *M* (*SD*)	Dominant (%) *M* (*SD*)
L1 Speakers	34 (8)	9 (4)
L2 advanced	46 (12)	9 (3)
L2 pre-intermediate	18 (6)	10 (3)

Note. Percentage of L1 out of 121. Percentages of L2 Advanced out of 281. Percentage of L2 Pre-Intermediate out of 82.

A preliminary analysis of normality using the Shapiro-Wilk normality test showed that the scores in all three groups were normally distributed (L1 speakers: $W = 0.96$, $p = 0.51$; L2 advanced; $W = 0.89$, $p = 0.07$; L2 pre-intermediate: $W = 0.85$, $p = 0.08$). A one-way ANOVA that did not show a main effect of groups (F (29) = 0.792; $p = 0.45$). Since the absence of dominant responses is a sign of lexical diversity and, in turn, lexical diversity is a sign of complexity, the analysis does not provide support for an effect of proficiency.

3.8.12 *Discussion*

The first research question concerned the reliability of the I-Lex stimulus words. The analysis was carried out on the L1 speakers' group and showed that the vast majority of the stimulus words generated a sufficient number of infrequent responses (i.e., words over the 50% threshold established by Meara & Fitzpatrick (2000)). Yet five stimuli fell short of meeting the frequency threshold: they were *cartolina* (*postcard*), *chiudere* (*close*), *obbligo* (*obbligation*/*duty*), *posto* (*seat*), and *speranza* (*hope*). Interestingly, none of the five stimulus words reported above appeared to increase the Cronbach's α values if removed (see Figure 3.4). Removing any of problematic stimulus words, *cartolina* (*postcard*), *obbligo* (*obligation*/*duty*), *chiudere* (*close*), *posto* (*seat*), and the critical *speranza* (*hope*) lowers the alpha level. This suggests that although they did not produce enough infrequent responses, the five stimulus words did not affect the internal reliability of the I-Lex task.

A key point to make is that, by tallying the absolute number of expected infrequent responses, the difference between four of the problematic words (*cartolina* (*postcard*): 47%; *chiudere* (*close*): 48%; *obbligo* (*obbligation/duty*): 43%; *posto* (*seat*): 48%;) and the other stimulus words is low. Considering the 36 participants that formed the L1 group, each of the stimulus words could potentially elicit 144 responses (i.e., four potential words for 36 participants), out of which at least 72 words (i.e., 50%) should be infrequent words. By changing the percentages of the problematic stimulus words into actual scores, four of them fell short of the expected threshold only by a few responses. Namely the stimulus word *cartolina* (*postcard*) fell short of 4 words (68 infrequent words against the threshold of the 72 required infrequent words), *chiudere* (*close*) fell short of 3 words (68 infrequent words against the threshold of the 72 required infrequent words), *obbligo* (*obbligation/duty*) fell short of 10 words (62 infrequent words against the threshold of the 72 required infrequent words), and *posto* (*seat*) fell short of 3 words (68 infrequent words against the threshold of the 72 required infrequent words). The only world that scored far from the expected threshold was *speranza* (*hope*) marking a 28 words difference (44 infrequent words against the threshold of the 72 required infrequent words). In this respect, only one word seems not to be well-suited to elicit infrequent responses. As will be discussed in the Limitations (see Section

3.8.13) further experimental analysis is needed to evaluate the problematic item more thoroughly.

Furthermore the analysis showed that there were no dominant responses that exceeded the recommended threshold of 25%, with only one item reaching the threshold without exceeding it (Meara & Fitzpatrick, 2000). Importantly, the five problematic items did not generate predominant responses, showing that, overall, they still could provide lexical diversity, although they did not provide enough infrequent responses. Overall, the L1 speakers' groups provided a reliable word association norm that met, almost entirely, the criteria already established for Lex30 (Meara & Fitzpatrick, 2000).

Concerning the second research question, the I-Lex task managed to replicate the effects that appeared in the studies carried out to validate Lex30. First, the Cronbach alpha showed that the internal validity of I-Lex was over .80 comparable with the value in Fitzpatrick & Clenton (2010). Second, the scores obtained by L2 speakers were significantly lower than the ones attained by L1 speakers (Fitzpatrick & Clenton, 2010; Fitzpatrick & Meara, 2004; Walters, 2012). Furthermore, the advanced L2 group scored significantly better than the pre-intermediate L2 group. Based on these results, it is safe to say that the I-Lex score increases along with proficiency with L1 speakers achieving the highest scores. A similar relationship between proficiency and productive vocabulary knowledge was found for Lex30 (Fitzpatrick & Clenton, 2010; González & Píriz, 2016; Moreno Espinosa, 2009; Walters, 2012). In some studies, proficiency was directly measured in different ways, such as improvement after instruction (Fitzpatrick & Clenton, 2010) and improvement caused by CLIL (González & Píriz, 2016; Moreno Espinosa, 2009). In the case of the present validation study, proficiency was not measured directly.

The results of the take and retake reproduced the same pattern found in Fitzpatrick & Meara (2004), with a non-significant difference between the first and the second administration of the task. This shows that the amount of variance did not differ between the take and the retake of the task, and the subjects gave the same proportion of infrequent words on both occasions. Furthermore, the analysis of the scores showed that roughly one-third of responses was shared, one-third was unique to the Take, and roughly one-third was unique to the Retake. The pattern coincides with the one found in Fitzpatrick & Meara (2004), suggesting that I-Lex, like Lex30, is capable of tapping into the same type of lexical knowledge on separate occasions. Another important point examined in the second research question concerned the difference between the scores with the two frequency lists. The comparison was between the scores generated by the I-Lex1000 list and those generated by the I-Lex2000 list. Hence, in addition to the I-Lex1000 list, I analysed the scores of the I-Lex2000 list, which contains the NVDB list with the 2000 most frequent Italian words. The results mirrored the scores achieved with the I-Lex1000 list.

The third research question analysed the lexical networks produced by each group. This type of analysis is different from the analysis of the scores. Each lexical network contains all the responses provided by each of the three groups; the resulting lexical networks reflect the lexical representations shared by the participants of each group. The assumption behind this analysis is that lexical networks, particularly those derived by word association, are sensitive to word-level lexical properties (De Deyne et al., 2013, 2019; De Deyne & Storms, 2008a; Meara, 2009; Steyvers & Tenenbaum, 2005; Wilks & Meara, 2002). The present study could only focus on very general measures about the density of the networks, as the numbers were extremely small to extract large-scale measurements. However, some important patterns appeared.

The first measurement taken on the lexical networks of the three groups counted the number of infrequent words. Chapter 2 showed evidence that the number of infrequent words indicates the complexity of the network. The most frequent words are at the centre of the network and are densely interconnected, while infrequent words are added to differentiate meaning relations (De Deyne & Storms, 2008a; Steyvers & Tenenbaum, 2005). This means that networks are meant to increase as a function of proficiency and L1 speakers should have more connections with infrequent words. These assumptions gained some partial support from the analysis of the infrequent words in the three lexical networks. The analysis showed that the number of infrequent responses in each network was significantly predicted by the level of proficiency of the groups, at least in the case of the L1 and L2 speakers as a whole. However, the difference within the L2 group was not significant. This might be due to the disparity of the two groups (30 versus 82) that could affect the dimension of the lexical networks generated by each group.

The next analysis carried out on lexical networks aimed to gauge the effects of the three groups on the absence of predominant responses. The analysis did not show any significant effect of proficiency. Each of the three lexical networks was able to produce a good level of lexical diversity. Dominant responses are nothing short of the number of the most frequent responses given to a stimulus word. Although in this study they did not reflect proficiency at all, they seem to be a sensitive measure. De Deyne et al. (2019) found that the most frequent responses showed high correlation with very fine-grained lexical measures, like a lexical decision, a picture naming task, and semantic decision task. It is quite possible that the three lexical networks constructed in this study were not large enough. Looking at the numbers, the rates of the predominant responses were extremely low compared to those of De Deyne et al. (2019).

The proportion of unique words can also be regarded as a close reflection of the complexity of lexical networks. Similar to the pattern found for infrequent words, the number of unique words partially correlated with proficiency, as the difference between the L1 speaker group and the L2 groups was significant. However, proficiency did not appear to be the main factor between the two L2 groups. Only in the case of participants with the widest lexical network, the L1 speakers, the effect of vocabulary knowledge appears to be a crucial factor behind the complexity of their lexical networks. As noted previously, the lack of significant difference between the L2 groups and the L1 group could be due to the numerical difference between the groups. The extremely high variation displayed by the L2 pre-intermediate group (*SD* = 17%) suggests that individual differences might have overshadowed the overall effects of the group variance. However, the significance of the infrequent words in predicting proficiency effects between the L1 and the L2 group reveals that examining small lexical networks formed by word association tasks can give some fresh insight into vocabulary knowledge.

Analysis of the results demonstrates that when the proficiency of L2 speakers is close to that of L1 speakers, the higher the number of infrequent and unique words that can be found in their lexical networks (Meara, 2009; Wilks & Meara, 2002). The role of infrequent words in lexical networks and the overall results produced by I-Lex clearly support the idea that frequency is a reliable predictor to differentiate between L1 and L2 speakers. It marks a clear difference when used to gauge the performance of the participants and when used to analyse the properties of lexical networks produced by each group. Since L2 speakers must deal with two languages at the same time, they rely on weaker frequency representations. Hence, their ability to recall words is going to be largely based on the words that they encounter - and use - more frequently, chiefly high-frequency words (Gollan et al., 2011). Second, the analysis of the lexical networks of the three groups shows that, overall, the lexical network of L2 speakers has fewer infrequent words and a lower degree of lexical diversity, as measured through unique words. This relation between frequency and aspects of vocabulary knowledge ties in with the Lexical Entrenchment Hypothesis described in

Chapter 2. The hypothesis argues that the effects of frequency depend on vocabulary knowledge (Diependaele et al., 2013).

One can speculate that, from the organisation of lexical networks to the I-Lex scores, what differentiates L1 and L2 speakers is the effect of frequency on the availability of lexical features. Unsurprisingly then the low-proficiency L2 speakers could not achieve high scores on the I-Lex and showed the least organised networks as a group, of course L1 speakers showed the reversed pattern. The results of the study show the importance of frequency as a measure of lexical representations. The fact that individual scores significantly reflected proficiency shows that the number of infrequent words known by L2 speakers increases with their proficiency. In addition, the results also show that frequency underpins the growth of lexical networks, facilitating lexical access and the subsequent activation of links between words. To deal with this properly, more detailed lexical representations are needed, though. This is what the next chapter aims to do.

3.8.13 *Limitations*

The present study has some limitations. First, due to restrictions in data collection, a parallel version of I-Lex could not be tested to add strength to its reliability of the I-Lex. This analysis, as part of the validation process, was instead carried out by Fitzpatrick & Clenton (2010) using a parallel version of Lex30. For the same reason, it was not possible to compare the I-Lex task with other productive vocabulary tasks in Italian. This comparison was indeed present in the first validation study (Meara & Fitzpatrick, 2000) and has been used since in other studies (Fitzpatrick, 2010; Fitzpatrick & Clenton, 2017; Fitzpatrick & Meara, 2004). In addition, five of the stimulus words turned out to be slightly short of the threshold set up by (Meara & Fitzpatrick, 2000), although not to a great extent. In effect, only one item was well below the threshold of 50% of infrequent responses. However, all the items performed well in terms of internal reliability. Scaling up the group of L1 speakers in a future study will help shed light on the issue, showing if a larger number of participants will improve their elicitation power of the five items. Two of the items were not directly translated but adjusted from the Lex30 stimulus list, based on the use of Frame Semantics, namely, *cartolina* (*postcard*) and *obbligo* (*obligation*/*duty*). A further study with more participants could help to determine whether the two words were not well chosen. Overall, the impact of these five items was relatively small, and the final decision was to keep them in the current version of I-Lex.

Although the normal distribution in each group showed homogeneity, their different sizes might have overshadowed more nuanced statistical effects. The analysis based on the lexical network formed by combining all responses of each group would have benefited the most from groups of the same size. In particular, the disparity between the low-intermediate and the advanced group might have affected the results. Overall, the expected patterns managed to appear, but adding a substantial number of participants in the three groups would certainly improve the results.

One problem in the study was that the proficiency of the L2 groups could not be tested independently. Clear information was still available, as both L2 groups had achieved a language certificate before carrying out the task. Nonetheless, the possibility of measuring other aspects of their linguistic knowledge would have made it possible to widen the reliability process. Several studies have found correlation between Lex30 and other linguistic abilities (e.g. Uchihara et al., 2020; Uchihara & Saito, 2019). However, two comparisons with on-line processing task, are going to be present in the next studies on language processing (Chapter 6 and Chapter 7). The results will be commented in those chapters.

3.8.14 ***Conclusions***

The chapter started with the hypothesis that frequency is an effective tool for measuring vocabulary knowledge. The results have supported this hypothesis. Some important evidence that the lexical networks can reflect other aspects of vocabulary knowledge has also appeared and is linked to frequency; see the significance of infrequent and unique words as predictors of proficiency. The ideas that were developed at the end of Chapter 2 about the importance of lexical knowledge and frequency have gained some important support. The next chapter seeks to give a solid theoretical background to importance of lexical knowledge for syntactic representations. It will examine a theoretical framework that is entirely based on lexical features and views syntactic processes as lexically driven.

Chapter 4 HPSG overview

4.1 Introduction

In Chapter 2 I have shown that lexical entries can encode a great deal of information spanning from meaning and form features to syntactic features. Features, in turn, may encode complex patterns of information. In Chapter 2 and Chapter 3 I have also introduced the notion of feature availability: when lexical features are not readily available, activating a word in the mental lexicon becomes difficult. Then I developed the frequency-based vocabulary measurement I-Lex with the aim of capturing the complexity of lexical representations in Italian. Taken together, these notions strongly suggest that (1) a L2 language processing model needs to integrate all the aspects that are inherent to lexical representations and that (2) those aspects can be measured. In this chapter I explore these two points by describing HPSG (Head Driven Phrase Structure Grammar), a theoretical framework that puts grammatical information into lexical entries, i.e. it is lexically driven (Abeillé & Borsley, 2019; Boas & Sag, 2010; Flickinger et al., 1983, 1985; Pollard & Sag, 1994; Sag, 1997, 2010; Sag et al., 2003). HPSG theory aligns with connectionist models, which assume that all the linguistic information is lexically encoded and becomes available incrementally following the input stream (e.g., MacDonald et al., 1994; McRae & Matsuki, 2013; Spivey-Knowlton & Sedivy, 1995). HPSG preserves the same assumptions as connectionist models, thus offering an exhaustive framework to formalise language descriptions.

The first section of the chapter deals with the origins and the implementation of the Head Driven Phrase Structure Grammar showing how the key features that underlie its architecture were developed and implemented whilst paying close attention to the characteristics of language processing. Section two presents the basic notion that underpins

HPSG outlining the key properties of the theory: it is feature-based, it is lexically driven, and it is based on unification of compatible features - i.e., features that carry non-conflicting information. Section three examines some practical applications of HPSG to some sentence processing phenomena for both the L1 and L2. The last section describes how HPSG handles filler gap dependencies. After a general analysis of the features needed to form relative sentences, it focuses on the type of filler gap dependencies used in the experiments described in Chapters 6 and 7, namely, cleft sentences and oblique relative clauses.

4.2 Processing and the Design of Grammar.

Head Driven Phrase Structure Grammar (henceforth HPSG) is a theory developed by Carl Pollard and Ivan Sag in the mid-1990s (Pollard & Sag, 1994). The theory resulted from the combined efforts of a group of linguists and computer programmers to develop a formalism that would be useful for both fields. Once the foundations of the formalism were set up, the group of linguists went ahead in their endeavour to write a grammar of English (see Flickinger et al. 2019 for an historical overview). In the course of this further implementation, the theory proved to be well suited to describe some major characteristics of language processing (Sag et al., 2003; Sag & Wasow, 2011, 2014).

In developing HPSG, its proponents especially focused on three key properties of language processing, namely that it can deal with partial information extremely quickly, it relies on different sources of linguistic information, and it depends on the information encoded into lexical entries (Sag et al., 2003; Sag & Wasow, 2011, 2014).Examining these characteristics is pivotal to gain insight into the significance of choices that are behind the architecture of HPSG, especially in its more recent developments (e.g., Sag, 2010; Sag et al., 2003).

In regard with the first characteristic of language processing, according to Sag & Wasow (2015) the information flood to which language users are exposed contains a great deal of ambiguity and it is often underspecified. For this reason, the interpretation of an utterance hinges on information that, most of the time, tends to be incomplete. Ambiguity may affect, for instance, the meaning of words or the structure of the sentence itself. In Chapter 2, we dealt with cross-linguistic homographs showing that they can be ambiguous for the same form refers to two unrelated meanings. This is not, of course, a phenomenon limited to bilingual representations, as it is widespread in language in general. The structure of sentences, even when it appears to be fairly simple, can also be a source of difficulty (Sag & Wasow, 2014). Example 4.1 shows a case of structural ambiguity: *good* might either refer to *red wine* or be the modifier of the verb form *tastes*.

4.1 I almost forgot how good red wine tastes

The second aspect of the 'information flood' is under specification, which refers to linguistic representations that have a low degree of specificity and need to be interpreted in relation to the context (Sag & Wasow, 2014). In Examples 4.2 and 4.3 the entire sentence must be processed to decide if the word *book* refers to a physical object or to the abstract information contained in it. In this case, the lexical representation of the word *book* is not specific enough to draw a clear difference between the two meanings associated with it. What is missing in the lexical representation of the word is provided by the context.

4.2 John could not understand the book because all its pages were torn

4.3 John could not understand the book because he disagreed with the theory

Language users are normally fast at navigating throughout the different levels of ambiguity and extremely skilled in dealing with under-specification (Sag et al., 2003; Sag & Wasow, 2014). The fact that potentially incorrect and inaccurate interpretations are quickly and easily rejected and that comprehenders rarely makes mistakes in resolving ambiguity and under-specification argues for a processing system that is able to successfully cope with partial linguistic information (Sag & Wasow, 2014).

The second property of language processing that HPSG integrates and implements into grammar is that processing a string of language requires dealing with different sources of information at the same time (Sag et al., 2003; Sag & Wasow, 2011, 2014). Apart from grammatical knowledge, factors such as discourse context, gestures, and other visual information along with general knowledge about the world and culture are crucial in processing sentences (Sag & Wasow, 2014). In Example 4.4 *that thing* can be interpreted only if the speaker is aware of what has been said in the previous discourse segments. Regarding general knowledge example 4.5 requires some familiarity with the western movies and one of its most iconic actors, *John Wayne*, in order to understand that the comparison highlights the strength and the toughness of *Andrew* or that he is dressed like a cowboy,

4.4 That thing you said before, I wouldn't bring it up now!

4.5 Andrew, you look like John Wayne!

This shows that the number of resources needed to interpret language stretches far beyond syntactic and general semantic information. The representations of linguistic expressions must be built to enable the encoding of as much information as possible.

The third property of language processing is that all the information needed to successfully comprehend a sentence lies in the words themselves. In HPSG words are regarded as the atoms of the grammar (Pollard & Sag 1994; Sag et al., 2003). The amount of information a word can contain may be extremely rich. In Chapter 2, we provided evidence that words encode information about which words they can combine with and which structure(s) they appear in. We have also shown that they also contain fine-grained information about their meaning and the variety of links with the meaning of other words, even words that belong to a different language (e.g., Van Hell & de Groot, 1998). Along the same lines, Sag & Wasow (2014) stated that "words have phonological and orthographic forms and literal meanings. They also have associated subcategorization information, that is, information about what types of expression they combine with." (p.53)

It is important to note that the properties of language processing that HPSG has implemented into the grammar have been singled out over the years by other frameworks as well. The fact that processing is rapid in integrating partial information is not specific to the HPSG. It is widely assumed that the interpretation of sentences occurs on the basis of the partial information made available within the input in an incremental fashion. Van Gompel (2013), after reviewing four decades of studies on L1 processing, argues that:

> the last 40 years or so have seen major advances in our understanding of how language users process sentences. It is clear that syntactic processing is highly incremental and that various non-syntactic sources of information generally have rapid effects on syntactic ambiguity resolution. Even many aspects of semantic interpretation are very quick. (p. 4)

Furthermore, a wide body of research has granted wide support to the assumption that readers and listeners use syntactic information along with additional sources of information such as plausibility, frequency, and lexical preferences (e.g., Ellis et al., 2014; Pickering &

Traxler, 2006; Reali & Christiansen, 2007; Tily et al., 2010; Traxler & Pickering, 1996). Discourse context, although associated with higher-level processing based on the individual's personal knowledge, is often applied automatically as syntactic processing (e.g., Altmann & Steedman, 1988). For example, Weissman & Tanner (2018) showed no difference in ERP effects between ironic emojis - regarded as pragmatic markers of discourse context - and literal utterances. However, the notion that all sources of information can be associated with lexical entries is rather controversial. Several theories posit that in addition to syntactic information encoded in the lexical entries, there are certain abstract syntactic features that are not part of lexical representations at all, but are rather linguistic universals or principles, e.g., recursion (Adger, 2003; Chomsky, 1995; see Müller, 2016 for review). However, on a general level, the assumption that syntactic and semantic information, discourse information, and general knowledge information are largely dependent on and encoded into words is widely agreed (Altmann & Steedman, 1988; MacDonald et al., 1994; McRae & Matsuki, 2013).

Despite differences on the specific type of lexically encoded information, HPSG assumptions about language processing are related with the main conclusions about language processing from previous research. Owing to this, HPSG has adopted the view that since every source of information can be exploited by the processing system, linguistic representations must be flexible enough to encode all kinds of information.

4.3 A Processing Compatible Grammar.

Since its inception HPSG has always focused on the implementation of grammatical descriptions that could also suit properties of language processing in the sense outlined in the previous section (e.g., Sag et al., 2003; Sag & Wasow, 2014; Wasow, 2019). In this respect Sag et al. (2003) explicitly stated that HPSG encompasses exactly the same components as those that are at the core of the constraint-based model developed by MacDonald et al. (1994). In the model, like in HPSG, all the components of the linguistic representations used by the processing system are lexically encoded.

HPSG has been extensively exploited in computational implementations (e.g., Bender & Emerson, 2021; Copestake, 2002; Müller, 2015). However, the implementations directly related to language processing phenomena have not been pursued to the same extent as those in computer science. Nevertheless, some notable efforts in this direction have been made in the last two decades, in terms of first language acquisition (Green, 2011), processing of long-distance dependencies (Chaves, 2013), analysis of spoken language (Lücking et al., 2019) and second language acquisition (Mellow, 2004). Before turning to the specifics of these studies, the three key characteristics that HSPG proponents singled out for a grammar to be processing compatible will be examined. A grammar firmly grounded in language processing has to be surface oriented, constraint-based, and strongly lexicalist (McRae et al., 1998; Sag et al., 2003; Sag & Wasow, 2011, 2014; Vosse & Kempen, 2000; Wasow, 2019).

The first of the three properties needed for a processing-compatible grammar is surface orientation. Surface orientation means that the grammar should be able to associate structures directly to words and string of words in the same order as they are heard or read. This account for the fact that processing is incremental and, most of the time, relies on partial information (Sag et al. 2003; Sag & Wasow 2011; Sag & Wasow 2015). As Sag & Wasow (2015) noted, “comprehenders begin determining the intended interpretations of expressions as soon as they are uttered, without waiting for the grammatical analysis of any subsequent material” (p.51). While comprehending a sentence, individuals build partial hypotheses that are evaluated in terms of how they fit into the segments already interpreted. HPSG constructs language representations that may provide partial information that can be effectively used to assemble words and expressions as they are encountered by the processing system. Example

4.2 above shows that, for instance, the initial interpretation of the lexical entry *book* starts as soon as the word is comprehended (although this may be prior to the word being completed (see cohort model of Marsen-Wilson (1987)) and, once the new segments of the sentence are processed, the interpretation is accordingly refined. A surface-oriented grammar must provide lexical representation of the word *book* that can be general enough but easy to make more specific as the new input requires (see Section 4.9 below for lexical representations).

The second property of a processing-compatible grammar relies on the notion of constraint. The term constraint refers here to bits of information that are encoded into a linguistic expression (McRae & Matsuki, 2013; Sag et al., 2003). A key property of constraints in HPSG is that the different bits of information they contain are directly determined by the type of linguistic expressions to which they are associated. For example, a verb will be constrained by factors such as number, person, tense, and the type of process it encodes, while tense and person will not be proper constraints for nouns. The constraint-based property of the grammar combines with the property of being surface oriented. Given that lexical information is normally made available following the order of the input, various sources of information encoded in lexical entries may not necessarily be used at the same time and in the same order. Hence, some constraints that are relevant in order to combine two linguistic expressions might not be as relevant when new expressions become available (MacDonald et al., 1994; McRae & Matsuki, 2013). More evidence is also provided by several studies using ERP methodologies. For example, studies have shown that, in speech production, the access of semantic and phonological information is not strictly ordered (with phonological information coming after semantic information) and both sources of information may even be consulted in parallel (Abdel Rahman et al., 2003; Abdel Rahman & Sommer, 2003; Guo & Peng, 2007). Two studies have also shown equally unordered and task-dependent access for morphological and phonological information for L1 (Shantz & Tanner, 2017) and L2 speakers (Shantz & Tanner, 2020).

The final property of a processing-compatible grammar is that it is strongly lexicalist. This means that all various sources of information needed to interpret and produce sentences are encoded into lexical entries. When constraints are entirely specified in the lexicon, a grammar only needs to account for which type of expressions a word combines with and which type of meaning relations arise between a word and the words that precede and follow it (Sag et al., 2003; Sag & Wasow, 2014). Hence, the processing system merely needs a few broad-scope grammar rules whose purpose is limited to assembling the right constraints as soon as they are integrated, without positing any predetermined structure.

In the following sections, I am going to show what types of features are encoded in lexical items and how the HPSG grammar is based on simple rules that simply combine lexical features into grammatical patterns. I will focus on the features needed for the HPSG analysis of the language processing phenomena, i.e. Filler-gap dependencies, that I will present in the empirical study in Chapter 6 and Chapter 7. Hence, the section is not intended as an exhaustive overview of HPSG (see Sag et al. (2003) for a complete introduction). However, while the number of types and associated features employed in HPSG is larger than the one I will show, the basic features of the grammar have been held constant over the years (Müller, 2016).

4.4 Basic Concepts in HPSG: Features and Types

HPSG breaks down every grammatical category into all its components and seeks to define how these components might generate grammatical expressions. The components are called features: each feature carries a piece of information that is technically referred to as value. Features are used extensively to cover a great deal of phenomena: there may be features for words, features for semantic representations, features for phrases, features for specific syntactic structures, and even features for describing rules (Sag, 1997, 2010; Sag et al., 2003). To arrange features, HPSG makes use of the notion of type. A type is the label given to a set of features that describe a specific property of linguistic expressions. The purpose of using types is to group features that are capable to describe all the aspects of each linguistic expression (Müller, 2016; Pollard & Sag, 1994; Sag et al., 2003). Having an architecture of this type at the basis of the grammar "amounts to the beginning of a linguistic ontology: the types lay out what kinds of linguistic entities exist in our theory and the features associated with those types tell us what general properties each kind of entity exhibits" (Sag et al., 2003, p. 59).

It is not the scope of this chapter to provide a complete and exhaustive arrangement of all the types commonly used in HPSG. However, a brief description of the most basic types and related features can provide a good insight into how they work. Some types are commonly used and they clump together features that describe very general properties, such as *word*, *phrase*, and part of speech like *noun*, *verb,* and *adjective,* and so forth. Two types, *syn-cat* and *sem-cat* are important in HPSG. The type *syn-cat* (syntactic category) is the label for all those features used to describe syntactic representations and structure-based relations between expressions, whereas the type *sem-cat* (semantic category) is the label given to the features used to describe semantic representations and meaning-based links between expressions (Sag et al., 2003). Types can also label features that describe syntactic operations. We focus on two types, *agr-cat* and *val-cat*. The former contains all the features that are needed for two expressions to agree, while the latter contains all the features that describe the sort of syntactic structure in which a word may appear.

Having features arranged into specific types captures the complexity and variety of all the specific properties displayed by linguistic expressions and in terms of differences and commonalities at the level of meaning, form, and syntactic structure (Müller, 2016; Sag et al., 2003). The next section provides an example of such a type-based arrangement of features, showing how it may work when put into practice.

4.5 A Sample of a Fully Operative Feature-Based Grammar

Having the most basic types to describe linguistic expression in place, it is necessary to add all the specific features for every type. Table 4.1 gives an overview of the types with their associated features. Before examining the details shown in the table, it is important to bear in mind that the information encoded in each feature is technically referred to as its 'value', and we will use the term with that meaning in the rest of the chapter.

Table 4.1

A simple HPSG Grammar

TYPE	FEATURES
word	ARG-ST
syn-cat	HEAD
sem-cat	RELN, MODE, INDEX, RESTR
agr-cat	PER, NUM, GEND
val-cat	SPR, COMPS, MOD
sign	PHON, SYN, SEM
phrasal-construction (p-cx)	MTR, DTR, HD-DTR

Note. Adapted from Sag et al., 2003, p. 273

The feature that is associated with the first type *word* is ARG-ST which stands for argument structure: it is the feature that encodes the combinatorial potential of every word.

As every lexical entry encodes information about what kind of other expression it may occur with, ARG-ST is associated by default with every word. The feature ARG-ST has a value that consists of a list of expressions. The customary way to show the value of a feature is to use angle brackets (< >) for lists and to directly associate the value when it is a single object. Therefore, the feature ART-ST of a verb like *take* will have, for instance, a list with two elements: the subject NP and the object NP. The notation for a transitive verb will be ARG-ST < NP, NP >. A preposition will have one single element with which it combines, so its ARG-ST list will be ARG-ST < NP >. In HPSG the list of the feature ARG-ST is ordered in terms of obliqueness, with the subject preceding the object and the object preceding all other roles. This order reflects the Keenan-Comrie Hierarchy (Keenan & Comrie, 1977) based on the assumption that subjects and objects are more important, that is less oblique, than indirect objects and various kinds of adjunct. The feature ARG-ST does not directly refer to linear order, although the information it contains and the order of the words in the sentence tend to overlap. The linear order is determined by the type of phrasal constructions, as Sections 4.6 and 4.8 show.

Both the type *syn-cat* and the related feature HEAD contain information about syntax. The feature HEAD specifies the category to which an expression belongs, thus providing key information about the meaning relations and the structural links between the head and its dependents. For instance, a phrase headed by a verb may describe complex events arranged in complex syntactic structures, while a phrase headed by a preposition describes basic relations between expressions (e.g. Kroeger 2005). Regarding the actual values to the feature HEAD, it will suffice to draw on the labels normally assigned to parts of speech: *adjective, preposition, adverb, conjunction, verb, noun,* and *determiners* (Sag et al., 2003).

More specific information about the syntactic structure an expression may be a part of is provided by the two features SPR (specifier) and COMPS (complement) associated with the type *val-cat*. The feature SPR encodes both the subjects of verbs and the determiners of nouns, while the feature COMPS encodes all the complements that a word can combine with. The other feature associated with the type *val-cat* is the feature MOD that encodes information about the expression it modifies. Modifiers, as much in HPSG as in other current grammatical theories, are expressions that carry ancillary meaning added to the meaning of the phrase as a whole (Sag et al., 2003). However, it is not always clear how to draw a line between complements and modifiers. It is understood that complements correspond to the dependents that are directly selected based on the event described by the verb or the entity referred to by the noun. Unlike complements, modifiers carry information that is not directly selected by the head. However, modifiers cannot occur freely and there are restrictions on what kind of expression a modifier can modify: an adjunct, for instance, can modify verbs but cannot modify nouns. The head word seems to exert some sort of influence on the possibility of taking modifiers, although it is generally weak and limited in scope (see Bouma et al., 2001 for an HPSG review). The type *agr-cat* (where *agr* stands for agreement) accounts for the fact that the head may agree with their dependents. In Italian, for instance, verbs agree with their subjects and nouns agree with their determiners, as shown by Example 4.6.

4.6	I	bambin-i	gioc - ano
	the-3rd-plural	child-3rd-plural-masculine	giocare-3rd-plural-present
	The children	play	

The form of the article *i* agrees with the plural form of the head noun *bambini* which, in turn, agrees with the verb *giocano* as marked by the suffix -*ano*. To account for this type of morphological information, the type *agr-cat* contains the features PER, NUM, GEND that indicate, respectively, person, number, and gender. The fact that the three features are used in

Italian does not mean, *per se*, that this trait is common among languages. English, for instance, makes use of the GEND feature only with some forms of personal pronouns, and the features PER and NUM are used only to mark the third person of the present tense.

Features that encode semantic information are grouped under the *sem-cat* type. The feature MODE describes the very general types of meaning that can be expressed with language. The values *proposition, question,* and *directive* (i.e. imperative sentences) model the general meaning conveyed by a sentence, while *reference* models the general meaning conveyed by nouns. The feature INDEX describe the link to the situation or the entity to which a linguistic expression refers. For, in principle, there is an index for every situation or entity referred to by a certain linguistic expression, the values that the feature INDEX may take are potentially unrestricted. However, the number of values is normally determined by the discourse context: every newly introduced referent will receive its own index and keep it throughout the discourse. In this respect, the feature INDEX may be likened to reference markers in theories that deal with discourse representations (Pollard & Sag, 1994). Example 4.7 shows how the feature can be used. Note that the INDEX values are represented by letters, as is normal practice in HPSG. Each lexical entry has its own feature INDEX with its own value. Since the pronoun *him* has the same referent as the NP *the teacher* they also share the same value for the feature INDEX, that is, they are *coindexed*.

4.7 The teacher	explained	syntax	while	the students	scoffed	at him
INDEX i	INDEX j	INDEX m		INDEX n	INDEX p	INDEX i

The third feature structure of the type *sem-cat* is the feature RESTR that encodes the semantic restrictions that are relevant to the meaning of the sentence. In HPSG semantic restrictions are related to several factors, such as “what properties some individual has, who did what to whom in some situation, when, where or why some situation occurred, and so forth" (Sag et al., 2003, p. 138).

The feature RESTR has the property of taking, as its value, a list of features that are, in turn, arranged into a type. The type is called *predication,* and the features in it describe what meaning relations are determined by linguistic expressions and the kind of participants involved in the relation. Meaning relations are shown by associating the written form of a word to the feature RELN (i.e. relation). The feature RELN for the verb *take*, for instance, is simply modelled as [RELN take]. Participants have explicit features that describe their specific roles such as DRINKER and DRUNK for the verb *to drink*, GIVER, RECIPIENT, and GIFT for the verb *to give,* and so on. Although this description seems specific enough in order to capture the complexity of meaning relations, in recent implementations of HPSG, the type *predication* has incorporated the notion of frame as developed in Frames Semantic (Boas & Sag, 2010; Fillmore & Baker, 2009). As outlined in Chapter 3, frames encode a more complex description than that provided be the type *predication*. Frames stand for cognitive structures that contain the set of knowledge, beliefs, and schemes of practice of individuals' experiences. An example of a frame for a verb is given in Figure 4.1. It is taken from the website FrameNet (http://framenet.icsi.berkeley.edu/) which has been built on the basis of the developments and implementations of Frame Semantics carried out over the last four decades (Baker et al., 2003; Fillmore, 1982; Fillmore & Baker, 2009).

Figure 4.1

An excerpt of the Frame Ingestion

Frame Element	Core Type
Degree	Peripheral
Duration	Peripheral
Ingestibles	Core
Ingestor	Core
Instrument	Peripheral
Manner	Peripheral
Means	Peripheral
Place	Peripheral
Purpose	Peripheral
Source	Peripheral
Time	Peripheral

Note. Taken from https://framenet2.icsi.berkeley.edu/fnReports/data/frameIndex.xml

The example shows the frame to which the verb *to drink* belongs. The assorted colours illustrate the semantic roles, called frame elements in the theory. A frame encompasses both arguments and adjuncts: arguments are called core elements, and adjuncts are called peripheral. The two core roles are Ingestor and Ingestible and correspond to, respectively, the subject and the object. Overall, swapping the type *predication* with the relative *frame* allows for finer-grained representations of meaning relations. For this reason, we are going to adopt the frame-based approach in modelling our lexical entries.

The last two types listed in Table 4.1 are *sign* and *phrasal construction.* Although they are related to the theoretical underpinnings of HPSG, they are important for they provide clear-cut differentiations between structural and meaning-related information. The *sign* type is derived from de Saussure's assumption that language is made of a special kind of objects that are a conventional association of a form with a meaning called signs (Sag, 1997, 2010; Sag et al., 2003). The notion of sign indeed applies to single words and to larger structures, such as phrases and clauses, resulting from combining several linguistic expressions and their phonological forms. The type *sign* has three features. The feature PHON provides the phonological form of any type of expression, be it a word or larger phrasal structure. Concerning its meaning, the type *sign* is constrained by the two broad-scope features, SYM and SEM. The feature SYN contains all sorts of combinatorial information, while the feature SEM contains all sorts of semantic information. The type *sign* is the basic type in HPSG in that it contains, on a very general level, all the features needed to describe linguistic expressions. Single words, phrasal constructions, and complex clausal constructions are all signs. For example, in the previous sentence, the word *constructions* and the noun phrase *single words* are both signs. Therefore, all other more specific features are integrated into the two features SYN and SEM. Since the feature SYN provides broad-scope information about syntax it incorporates the more specific features HEAD, SPR, COMPS and MOD. The feature SEM incorporates the features MODE, INDEX and RESTR that provide finer-grained information about meaning.

Finally, the type *phrasal-construction* (abbreviated as *p-cx*) encompasses three features DTRS, MTR, and HD-DTR. The feature MTR, which stands for *mother*, encodes the information carried by the entire phrase. The feature DTRS, which stands for daughters, encodes the information provided by all the smaller components that combine to form the phrase. The feature HD-DTR stands for head daughter and marks the special status of all the heads amongst the other components of a phrase. For example, the last segment from the previous sentence, *a phrase*, is a phrasal construction. The features MTR will contain the entire phrase, in the form MTR < *a phrase* >. The feature HD-DTR will contain the head noun of the phrase and the feature DTRS the specifier. As a result, the two features will have the form DTRS < *a* > and HD-DTR < *phrase* >.

4.6 How Grammar Works: Unification

To enforce how features are assembled into meaningful and grammatical sentences, HPSG relies on a single procedure called feature unification (Sag et al., 2003; Shieber, 1986). In HPSG, if the features associated with two lexical expressions do not carry conflicting information, they can be combined. The process is called unification and works in a straightforward manner: it combines two feature structures into a new feature structure that contains features from each of them and nothing more. Suppose that we have two Italian linguistic expressions, the noun phrase *un uomo* (*a man*) and the finite verb *cammina* (3rd person singular: *walks*). Both lexical expressions can be described by the underspecified feature structures shown in Figure 4.2.

Figure 4.2

Two non-conflicting Underspecified Feature Structures

$$\begin{bmatrix} \text{PHON} & <\text{un uomo} \\ & \begin{bmatrix} \text{PER} & \text{3rd} \\ \text{NUM} & \text{singular} \\ \text{GEND} & \text{masculin} \end{bmatrix} \end{bmatrix} \qquad \begin{bmatrix} \text{PHON} & <\text{cammina}> \\ \text{NUM} & \text{sing} \\ \text{PER} & \text{3rd} \end{bmatrix}$$

The lexical representation for the linguistic expression *un uomo* is made up of the feature PHON, and the morphological features PER, NUM and GEND that mark the agreement (i.e. they are features of the type *agr-cat*). The representation of the verb *cammina* (*walks*) has the feature PHON and the morphological features NUM and PER. The values of the feature NUM and PER are the same in both feature structures of the two expressions.

Although the feature GEND is not associated with the verb, it does not carry conflicting information since, in Italian, it is acceptable that nouns mark their gender while finite verb forms do not. Overall, the feature structures of the two lexical representations carry compatible information and they can be unified giving rise to a new expression. The result is shown in Figure 4.3.

Figure 4.3

An Example of Unification

$$\begin{bmatrix} \text{PHON} & \text{< un uomo cammina >} \\ \begin{bmatrix} \text{PER} & \text{3rd} \\ \text{NUM} & \text{singular} \\ \text{GEND} & \text{masculin} \end{bmatrix} & \end{bmatrix}$$

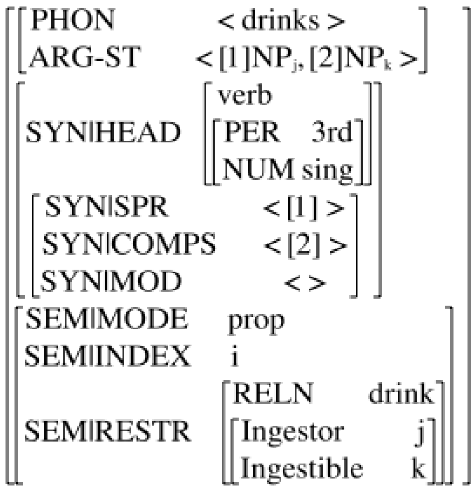
Unification fits well with the view adopted by HPSG about the fact that language processing is incremental. Adopting unification as the basic operation leads to the view that sentences are processed through the incremental matching of features, based on the information that they encode as it becomes available when new segments are integrated into the sentence structure already built (Sag & Wasow, 2011, 2014; Wasow, 2019).

4.7 Lexical Representations as the Building Blocks of Grammar

After describing the details of combinatorial, word form, and meaning features and the operation of unifying them, it is easy to model a more detailed representation and provide an example of the degree of complexity that can be accounted for by HPSG representations. The lexical item is the verb form *drinks*. To improve the readability, all the syntactic features are preceded by the notation SYN| and the notation SEM| precedes all the semantic features. The detailed representation is shown in Figure 4.4.

Figure 4.4

A Lexical Representation with the Features of HPSG

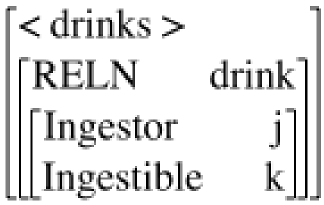


The first point to note is that there are several features that account for combinatorial information. The features COMPS, SPR and MOD encode the information on syntactic structures. The feature ARG-ST encodes the information about combinatorial potential of a lexical entry. The feature RESTR encodes a fully specified description of the meaning relations between linguistic expressions. There is obviously a connection between the information expressed in the features SPR and COMPS and in RESTR which is determined by the properties of the verb *drink*. The subject (i.e. SPR) is linked to the Ingestor role, and the object (i.e. COMPS) of *drink* is linked to the Ingestible. The key role of the feature ARG-ST is to map semantic information into syntactic information (Müller & Wechsler, 2014; Wechsler, 2020). The mapping occurs in two directions. The first mapping links the participant roles encoded in the feature RESTR to the linguistic expressions in the list of the feature ARG-ST. The indices *j* and *k* shared between the NPs in the ARG-ST list and the indices of the semantic features Ingestor and Ingestible determine the first type of semantics-based mapping. Syntax-based mapping combines the linguistic expressions in the ARG-ST list with the elements in the list of the feature COMPS and SPR (Müller & Wechsler, 2014; Wechsler, 2020). The meaning of the shared tags [1] and [2] is that the elements in the ARG-ST list are unified with the elements in SPR and COMPS.

One notion related to language processing that is directly modelled into the feature-based representations is underspecification. In HPSG an underspecified expression is an expression described by a subset within the features of the entire grammar. If the features are appropriately associated to the lexical entry, then the lexical entry will be grammatical. An underspecified example of the verb form *drinks* are shown in Figure 4.5. The verb is described through the RESTR feature structure (that contains the feature RELN) and the value of the feature PHON.

Figure 4.5

An Underspecified Lexical Representation for the Verb Drink



Underspecified representations should not be taken as a direct model of cognitive process. The aim of the two underspecified representations is to represent the fact that, depending on factors related to the communicative event, some lexical features may be activated and accessed more quickly or earlier than others. However, even subsets of features can give the processing system enough information to interpret a sentence.

The representation in Figure 4.4 helps clarify how a processing-compatible grammar works. Once the processing system encounters a lexical entry, all the features needed to build a meaningful segment of a sentence are already available. The lexical entry can combine with other lexical expressions, provided that they follow the requirements of the RESTR feature structure and that they may be unified with the elements encoded in the features SPR and COMPS. This is in line with the strong lexicalism advocated by HPSG and opens the possibility of exploiting the information encoded in lexical items while doing without both phrasal and movement rules. In formal syntactic frameworks based on transformational approaches, the steps leading from words to phrases have been encoded in broad-scope rules or in movement of linguistic expressions across specific structural positions (Adger, 2003; Chomsky, 1995). Since its start HPSG used a different mechanism based on so called Immediate Dominance Schemata (Pollard & Sag, 1994). Schemata were set of constraints associated to certain type of phrases. Sag (1997) further enhanced this approach, considering schemata like any other feature structure, not different from lexical representations. He notices that "just as lexical entries are descriptions of (or constraints on) feature structures of type word, schemata are descriptions of feature structures of type phrase." (Sag, 1997, p. 437).

Using type of phrases instead of rules is a straightforward process. The basic types of phrase are determined by the three features SPR, COMPS, and MOD, which capture the links between the head and the element that combine with it based on its meaning (i.e. the features MOD and RESTR) and its combinatorial potential (i.e. the feature ARG-ST). The grammar has already features for the components of phrasal structures, namely HD-DTR, DTRS and MTR classed under the specific type *phrasal-construction* (see Table 4.1). The three type phrasal constructions are, respectively, *Head Specifier Phrasal-construction, Head Complement Phrasal-construction,* and *Head Modifier Phrasal-construction*. The three types are given in Table 4.2 (to make the interpretation simple, only phonological forms, listed in the feature PHON, have been used).

Table 4.2

The Feature-based Phrasal Constructions

Type of phrase	Feature Structure
Head Specifier Phrasal-construction	MTR [PHON < a man walks >] HD-DTR [PHON < walks > ; SPR [PHON < a man >]] DTRS < [PHON < a man >] , [PHON < walks >] >
Head Complement Phrasal-construction	MTR [PHON < walks to Swansea >] HD-DTR [PHON < walks > ; COMPS [PHON < to Swansea >]] DTRS < [PHON < walks >] , [PHON < to Swansea >] >
Head Modifier Phrasal-construction	MTR [PHON < walks quickly >] HD-DTR [PHON < walks >] DTRS < [PHON < walks >] , [PHON < quickly > ; MOD [PHON < walks >]] >

The *Head Specifier Phrasal-construction* constrains how NPs and VPs combine with, respectively, determiners and subjects. The specifier *a man* is the list of the SPR feature of the head verb *walks*. It is also encoded into the feature DTRS along with the head. The *Head Complement Phrasal-construction* puts the lexical expression *to Swansea* into the COMPS list of the head verb *walks* and in the list of the feature DTRS together with the head. Finally, the *Head Modifier Phrasal-construction* puts the verb head *walks* into the list of the feature MOD associated to the modifier *quickly*. As in the previous cases, the modifier is also encoded in the feature DTRS.

The grammar developed so far has feature structures for every linguistic object, ranging from words to fixed phrasal constructions. In terms of lexical feature-based representations have come a long way: still there is one aspect of language processing that needs to be accounted for in finer-grained detail, that is, incrementality. The next section gives an example of how lexicalism and incrementality work together to build sentences.

4.8 A Simple Processing Model

In this section, we show how the feature-based grammar can describe the process of assembling a simple sentence in an incremental way, simply by unifying the features of linguistic expressions as they become available (Sag & Wasow, 2014). The proposed description does not make a difference between comprehension and production. Of course, in the case of production, features need to be accessed first, while in comprehension they only need to be activated. It is worth mentioning that activating and accessing lexical features overlaps to a great extent with the two notions, commonly applied in vocabulary research, of recognising and recalling lexical items (e.g., Nation, 2013; Schmitt, 2010). However, despite the different processes involved, the mechanism of unifying features according to the type of phrasal constructions holds for both processes. The sentence fragments (a) to (e) shows how these representational conventions are used for the sentence *a man walks quickly to Swansea.* For simplicity, the example is based on the transitive version of the verb walk.

a) a man →

b) a man walks → [*Head Specifier Phrasal-construction*]

c) a man walks quickly → [*Head Modifier Phrasal-construction*]

d) a man walks quickly to Swansea → [*Head Complement Phrase*]

e) a man walks quickly to Swansea → [*Sentence*]

The arrow → points to the newly constructed syntactic structure given in terms of the types of phrasal constructions showed in Table 4.2. The components of the phrasal constructions that are normally encoded in the feature DTRS (Table 4.2) are backgrounded in a light grey colour to help interpretation. The analysis starts from Fragment (a) which shows the expression *a man* firstly encountered by the processing system when interpreting the target sentence. When the head verb *walks* becomes available to the processing system, the feature ARG-ST contains the elements that the verb combines with. This means that the verb form *walks* needs a subject and a prepositional complement that may be encoded in the features SPR and COMPS. Fragment (b) shows how the first phrase formed is the *Head Specifier Phrase*: the process occurs by unifying the NP encoded by the feature SPR with the expression *a man*. In Fragment (c) the adverb *quickly* enters the input and the head verb is unified with the expression encoded in the feature MOD. The resulting phrase formed by the modified verb and the modifying adverb is the *Head Modifier Phrase*. In Fragment (d), after encountering the expression *to Swansea,* the processing system can unify it with the PP encoded in the feature COMPS. The resulting structure is the *Head Complement Phrase*. Fragment (e) shows that the full sentence is complete according to the combinatorial potential encoded in the head.

4.9 HPSG and the Lexicon

Because of the strong lexicalism advocated by HPSG, lexical entries are extremely informationally rich. As all the essential components of the grammar are encoded into lexical items, it is crucial to establish what sort of lexical organisation is necessary. The lexicon cannot be regarded as a simple list of lexical entries with their own feature structures. In fact, all the features of the grammar described so far do not refer to single words, but capture general properties shared by types of lexical items (Davis & Koenig, 2019; Sag et al., 2003). Combinatorial features such as COMPS and SPR, for instance, allow us to generalize over all types of verbs: *take, grab, grip, hold,* but also *drink, eat, nibble, sip* share the same value for both features as they all have a subject and an object. Hence, the feature structure [COMPS < NP >] can capture a property shared by verbs that differ greatly in other terms. Semantic features lead to draw similar generalisations: the RESTR features structure shown in Figure 4.6 reflects that *drink*, *eat*, *nibble* and *sip* describe the same kind of event where an entity (Ingestor) consumes food or drinks (Ingestibles).

Figure 4.6

An Underspecified Lexical Representation for the Verbs Drink, Eat, Nibble and Sip

```
[RELN    "consume food
           or drinks"    ]
[[Ingestor      j]       ]
[[Ingestible    k]       ]
```

Meaning features can capture general shared properties even where the degree of similarity is exceedingly low. For example, the verbs *amuse* and *change* do not seem to share a great deal of meaning. However, while the entity that undergoes the event acted by the agent is described by two unique features the first of their argument overlaps and is encoded by the shared feature Agent (Figure 4.7).

Figure 4.7

Two Underspecified Lexical Representation for the Verbs Amuse and Change.

$$\begin{bmatrix} \langle \text{amuse} \rangle \\ \begin{bmatrix} \text{RESTR|Agent} & j \\ \text{RESTR|Experience} & k \end{bmatrix} \end{bmatrix} \begin{bmatrix} \langle \text{change} \rangle \\ \begin{bmatrix} \text{RESTR|Agent} & j \\ \text{RESTR|Entity} & k \end{bmatrix} \end{bmatrix}$$

As shown in the few examples reported above, lexical relations may display a high degree of complexity. The effort, within HPSG, has been directed towards modelling lexical relations in a simple way, holding on to the type-based system and the versatility afforded by it (Davis & Koenig, 2019; Sag et al., 2003).

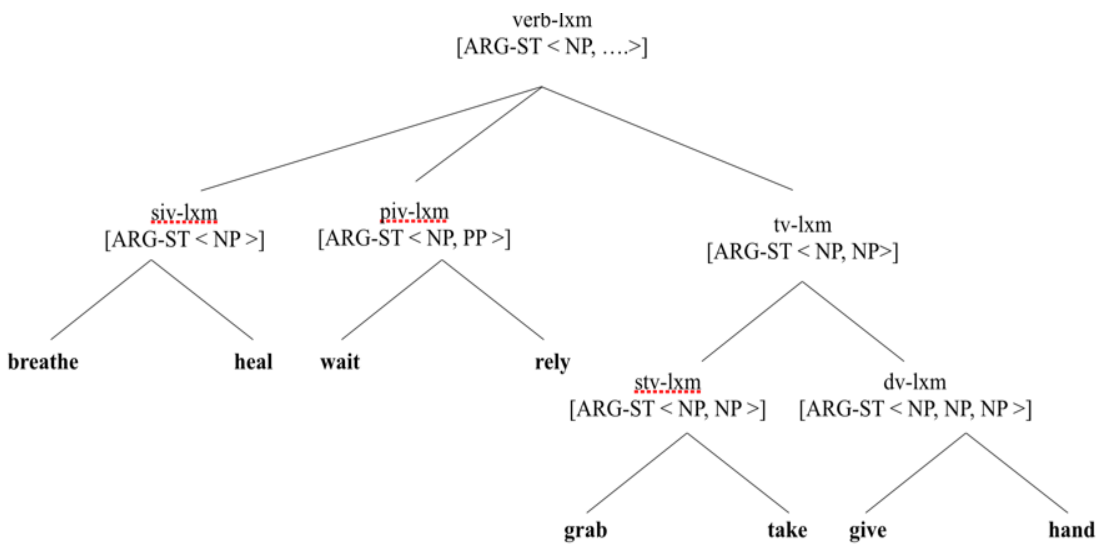
In dealing with lexical organisation, the approach pursued by HPSG has been that of arranging linguistic expressions into hierarchies. Hierarchies are organised based on the amount of information associated with lexical entries (Briscoe et al., 1993; Flickinger et al., 1985; Sag et al., 2003). They aim to capture the information shared by classes of lexical entries and to express generalisation at different levels of lexical representation (Davis & Koenig, 2019).

Suppose that we want to construct a hierarchy for the linguistic expressions of the type *verb.* A straightforward way might be to rely on the combinatorial potential of verbs. The combinatorial information, besides being encoded separately into the features SPR and COMPS, is entirely encoded into the feature ARG-ST. Thus, constructing a lexical hierarchy for verbs will depend on the different content of the ARG-ST list (see Sag et al. (2003) and Davis & Koenig (2019) for a similar approach). The hierarchy may be arranged into type and subtypes, based on the specificity of combinatorial information encoded in the feature ARG-ST. Lexical entries which encode broad-scope information are at the top of the hierarchy, while lexical entries that encode more specific information appear at the lower levels of it.

The first important generalisation that the hierarchy needs to capture is that all types of verbs must have a subject. The notion corresponds to having a non-empty value for the feature ARG-ST - recall that the subject is the first element in the ART-ST list (Sag et al., 2003). The next lower level of the hierarchy encompasses intransitive verbs and transitive verbs. Intransitive verbs have subtypes for intransitive verbs that do not take any complement (*siv*: *strictly intransitive verb*) and a subtype for intransitive verbs that take a prepositional complement (*piv*: *prepositional-intransitive verb*). Transitive verbs have a subtype for common transitive verbs that have a subject and an object (*tv*: *transitive verb*) and subtype for ditransitive verbs (*dv*: *ditransitive verb*), which take two complements - like *give* or *hand* in sentences like: *John handed/gave Bill the stapler*. The hierarchy for the lexeme of the verb type is shown in Figure 4.8.

Figure 4.8

The Hierarchy for the Verb Lexeme



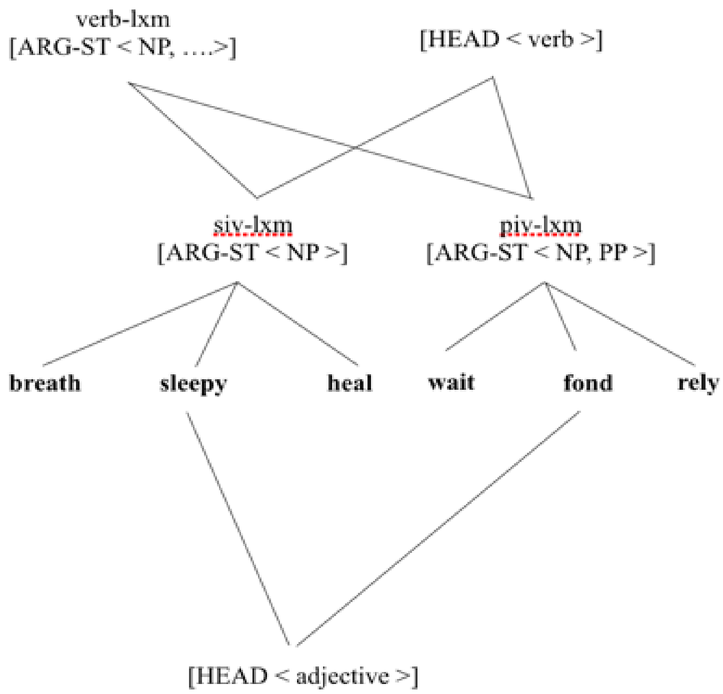
An important concept related to hierarchies is that less specific information is inherited by all lexical entries. HPSG calls this notion default inheritance and is the customary way to pass bits of information throughout lexical hierarchies (Davis & Koenig, 2019; Sag et al., 2003). At the same time, lexical entries that encode more specific information can override those that encode less specific information. Thus, the generalisation that all verbs have a subject is reflected by the fact that every verb inherits, by default, the feature ARG-ST < NP, … >. While this piece of information can be easily passed on to intransitive verbs as a default, it is overridden by the more specific information provided by the feature ARG-ST of transitive verbs that need and object to complete their meaning.

The features employed in HPSG are shared across different lexical entries, so that the same hierarchy works for multiple types of lexical entries, without positing any further theoretical device. Therefore, although the hierarchy given in Figure 4.8 offers a representation for verbs only, the same hierarchical setup can be extended to incorporate different types of lexical entries. Since the feature ARG-ST is present in every word, it may be exploited to describe all types of lexical entries in the same hierarchy.

Suppose that one wants to add the two adjectives *sleepy* and *fond* to the hierarchy implemented for verbs shown in Figure 4.8. The list of the feature ARG-ST for *sleepy* contains the specifier but no other elements (i.e. the determiner in a phrase such as *the sleepy cat* or *the sleepy baby*). In contrast, the list of the feature ARG-ST for *fond* contains a determiner and one more element, the prepositional phrase headed by the preposition *of* (i.e. *fond of his wife*). The two ARG-ST overlap perfectly with those associated with the two verb types *strictly intransitive* and *prepositional intransitive lexeme*. The hierarchy needs a slight adjustment to signal that the same features are shared between two different parts of speech. This is easily achieved by adding the feature HEAD that encodes the category to which the lexical entries belong (see Table 4.1 above). The rearranged lexical hierarchy is shown in Figure 4.9.

Figure 4.9

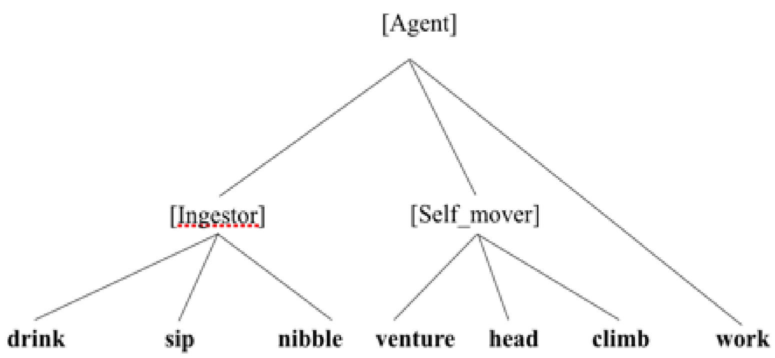
The Multiple Hierarchy for the Intransitive Verb and Noun Lexeme



The examples in Figure 4.8 and Figure 4.9 are limited to the features ARG-ST and HEAD. Using the same default-based model, it is possible to construct a meaning-based hierarchy that contains semantic features modelled in terms of frame elements. The underlying principles do not change. Semantic features that encode broad-scope information are inherited by default, and they can be overridden in case features encoding more specific information are added to the hierarchy. Features that can be overridden are called defeasible features in HPSG. The meaning-based hierarchy is shown in Figure 4.10: the features used refer only to the meaning encoded into the subject role.

Figure 4.10

A Meaning-based Hierarchy for The Verb Lexeme

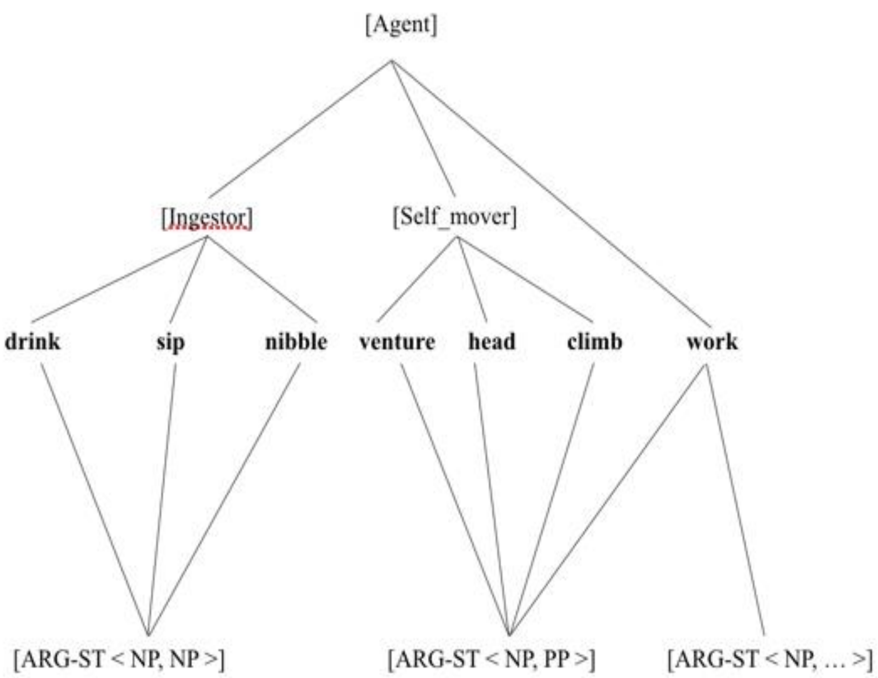


The abstract meaning encoded by the feature Agent is inherited by all lexical items by default but overridden by those verbs which need a more specific description to be associated to the role of subject.

Meaning representations can also incorporate other types of information. Looking carefully at the hierarchy in Figure 4.10 it appears that *drink, sip,* and *nibble* are transitive verbs, *venture, walk,* and *climb* are prepositional intransitive and *work* is an intransitive verb that can also take prepositional content. It is a straightforward process to add the feature ARG-ST to the meaning-based network. The resulting hierarchy is shown in Figure 4.11. At this point, it is clear that the hierarchy starts to visually resemble a lexical network. Indeed, hierarchies are a special type of network. Unlike networks, wherein connections between nodes do not express much information, a hierarchy assigns meaning to each link (Smith, 1991).

Figure 4.11

A Meaning- and Syntactic-based Hierarchy for the Verb Lexeme

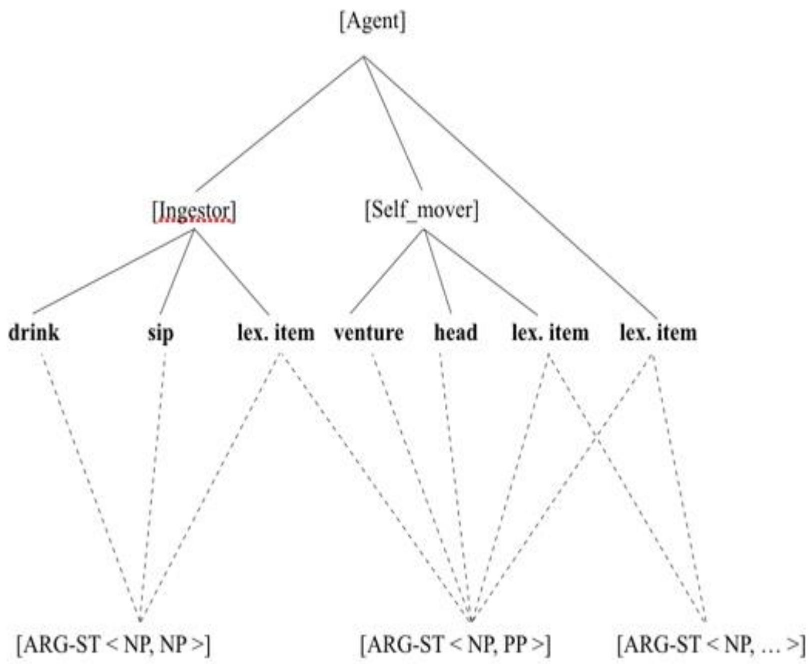


An additional aspect of lexical hierarchies relates to the way lexical information is accessed and lexical entries produced online. Hierarchies such as those described so far have fully specified lexical entries. Although it might work as a theoretical device, a fully specified lexical hierarchy seems unlikely when compared to the way mental lexicon works.

There is evidence that features are indeed attached on-line at the time of interpreting or producing an utterance (e.g., Evans & Levinson, 2012; Koenig & Jurafsky, 1995). Hence, for instance, the fact that the verb *to work* can inherit two different values for the feature ARG-ST does not need to be encoded twice. Which feature ARG-ST the verb takes can be determined online based on the meaning that the speaker intends to convey. This type of approach is called online type creation (Davis & Koenig, 2019; Koenig & Jurafsky, 1995). According to the approach, lexical entries are stored as fundamentally underspecified in the grammar. Therefore, some of the types and features are not encoded in words or represented as links in the hierarchy. In contrast, the links that associate some of the features with lexical entries are built on-line. This can be simply shown by a slight modification of the hierarchy shown in 11. The revised hierarchy is in Figure 4.12.

Figure 4.12

An On-line Type Meaning- and Syntactic-based Hierarchy



Note. The Hierarchy is loosely based on Koenig & Jurafsky (1995), p. 4

The links between the feature ARG-ST are now marked with a dotted line showing that they may not be part of the hierarchy, but they are built on line. This opens the possibility of adding many more lexical entries to the hierarchy, as shown by the new general label *lex. items* (i.e. lexical items). As a result, when new lexical items are added, new links are also formed. Verbs that encode the semantic feature Ingestor like *eat* and *dine* can be transitive and prepositional intransitive at the same time. Verbs that encode the semantic feature

Self_Mover like *run*, *stroll*, and *scuttle* might be prepositional intransitive but also intransitive. This sort of approach allows one to relax the complexity of the hierarchy and to pass it onto the processing system (Koenig & Jurafsky, 1995).

4.10 Models for Language Processing.

Building on the assumptions HPSG makes about the relationship between language processing and grammar design, two complex syntactic phenomena have been modelled using feature-based approaches, namely long-distance dependencies (Chaves, 2013) and reduced relative sentences (Hana et al., 2002).

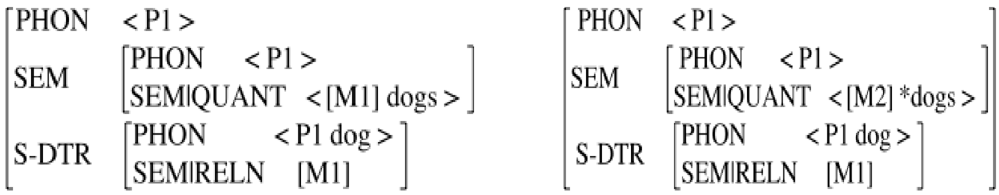
Furthermore, a fully formalised and feature-based description of interaction and conversation has also been provided (Ginzburg, 2012a, 2012b; Lücking et al., 2019). Although not related to language processing, the study of conversation and the effort to formalise it in a feature-based grammar can tell a great deal about the traits and idiosyncrasies of spoken language, as it occurs in natural settings. Indeed, factors such as incrementality and surface orientation that underlie competence grammar within HPSG are widespread in conversations. First, Ginzburg (2012) and more recently, Lücking et al. (2019) have extensively worked on formulating a formal description of so-called non-sentential utterances, based on HPSG and a closely related framework called Type Theory with Records (Lücking et al., 2019). These endeavours are important in outlining the flexibility of feature-based grammatical theories like HPSG. However, since the fragments of the grammar that they modelled are not the focus of this study, I am not going into further details.

Closely related to the present research is the work carried out by Mellow (2004) in which he modelled the L2 language representations focusing on form-meaning mappings. Quite often L2 speakers tend to generate false associations between what should be the suitable form and the intended meaning (Corver, 2003; Mellow, 2004). This results in a mismatch between the content of the PHON feature and the other features associated with lexical representations, namely SEM and SYN. This is clear in forms that carry little or redundant meaning, such as the suffix *-s* that marks subjects in the third person. In many occasions, the subject is easily understood through the context, and the morphological marker may either go unnoticed or not realised by L2 speakers (Vanpatten, 2015; VanPatten, 2004). Research has shown that L2 speakers tend to process content words first to interpret sentence meaning without extensively relying on grammatical information (e.g., Clahsen & Felser, 2006; Vanpatten, 2015; VanPatten, 2004). These factors are all tightly interwoven with proficiency: as the level of proficiency develops, L2 speakers grow less susceptible to the effects of input, L2, and task type so that supplying correct forms increases over time (e.g., Clahsen & Felser, 2006; Dijkstra & van Heuven, 2002). A patchy mapping of form-meaning in L2 speakers shows that supplying the correct forms depends on the strength of connections between various meaning and word form features, both encoded in lexical items. Mellow (2004) took this notion further, assuming that the mapping between forms and meaning can be couched into a connectionist framework.

Stronger connections between grammatical forms and meaning give rise to low variability in supplying the correct forms, whereas weak and poor connections give rise to high variability in supplying the incorrect forms. The formalisation he proposes translates form-meaning mapping into the linkage between the feature RELN and the feature PHON. An example is given in Figure 4.13: on the left-hand side is the representation of the successful mapping between the plural meaning of the lexical item *dog* and the plural form *dogs*.

Figure 4.13

Correct (left) and Incorrect (right) Form-Meaning Connection for the Word Dogs



Note. Adapted from Mellow (2004) p. 148

The representations devised by Mellow (2004) are slightly different from the ones described so far. Every *sign* type has three features. The PHON feature encodes the phonological form. The S-DTR feature structure represents the phrase and its overall meaning (through the feature RELN), while the SEM feature structure represents the specific meaning of the phrase: in the example shown in Figure 4.13 it is encoded into the feature QUANT (i.e. quantification). The actual mapping is represented through the association of the meaning feature QUANT (i.e., quantification) to the feature PHON. The value of the features PHON, RELN and QUANT are given as a twofold label where the letters P and M stand for Phonology and Meaning, while the number is used, in the usual manner, as tags to mark feature unification.

Looking at the representation on the left-hand side of Figure 4.13 the meaning of the singular form *dog* is mapped onto the right form: the plural meaning M1 encoded in the feature QUANT is correctly unified with the phonological form P1. An ill-formed case is on the right-hand side. The mapping between the meaning of plurality encoded in the feature QUANT and the plural phonological form is carried out incorrectly, so the plural form -*s* does not appear in the phonological representation encoded in the feature PHON. This ill-formed mapping is shown by the different tags for the meaning label M2 and the phonological label P1.

The type of representation developed by Mellow (2004) draws upon the notion that every linguistic expression can be regarded in terms of meaning-form mapping. This approach is related to how HPSG view linguistic expressions. However, the representations designed by Mellow (2004) are better suited to describe production. Things can be different when describing the shortcomings of L2 speakers in comprehending L2 expressions. Although the lexical features that describe productive and receptive processes are the same, the form-meaning mapping is slightly different. In receptive tasks form-meaning mappings are already available in the input, so that processing issues arise when L2 speakers do not activate the correct features. In fact, this will not result in a mismatch between meaning features and phonological features, but it will appear either as an incorrect interpretation or in the form of processing slowdowns. For this reason, the type of representation adopted in the following chapters is not based on form-meaning mappings. They will draw on the notion of feature availability that is well suited to account for both form-meaning mismatches in production and faulty interpretation in comprehension.

4.11 Filler Gap Dependencies in HPSG: Relative Clauses and Cleft Sentences

This section introduces a simplified analysis of the relative clauses as they are modelled within HPSG. Relative clauses have some specific properties: the filler is a relative pronoun, also called the wh-element, and the entire clause serves as a modifier of a noun[2].

[2] At this point we are ignoring the relative sentences formed with that. The reason is that in Italian *that* and *who* are homophones and uses the same form *che*, so that the boundaries between the two are difficult to show (Salvi & Vanelli, 2010).

HPSG handles all these properties using lexical features. The feature MOD describes the fact that relative clauses modify NPs: its value corresponds to the expression that it modifies. This is shown in Example 4.8 and the feature description in Figure 4.14. The feature MOD is attached to the head verb of the relative clause, and its value corresponds to the NP *il bambino*. The relationship is straightforwardly formalised using a shared tag [1], which shows the unification between the two expressions.

4.8 Il bambino con cui la maestra ha parlato
The child with whom the teacher has spoken
The child with whom the teacher played

Figure 4.14

Feature Description for Example 4.8

[1] il bambino	con cui PP[REL]	la maestra	ha parlato MOD [1]
[1] the child	with whom PP[REL]	the teacher	has spoken MOD [1]

Besides the feature MOD relative clauses are constrained by two specific features. The first is the feature REL encoded in the relative pronoun, the second is the feature GAP encoded in the word that is “filled” by the filler. The feature GAP signals where the dependency must be resolved.

The feature REL is directly associated with the relative pronoun. Hence, in Example 4.8, the feature description of the relative is < cui [REL] >. Relative pronouns may also be part of a larger phrase, called wh-phrase. In Example 4.8 the filler is a PP that indeed has the preposition *con* (*with*) followed by the relative pronouns *cui* (*whom*). Stipulating that the value of the feature REL of a phrase is the sum of all the feature REL of its components ensures that the information encoded in the pronoun *cui* appears in the entire phrase (Arnold & Godard, 2019; Sag, 1997). This rule might seem trivial, as there are hardly cases with a wh-phrase having two or more pronouns. However, it is important because it avoids involving any structural mechanism: If some expression encodes the feature REL, the phrase that is part of will also have the feature REL (Arnold & Godard, 2019).

The other key feature GAP models the link between the wh-phrase and the relative verb. In line with the lexicalism of the theory, HPSG regards the feature GAP as a lexical feature that stores the information missing from a certain type of expression (Arnold & Godard, 2019; Sag, 1997, 2010; Sag et al., 2003). An updated feature description of Example 4.8 is provided in the feature structure in Figure 4.15. The feature GAP is encoded in the lexical entry of the verb *giocare* (*to play*) and takes as its value the missing PP complement *con cui*, exactly the wh-phrase that is understood to be the complement of the relative verb. The structural link between the wh-element and the relative verb is formed through the unification between the content of the feature GAP and the relative pronoun. If the wh-phrase has information that is compatible with that encoded in the feature, GAP unification can occur. In Example 4.8 the information in the feature GAP is equivalent to GAP <PP>, which corresponds to the type of the PP wh-phrase *con cui*. The shared tag [2] mirrors the process of unification.

Figure 4.15

Updated Feature Description for Example 4.8

[1] il bambino	[2] con cui PP[REL]	la maestra	ha parlato MOD [1] GAP [2]
[1] the child	[2] with whom PP[REL]	the teacher	has spoken MOD [1] GAP [2]

Unlike other features, the feature GAP of the relative verb *ha parlato* in 4.8 encodes the information about a non-adjacent dependent that is not in the relative sentence. A main concern in this approach is that all the dependents of a head are in the list of the feature ARG-ST and must appear either as a value of the feature SPR of COMPS. The features SPR and COMPS keep track of elements that are next to their head, while the GAP feature does not. To deal with the issue, HPSG has made use of lexical rules (Sag, 1997) and phrasal constraints (Bouma et al., 2001). The phrasal approach simply posits that, in every phrase, an expression in the ARG-ST feature list may appear in the GAP feature list rather than in the COMPS feature list. The lexical rule posits that an expression is removed from a verb's COMPS list and it is added to the list of the GAP list.

Using either a specific phrasal construction or a lexical rule is not only a theoretical matter, but it also reflects that verbs that have the feature GAP combine with their dependents in a distinct way. As outlined in Section 4.4 the feature ARG-ST maps the lexical information about the subcategorization frame into the syntactic features SPR and COMPS. In the case of filler-gap dependencies, the mapping needs also relying on the feature GAP. The extra feature involved in the syntactic mapping mirrors the slowdown effects often observed when the processing system interprets relative clauses.

The structure formed by the relative clause and the modified NP is used, in effect, as one phrase, as English examples 4.9 to 4.11 show. Following the approach of Sag et al. (2003) and Sag (2010) this type of phrase is called *Head Relative Construction*. Relative clauses, despite having a complex internal structure, work, to all effects, as adjuncts (e.g., Arnold & Godard, 2019; Kroeger, 2004).

4.9 The child with whom the teacher has spoken went into the classroom.

4.10 Do you know the child with whom the teacher has spoken?

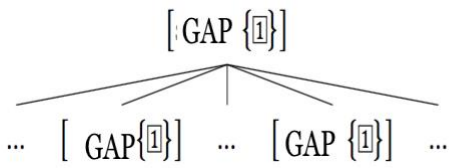
4.11 I met the child with whom the teacher has spoken.

In the *Head Relative Construction,* the feature GAP is situated at some distance from the filler further down the relative clause. However, as it occurs with all phrasal constructions in HPSG, the information encoded in each lexical entry must also appear in the feature structure of construction itself (Sag et al., 2003). This is not a mere structural concern, but it reflects that when the processing system encounters a relative pronoun, it can recognise which construction is part of. Then, it can work out properly and incrementally all the information needed to build the phrasal construction (Sag & Wasow, 2014).

The way to pass the information of the relative pronoun to the entire phrase is straightforward. The GAP value of an entire phrase results from adding up all the GAP values of the dependents contained in it (Arnold & Godard, 2019; Sag, 1997; Sag et al., 2003). The process is modelled into the Gap Principle, which says that the GAP value of any construction is the union of the GAP features of all its components (Sag et al., 2003). The principle ensures that if a phrase is missing something, no matter how deeply embedded, it is missing the same element as each of its smaller components. The principle is formalised in Figure 4.16

Figure 4.16

The Gap Principle



Note. adapted from (Sag et al., 2003, p. 437)

The Gap Principle assumes that any expression has its own feature GAP and that any expression has the potential of encoding missing elements. This assumption might seem far-fetched as, in fact, all words in a relative sentence, apart from one, have empty values for the feature GAP.

4.12 Who did you say Maria loved and John hated?

4.13 È	il ragazzo	che	Maria ama	e	John odia
PRO is	the guy	who	Maria loves	and	John hates

This is the guy Maria Loves and John hates

While this is the case in many filler-gap dependencies, there are notable exceptions in English and Italian. Indeed, both English and Italian can form sentences with more than one feature GAP linked to the same relative pronoun, as shown by the tag [1] in examples 4.12 and 4.13 and the feature description shown in Figure 4.17 (the English example 4.12 is adapted from Chaves (2012), p. 472).

Figure 4.17

Feature Description for Examples 4.12 and 4.13

Ex.4.12	[1]Who	did you say	Maria	loved GAP [1]	and John	hated GAP [1]
Ex.4.13	È il ragazzo	[1]che	Maria	ama GAP [1]	e John	odia GAP [1]

Thus far, the features REL and GAP have accounted for the syntactic links within the relative clauses. However, there need to be features to model the semantic relation between the relative clause and the NP it modifies and between the gap and the filler. In Example 4.8, the meaning link that the wh-phrase *con cui* has with the modified NP *il bambino* is primarily determined via co-indexation. Co-indexing the wh-phrase and the modified NP formalises the fact that the relative pronoun is linked to the same referent as the co-indexed noun. In terms of formal representation, the feature REL takes, as its value, the same index as that of the feature INDEX encoded in the modified noun. Figure 4.18 shows the updated feature structure for Example 4.8 with all the features that undergo the co-indexation process. The value *j* of the feature INDEX of the NP *il bambino* is the same as that of the feature REL – of course, the modified NP encoded in the feature MOD is also co-indexed with the feature REL.

Figure 4.18

Updated Feature Description for Example 4.8

[1] il bambino INDEX j	[2] con cui PP[REL j]	la maestra	ha parlato MOD [NP_j] GAP [2]
[1] the child INDEX j	[2] with whom PP[REL j]	the teacher	has spoken MOD [NP_j] GAP [2]

Co-indexation accounts for some aspects of semantic relations within the relative clause and the modified NP. Yet, the *Head Relative Construction* also describes a complex meaning unit, so the co-indexing of multiple lexical items covers only a part of the meaning relations of the construction. To account for this, HPSG makes use of the feature RESTR and the Semantic Compositional Principle in Sag et al. (2003). The principle ensures that the meaning of a construction is the result of a simple compositional process. Thus, the feature RESTR of every construction is the sum of the features RESTR of every word in the construction. Figure 4.19 shows how the Semantic Compositional Principle and the feature RESTR operate in the *Head Relative Construction*.

Figure 4.19

The Head Relative Construction

$$\begin{bmatrix} \text{MTR} & \begin{bmatrix} \langle \text{il bambino ... ha parlato} \rangle & \\ \text{SEM|INDEX} & j \\ \text{SEM|RESTR} & [1] \oplus [2] \end{bmatrix} \\ \text{HD-DTR} & \left\langle \begin{bmatrix} \langle \text{il bambino} \rangle & \\ \text{SEM|INDEX} & j \\ \text{SEM|RESTR} & [1] \end{bmatrix} \right\rangle \\ \text{DTRS} & \left\langle \begin{bmatrix} \langle \text{con cui ... ha parlato} \rangle & \\ \text{SEM|RESTR} & [2] \end{bmatrix} \right\rangle \end{bmatrix}$$

The feature structure in Figure 4.19 shows that the value of the feature RESTR of the *Head Relative Construction* is the sum of the features RESTR of the modified NP and of the VP of the relative clause. For the sake of simplicity, the content of RESTR is simply shown by a numeric tag. The Semantic Compositional Principle holds for any phrasal construction, so it also includes the VP *la maestra ha parlato* in the relative clause. Although this is not explicit in Figure 4.19, the Semantic Compositional Principle can be applied within the VP whose feature RESTR is the sum of the features RESTR of the verb form *ha parlato* and the NP *la maestra*.

A final remark on how HPSG deals with subject relative clauses. Simple relative clauses such as 4.14 were initially not regarded as real filler-gap dependencies (Pollard & Sag, 1994).

4.14 Il bambino che ha parlato con la maestra
The child who has spoken with the teacher
The child who spoke with the teacher

Subject relative clauses form a local dependency, in that the relative verb immediately follows the subject relative pronoun. For this reason, Sag (1997) described subject relative clauses in the same way as any other type of subject clauses, where the ARG-ST list has the subject as the first element. The only difference is that the element that must appear in the SPR list is the relative pronoun instead of an NP. Under this account the feature GAP can be excluded from subject relative clauses as it encodes the same information as the feature SPR. Hence, Example 4.14 would have the feature description shown in Figure 4.20.

Figure 4.20

Feature Description for Example 4.14

[1] il bambino INDEX j	[2] che [REL j]	ha parlato SPR [REL j] MOD [NP_j]	con la maestra
[1] the child INDEX j	[2] who [REL j]	has spoken SPR [REL j] MOD [NP_j]	with the teacher

4.11.1 *Filler Gap Dependencies Used in the Present Research Project*

In this last part, I will describe cleft sentences and oblique sentences, the two filler gap dependencies that I will investigate in the experiments described in Chapter 6 and Chapter 7. The first type, cleft sentences, are sentences that have a focused element modified by a relative clause (Kroeger, 2004; Pollard & Sag, 1994). A cleft consists of a main clause headed by the copular verb form *was* and the dummy subject *it*, and a relative clause which modifies the predicative argument of the copular verb (Kroeger, 2004). A cleft is shown in Example 4.15.

4.15 It was the girl who bought the car

The focused constituent *the girl* contains information that is usually new in the discourse context, while the relative clause *who bought the car* contains information that is already available in the discourse (Kroeger, 2004). The way HPSG handles cleft sentences is straightforward. The relative segment is not different from subject and object relative clauses. The difference is that the copular *be* is regarded as a lexical entry that has some specific properties (Pollard & Sag, 1994). The lexical entry is given in Figure 4.21.

Figure 4.21

Lexical Entry for the Verb to Be in Cleft Sentences

$$\begin{bmatrix} < \text{be} > \\ \text{SYN|SPR} < \text{it} > \\ \text{SYN|COMPS} < \begin{bmatrix} < \text{the girl} > \\ [\text{FOCUS} +] \end{bmatrix}, \text{S} \begin{bmatrix} \text{MOD} < \text{the girl} > \\ \text{GAP} < > \end{bmatrix} > \end{bmatrix}$$

The copular *be* has two complements, the predicative complement *the girl* and the relative clause. The new feature FOCUS highlights the discourse status of the modified NP, and the relative clause contains the features MOD and GAP. This lexical entry ensures that Example 4.15 has the proper feature description showed in Figure 4.22.

Figure 4.22

Feature Description for Example 4.15

[1]It	was SPR [1] COMP < [2], S >	[2] the girl FOCUS + INDEX j	[3] who [REL j]	bought SPR < [REL j] > MOD < NP_j > GAP [3]	the car

As can be seen, the relative segment has the same feature of all relative clauses: the features REL, MOD and co-indexation works in the usual way. Since Example 4.15 describes a subject cleft, the subject relative clause has the REL feature in the SPR list. Example **4.16** shows an object cleft, where the relative clause has its own subject and the relative pronoun is the object.

4.16 It was the girl who the man greeted

As shown by the feature description Figure 4.23 the feature for the relative segment is the same used for relative clauses and the focused segment, like subject clefts.

Figure 4.23

Feature Description for Example 4.16

[1]It	was SPR [1] COMP < [2], S >	[2]the girl FOCUS + INDEX j	[3] who [REL j]	the man	greeted MOD < NP_j > GAP [3]

The feature descriptions given in Figure 4.22 and Figure 4.22 work in the same way for Italian, except that Italian is a pro-drop language. Subject pronouns are optional, unless they mark a precise discourse function (Schwarze, 2009). The only difference in the lexical entry for the verb would be that, under the assumption that HPSG does not use empty categories, the feature SPR would have an empty value, appearing in the form SYN|SPR < > (Pollard & Sag, 1994).

The last type of filler-gap dependency that I will describe is relative sentences wherein the filler has the role of the adjunct of the relative verb. Following Kroeger (2004) I call this type of relative oblique relative clauses. An oblique relative clause is shown in Example **4.17**.

4.17	Ho visto	il ragazzo	con cui	Alice	studiava italiano
	I saw	the guy	with whom	Alice	studied Italian

I saw the guy with whom Alice used to study Italian

In Example **4.17** the wh-element *with whom* (*con cui*), and the modified NP *il ragazzo* (the guy), are the adjunct in the relative clause, for the verb *studiava* (*studied*) has its subcategorization frame already fulfilled by the subject *Alice* and the object *italiano* (*Italian*).

HPSG handles the oblique relative using the basic features shown so far for relative clauses, although some details need to be added. In fact, since the adjunct-modifying filler is neither part of the COMPS nor of the ARG-ST list, HPSG uses a feature named DEPS that contains the list of the feature ARG-ST and all the adjuncts that are part of the sentence

(Bouma et al., 2001). Since the feature DEPS has all the elements that cooccur with a verb, if one of the adjuncts is encoded in the feature GAP, it is not “extracted’ from the list of the feature COMPS. It simply appears in the DEPS list, and no specific phrasal construction is needed (Bouma et al., 2001). The result is given in the feature description shown in Figure 4.24.

Figure 4.24

Feature Based Description for Example 4.17

I saw	the guy INDEX j	[3] with whom REL j	[1] Alice	studied SPR < [1] > COMPS < [2] > GAP < [3] > ARG-ST < [1], [2] > DEPS < [[1], [2], [3] >	[2] Italian

In this last section I will focus on how the processing-oriented approach of HPSG can easily account for the way the processing systems handles filler gap dependencies. As a framework, HPSG is surface oriented and relies on lexically encoded information and a small set of phrasal constructions (Sag & Wasow, 2011, 2014; Wasow, 2019). Hence, like for any other type of syntactic structure, the processing system interprets relative clauses as soon as it encounters the lexical features encoded in each lexical item. First it interprets the relative pronoun and the feature REL, associating it to the modified NP based on the same shared referent (see example 4.8). When the processing system interprets the verb in object and oblique relative clauses, it needs to access the information encoded in both features COMPS and GAP. The feature COMPS reflects the process of accessing the information encoded into the subcategorisation frame of the verb and of linking it to the lexical items that are adjacent to it. The feature GAP reflects the process of linking the information encoded into the subcategorisation frame to lexical items that are not adjacent to verb, namely the modified NP and the relative pronoun.

The process of dealing with the further piece of information encoded in the feature GAP also reflects the processing cost normally associated to the interpretation of object and oblique relative sentences (e.g., Felser & Roberts, 2007; Juffs & Rodrigues, 2015; Miller, 2014a; Traxler et al., 2002). In contrast, subject relative clauses do not have the feature GAP since the processing system does not need to access further lexical information that encodes the links to non-adjacent lexical items. This reflects that subject relative clauses are normally not costly in terms of language processing (e.g., Juffs & Rodrigues, 2015; Traxler et al., 2002). In this respect Chaves (2013) assumes, for example, that all subject relative clauses have an empty value for the feature GAP by default, defining the approach as "a parsing rule completely separate from the grammar" (p. 32).

Overall the feature-based system used within HPSG makes predictions about syntactic complexity in terms of the amount and the type of lexical information. Under this approach, HPSG views relative sentences as segments constrained by a certain type of features like any other lexical item. If the processing system can gather lexical features that are easily available, the interpretation of the relative clause becomes less costly. I will expand and explain this assumption in Chapter 5 and in the two studies described in Chapter 6 and Chapter 7.

4.12 Conclusions

The present chapter described the main properties of HPSG, a lexicalist and feature-based grammar framework. The framework handles all levels of linguistics descriptions in terms of lexically encoded information. This approach applies to different types of phenomena: basic syntactic processes, the way lexical relations hold in the lexicon, and more complex syntactic structures such as filler-gap dependencies. HPSG is well suited to investigate syntactic and lexical processing under a unified approach. In the next chapter, I will examine whether a constraint-based and lexicalist framework such as HPSG can be adopted to reconcile L2 lexical and syntactic processing. I will then test the hypothesis in two experiments in Chapter 6 and Chapter 7.

Chapter 5 Feature Availability and Language Processing

5.1 Introduction

In previous chapters, I described the complexity of vocabulary knowledge (Chapter 2) and how frequency can be used to capture various aspects of it. In the I-Lex study on productive vocabulary knowledge reported in Chapter 3, L2 speakers have been shown to know fewer infrequent words than L1 speakers. The study revealed that frequency is a reliable predictor to differentiate between the vocabulary knowledge of L1 and L2 speakers. Frequency is related to feature availability, that is, the number of features that the processing system can use to access words and form meaningful structures.

The present chapter aims to evaluate the role frequency plays in current models of sentence processing and to propose a novel approach to handle lexical and syntactic processing. First, I present two commonly used models of L1 and L2 sentence processing that have a syntax-first setup, the Garden Path Theory and the Shallow Structure Hypothesis. The phrase "syntax-first" means that both models assume that sentence processing occurs in two stages, with a first stage that relies on syntactic information, followed by a second stage where other types of information are used. Processing problems often shown by L2 speakers are attributed to their difficulty in handling the first syntactic stage (Clahsen & Felser, 2006, 2018). Next, I am going to describe models that incorporate frequency effects and do without a clear-cut distinction between lexical and syntactic processing, namely constraint-based and usage-based models of language processing. In these models, the processing problems of L2 speakers are ascribed to their difficulty in dealing with different sources of lexically encoded information (e.g., Bialystok et al., 2008; Ellis et al., 2014; Sullivan et al., 2018). As I will explain, those models are based on the same assumptions underlying HPSG described in Chapter 4, namely lexicalism and constraint-based design.

In the second part, building on these models and the nature of lexical representations described in Chapter 2, I will explain how the effects of lexical processing can exhaustively describe sentence processing without positing separate stages. I am going to describe how lexical processing interacts with vocabulary knowledge and use the notion of feature availability to reconcile lexical and sentence processing. I will briefly describe models of bilingual lexical processing focusing on how the interaction of distinct types of lexical features underlie lexical access. Finally, I will explain how the effects of lexical processing can account for the effects of syntactic processing found in L2 speakers. The claim is based on the notions of lexicalism that underpin HPSG and constraint-based models.

The last section lays the groundwork for an account of sentence processing entirely based on the notions of feature availability, frequency, and feature-based lexical representations. Drawing on the precision of the HPSG representations described in Chapter 4, I propose a first version of a new unified model of lexical and structural processing. The model is then going to be tested in two online experiments in Chapter 6 and Chapter 7 and fully formalised in the discussion in Chapter 8.

5.2 Syntax-First Models

Many of the current, commonly used models of sentence processing are based on Minimalism and its previous incarnation, the Principles and Parameters theory (e.g., Chomsky, 1981, 1995). In recent years, some technical aspects of Minimalism (Chomsky, 1995) have also been introduced into theories of L2 language processing (van Hout et al., 2003). Both Minimalism and Principles and Parameters theory have argued that the bulk of features needed to process language is in the lexicon, but they draw a clear-cut line between syntactic features and semantic features (Adger, 2003, 2010; Adger & Svenonius, 2011; Chomsky, 1995).

These properties have been implemented in the Garden Path Model, a model of sentence processing that relies on syntax as the first and main source of information (Frazier, 2013). The model adopts a distinction between syntactic and semantic features, assuming that the processing system works in two stages. First, it forms the syntactic structure of the sentence using abstract grammatical features, then it uses semantic features to support or to revise the first structurally driven interpretation (Frazier, 2013; Frazier & Clifton, 1997; Juffs & Rodrigues, 2015). In the first stage, the processing system first retrieves the syntactic category to which a lexical item belongs. This unveils the type of phrases allowed by the grammar: for instance, a verb may combine with its complement and form a VP, or a determiner can combine with a noun and form an NP, and so forth. This type of information is crucial in determining and limiting the number of constituents that can be merged. The tendency to create syntactic structures with the fewest number of constituents is called Minimal Attachment and reflects the universal tendency of the processing system to form the simplest syntactic structure (Juffs & Rodrigues, 2015). Minimal Attachment predicts that the processing system, in Examples like **5.1**, will interpret *the spy* as the one who is holding the *binoculars.*

5.1 The officer saw the spy with the binoculars

In addition, processing a sentence also entails how to interpret new incoming material. The Garden Path Model argues that the processing system will usually add as little material as possible into the sentence already being built, without postulating any new structure. This is ensured by a principle called Late Closure, which guarantees that any phrasal structure is "closed" as soon as it is allowed by grammar (Frazier & Clifton, 1996). Hence, in Example

5.2, Late Closure leads the processing system to initially interpret *the book* as the object of the verb *reading*.

5.2 While Amanda was reading the book fell of the desk

When, as Example **5.2** shows, the structure appears not to be the correct one, the processing system revises the first interpretation based on the new available information encoded in the verb fell and builds a new syntactic structure (e.g., Frazier & Clifton, 1997; Juffs & Rodrigues, 2015).

5.3 Syntax-First Processing in L2 Speakers

The principles developed within the Garden Path theory have been extended to the study of L2 processing. This has led Clahsen & Felser (2006) to propose a model that could explain why L2 speakers are more sensitive to syntactic ambiguity and slower in the reanalysis process than L1 speakers. Clahsen & Felser (2006) argue that the processing disruptions shown by L2 speakers occur because they fail to apply basic syntactic parsing rules like the Minimal Attachment and the Late Closure (e.g., Clahsen & Felser, 2006, 2018; Felser et al., 2003; Juffs & Harrington, 1995; Juffs & Rodrigues, 2015; Marinis et al., 2005). According to Clahsen & Felser (2006), L2 speakers are less likely to rely on the abstract phrasal operations that are needed to achieve a native-like syntactic processing. Instead, they rely mainly on semantic information to build syntactic structures. For this reason, their syntactic parsing is "shallower" than those of the L1 speakers (Clahsen & Felser, 2006, 2018).

However, the evidence for the claim that L2 are less capable of constructing native-like syntactic structures is far from conclusive. Some studies have shown substantial differences between syntactic processing of L1 and L2 speakers (Felser et al., 2003b, 2012; Marinis et al., 2005; Papadopoulou & Clahsen, 2003), while some others have failed to show any difference at all (e.g., Miller, 2014; Omaki & Schulz, 2011; Pliatsikas & Marinis, 2013; Witzel et al., 2012).

Most of the studies that have considered how L2 processing works have used relative sentences that are a type of filler-gap dependency (see Chapter 4, Section 4.11). The experimental design is based on the comparison of different types of filler-gap dependency. The filler and the gap can be close, like in subject relative sentence **5.3**, or far from one another, like object relative sentence **5.4**. Fillers can also be ambiguous as to which NP they refer to, like Example **5.5** where the relative *who was in the garden* can be the modifier of both nouns *the servant* (high attachment) and *the conductor* (low attachment).

5.3 I have met the doctor who insulted the patient with a broken arm.

5.4 I met the doctor who the patient with the broken arm insulted.

5.5 The man saw the servant of the conductor who was in the garden.

Some studies have shown that when the filler and the gap are distant L2 speakers show reading times slower than L1 speakers and do not seem to benefit from the structural cues present in the sentence (Marinis et al., 2005). However, if L2 speakers have enough exposure to the language, they show the same reading time patterns as L1 speakers (Pliatsikas & Marinis, 2013). Other studies have looked at the type of attachment (Example **5.5**), which can be language-specific. Speakers of different languages usually show a preference for either type (Cuetos, 1988; Felser et al., 2003b; Papadopoulou & Clahsen, 2003). Some experiments showed that, unlike L1 speakers who show consistent preferences for the type of attachment of their native language, L2 speakers do not show any preference for either type of attachment (e.g., Felser et al., 2003; Papadopoulou & Clahsen, 2003).

Evidence arguing against these results also emerged. For example, Witzel et al. (2012) have shown that L2 speakers show the same attachment patterns as L1 speakers when the high- or low-attachment interpretation can be disambiguated by meaning information. Pan et al. (2015) found that, in L2 speakers, relative attachment depends on the preceding discourse context. The interpretation of the modified NP between two potential candidates may be determined by whether it is mentioned in the preceding context. Finally, several studies have found that L2 speakers show the same sensibility as L1 speakers in filler-gap dependencies that contain grammatical violations (Cunnings et al., 2010; Felser et al., 2012; Kim et al., 2015; Omaki & Schulz, 2011). Sensitivity to grammatical violation suggests that both L1 and L2 speakers can achieve similar levels of structural complexity during parsing (Cunnings, 2017).

Some accounts do not draw a clear line between syntactic and semantic information to explain the difference between L1 and L2 processing, but adopt a more nuanced approach. For example, Cunnings (2017) also claims that there is evidence suggesting that L2 speakers are more sensitive to use discourse-level and semantic information than L1 speakers. However he argues that “such findings do not in themselves indicate shallow parsing and instead suggest L2 learners are more sensitive to constructing pragmatically-appropriate parses” (Cunnings, 2017, p. 672). Cunnings (2017) claims that the difference between L1 and L2 speakers can be explained by how the memory of L2 speakers retrieves non-syntactic lexical information they especially rely on to process sentences correctly. Cue-based models have focused on how the memory system can be hindered when accessing the lexical information that is necessary to build syntactic structures (e.g., Lewis et al., 2006). In cue-based models, lexical features that encode, for instance, semantic or discourse-based information, work as *cues* for the processing system to integrate lexical items into the right syntactic structure (e.g., Lewis et al., 2006; Van Dyke & Johns, 2012; Van Dyke & McElree, 2011). Because L2 speakers do not have full knowledge of all necessary lexical features, they often fail to use them as appropriate cues for the processing system (Cunnings, 2017).

Other models do not make any difference between the various sources of information accessed by the processing system. Information about frequency, semantics, and discourse context is exploited at the same time. These models are the focus of the next section.

5.4 Frequency-Based and Constraint-Based Approaches to Sentence Processing

Research carried out in usage-based models has shown that L1 and L2 speakers know, as a part of their linguistic representation, a great amount of frequency information which they use when processing language (e.g., Bybee, 2008; Ellis, 1996, 2013; Ellis et al., 2014; Ellis & Ferreira-Junior, 2009). According to usage-based approaches, the effects of frequency are prominent on learning, memory, and perceptions: the more individuals experience something, the more the memory of it becomes entrenched and easy to access (Ellis, 2013; Ellis, 2002). As Ellis (2013) puts it: "The more times we experience conjunctions of features, the more they become associated in our minds, and the more these subsequently affect perception and categorisation" (p.195).

Frequency is a key factor in implicit language learning. Usage-based approaches view language knowledge of L2 and L1 speakers as the result of a statistical learning of form-meaning patterns (Bybee, 2006, 2008; Ellis, 2013; Ellis, 2005). The most frequent and prototypical regularities in the input are those that drive the process of learning the regularities in the grammar, based on the distributional properties of the language (Bod et al., 2003; Ellis, 2013; Ellis, 2005). When learning another language, L2 speakers already have clear frequency-based knowledge of their first language, at all levels of linguistic representations (Dijkstra & van Heuven, 2002; Ellis, 2013). Being able to rely on L1 frequency patterns can lead to transfer from the L1 to the L2, thus biasing the expectations about the L2's frequency (Ellis, 2013; Ellis, 2002).

According to usage-based approaches, L2 speakers rely on frequency when processing the subcategorization information encoded in the verb, using the most frequent, prototypical, and generic exemplar within each subcategorization frame (Ellis, 2002; Ellis et al., 2014; Ellis & Ferreira-Junior, 2009). This has been shown by, for example, Ellis et al. (2014) who investigated VAC (Verb-argument Constructions), formed by verbs followed by a preposition. Participants, both L1 and L2 speakers, were presented with a phrase missing a verb like *it __ into the* ... and were asked to fill the blank with the first verb that came to their mind. Ellis et al. (2014) found that the frequency of the verbs produced in each construction correlated with their corpus frequency. The correlation figures were approximately .60 for the L2 groups examined in the study and .70 for the L1 group. The study also showed that factors related to verb meaning affected the results: the verbs with a meaning that was most typical of Verb-argument Constructions were provided more frequently as responses in the task. The interaction between frequency and semantic factors is a pattern that also occurs in lexical processing, where frequency and meaning features interact and affect reaction times in various lexical processing tasks (e.g., Dijkstra et al., 2019; Lemhöfer et al., 2008; Lemhöfer & Dijkstra, 2004). To summarise, frequency facilitates L2 speakers to form abstract representations from the most prototypical and frequently used patterns, thus improving their implicit knowledge (e.g., Denhovska et al., 2018; Ellis, 2013; Ellis & Ferreira-Junior, 2009).

While usage-based models focus on the impact of frequency on processing and learning, constraint-based models focus on the interaction between frequency and lexical and syntactic information. First, they assume that lexical and syntactic processing are affected by the same factors and must be regarded as part of the same process (e.g., MacDonald et al., 1994; Trueswell, 1996; Trueswell & Tanenhaus, 1994). As a consequence, constraint-based models also assume that frequency affects both lexical and syntactic representations. For example, some verbs take different subcategorization frames that have different frequencies and tend to occur in specific syntactic structures. Some verbs tend to be followed by certain types of phrases more often, and some verb occur more often in certain structure based on their morphological form (MacDonald et al., 1994; Trueswell, 1996). Experimental results have, indeed, highlighted that the relative frequency of different subcategorization frames plays a crucial role in processing ambiguous sentences (MacDonald et al., 1994; McRae & Matsuki, 2013; Trueswell et al., 1993, 1994; Trueswell, 1996).

McRae & Matsuki (2013) identified subject-verb-object bias as one of the most frequent patterns encoded in the verb subcategorization frame. The pattern primes the processing systems to interpret the NP after a verb as the default object. The subject–verb–object bias operates in an analogous way to the Minimal Attachment, as it produces the simplest structure. However, in constraint-based models, the subject–verb–object bias it is affected by frequency and other lexical factors (i.e., lexically encoded features). Several studies have shown that the bias is a lexically specific syntactic bias. For example, Trueswell et al. (1993) investigated the difference between verbs that are more frequently followed by an object NP and verbs more frequently followed by a sentential complement. They looked at sentences like **5.6** and **5.7**.

5.6 The student forgot the solution was in the back of the book

5.7 The woman hoped the address was in the directory.

According to the subject–verb–object bias, in example **5.6** the NP *the solution* receives the object interpretation although it is the subject of the reduced relative *the solution was in the back of the book*. Trueswell et al. (1993) found that, after verbs with an NP bias like *forgot*, participants were slow in interpreting it as the relative clause subject, while they could easily do so when the verb had a sentential complement bias like *hoped*. Dussias & Cramer Scaltz (2008) tested the same subject–verb–object bias with L1 and L2 speakers. They found that, after a verb with a sentential complement bias, such as example **5.7**, L2 speakers did not show any difficulty in assigning to the post verbal NP the role of the subject in the following clause. The study by Dussias & Cramer Scaltz (2008) showed that L2 learners can use the statistical frequency information during sentence processing to the same extent as L1 speakers.

In addition to subject-verb-object bias, McRae & Matsuki (2013) highlight the importance of several lexical factors that determine subcategorization preferences. For example, the verb sense is related to the type of syntactic structure in which it occurs. With a verb such as *admit*, the meaning of "let in" has a strong direct object bias, while the meaning of "acknowledge" has a bias toward the sentential complement. Some studies have found that this can affect syntactic processing (e.g., Hare et al., 2003; Roland & Jurafsky, 2002). Another lexical factor related to the subcategorization bias is the morphological form of the verb. Different verb forms, for instance, occur with different frequencies, and they can prime one syntactic structure over another (Bybee, 2008; Trueswell, 1996). Subcategorization bias may be susceptible to the effects of thematic roles. Research has shown that individuals are sensitive to the semantic relations encoded in the subcategorization frames of verbs, which, in turn, can affect syntactic processing (MacDonald et al., 1994; McRae et al., 1998; McRae & Matsuki, 2013; M. J. Spivey-Knowlton & Tanenhaus, 1998).

Constraint-based models do not posit separate stages in the access of lexical information. The processing system can exploit any type of lexical feature as soon as it becomes available. Since constraint-based models have focused on specific phenomena, they have developed linguistic descriptions that are extremely accurate, but tend to be of limited scope and are difficult to generalise. However, a way to generalise the main assumptions of constraint-based frameworks is that of using the formal machinery developed within HPSG. As I have explained in Chapter 4 (see Section 4.3) the HPSG proponents have pointed out how their framework can account for the main properties of language processing, highlighting that HPSG is surface oriented, constraint-based, and strongly lexicalist (Sag & Wasow, 2011, 2014; Wasow, 2019). In this respect HPSG has expanded the assumptions of constraint-based models, developing a grammar that relies solely on lexically encoded information and small a formal setup of principles and rules (Pollard & Sag, 1994; Sag et al., 2003). As a result, unlike constraint-based models, HPSG can be generalised to all sorts of syntactic phenomena (Sag & Wasow, 2011, 2014; Wasow, 2019). The mapping between the syntactic and semantic levels of representations can easily explain how structural and meaning relations interact without positing separate processing stages (Müller & Wechsler, 2014; Sag et al., 2003). Crucially, the hierarchical organisation of the lexicon, enriched by the on-line type construction mechanism (Davis & Koenig, 2019; Koenig & Jurafsky, 1995) places lexical and syntactic representations under a unified theoretical approach.

The overview of the models that I have presented so far shows some key points about the role of frequency in sentence processing. First, frequency information encoded in the verb subcategorization frame is crucial in helping the processing system interpret sentences correctly. In addition, when verb frequency is controlled for in sentence processing, the overall effects change dramatically, arguing for sizeable and very quick interaction effects between syntax and lexical frequency. Since frequency plays a key role in how lexical information is accessed and used to form syntactic structures, its effects should appear in the first stages of language processing. This hypothesis was tested by Staub (2011). Using an eye-tracking design, in a series of experiment on subject-object ambiguities, he found that frequency affected the early measurements of eye fixations, while effects of structural syntactic integration appeared separately in later measurement. Based on the pattern of results, Staub (2011) drew the conclusion that the findings of the study argue for a stage of extensive lexical processing that precedes syntactic integration.

All the evidence examined so far strongly suggest that a more comprehensive explanation for the gap between L1 and L2 processing should consider all the factors that affect L2 lexical processing and lexical representations. Therefore, in the following sections, I will bring together the evidence on the effects of frequency and lexical information on language processing. The aim is to incorporate them in a unified model of lexical and syntactic processing, focusing on L2 speakers.

5.5 Feature Availability and Language Processing

A useful notion to explain how lexical processing works is feature availability. I have briefly introduced the notion of feature availability in Chapter 3. Feature availability refers to the number of features that the processing system can use to activate links between words in the lexical network and form meaningful structures. In this respect, the meaning of feature availability overlaps fully with that of vocabulary knowledge. The only difference is that it links the notion of vocabulary knowledge to the feature-based lexical representations used in HPSG.

5.5.1 *Frequency and Feature Availability*

As explained in Chapter 2, Diependaele et al. (2013) argue that L2 lexical representations are not as “entrenched” and lack the same depth and breadth as those of the

L1 speakers. In addition, L2 vocabulary knowledge is especially strained by low-frequency words, which are more difficult to process for L2 speakers. The result of the I-Lex study reported in Chapter 3 provided evidence that frequency interacts with vocabulary knowledge. It appeared that the L2 speakers had problems organizing their lexical representations and retrieving and accessing low-frequency words as effectively as the L1 speakers.

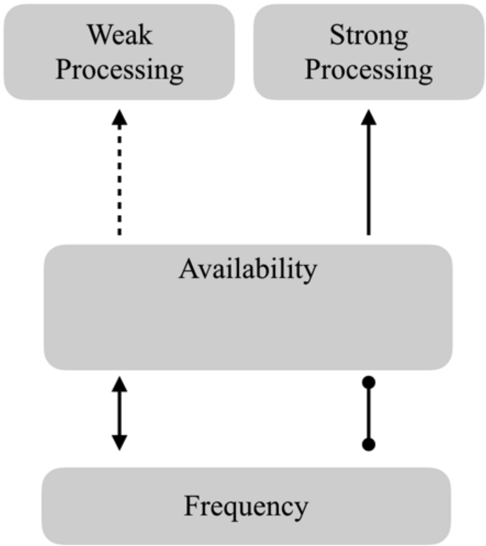
The relationship between vocabulary knowledge and frequency effects is not limited to the processing of words in isolation. It also appears when L2 speakers are processing words in ordinary reading tasks where the discourse and sentence context are controlled for. Cop et al. (2015) analysed L1 Dutch speakers of L2 English while reading paragraphs. The design of the study was based on an eye tracking measurement and had a receptive vocabulary task. The results showed that reading in the L2 produced greater frequency effects on fixation times. However, when participants read in their respective L1s, the frequency effects were of the same size in both the L1 and L2 groups. Moreover, vocabulary knowledge significantly interacted with frequency, in that higher L1 vocabulary knowledge speeded fixation times. Using a similar design with an eye tracking measurement and a reading task, Whitford & Titone (2012) found that gaze durations were expectedly longer when reading in the L2. Greater effects on reading times for low-frequency words and smaller effects for high-frequency words were found for both the L1 and L2 groups. However, the difference between high- and low-frequency L2 words lessened for the L2 speakers who had greater exposure to their second language. Although exposure was measured with a self-report questionnaire, Whitford & Titone claim that it can be regarded as a reliable proxy for vocabulary size, arguing that vocabulary knowledge is determined by the amount of exposure to L2. Since exposure is extremely attuned to the frequency patterns found in the language, high-frequency words that are more often encountered are processed more easily than low-frequency words (Gollan et al., 2008, 2011)

To summarise, frequency and vocabulary knowledge interact so that frequency effects are larger when vocabulary knowledge is low, both in L1 and L2 speakers. This occurs in the lexical processing of single words and, importantly, in reading tasks where no specific syntactic structure is used and words are accessed in a natural discourse context (Cop et al., 2015; Whitford & Titone, 2012). Furthermore, as the overview of usage-based and constraint-based models showed, frequency interacts with syntactic processing using lexical information in a similar way to lexical processing. The interaction is multilayered in that it depends on the interaction between frequency and lexical properties, such as, for instance, the subcategorization bias of a verb and its morphological and meaning features (see Section 5.4).

To describe the relationship of frequency and feature availability and how it underpins processing, I propose a simple model shown in the flow chart in Figure 5.1. The first two levels show the interaction between frequency and feature availability. The arrow with the round ends represents the effects of high-frequency words, while the arrow with pointed ends represents the effects of low-frequency words. With low-frequency words, lexical features are less readily available and more difficult to access. As a result, the processing system may fail to retrieve the features it needs to combine words into a certain syntactic structure. With high frequency words, lexical features are easier to access and more readily available, so that the processing system can retrieve the lexical features needed to combine words into the correct syntactic structure. The process is reflected in the third level of the model in Figure 5.1, which represents the two possible effects on processing caused by a high or low degree of feature availability. While fewer features lead to weak processing, the ability to access the right number of features leads to strong processing.

Figure 5.1

A Feature-based Lexicalist Model of Language Processing



5.5.2 *Bilingual Representations and Feature Availability*

So far, feature availability has been described merely in terms of vocabulary knowledge and frequency. However, in Chapter 2 two other factors, non-selectivity and lexical networks, were a key determinant of how easily lexical features can be available.

Non-selectivity is a key component of lexical representations in L2 speakers and affects how lexical features are stored and accessed. The research on non-selectivity has helped implement models of L2 lexical processing and lexical representations (de Groot & van Hell, 1998b; Dijkstra et al., 2019; Dijkstra & van Heuven, 2002; Grosjean & Byers-Heinlein, 2018; Shook & Marian, 2013). Although based on evidence drawn from experiments on specific lexical items like cognates or homographs, the setup of the models is flexible enough to explain how L2 speakers process lexical features on a broader level. Thus, the key features of the models can be regarded as inherent to L2 lexical processing, with the overlap between the L1 and L2 making them more prominent.

The first general claim of the models is that, unlike L1 speakers, L2 speakers need a higher degree of cognitive control to handle the potential competition arising from shared lexical information (Bialystok et al., 2008; Schwartz & Kroll, 2006b). Models also make claims about the way L2 lexical processing of semantic and word-form features occurs. In terms of meaning features there is an asymmetry between L2 and L1 conceptual representations (e.g., Kroll et al., 2010; Tokowicz, 2014). Specifically, the number of meaning features that a person knows in their L1 leads to more detailed lexical representations, whereas the lower number of meaning features known in the L2 leads to coarser-grained lexical representations (Finkbeiner et al., 2004; Kroll & Stewart, 1994; Tokowicz, 2014). Additionally, there is evidence that some meaning features are represented in a different way in L1 and L2 speakers. For instance, the Distributed Feature Model predicts that L2 speakers tend to access concrete words more quickly and more effectively than abstract words (de Groot & Brink, 2010; de Groot & van Hell, 1998a).

In terms of word-form features, phonological and orthographical features are activated based on the similarity of the input word to the words stored in the mental lexicon. Backward activation and inhibition caused by semantic features or lexical representations already activated may also change the activation level of word-form features (Dijkstra & van Heuven, 2002; Shook & Marian, 2013).

The models of bilinguals' lexical representations clearly suggest that lexical access in L2 speakers has specific traits. The L2 processing system must be more attuned to the overlap between the two languages, in order to respond to effects that can be either facilitatory or require inhibition (Kroll et al., 2008; Poort & Rodd, 2019).

In addition to the models of bilingual lexical processing and representations, in Chapter 2 I have examined the basic architecture of L2 speakers' lexical networks. An important aspect of lexical networks is that they are linked to frequency and feature availability. As noted, frequency underlies how lexical networks are built, with low-frequency words added to the more frequent and more connected words in the networks to differentiate meaning (De Deyne & Storms, 2008a; Steyvers & Tenenbaum, 2005). This growth process appeared more thoroughly in wide association networks assembled in large-scale studies (De Deyne et al., 2019; De Deyne & Storms, 2008b; Vitevitch, 2008). However, as word association tasks offer a reliable excerpt of lexical representations, the way lexical networks grow can be considered a universal property that has a central role in the human semantic system (De Deyne & Storms, 2015).

The results obtained by word association tasks in small-scale experiments are also relevant in terms of feature availability. The links between stimuli and associates reflect the type of lexical information available to L1 and L2 speakers (De Deyne & Storms, 2015; Fitzpatrick & Thwaites, 2020). Choosing one connection over another might be indicative of the type of lexical features available. As a result, if fewer feature are available, fewer links can be built, and this appears in a less dense lexical network (Meara, 2009). Low-frequency words are more problematic here, as their lexical features are less available. The results of the I-Lex validation study in Chapter 3 have supported this notion. The study has shown that L1 speakers provided more low-frequency words than both L2 groups, because their vocabulary knowledge was deeper, and they could rely on many more available lexical features. In contrast, the number of lexical features that L2 speakers can access is reduced, so that the number of links that can be activated among the words in the network is significantly lower. This effect is greater in the case of low-frequency words so that the links between them and the rest of the network might not be activated at all (Meara, 2009).

5.5.3 *Feature Availability and Sentence Processing*

All the evidence examined so far strongly suggest that a more comprehensive explanation for the gap between L1 and L2 processing should consider all the factors that affect L2 lexical processing and lexical representations.

The first lexical effect that might be crucial in determining the difference between L1 and L2 sentence processing is the interaction of feature availability with frequency. This claim, which resonates with the main assumptions of constraint-based models, can be easily translated using the feature structures of HPSG described in the previous chapter. For example, if the syntactic features SPR, COMPS, and MOD are not available, the processing system may fail to link a word to its complement or to its subject/determiner. In contrast, high-frequency words can improve the availability of lexical features. When the processing system can easily activate the syntactic features SPR, COMPS, a word can be linked to its complement and its subject/determiner. The example is limited to syntactic features, but, as constraint-based models show, frequency might affect other types of lexical information. For example, frequency can affect the availability of morphological features that encode information about the tense or features that encode information about thematic roles. Both types of features can play a key role in how successfully the process system performs syntactic processing. Again, HPSG can handle this type of effects using the feature PHON and morphological features such as NUM or PER that underlie various agreement processes (see Chapter 4, Section 4.4). I will further explain how HPSG models how words are linked together in lexical networks in the last section of the chapter, where I will set up a unified model of lexical and syntactic processing.

The frequency effects on the processing system are in contrast with the claim of syntax-first models, where frequency is not regarded as a proxy for syntax (Frazier, 2013). Procedures such as the Minimal Attachment occur automatically and are independent of the frequency of lexical entries (Frazier, 2013). The Shallow Structure Hypothesis does not make specific claims about frequency and, in principle, frequency could be added into the framework. In contrast, the interaction between vocabulary knowledge and frequency plays a prominent role in L2 speakers: frequency affects the activation of all kinds of lexical features and its effects have been found in several processing tasks based on distinct types of lexical information (Brysbaert et al., 2017, 2018; Diependaele et al., 2013).

Furthermore, a key factor in bilingual lexical processing models is that lexical features interact during the entire lexical access process. Phonological features are the first source of information to be accessed to recognise a word, but this process can be extremely difficult for L2 speakers. Indeed, they have to handle phonological features of both languages that may carry conflicting information, so the process needs more cognitive resources (Bialystok et al., 2008; Poort & Rodd, 2019). In addition, while the processing system is handling phonological features, activation may spread to meaning features which, in turn, might cause backward inhibition or facilitate phonological processing (Dijkstra et al., 2019; Dijkstra & van Heuven, 2002). The interaction between meaning and word-form features occurs quickly. Such a level of complexity argues against an ordered process of lexical access that starts with syntactic features. In fact, a great deal of lexical features must be accessed concurrently or even before syntactic features are available, and the process cannot always operate orderly.

The fact that accessing features is a complex process that might not be ordered and might proceed based on the task and the information available has gained support from several studies that have used event-related potentials (ERPs). For example, Guo & Peng (2007) showed that, in L2 production, semantic features are retrieved earlier than phonological features. Shantz & Tanner (2020) found that L2 speakers, not differently than L1 speakers, can retrieve grammatical gender and phonological features in a non-ordered fashion, depending on the type of task. Although limited to L1 speakers, there is also evidence that the retrieval of phonological features can occur concurrently with the retrieval of semantic features (Abdel Rahman & Sommer, 2003). Furthermore, in production tasks, phonological retrieval is not rigorously dependent on earlier semantic feature processing (Abdel Rahman et al., 2003).

The studies described above show that lexical features are accessed opportunistically, with no determined paths. All this reflects a multilayered picture in which lexical features interact in complex ways. Therefore, it seems challenging to assume that the interaction between meaning and form features can occur without affecting the access of syntactic features. Moreover, incorrect processing in L2 speakers may be caused by not associating the correct grammatical features to a lexical representation or, in contrast, by associating the incorrect ones (Corver, 2003; Hawkins & Liszka, 2003). In relation to this, Corver (2003) argued that all the features that are used to build syntactic structures in a sentence are the same features as those at the word level. Given this hypothesis, difficulties in accessing syntactic features encoded in words are also reflected at the structural level. He highlights how the problems showed by L2 speakers "can now be due to mis-categorisation (i.e. 'mis-' from the perspective of the target-language): an incorrect categorial feature is associated with some sound-meaning pair." (p. 52-53).

To summarise, if building the correct syntactic structure requires the processing system to find the information in lexical items, all the complex intertwining between word-form and meaning features is expected to play a key role. This is likely to affect the L2 lexical representations even more, as they are not at the same level of complexity as those of L1 speakers. This might not be straightforward to reconcile with syntax-first approaches, where syntactic features are available separately from other lexical features and have little in common with the rest of lexical information.

The importance of L2 lexical access in building syntactic structures comes from a series of studies that have examined how frequency can affect sentence processing in L2 speakers. For instance, Hopp (2016) investigated the filler-gap dependencies, finding that the L2 speakers problems were worse with low frequency verbs but improved with high frequency verbs. Other studies have found that vocabulary size and lexical representations interact with the ability of L2 speakers to handle filler-gap dependencies. Kawaguchi (2016) found that vocabulary size predicted the ability of L2 speakers to form complex wh-questions and Miller (2014a) and Hopp (2017) found that the use of cognates reduced the effects of complex syntactic structure on L2 speakers. In the next two chapters, I will describe some of those studies in detail, introducing my own studies on the effect of frequency and vocabulary on syntactic processing. As for now, the results of the studies lend themselves to re-examine the role of frequency and vocabulary more thoroughly.

Finally, high frequency and the use of cognates are not the only factors that can reduce the disadvantage of L2 speakers in syntactic processing. As shown in Chapter 2, syntactic features are sensitive to non-selectivity to the same extent as other lexical features. When syntactic features of the two languages overlap, syntactic processing of the L2 is facilitated. Syntactic facilitation is also enhanced by other types of lexical information: when the head of the syntactic structure in the L1 shares word-form features and meaning features with the L2, syntactic processing is facilitated (Hartsuiker & Bernolet, 2017). Yet, the syntactic overlap between the L1 and the L2 is dependent on how L2 speakers access other lexical features, which is affected by the difficulties described above. Despite a degree of facilitation for overlapping syntactic patterns, cases of mis-categorisation tend to be frequent in syntactic processing of L2 speakers.

Considering the multi-layered nature of lexical representations, it seems that feature availability can be a reliable notion to explain both lexical and sentence processing. This approach also aligns with the claims that underlie HPSG representations. A processing compatible grammar needs to be lexicalist, feature-based, and make use of both syntactic and non-syntactic features to combine words into well-formed structures (Sag & Wasow, 2011, 2014; Wasow, 2019). Hence, in the next section I am going to set up a model, based on HPSG, that unifies the effect of feature availability on lexical representations and sentence processing.

5.6 A New Model: The Feature-Based Lexical Network Model

This section aims to incorporate the key points about L2 lexical representations outlined so far into the feature-based grammar developed in HPSG. The aim is to set up a model that can account for the multilayered nature of lexical items, the setup of lexical networks, and how they underpin sentence processing.

Concerning the first factor, words encode a considerable amount of information, and it is crucial to put in place a set of principles that can describe how the different bits of information interact. For example, Fitzpatrick (2006) proposed three labels to describe the lexical information that underpins how words can be associated. The first label relates to word-form features, the second to meaning features, and the third to combinatorial features. This aspect is not difficult to accommodate into HPSG, for the three primary features PHON, SYN and SEM encode exactly the same type of phonological, syntactic and semantic information proposed by Fitzpatrick (2006). In addition, within HPSG, the three labels also reflect that the processing system can simultaneously use diverse sources of lexical information to combine lexical items.

Concerning lexical networks, a key point is that, since words can encode multiple features, multiple types of links may exist simultaneously within networks. According to Cong & Liu (2014) the presence of multiple links might be described using diverse types of subnetworks that can work separately based on the type of links they contain. An inventory of the various networks will contain at least a semantic network to model meaning relations, a dependency network to model syntactic relations, and a co-occurrence network to model the linear ordering of linguistic expressions (Cong & Liu, 2014). However, using subnetworks introduces an unwieldy level of complexity, since it would be difficult to work out how multiple networks operate at the same time, which one is used first, and how they interact.

A workable approach is to set up, for every word, a list that contains all the lexical features which the processing system may use to activate a link to another word. Such an approach makes the entire organisation of lexical networks feature-based. The grammar developed by HPSG can straightforwardly describe how the features of the network work and what type of information they encode. In a feature-based network, meaning features are most likely to guide how words form links to one another. A straightforward piece of evidence comes from word association tasks, where meaning-based association tend to be prevalent (e.g., De Deyne & Storms, 2015; Fitzpatrick & Thwaites, 2020). Combinatorial features are employed to encode structural relations between two words, and the processing system normally activates them in addition to meaning connections. These processes fall under the scope of the features SPR, COMPS, MOD and RESTR, and more specific features can be easily added via Frame Semantics. For example, in a simple transitive verb phrase, the link between the verb and the noun is based on combinatorial and meaning features. The combinatorial feature COMPS ensures that the noun is linked to the verb as the object. The feature RESTR, which describes the meaning restrictions of the verb, determines which nouns are more likely to be part of the transitive structure.

Word-form features can be regarded as an inventory of linguistic units that are activated when word meaning is connected to a form, reflecting either the first stage of language comprehension or the last stage of language production (e.g., Evans & Levinson, 2012; Levelt & Meyer, 1999). When two words have similar form features, a link between them may be activated. A subset of word-form features is related to morphological forms that depend on a series of linguistic phenomena occurring during language processing, such as agreement, tense marking, and passive-active alternations (Adger & Svenonius, 2011; Sag et al., 2003). Without going into further details about morphological phenomena, one might simply assume that they work like combinatorial features, thus encoding general type of structural relations. Links based on word-form features are encoded in the feature PHON that captures not only the form of single words but also of larger constructions. The process of online type creations described in Chapter 4 (see section 4.9) can handle morphological processing in a straightforward manner, without the need of introducing specific features (Davis & Koenig, 2019).

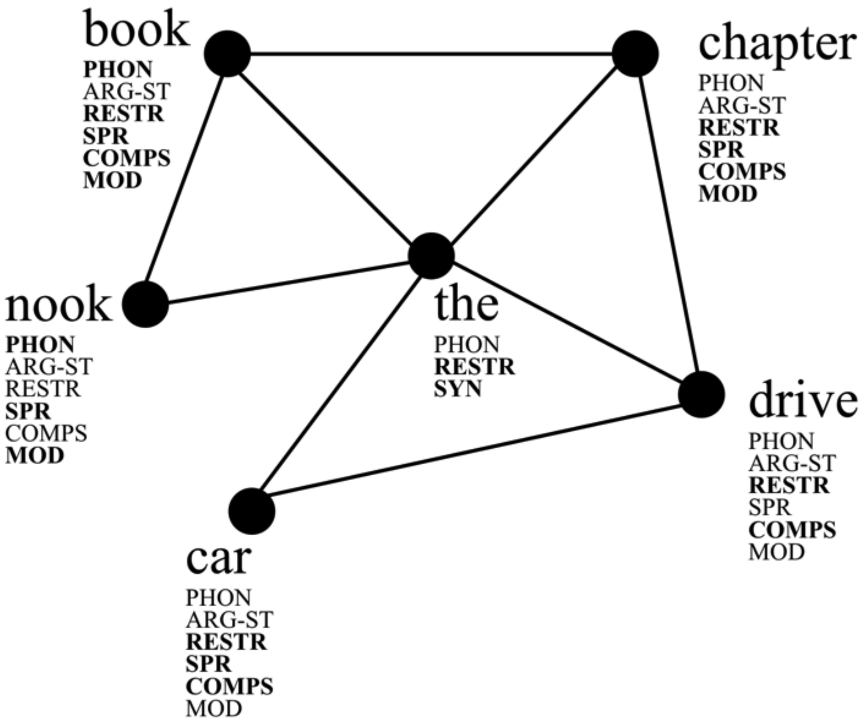
A feature-based network also needs to explain how words are combined online. Link activation is the main process. First a word is retrieved or recognised along with the features it encodes, then one link towards another word is activated. However, the processing system cannot simply add features until it has enough resources to switch on a link between two words. The small number of rules used in HPSG to form phrasal constructions can model the activation of links between words in a lexical network. Rules are extremely simple and are based only on a single underlying device: features that are compatible can be unified. In network terms, this means that, when two words have features that do not carry conflicting information, the link between them can be activated. Importantly, unification is not ordered, and it can proceed with the features available when a word is encountered. Since HPSG rules are nothing short of feature structure describing complex lexical bundles, they can be considered as lexical rules. This adds theoretical ground to the idea that the lexical information is enough to build sentences, without using different sets of rules and principles.

Importantly, the feature-based network assumes that the strength or weight of links is determined by the type and the number of features available for every word, arranged into the phrasal constructions that are part of the grammar. As more features become available, links increases its strength. Links based on meaning features are expected to be the strongest, although there are cases where other types of link can supply enough information to set up strong connections. The notion of associating a value to a link is useful as it allows us to add further information. A straightforward way to look at it is to compare a lexical network to a road map. The roads that connect cities can be labelled according to the distance or traffic. Both factors can facilitate or hinder travel from one city to another.

Frequency also plays a role, in that features encoded in more frequent words are more easily available to activate links between words, in terms of lexical and structural processing (Gollan et al., 2008, 2011; Hopp, 2016; Tily et al., 2010). The fact that lexical features determine the strength of connections does not affect the main properties of the network, which are determined by the number of links and nodes contained in it (Meara, 2009). The strength of the links determines the ease of processing, in that strong links are activated more easily, whereas weak links require more cognitive resources (e.g., MacDonald et al., 1994; McRae et al., 1998). The feature-based lexical network is depicted in Figure 5.2.

Figure 5.2

An HPSG feature-based lexical network



Each word in the network has rich lexical information about form, syntax, and meaning. The features that contribute to activating a link from one word to another are in bold. Every link is activated by specific features combined in a specific rule or phrasal construction. The words *drive, car, book,* and *chapter*, have meaning-based connections encoded in the feature RESTR. Combinatorial links underpin the segments *the book, the chapter, the book chapter, the car, the nook, book nook* and for VP *drive the car.* All combinatorial links are determined by syntactic features. The links to form the NPs *the book, the chapter, the car, the nook,* are activated by the features RELN, RESTR, SPR combined in the Head Specifier Phrasal-construction. The links to form the NPs *the book chapter* and *the book nook,* are activated by the features RELN, RESTR, SPR, and MOD combined in the

Head Specifier Phrasal-construction and in the Head Modifier Phrasal-construction. The links to form the clause *drive the car* are activated by the features RELN, RESTR, SPR, COMPS combined in the Head Complement Phrasal-construction.

In the following chapters, the model is going to be applied to filler-gap dependencies, in two studies that make use of the I-Lex task to measure vocabulary and frequency to weigh the role of feature availability in sentence processing.

5.7 Conclusions

The chapter has described three different models of sentence processing, each relying on a different source of information. Constraint-based models are better suited to reconcile the role of frequency and vocabulary knowledge in sentence processing. With the support of the lexicalist grammar developed by HPSG, the constraint-based model can be formalised with greater precision. Considering the pervasiveness of lexical effects, I have proposed a feature-based lexical network model in which lexical and syntactic processing can be unified. The model also incorporates the notion of feature availability. In the next two chapters, I will test the model and the notion of feature availability by measuring the effects of frequency and vocabulary knowledge on two on-line tasks. The two experiments aim to extend the results of the studies by Hopp (2016) and Tily et al. (2010) briefly described in section 5.5.3. The two studies have found that, in some type of filler-gap dependencies, frequency interacts with syntactic processing in L1 and L2 speakers. Moreover, the experiments I will present in the next two chapters will test whether the feature-based lexical network model can be used on L2 speakers of Italian. This will also offer the opportunity to extend the theoretical machinery of HPSG to the study of the L2 and test its flexibility on real linguistic data.

Chapter 6 Study 1

6.1 Introduction

In the previous chapters, I have described the importance of frequency as a reliable tool to measure lexical representations. Chapter 2 showed that, according to the Lexical Entrenchment Hypothesis, frequency effects become more prominent when lexical knowledge is low (e.g., Brysbaert et al., 2017; Cop et al., 2015; Diependaele et al., 2013). This pattern appeared in the study describe in Chapter 3. The results of I-Lex showed significantly higher scores for L1 than for L2 speakers. In terms of lexical representations, L1 speakers build complex lexical networks that have more links towards low-frequency words. In contrast, L2 speakers, because of their limited exposure to the language, build less complex networks with fewer links to low frequency words. The interaction between frequency and vocabulary knowledge can be shaped in terms of features availability. Frequency determines the number of features available to form new links in the network.

In addition, research on vocabulary knowledge and lexical processing has shown how each word encodes several types of information: syntactic features, meaning features, word-form features, and information about frequency (see, e.g., Dijkstra et al., 2019; Dijkstra & van Heuven, 2002; Fitzpatrick, 2006; Nation, 2013). This suggests that a theoretical framework where lexically encoded information drives any other linguistic operation might be well suited to explain both lexical and sentence processing effects (e.g., MacDonald et al., 1994; Sag & Wasow, 2014). The HPSG theoretical approach described in Chapter 4 meets all the requirements: it is lexically driven and it does not posit any specific domain outside the lexicon (Pollard & Sag, 1994; Sag et al., 2003). The differences between language objects like words and phrases are entirely due to the type of information encoded into lexical features (Boas & Sag, 2010; Sag, 1997; Sag et al., 2003). This approach aligns with experimental results showing how frequency affects the way the processing system interprets syntactic structures (Ellis, 2013; Ellis et al., 2014; Trueswell, 1996; Trueswell & Tanenhaus, 1994). In a lexically driven and feature-based theoretical approach, frequency effects are unproblematic. They stem naturally from the fact that frequency hampers the availability of the lexical features needed to form a specific syntactic structure. The simple feature-based lexical network shown in Chapter 5 can accommodate these effects in a straightforward manner. Lexical features activate links between words in the lexical networks. The HPSG feature-based grammar offers explicit information on the type of features and the type of links.

This chapter presents a study that examines the relationship between frequency, vocabulary knowledge, and sentence processing. The study focuses on the processing of a specific syntactic structure, cleft sentences, a type of filler-gap dependency that has a focused constituent (e.g. *it was John*) followed by a relative clause (*who drunk the wine*). The first section describes how the processing system tends to be slower when interpreting object relative clauses, analysing the various explanations that have been brought in to explain it.

Next, the chapter describes some pivotal studies that have shown how vocabulary knowledge interacts with syntax (David et al., 2009; Rogers et al., 2018). It presents studies that have investigated the role of frequency (Hopp, 2016; Tily et al., 2010) and vocabulary size (Kawaguchi, 2016) in filler-gap dependencies. Two studies have specifically investigated the difference between object and subject relative clauses embedded into cleft sentences (Hopp, 2016; Tily et al., 2010). Kawaguchi (2016) has focused on the relationship between receptive vocabulary knowledge and the ability of L2 speakers to provide certain types of syntactic structures (Kawaguchi, 2016). The main section of the chapter presents the study on sentence processing. The study is based on the same item design as Hopp's (2016), but it involves a sentence repetition task instead of Hopp’s SPR and I-Lex as a measure of vocabulary knowledge. The formal machinery is based on the feature-based lexical network and the HPSG features structures.

In the discussion section I will apply the feature-based lexical network model developed in Chapter 5 to the result of the experiment. The model is expected to reflect the interaction between the availability of lexical features and syntactic processing in terms of the activation of structural links between words in the lexical network.

6.2 Object and Subject Relative Clauses in Language Processing

Consistent experimental evidence has shown that object relative clauses are processed more slowly than subject relative clauses. This has been observed in L1 speakers (Gibson et al., 2005; Lau & Tanaka, 2021; Santi et al., 2019; Staub, 2010; Traxler et al., 2002, 2005) and L2 speakers alike (see Juffs & Rodrigues, 2015 for review). In contrast, a processing advantage for subject relative clauses has consistently emerged from research. Most of the findings show that subject relative clauses are acquired earlier and are easier to produce and to process than object relative clauses in many languages (Lau & Tanaka, 2021). The learning process mirrors the noun phrase accessibility hierarchy of Keenan & Comrie (1977), according to which subject relative clauses are the easiest type of relative clause, followed by object and indirect object relative clauses.

Researchers have proposed various explanations to account for the difference, based on discourse factors (e.g., Gordon et al., 2004; Mak et al., 2008; Roland et al., 2012), structural properties (e.g., Clifton & Frazier, 1989; Juffs & Rodrigues, 2015), and memory-based accounts (e.g., Gibson, 1998, 2000; Gordon et al., 2002; Gordon & Lowder, 2012; Parker et al., 2017; Van Dyke & Johns, 2012).

Concerning discourse factors, the difference between object and subject relative clauses can be mitigated by discourse context (e.g., Altmann & Steedman, 1988; Mak et al., 2008;

Roland et al., 2012; Warren & Gibson, 2002). Mak et al. (2008), for example, noted that when the subject of the object relative clause is the topic in the previous discourse context like *the burglar* in Example **6.1**, there is no difference in reading times between subject and object relative clauses.

6.1 The burglar has been arrested during a burglary in a large villa. The police have told the occupants, whom the burglar has knocked down, that the man has committed more crimes.

Another discourse-based factor that reduces the disadvantage of object relative clauses is the referent of their subject. Several studies have shown that when object relative clauses contain a personal pronoun, they become easier to process (e.g., Gibson & Warren, 2004; Kaan, 2001; Roland et al., 2012).

Structural accounts have explained the difference between subject and object relative clauses in term of syntactic structure (Clifton & Frazier, 1989; Pickering & Traxler, 2006). The Active Filler Hypothesis assumes that if there is a filler that is active, i.e., it has not been attached to its gap, the processing system seeks to locate it at the first available segment (Clifton & Frazier, 1989). In object relative clauses like Example **6.2** (taken from Traxler et al. (2002), p. 85), the processing system tries to insert the subject of the main sentence, *the businessman,* as the subject of the relative clause. This leads to misanalysis and forces to rebuild the structure of the sentence, assigning to the NP *the secretary* the role of subject of the relative clause. This further processing stage makes object relative clauses slower to interpret (Clifton & Frazier, 1989; Traxler et al., 2002).

6.2 The businessman that the secretary married invited the bookkeeper to the party.

The reason processing system tries a subject interpretation of the relative clause is related to the principles of syntax-first models. As outlined in Chapter 5, those models predict that the processing system first builds the structure with fewer constituents (Frazier, 2013).

Memory-based accounts focus on the cost to the processing system to keep the filler encoded in memory until the gap is encountered (Gibson, 1998, 2000). The cost of memory encoding may be caused by the integration of new discourse referents between the modified NP and the relative verb. Since object relative clauses put an additional referent between the relative verb and the modified NP they need a costlier integration process (Gibson, 1998, 2000). Alternative memory-based accounts focus on the interference that might occur between the modified NP and the subject NP of the object relative clause (Gordon et al., 2001, 2004; Gordon & Lowder, 2012; Van Dyke & McElree, 2011). If the two NPs have distinct representations reliant on separate lexical features, the cost of keeping them distinct is low. If the representations are similar, to distinguish between similar lexical features is costlier for the processing system (Parker et al., 2017; Van Dyke & Johns, 2012; Van Dyke & Lewis, 2003). In subject relative clauses, the modified NP is also the subject of the relative so that there are no other competitor NPs.

Furthermore, a factor that can affect the difference in how the processing system handles subject and object relative clauses is related to the meaning features of both the modified NP and subject NP in the object relative clause. Traxler et al. (2002) found that meaning features like agentivity and animacy could reduce the difference between object and subject relative. De Vincenzi (1991), analysed object and subject direct wh-questions in Italian, finding that subject relative clauses with the animate filler were read faster than object relative clauses, but the same difference did not appear in relative clauses with an inanimate filler.

Some studies have started looking at the effects generated by the relative frequency of the verb in the relative clause, finding that frequency interacts with the differences between subject and object relative clauses (Hopp, 2016; Reali & Christiansen, 2007; Tily et al., 2010). These studies that lay the basis for the present experiments are described in the following section.

6.3 Relative Clauses in L2 Speakers

The first studies that investigated the way L2 speakers handle relative clauses focused on how they were learned. One first result was that the learning process mirrored the noun phrase accessibility hierarchy of Keenan & Comrie (1977). In both instructional context and natural acquisition, L2 speakers start processing subject relative clauses and then move along the hierarchy towards more complex forms (Eckman et al., 1988; Gass & Ard, 1984; Lau & Tanaka, 2021). Other factors have been addressed in addition to the hierarchy to explain how L2 speakers learn relative clauses, based on the position of the relative sentence within the main sentence (Izumi, 2003). Looking at studies based on online processing tasks, the disadvantage caused by object relative clauses concerns L2 speakers as well. For example, Juffs & Rodrigues (2015) reported a study in which they found that both L1 and L2 showed slower reading times for object relative clauses.

Several studies have adopted a more specific approach specifically looking at the relation between vocabulary size and morphosyntactic development, comparing vocabulary measures and specific syntactic structures (David et al., 2009; Rogers et al., 2018). In the first study, David et al. (2009) investigated the relationship between lexical diversity and various grammatical features in speakers of L2 French. In the study, lexical diversity was measured with the Guiraud index (Guiraud, 1954), which is an alternative reliable measurement based on the ratio between types and tokens. Various syntactic categories were also examined and compared with the Guiraud index: gender, the ability to produce VPs, the ability to use subject clitics, and to form embedded clauses. In the most recent versions of transformational approaches, all these categories are determined by grammatical features that are lexically encoded (Adger, 2003; Adger & Svenonius, 2011; Chomsky, 1995). Hence David et al. (2009) reasoned that a correlation of such syntactic categories with a measure of lexical complexity could be expected to appear. For some of the categories examined, it was indeed the case: lexical diversity correlated with the production of VPs and embedded clauses (CPs); however, no correlation was found with either gender and the use of clitics. In grammatical terms, the use of clitics was intended to measure the ability for the L2 speakers to build the specific functional category TP (Tense Phrase). The TP is the phrasal configuration that accounts for various grammatical properties, like the presence of the subject and the tense features of the entire sentence, modal verbs and negation, and the clitics (e.g., Adger, 2003).

To evaluate whether L2 French speakers had acquired the grammatical features to form a TP, Rogers et al. (2018) used a different approach from David et al. (2009). They examined the position of the negation and adverbs as a foothold to evaluate the presence of the syntactic features needed to build the TP. In the study, the comparison was between an oral production task and a grammaticality judgment task and a receptive vocabulary test. They found a significant high correlation between receptive vocabulary size and both the oral task and the grammaticality judgement task. These studies are among the first to find that some specific vocabulary measurements correlate with the ability to form complex syntactic structures. They resonate with the claim that the specific properties of the lexical representations of L2 speakers argue for an interaction between vocabulary size and the ability to process complex syntactic structures.

Given the notions laid out in Chapter 5 and it would be crucial to see whether lexical effects, framed as interaction between frequency and vocabulary knowledge, could appear in complex syntactic structure like filler-gap dependencies. Traditionally, relative sentences have been regarded as structurally-driven phenomena, relatively immune to lexical effects. However, one study showed that vocabulary and syntactic development in wh-questions progress conjointly (Kawaguchi, 2016), while two studies (Hopp, 2016; Tily et al., 2010) shown that frequency interacts with the effects of syntactic structure in cleft sentences. In the next section, these three studies are going to be examined. The rest of the chapter presents a study based on the same item design as Hopp's (2016) with a different processing task and a focus on a different L2, namely, Italian.

6.4 Vocabulary Size, Frequency and Lexical Access in Filler-Gap Dependencies

Kawaguchi (2016), working within Processability Theory (Pienemann, 1998), investigated whether vocabulary size interacted with the language development of the L2 speaker and their ability to produce correct wh questions. The study used The Vocabulary Size Test (Beglar & Nation, 2013) to measure receptive vocabulary size. Based on vocabulary scores, L2 speakers were divided into three groups. A low group with an estimated receptive vocabulary knowledge of the first 5000 English words, a mid-group with a knowledge of the first 6000 and a high group with a knowledge above the 10,000-word family level. To compare vocabulary knowledge and the ability of L2 speakers to produce wh-questions, a spot the differences task was used. In the task, a participant and an interviewer must find differences in similar pictures by asking questions to each other. L2 speakers with a small vocabulary were less accurate in producing wh questions. Those with a mid-vocabulary size could produce wh questions using a variety of forms (which, how long, how many, etc.), still showing some inaccuracies. The high vocabulary group could also produce a variety of wh questions completely avoiding ungrammatical forms. The results of

Kawaguchi (2016) offered a reliable contribution to the hypothesis that vocabulary size is related to the syntactic development of L2 speakers.

Frequency effects in the processing of filler-gap dependencies were found for L1 speakers in a study carried out by Tily et al. (2010) using subject and object cleft sentences that contained either a low- or a high-frequency verb, producing quadruplets like the following one (taken from Tily et al. 2010, p. 927). Examples **6.3** and **6.4** are subject and object clefts with a high frequency verb (*pleased*), while **6.5** and **6.6** are subject and object clefts with a low frequency verb (*assuaged*).

6.3 It was Brendan who pleased Amanda with a radiant gift.

6.4 It was Brendan who Amanda pleased with a radiant gift.

6.5 It was Brendan who assuaged Amanda with a radiant gift.

6.6 It was Brendan who Amanda assuaged with a radiant gift.

The critical regions examined in the analysis were the VP after the relative pronoun (*pleased Amanda* vs. *Amanda pleased*), which may be interpreted as the beginning of either a subject or an object cleft and the segment of the sentence formed by the two following words (the preposition and the following determiner, *with a*). The study used a self-paced reading task and had frequency as the main predictor. Tily et al (2010) used two models in their analysis. In the first, cleft verbs were transformed into a binary predictor and considered high- or low-frequency. In the second, the frequency was added as a continuous variable, so that each cleft verb had its own absolute frequency. The two models produced different results.

The results of the first model showed that, in the cleft region, clefts with high-frequency verbs and subject clefts were both read more quickly. In addition, frequency and cleft type showed a significant interaction so that, in clefts with high frequency verbs, subject clefts yielded faster reading times than object clefts. In contrast, no significant difference between object and subject cleft appeared for clefts with low frequency verbs. At the post-cleft region, the analysis of reaction times showed faster reading times for high-frequency verbs and subject clefts. The interaction between frequency and cleft type was not significant, but the data showed that the larger difference in reading times was between subject and object clefts with low frequency verbs.

The second model, with frequency as continuous predictor, replicated the same reading times in the cleft region. At the post-cleft region, a significant interaction between frequency and cleft type appeared for low-frequency verbs, producing faster reading times for subject clefts. The results show that frequency can hinder or delay the automatic building of the syntactic structure. As Tily et al. (2010) note:

> [w]hen the verb is high frequency, lexical retrieval is fast, and the structural operation is performed during the verb region; when the verb is low frequency, retrieval is slow, and the structural operation is delayed until the post cleft region. (p. 919).

More importantly, results seem compatible with the notion of feature availability (see Chapter 5, Section 5.5). When the processing system manages to collect enough features encoded into the cleft verb it can easily use them to build the right syntactic structure. If availability is disrupted by frequency, it takes time to make up for the missing features so that the effects spill over into the post cleft region.

Hopp (2016) replicated the study by Tily et al. (2010), adding a group of L2 speakers. In the study sentences were arranged in the same way as Tily et al (2010). Examples **6.7** and

6.8 show subject and object clefts with a high-frequency verb (*scared*), while **6.9** and **6.10** show subject and object clefts with a low-frequency verb (*frightened*).

6.7 It was Amanda who scared Sulena with a frightening look.

6.8 It was Amanda who Sulena scared with a frightening look.

6.9 It was Amanda who frightened Sulena with a frightening look.

6.10 It was Amanda who Sulena frightened with a frightening look.

The study also used a self-paced reading task to investigate the effect of frequency on lexical access and its impact on the building of the relative segment inside the cleft construction. A group of advanced German English learners and a group of native English speakers participated in the experiment. As in the study by Tily et al. (2010), two critical regions were examined in the analysis: the VP after the relative pronoun (*Sulena scared* vs. *scared Sulena*) and the two following words (the preposition and the following determiner: *with a*). Every object and subject pair had a low frequency verb such as *to frighten* and a high frequency verb such as *to scare*.

The analysis of the L1 speaker group showed no effect of frequency in the cleft region and a significant effect of cleft type whereby the subject clefts were read faster. The interaction between frequency and type was not significant though. A post hoc analysis showed that there was still a highly significant effect of syntactic type for the low-frequency verbs, with subject cleft sentences yielding faster reading times. In the post cleft segment, the pattern remained unaltered: significantly faster reading times for subject cleft sentences, no effect of frequency and no interaction between the two. Post hoc analysis showed that low frequency significantly affected syntactic structure in that object cleft sentences were read more slowly than subject clefts.

Concerning the L2 group, a different pattern arose from the analysis of reading times. In the cleft region, the L2 speakers showed a significant effect of frequency and a significant interaction of structure and frequency. The post hoc analysis showed that there was a significant effect of structure for the high-frequency verbs, with subject clefts read more faster than object clefts. In the post cleft region there was a significant effect of structure yet no effect of frequency and a marginally significant interaction of structure and frequency, since L2 speakers showed a significant effect of structure only with low frequency verbs. Further analysis carried out using the frequency of cleft verbs as a continuous variable strengthened the previous pattern. The reading times of L2 speakers increased for object cleft sentences with high frequency verbs at the cleft segment while there was no structural difference between object and subject clefts for low frequency verbs. In the post-cleft segment, low-frequency verbs produced significantly longer reading times for object cleft sentences, while the difference between object and subject clefts was only marginal for high-frequency verbs.

Hopp (2016) highlights that, in the L2 group, there is a linear relationship between frequency effects and the region where differences in reading times arise. When the verb frequency decreases, the processing difficulty of building an object cleft sentence progressively moves from the cleft to the post-cleft region. In short, when L2 speakers had problems in accessing the verb because of its low frequency, the effects of syntactic processing arise only after the processing system has gone past the verb region. Therefore, according to Hopp (2016), the differences between L1 and L2 speakers in syntactic processing may not be caused by lack of precision in assembling the sentence structure. Instead it may simply result from the slowdowns in lexical access caused by the difficulty showed by L2 speakers in lexical access. As a result, this affects syntactic processing, in that it is slowed by the initial ‘lexical delay’. Hopp (2016) labels the relation between frequency and syntactic processing as the lexical bottleneck hypothesis, which assumes that, in L2 speakers, 'difficulties in lexical processing that subserve parsing can lead to nonnative-like sentence processing' (p. 21).

6.5 The Study

Hopp's (2016) study is the first that compared effects of syntactic complexity in the L2 processing of filler-gap dependencies with the slowdowns in lexical access shown by L2 speakers. As predicted by the Lexical Entrenchment Hypothesis, the effects of frequency on L2 lexical processing become more prominent when lexical knowledge is low (Brysbaert et al., 2017; Diependaele et al., 2013). Additionally, the fact that this initial 'lexical delay' is caused by low frequency verbs ties in with the results of I-Lex analysed in Chapter 3 that showed how frequency underlies the different degree of complexity of L1 and L2 lexical representations. In the study, the significantly higher scores of L1 speakers are due to their more complex lexical networks that have more links towards low-frequency words. In contrast, L2 speakers rely on less complex networks with fewer links to low-frequency words (Meara, 2009; Wilks & Meara, 2002).

Given the interaction between the effects of frequency, vocabulary knowledge, and lexical access described in Chapter 2, examining the effects of vocabulary knowledge on syntactic processing would shed more light on the extent of Hopp's lexical bottleneck hypothesis. For these reasons, the present study has the same item design as Hopp's (2016), using the same low- and high-frequency setup for the items. However, unlike Hopp's (2016) study, it makes use of an elicited oral imitation task to measure syntactic processing. In addition, the experimental design includes I-Lex to measure the effects of productive vocabulary knowledge. To evaluate whether the lexical bottleneck hypothesis can be extended to another language than English, the experiment is carried out with speakers of L2 Italian.

The I-Lex task is well suited to measure vocabulary knowledge and its relationship with frequency. Adding it as a further measurement can certainly help describe effects of lexical access with more precision. In addition, I-Lex is a productive task that encompasses the twofold process of retrieving and producing words based on a stimulus list. Coupling it to a sentence processing task where a certain structure is produced based on a stimulus list can make the relation between lexical access and syntactic processing more accurately defined.

6.5.1 *Research Questions*

The study shares Hopp's (2016) assumption that "higher demands in L2 lexical processing can lead to differences between native and non-native syntactic processing" (p. 18). What Hopp (2016) defines as higher demand can be straightforwardly translated into reduced availability of lexical features. The results of the I-Lex study and the feature-based lexical network model proposed in Chapter 5 offer a more precise framework, in which frequency interacts with the depth and width of lexical representations. In addition, the HPSG approach lends itself to shape frequency effects on sentence processing more rigorously, defining which features are harder to access when lexical processing is under 'high demand'.

Research question (1): To what extent will the target syntactic structure and the verb frequency interact in the L1 and the L2 group?

In line with the research outlined so far, object clefts are expected to be more difficult to process than subject clefts. As explained in Chapter 4 (see Section 4.11.1) the processing system must link non-adjacent lexical items in order to interpret the meaning of the object relative clause. This adds a cognitive cost, which is reflected by the process of accessing the additional information encoded in the feature GAP. On the contrary, in the case of subject relative clauses, the relative verb is available straight after the relative pronoun, encoded in the feature SPR like any other segment containing a subject followed by a verb. This is in line with the results in Hopp (2016) and Tily et al. (2010), where both the L1 and the L2 groups showed a disadvantage in interpreting object clefts. Additionally, low-frequency verbs are going to further reduce the number of features available to the processing system to form a cleft sentence. This is expected to result in an even more marked difference between subject- and object-clefts with low-frequency verbs. In contrast, high frequency does not affect the availability of lexical features to the same extent, so that the difference between subject and object clefts should be smaller. Note that in models that posit a first processing stage based on syntax, frequency effects are not expected to interact with syntax. On the contrary, based on results of Hopp (2016) and the notion of feature availability presented in Chapter 5 (see Sections 5.5.3 and 5.6), frequency is expected to play a prominent role in preventing the processing system to access syntactic features. Moreover, in line with Hopp (2016), L1 speakers are expected to show very small effects in low frequency cleft sentences, for their vocabulary knowledge is wider and their lexical access less prone to be affected by frequency. On the contrary, L2 speakers should show prominent effects when processing cleft sentences with low-frequency verbs, since the quality of lexical representations of L2 speakers is more apparent with low-frequency words (Brysbaert et al., 2018).

Research question (2): to what extent is there a relationship between productive vocabulary knowledge, measured through the I-Lex task, and syntactic processing?

The notion of feature availability predicts higher vocabulary knowledge naturally entails more features available at the level of meaning, word form, and syntax. Under the HPSG lexicalist approach, all kinds of feature are encoded in each lexical entry, so that low frequency is expected to reduce the availability of word-form, semantic, and syntactic features at the same time. Furthermore, as previously shown in Chapter 2 and Chapter 5, frequency effects and vocabulary knowledge interact in lexical processing (e.g., Brysbaert et al., 2016; Diependaele et al., 2013). As a result, the predictions are vocabulary size is expected to correlate to the ability of processing the target syntactic structure. In fact, L1 and L2 speakers with a high vocabulary knowledge should be able to access more lexical features and be able to use them in their syntactic processing.

6.5.2 *Design of the Study*

To address these research questions, two tasks and a background questionnaire were administered to participants (see Appendix 2). The first task aims at measuring productive vocabulary knowledge, while the second task aims at measuring the accuracy in sentence processing.

The vocabulary task is I-Lex and the version used here is the same used in the validation study in Chapter 3. As outlined in Chapter 3, I-Lex is a word association task, with 30 stimulus words to which participants can associate up to four words. Every word that is not among the first 1000 most frequent Italian words is scored one point. I-Lex captures vocabulary knowledge of both L1 and L2 speakers by measuring their ability to access and retrieve low frequency words; it accounts for the difference between L1 and L2 lexical networks in terms of lexical diversity; it has shown to be well suited to capture the relation between lexical knowledge and the level of proficiency.

The second task is an Elicited Oral Imitation Task (Borro & Luoni, 2015; Erlam, 2006; Jessop et al., 2007; McManus & Liu, 2020; Tracy-Ventura et al., 2014; Wu & Ortega, 2013). An Elicited Oral Imitation Test (henceforth EI) requires participants to listen to a stimulus and to repeat it as closest to the original as possible: the underpinning rationale is that it is possible for speakers to repeat accurately only the utterances they have processed and comprehended effectively (Erlam, 2006; Tracy-Ventura et al., 2014).

The EI task has been successfully used as a tool to measure the overall oral proficiency of L2 speakers but also to distinguish among level of proficiency with fine-grained precision (e.g. McManus & Liu, 2020; Wu & Ortega, 2013). The EI task appears also to correlate with other measures of language proficiency (Erlam, 2006; Tracy-Ventura et al., 2014). EI is well suited to differentiate between implicit knowledge resulting from exposure to communicative language and explicit knowledge resulting from learners focusing on formal practice and explicit instructions (Borro & Luoni, 2015; Erlam, 2006). Erlam (2006) found a high correlation between EI scores and the speaking and listening score on the IELTS test, but a low correlation with writing and reading. Erlam (2006) reasoned that the higher correlations with the IELTS's tasks wherein learners must process language in real time "is suggestive evidence that the elicited imitation test presented in this study may be accessing implicit language knowledge." (p.488). The ability of the EI task to tap into implicit knowledge and automatic processing was also found by Borro & Luoni (2015) with L2 speakers of Italian. They compared the EI scores with two tasks that were not time-constrained, a grammatical judgement task and a metalinguistic knowledge task. In the grammatical judgement task, participants have to find whether a sentence is correct or incorrect, while in the metalinguistic knowledge test, participants had to correct wrong sentences and verbalise the error. Borro & Luoni (2015), arguing that both tasks tapped into explicit knowledge, found that their scores did not correlate with the EI scores. They explained the results as evidence that the EI task can capture implicit and automatic knowledge and is not based on metalinguistic knowledge.

The EI task is expected to tap into a type of knowledge similar to that used by Hopp's (2016) participants when engaging in the self-paced reading task. In fact, the use of implicit knowledge, automatic processing, and the ability to rely on working memory are key features of both self-paced reading (Keating & Jegerski, 2015; Marsden et al., 2018) and the EI task (Erlam, 2006). The EI task also taps into both stages of lexical processing, recognition and recall (Schmitt, 2010), adding a more thorough measure of lexical access. In Hopp's (2016) study, only the recall stage was used to assess the effectiveness of L2 lexical access. In addition, the EI task also relies on priming effects in the production stage. Priming is determined by the aural input and its effects are reflected on the accuracy that participants show in repeating the stimulus sentence. In this respect priming has resulted reliable in measuring syntactic processing in L1 and L2 speakers (Branigan & Pickering, 2017; Hartsuiker et al., 2004).

6.5.3 *Materials*

The EI task consisted of 24 fillers and 16 subject and object cleft sentences that were the target items (see Appendix 4). For the data collection process, I had to rely on an opportunistic sample wherein students could be invited but not compelled to take part into the study. For this reason, I evaluated the effects of having very few participants and how to counterbalance it with the design of the items. I first considered a Latin square design based on the type of items in the study, where each of them had four conditions. A problem with such design was that, if participants turned out to be fewer than predicted, there would be the possibility of having a fairly high number of conditions that cannot be paired together. On the other hand, opting for a conservative approach with a Latin square design containing only a few items, could lead to having some conditions repeated more than others. Hence, I decided to prepare a number of items, reasonably small enough to avoid having unpaired conditions but not too small to have effects caused by specific items. Importantly, as I will explain when analysing the limitations of the study (see Section 6.5.11), I could control for the small number of items by adding them into the statistical analysis as random effects.

Cleft sentences were almost entirely translated into Italian from Hopp's (2016) items; unlike the original items, I added an adjunct at the beginning of each cleft, to avoid that participants could familiarise with the type of target items. In order to replicate Hopp's

(2016) design, the items were arranged in the same 2 x 2 design. The two critical factors were the frequency of the cleft verb and the type of syntactic structure. Like Hopp (2016), this 2 x 2 design gave rise to quadruplets such as the following:

6.11	Al lavoro	era	Amanda	che spaventava	Filippo	con i richiami
	At work	was	Amanda	who scared	Filippo	with the rebukes

At work it was Amanda who scared Filippo with the rebukes

6.12	Al lavoro	era	Amanda	che intimoriva	Filippo	con i richiami
	At work	was	Amanda	who frightened	Filippo	with the rebukes

At work it was Amanda who frightened Filippo with the rebukes

6.13	Al lavoro	era	Amanda	che Filippo	spaventava	con i richiami
	At work	was	Amanda	who Filippo	scared	with the rebukes

At work it was Amanda who Filippo scared with the rebukes

6.14	Al lavoro	era	Amanda	che Filippo	intimoriva	con i richiami
	At work	was	Amanda	who Filippo	frightened	with the rebukes

At work it was Amanda who Filippo frightened with the rebukes

All target sentences started with an adjunct phrase (e.g., *al lavoro*) followed by the focused segment (e.g., *era Amanda*), the relative subject or object clause (e.g., *che spaventava Filippo/che Filippo spaventava*) and the modifier of the relative clause verb (e.g., *con i richiami*). Every sentence exemplifies one condition. Sentence (11) is a subject cleft with a high frequency verb, sentence (12) is a subject cleft with a low frequency verb, sentence (13) is an object cleft with a high frequency verb, and sentence (14) is a subject cleft with a low frequency verb. To control for potential effects of memory load, half of the stimulus sentences were shorter, containing, on average 50.5 letters, while the other half had, on average, 73 letters.

High-frequency verbs were adopted from Hopp (2016). The Italian translation equivalents of Hopp's items were checked against the NVDB list (Chiari & Mauro, 2010, 2015). The items whose Italian translation equivalents were in the first 2000 most frequent words of the NVDB were selected as high-frequency items. The assumption was that, given that the NVDB is based on both frequency and coverage (Chiari & Mauro, 2010, 2015), L2 speakers with a high proficiency level would be familiar with the high frequency items. The low-frequency verbs were not in any of the three bands of the NVDB list, meaning that their frequency was lower than the first 5000 most frequent Italian words. The reasoning was that the EI starts with a comprehension stage, where both phonological and meaning features have a key role in interpreting the low-frequency word. The concern was that the gap between the first and the second frequency bands of the NVDB (first 2000 vs. second 2500) was too small to capture any difference in L1 speakers. For this reason, I decided to use low frequency verbs beyond the first 5000 most frequent Italian words. I relied on the NVDB to have the same frequency benchmark for both the EI task and the I-Lex task which is based on the NVDB list. However, to make sure that the frequency patterns of the items were reliable, both high and low frequency verbs were also checked against the itTenTen corpus (Jakubícek et al., 2010). High frequency verbs had a mean frequency of 59.29 per million while that of low frequency verbs was 3.27. The high and low frequency verbs that were used are shown in Table 6.1.

Table 6.1

List of the Verb used in the Cleft Sentences

High Frequency Verbs		Low Frequency Verbs	
Target item	Freq. per million	Target item	Freq. per million
impressionare (impress)	5.76	sconcertare (astonish)	1.85
spingere (force)	70.89	istigare (compel)	1.09
spaventare (scare)	14.9	intimorire (frighten)	2.19
pagare (pay)	146.42	liquidare (refund)	7.96
Average frequency	59.49	Average frequency	3.27

As Table 6.1 shows, when the NVDB is compared with the itTenTen corpus, some unevenness appears within the high- and low-frequency lists. This is caused by the different criteria used to construct the itTenTen corpus and the Italian corpus on which the NVDB list is based. However, the frequency pattern is preserved in the itTenTen as well, with all the low-frequency verbs showing smaller figures than high-frequency verbs. Furthermore, the use of the NVDB is based on the need of having a balanced reference corpus to match that used in Hopp's and Tily's studies, namely the SUBTLEX-UK (Van Heuven, Mandera, Keuleers, & Brysbaert, 2014). Although the SUBTLEX-UK is based on the subtitles of British television programmes and the NVDB is based on a wider collection of texts, both corpora went under a thorough norming process. Hence, I relied on the NVDB only, using the itTenTen corpus to double check that the frequency pattern high against low was preserved in a different corpus.

To prepare the EI stimulus words, the test items were recorded with an mp3 SONY recorder and then imported into a MacBook laptop using the software GarageBand. The whole track was filtered to cut off interference noise from the background and the format was improved from mp3 to ACC. A key point in designing the I-Lex is to control for the time between hearing the stimulus sentences and repeating it. Indeed, previous research has singled out the risk that EI might only involve the mere processing of the phonetic form (Erlam, 2006; Tracy-Ventura et al., 2014). To avoid the effect Borro & Luoni (2015) and Erlam (2006) added a grid to be completed straight after hearing the sentence. The grid asked participants to agree with the content of the stimulus sentences. The time spent on the grid after hearing the stimulus sentence, ensures that participants focus solely on the meaning, thus avoiding parroting. However, based on Tracy-Ventura et al. (2014), in the EI task used in this experiment, I inserted, after the stimulus sentences, a 3 seconds lag followed by a 100 ms beeping sound to signal the start of the repetition. I opted for a 3 seconds lag that was the equivalent of the time it took participants to complete a grid in Borro & Luoni's (2015) study. Like the grid, the lag also aimed to prevent participants from parroting the stimulus sentences.

6.5.4 *Participants*

A subset of the I-Lex study in Chapter 3 participant took part in the present experiment. Therefore, the L2 group (N = 21) comes from the 82 advanced L2 group of the I-Lex study. The same factors noted in Chapter 3 about the group are briefly summarised below:

- Participants have at least a B2 level of proficiency established through language certification.
- Participants, at the time of the experiment, were attending a C1 level Italian language course for 4 hours every morning, from Monday to Friday.
- Participants, at the time of the experiment, were attending, every afternoon, from Monday to Friday, a two-hour conference on subjects related to Italian culture.

Likewise, the L1 speakers (N=12) participants are part of the same L1 speakers' group that took part into the I-Lex study. Given the time constraints of my research project, I did not manage to recruit the L1 participants in Italy at the same time as I was collecting the L2 ones. I managed to start the recruitment of L1 speakers after the two studies on the L2 speakers were completed, in the final stage of my research. I collected L1 participants amongst the Italian students that were enrolled in Swansea University. The recruitment process turned out to be difficult for I needed to match the L1 education and age profiles to that of the L2 group, to have a fair comparison (see Table 6.2). For these reasons, I was not able to obtain a large group of L1 speakers. However, as I will explain when analysing the limitations of the study (see Section 6.5.11), adopting a mixed-effects statistical model could partially control for the small number of participants. Both groups completed a consent form and a questionnaire with questions about their age, their education and their L2s (see Appendix 2). Data collected from the questionnaire are provided in Table 6.2 below:

Table 6.2

Data of L1 and L2 Speakers as Collected in the Questionnaire

	Age	**Gender**	**Education**	**Proficien**	**First language**
L1 (N=12)	*M* = 28.2 *SD* = 9.4	F = 3 M = 9	BA = 7 PhD = 5	NATIVE	Italian (12)
L2 (N=21)	*M* = 28.2 *SD* = 5.1	F = 18 M = 3	BA = 14 MA = 7	B2-C1	Chinese (3), Portuguese (3), Russian (3), Spanish (3), English (2), Polish (2), Finnish, Georgian, German, Dutch and Serbian (1)

6.5.5 *Procedure*

Before completing the EI and the I-Lex, all participants received the same consent form, the same brief about the purpose and the method of the study, and the same questionnaire (Appendix 2, see Appendix 1 for Ethics Assessment Status). They were invited to read them carefully, fill all the parts and ask questions if anything would not be clear.

The I-Lex task was carried out in one session by the entire L2 speaker group, in pencil and paper form, and it lasted approximately 15 minutes. As mentioned in Chapter 3 the L1 speakers completed the task directly on a spreadsheet in one session of the same duration as L2 speakers.

The EI was carried out by L2 speakers individually, in four consecutive days after they had attended the afternoon conferences. Each participant received a printed note showing the time the task would take place. The procedure followed for the EI was the same in both groups. In the EI task, participants sat in front of the experimenter in a silent room, one at the time. A Macbook Pro laptop was in front of the experimenter and a small Sony mp3 recorder was in front of each participant. Participants were told that the EI task was a general proficiency test on various grammatical structures and, at no time, they were made aware of the structure under scrutiny. Participants first read instructions about the task written in English by the experimenter, who also explained the procedure orally to make sure everything was clear. Participants were given six practice items and invited to ask any remaining questions before the experiment started. The audio file contained all the items followed by the three-second lag and the beeping sound. Since the participants' responses were not recorded on the same device through which they were played, the experimenter would stop the audio file at the end of every stimulus and play a new one after the sentence was repeated. All participants wore headphones to minimize outside noise. After listening to the stimulus sentences and hearing the beeping signal, participants repeated the sentence as close as possible to the stimulus, in terms of meaning and in terms of form. Each EI session took approximately 20 minutes. In both the L1 and L2 groups, each participant performed the EI before completing the I-Lex. At the end of every session each file was imported into a Mac Pro laptop, and the format was changed from mp3 to ACC to improve the quality of the audio.

6.5.6 *Scoring of the Task*

In preparation, all the items produced by each participant were transcribed using the CHAT conventions (MacWhinney, 2000). As the experimental stimulus were prerecorded and did not change over the task, they were not part of the transcriptions so that each file contained only the sentences produced by participants. As for cleft sentences, I arranged a four-level frame to keep track of every syntactic variation. Table 6.3 shows the details of the scoring criteria.

Table 6.3

Criteria for the EI's Scores

Code	Score	**Criterion**
a	**1**	cleft correctly repeated
b	**0**	cleft not repeated
c	**0**	cleft repeated incorrectly
d	**1**	similar focused constituent with different syntactic structure: *c'è X che (there's X who), è stato X che (it was X who)*

Note. The English translation of Italian construction *è stato X che (it was X who)* seems to be the same as the usual cleft sentences. This is because the Italian verb *stay* used here and the verb *be* can be interchangeable in some contexts.

As shown in Table 6.3 to score 1 point, both the cleft structure and the overall meaning of each cleft sentence had to be correctly repeated. For a cleft structure to be considered correctly repeated, it needed to have the focused segment and the relative clause segment (i.e., *it was Filippo who pushed Amanda/ it was Filippo who Amanda pushed*) equal to the stimulus. When the verb or the NP after the relative pronoun were missed, the cleft structure was considered incorrect. This occurred in the L2 group, where some participants could repeat the cleft structure until the relative clause but could not recall the verb or noun and halted, as in Example **6.15** produced by participant 11.

6.15 Al lavoro era Filippo che qualcosa.
At work was Filippo who something
At work it was Filippo who something

The ability to repeat the focused segment, but not the following verbs or nouns, means that the structural link between the modified NP and the relative clause was not established. For the first and last adjuncts, synonyms that would fit the meaning of the sentence were considered correct. Finally, two versions of the target cleft were also accepted, as they were two specific variants of the syntactic structure that, in Italian, can be used as equivalents (Table 6.3, point d). The decision is based on the fact that, in both constructions, the pragmatic and syntactic features FOCUS, and GAP are in the same structural slots. There were such five cases out of the full set of items, all in the L2 group.

The I-Lex task participants received one point for every infrequent word out of a total of 120 available responses. The procedure carried out to tallying and reporting the results of the I-Lex is the one shown in Chapter 3. In the present experiment the analysis was conducted based on the I-Lex1000 frequency list.

6.5.7 ***Results***

6.5.8 ***Statistical Model for the EI***

The present study investigates frequency, a factor that is related to individuals' lexical representations (Ellis, 2011). Frequency effects are caused by individual differences in vocabulary knowledge, whose lexical representations are less entrenched (Brysbaert et al., 2017; Diependaele et al., 2013). Frequency itself, as the I-Lex results in Chapter 3 showed, underpins the organization of lexical networks of individuals (De Deyne & Storms, 2008a; Steyvers & Tenenbaum, 2005). To estimate the fact that frequency reflects individual differences in lexical representations, mixed-effects models seem a suitable type of analysis, as they can incorporate in the computation the random effects generated by participants (Baayen et al., 2008). Furthermore, including also random effects for items in a single analysis can indeed provide a suitable level of understanding that would otherwise be undermined by averaging over subjects or items (Cunnings, 2012).

The EI scores are categorical, for they fall into ether the correct and incorrect outcome, so that the distribution of the scores was binomial. As a result, the analysis on the data was carried out using a generalized linear mixed effects model with the lmer package available on R, version 3.6.3 (R Core Team 2013; Baayen et al., 2008). The statistical model set up for the analysis had reaction times, cleft type (subject vs. object), frequency (low vs. high) and language (L1 vs. L2) as fixed effects and subject and item intercepts as random effects. Although holding only random intercepts for both subjects and items intercepts, might increase the risk of having a Type I error (Barr et al., 2013), adding random slopes in the model caused a large eigenvalue ratio, showing low correlation between predictors (A Field et al., 2012). Assessing the importance of each predictor is hardly reliable with a large eigenvalue ratio, so random slopes could not be included in the final model (A Field et al., 2012)

6.5.9 *Analysis of the EI's results*

The descriptive results for the EI task are given in Table 6.4. Overall L1 speakers were more accurate than L2 speakers in repeating both the target cleft sentences and the fillers.

Looking at the L2 group, overall, subject cleft sentences produced more correct repetitions. The figures show a marked effect of frequency: the number of correct cleft sentences with low-frequency verbs is extremely low. There is a striking difference between subject cleft with high frequency verbs and object clefts with low frequency verb.

In contrast, the L1 speakers produced a remarkably high number of correct cleft sentences. There is almost no difference between clefts with high and low frequency verbs, but there seems to be a slight advantage for subject clefts. Overall, the gap between L1 and L2 speakers is more pronounced in cleft sentences with low frequency verbs, suggesting a strong effect of frequency.

Table 6.4

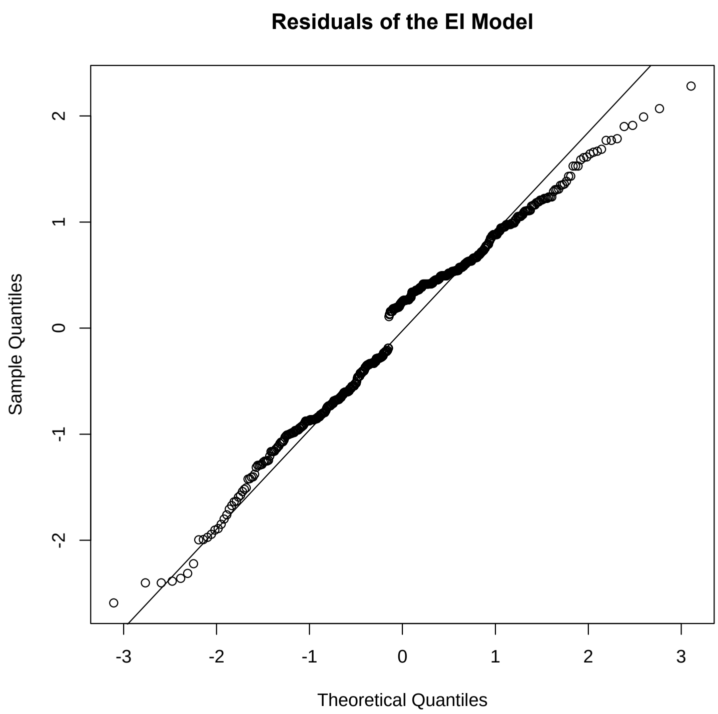
Clefts Correctly Repeated by Type and Frequency

	L1 Speakers		L2 Speakers	
	M(*SD*)		*M*(*SD*)	
Subject Clefts (out of 8)	7.1	0.7	3.6	2.1
Object Clefts (out of 8)	6.1	2.5	2.7	2.1
High Freq Subj (out of 4)	3.7	0.4	2.5	1.1
High Freq Obj (out of 4)	3.1	1.2	2.0	1.4
Low Freq Subj (out of 4)	3.4	0.6	1.1	0.7
Low Freq Obj (out of 4)	3.0	1.3	0.7	1.1

Before carrying out the statistical analysis, the EI data were checked for outliers in both the L1 and the L2 groups. No outliers were found that were greater than +/- 2SDs from the mean. Importantly, the distribution of the model residuals was first examined: mixed-effects models are required to show a residual distribution as close as normal as possible (but see Schielzeth et al., 2020). Figure 6.1 shows the quantile-quantile plot of the residuals of the overall model.

Figure 6.1

Quantile-quantile Plot of the EI's Model Residuals



The Q-Q plot compares the values of normal distribution against the real values of the model and calculate a regression line. The more the observations align with the line, the closer their distribution is to the normal one (A Field et al., 2012; Levshina, 2015). Since the EI data set has a huge number of observations, quantile-quantile plot representations are more reliable than tests that measure the normality of distribution, like the Shapiro-Wilk, which are heavily influenced by sample size (Levshina, 2015).

The interpretation of the residuals of the model in Figure 6.1 is based on Schielzeth et al. (2020), who have recently found that mixed-effects models are quite robust when the normality of distribution is violated on the variance of residuals. In their study, they found that, although the estimates of the model lost precision, they remain unbiased. They concluded that "our results should be viewed as encouraging and allow users of mixed-effects models to proceed with confidence. We conclude that mixed-effects models are largely robust even to quite severe violations of model assumptions (Schielzeth et al., 2020, p. 1150)". Figure 6.1 shows that there are few points at both tails that deviate from the regression line, but the residuals have a distribution fairly close to normal. Based on Schielzeth et al. (2020) the statistical model was then applied to the results of the study.

Research question (1): To what extent will the target syntactic structure and the verb frequency interact in the L1 and the L2 group?

The analysis of the EI task to answer to the first research question, examines first the overall model to investigate the effects of language. Then, using two different statistical models, the analysis examines how the target syntactic structure interacts in the L1 and L2 groups. Before describing the analysis, it is important to mention that in all the models, when sentence length was introduced, it was neither significant nor significantly interacted with any other predictor. Therefore, it was excluded from all subsequent analyses.

The analysis carried out with the general model did not show a significant effect of frequency (ß = - 0.31, *SE* = 0.64, *t* = - 0.49, p = 0.62), while the effects of cleft type (ß = 1.72, *SE* = 0.83, *t* = 2.07, *p* = 0.03) and language were significant (ß = - 1.78, *SE* = 0.71, *t* = - 2.49, *p* = 0.01). There was a significant interaction between language and frequency in that L2 speakers (*M* = 1.85, *SD* = 1.85) produced fewer correct cleft sentences than L1 speakers (*M* =

6.41, *SD* = 1.29) with low frequency verbs (ß = -1.71, *SE* = 0.71, *t* = - 2.40, *p* = 0.01). A Tukey post hoc test (Table 6.5) revealed that L2 speakers produce fewer correct subject clefts with high frequency verbs (ß = 2.66, *SE* = 0.88, *z* = 2.99, *p* = 0.05), fewer subject clefts with low frequency verbs (ß = 3.76, *SE* = 0.77, *z* = 4.83, *p* < 0.0001) and fewer object clefts with low frequency verbs (ß = 3.49, *SE* = 0.73, *z* = 4.72, *p* < 0.0001). In contrast, L1 speakers did not differ from L2 speakers with object clefts with high-frequency verbs (ß = 1.78, *SE* = 0.71, *z* = 2.43, *p* = 0.11).

Table 6.5

Post-hoc Analysis for the Overall Model

	L2 High-Freq Subj	L2 High-Freq Obj	L2 Low-Freq Subj	L2 Low-Freq Obj
L1 High-Freq Subj	0.05	0.005	< .0001	< .0001
L1 High-Freq Obj	0.9	0.19	.001	.0001
L1 Low-Freq Subj	0.58	0.07	< .0001	< .0001
L1 Low-Freq Obj	0.93	0.54	.006	.0001

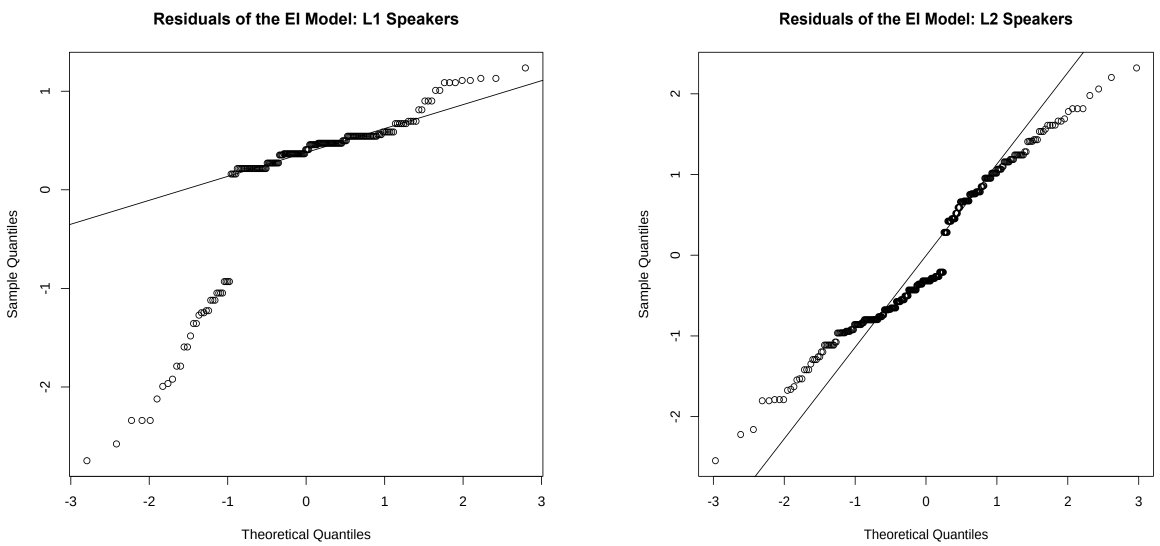
Note. Only p-values are given

The results did not show other significant interaction: unlike the Hopp (2016) study, the type and frequency of cleft sentences did not interact (ß = -0.83, *SE* = 1.06, *t* = -0.77, *p* = 0.43). There was no significant interaction between language and cleft type (ß = -0.87, *SE* = 0.85, *t* = -1.02, *p* = 0.3) nor the three-way interaction between language, frequency and cleft type (ß = 0.60, *SE* = 0.12, *t* = 0.53, *p* = 0.59).

Given the significant interaction between language and frequency, a separate analysis was carried out in both groups to find out if additional effects were present within each group. When the data were split by groups, the new model had subject and item intercepts as random effects and structure (subject cleft vs. object cleft) and frequency (high vs. low) as fixed effects. The distribution of the residuals in both models was not normal, as shown in the quantile-quantile plots of Figure 6.2. Despite the results of Schielzeth et al. (2020), the deviation from the normal distribution was too wide to ensure a reliable analysis. As a result, the analysis by group was carried out using the Kruskal-Wallis test and a post hoc Dunn test with the FSA Package from the software R (R Core Team 2013). Post hoc analysis was performed using the Dunn post hoc test with the Benjamini–Hochberg correction. The Benjamini–Hochberg correction has more statistical power than the Bonferroni correction for it is well suited to control for the false discovery rate, that is the number of falsely rejected null hypotheses over the total number of rejected null hypotheses (see Field et al., 2012, pp. 430-431).

Figure 6.2

Quantile-quantile Plots for the Residuals of the EI's L1 and L2 Models



The results for the L1 group are given in Figure 6.3. No main effect of cleft type was found ($H(3) = 2.67$, $p = 0.44$) nor was there a main effect of frequency ($H(1) = 1.17$, $p = 0.27$). The Dunn post hoc analysis on the L1 speakers (Table 6.6) results did not reveal any significant differences between cleft type or any interaction with frequency.

Table 6.6

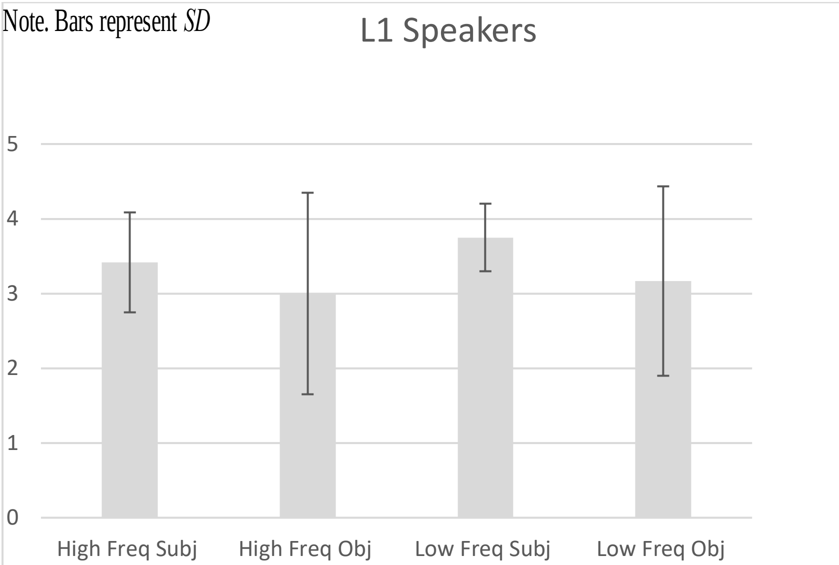
Post-hoc Analysis for the L1 Speakers

	High-Freq Subj	High-Freq Obj	Low-Freq Subj	Low-Freq Obj
High-Freq Subj	–			
High-Freq Obj	1.0	–		
Low-Freq Subj	1.0	1.0	–	
Low-Freq Obj	.73	1.0	1.0	–

Note. Only p-values are given

Figure 6.3

Graph of L1 Speakers Scores in the EI Task

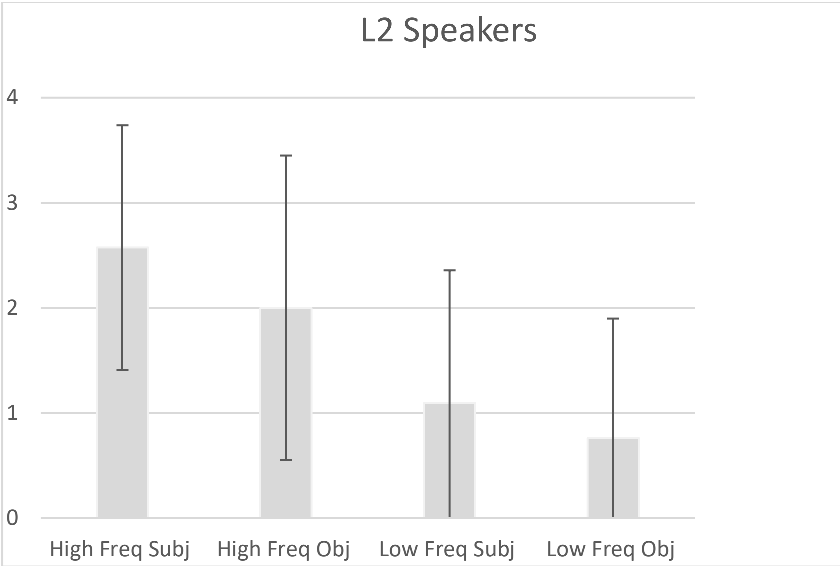


Note. Bars represent *SD*

The results for the L2 group are shown in Figure 6.4. Analysis of the L2 speakers group revealed a main effect of cleft type ($H(3) = 22.043$, $p < 0.005$) along with a main effect of frequency ($H(1) = 19.607$, $p < 0.005$), characterised by a significantly higher number of correct subject clefts and a significantly lower number of clefts with low-frequency verbs.

Figure 6.4

Graph of L2 Speakers Scores in the EI Task



Note. Bars represent *SD*

The Dunn post hoc analysis (Table 6.7), showed that all the significant differences were driven by frequency effects. First, L2 speakers repeated more correct subject clefts with high-frequency verbs than object clefts with low-frequency verbs ($z = 4.20$, $p < 0.005$) and more correct object clefts with high-frequency verbs than subject clefts with low-frequency verbs ($z = 2.05$, $p = 0.05$). The number of correct subject and object clefts with high-frequency verbs was higher than their low-frequency counterparts (high frequency subject vs low-frequency subject: z =3.38. p < 0.005; high frequency object vs low frequency object: z = 2.88, p < 0.05). Although the expected interaction between syntactic structure and type did not surface in the L2 speakers, they showed a more complex pattern than the L1 speakers.

Table 6.7

Post-hoc Analysis for the L2 Speakers

	High-Freq Subj	High-Freq Obj	Low-Freq Subj	Low-Freq Obj
High-Freq Subj	–			
High-Freq Obj	.2	–		
Low-Freq Subj	.002	.05	–	
Low-Freq Obj	.0001	.007	.4	–

Note. Only p-values are given

Overall the group analysis revealed that the expected interaction between frequency and cleft type did not surface. The ability of the L1 speaker to comprehend and correctly repeat cleft sentences was affected by neither frequency nor type. Furthermore, no effects were found in the L1 speakers' groups, contrary to both Hopp's (2016) and Tily et al.'s (2010) studies.

The L2 speakers results show some similarity with Hopp's (2016) study. Their poor results with low-frequency verbs in the EI task partially reflect the results that Hopp (2016) found in the post-verb region where the strongest effects were caused by low-frequency verbs. The interaction that he found between frequency and cleft type in the L2 groups is absent in this study.

Research question (2): to what extent is there a relationship between productive vocabulary knowledge, measured through the I-Lex task, and syntactic processing?

This section answers the second research question. It shows the analysis of the relationship between vocabulary productive knowledge and EI results, by measuring the correlation between I-Lex scores and EI scores. Before showing the results, it is appropriate to explain the choice of correlation analysis over that of entering the vocabulary scores in the main model used for the EI task. The experiment aims to find out whether the number of correct EI items could be predicted from the I-Lex scores. Studies that have looked at the relationship between Lex30 and other linguistics abilities (e.g., González & Priz, 2016; Uchihara & Saito, 2019) have also used correlational analysis. A different approach would be using the I-Lex scores as predictors in a mixed-effect model to evaluate to what extent both vocabulary and EI scores improve the fit of the model. This is common practice in studies that have compared the size of the vocabulary with the reaction times produced by lexical processing tasks (Diependaele et al., 2013; Lemhöfer et al., 2008). However, the scores of the EI task are different since they do not measure reaction times but take a binary (i.e., Boolean) value, so that the patterns of variation captured by the EI task and the I-Lex scores are measured in a distinct way. In fact, the statistical model with cleft type (subject vs. oblique), frequency (low vs. high) and vocabulary scores as fixed effects and subject and item intercepts as random effects did not converge. Convergence was not achieved even when the I-Lex scores were rescaled.

The I-Lex scores are given in Table 6.8. For the correlation analysis, the scores based on the I-Lex1000 frequency list were chosen over the I-Lex2000. The reason is that the I-Lex1000 is the list on which the concurrent validity between I-Lex and Lex30 is based, so it appears to be more reliable and robust.

Table 6.8

I-Lex Scores, Absolute (abl) and Percentage (%)

	I-Lex1000 (abl) *M*(*SD*)	I-Lex1000 (%) *M*(*SD*)
L1 Speakers (N = 12)	75 (8)	66 (7)
L2 Speakers (N = 21)	55 (12)	59 (7)

In addition, absolute and percentage scores were used in the analysis based on the results of Uchihara & Saito (2019) who found that the correlation between the scores of Lex30 and oral ability appeared with absolute scores but not with the percentage ones. The choice of including both scores also hinges on the similarity between the EI and the task used by Uchihara & Saito (2019). They used a timed picture description task, in which L2 speakers had to talk about a picture that had three cue words with a five seconds preparation. In effect, it is an elicited and online oral task that resembles the EI, although it gives far more freedom to participants.

The I-Lex results show the expected pattern with the L1 speakers outscoring the L2s. To see whether the difference was significant, a preliminary Shapiro-Wilk normality test showed that the scores of the two groups were normally distributed (L1: $W = 0.87$, $p = 0.09$; L2: $W = 0.96$, $p = 0.72$). A Welch Two Sample t-test showed that the difference between the L1 and the L2 group was significant ($t = 5.62$, $p < 0.005$). These results are in line with those found in the I-Lex study in Chapter 3, as they show that L2 speakers' vocabulary size is smaller than that of the L1ers.

The analysis of the EI scores shows that, in both groups, there was not a significant difference between cleft type or an interaction between cleft type and frequency. For this reason, a correlational analysis was performed between the I-Lex scores and the EI scores on the total number of clefts with high- or low-frequency verbs. A non-parametric Kendall's rank correlation test was used because of its robustness with low samples, as it was the case of the L1 group (A Field et al., 2012). The correlations for the L1 speakers are shown in Table 6.9 and those of the L2 speakers in Table 6.10. The correlation coefficient for the Kendall's rank correlation test is expressed as *tau.* To better evaluate the strength of the strength of the *tau* coefficient, all significant correlations were converted into Pearson's coefficient *r* applying the formula developed by Walker (2003).

Table 6.9

Correlation between I-Lex and EI scores: L1 Speakers

	High Freq Clefts	Low Freq Clefts
I-Lex scores (abl)	*tau* = - 0.03 (*p* = 0.94)	*tau* = 0.03 (*p* = 0.88)
I-Lex scores (%)	*tau* = - 0.27 (*p* = 0.24)	*tau* = 0.23 (*p* = 0.33)

As can be seen in Table 6.7 and Table 6.8 there was no difference between the absolute and percentage scores, as they both yielded the same pattern of results. Although the L1 speakers did not show any significant correlation, the L2 speakers showed a positive correlation between their I-Lex and EI scores in cleft sentences with low-frequency verbs. Once the *tau* values were converted into the Pearson's *r*, the correlation was high, .68 and .64 for the percentage scores.

Table 6.10

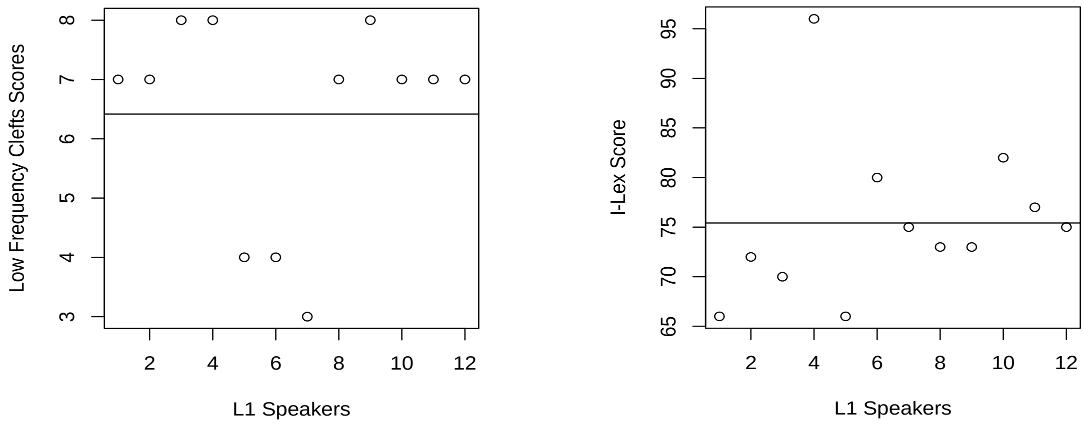
Correlation between I-Lex and EI scores: L2 speakers

	High Freq Clefts	Low Freq Clefts
I-Lex scores (abl)	*tau* = 0.24 (p = 0.14)	*tau* = 0.58 (p < 0.005) (r = 0.68)
I-Lex scores (%)	*tau* = 0.20 (p = 0.19)	*tau* = 0.45 (p < 0.01) (r = 0.64)

These results show that, with L2 speakers, the effects on processing syntactic structure with low frequency verbs can be predicted from L2 vocabulary knowledge. On the contrary, the knowledge of the vocabulary of the L1 speakers does not affect their syntactic processing. The L1 group results may have been caused by the fact that the scores of almost all participants were at ceiling in the EI task. Showing at ceiling scores in EI is not uncommon in L1 speakers: in Erlam (2006), for example, the L1 group showed a remarkable 97% accuracy. For this reason, the variance was too low to be reliably compared with the variance in the I-Lex. This can be seen in Figure 6.5, where the L1 speakers scores are compared in both tasks. The difference between the pattern of the EI scores with low frequency clefts and that of I-Lex scores is clear. Although the I-Lex scores show a varied pattern, the scores in low-frequency clefts are all almost at ceiling.

Figure 6.5

EI and I-Lex scores: L1 Speakers

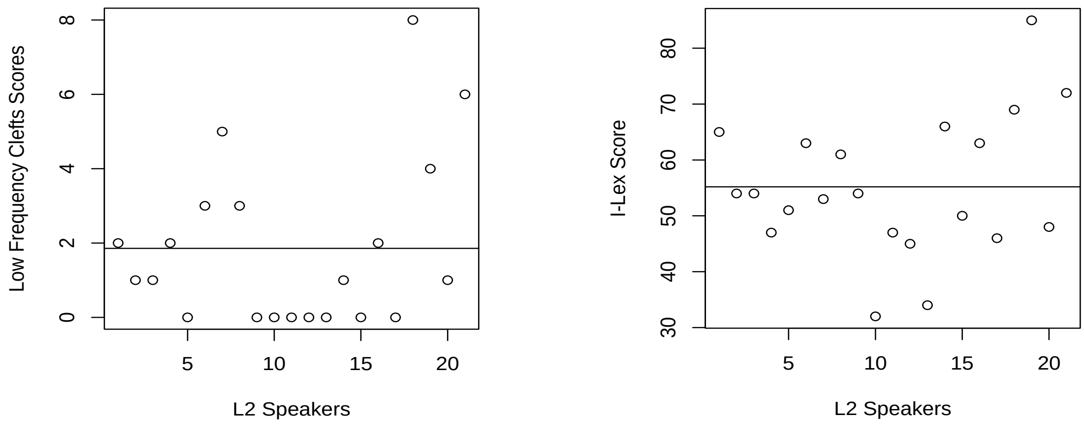


Note. The straight line represents the mean

Overall, the frequency of EI stimuli does not seem to have affected L1 speakers at all in this study, and this could have overshadowed the interaction between vocabulary and sentence processing. On the contrary, in the L2 speakers, the pattern of the EI scores with low frequency clefts and that of the I-Lex scores were similar, revealing a comparable degree of variance, as shown in Figure 6.6.

Figure 6.6

EI and I-Lex scores: L2 Speakers

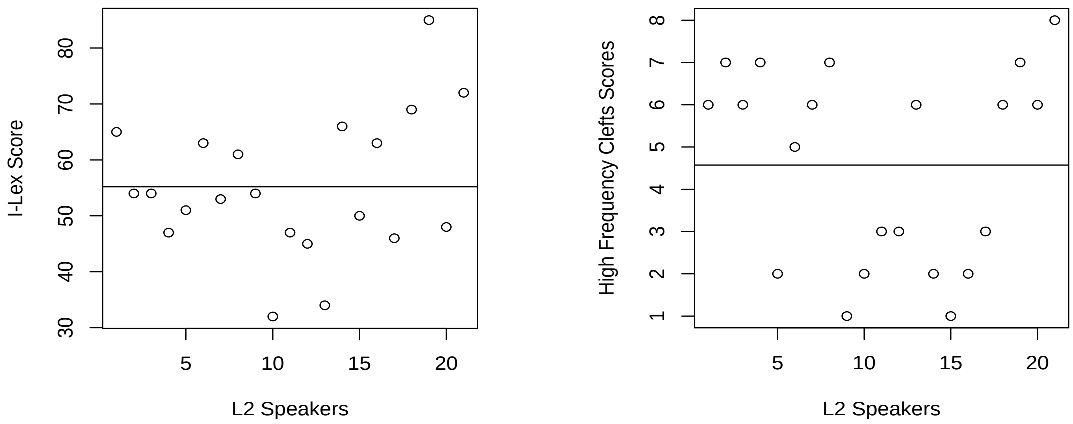


Note. The straight line represents the mean

Since I-Lex gives an estimate of vocabulary knowledge measuring the number of low frequency words, it might seem straightforward that it correlates with the processing of low frequency words. However, L2 speakers also had difficulties with clefts with high-frequency verbs that were not caused by the syntactic structure, so vocabulary knowledge might be a suitable predictor thereof. In fact, the analysis did not find any significant correlation between cleft with high frequency verbs and the I-Lex scores. The lack of significance may be attributed to the distribution of the EI scores in the L2 group. Figure 6.7 shows the scores of I-Lex and those of EI with clefts with high-frequency verbs. The distribution of the EI scores looks almost binomial: the scores of 10 L2 speakers hover around 6 and 7 while 7 of them hover around 2 or 3. The shape of the variance is clearly different form that of the I-Lex scores and this discrepancy might have caused the lack of significant correlation.

Figure 6.7

Scores of the I-Lex and of Clefts with High Frequency Verbs in L2 speakers



Note. The straight line corresponds to the mean.

6.5.10 *Discussion*

The results of the present experiment only partially replicated the effects of frequency on L2 speakers found in Hopp (2016), as the L2 group did not confirm the interaction between frequency and type. However, a hint of interaction was still present in the results of the L2 speakers. The number of correct subject clefts with high-frequency verbs was significantly larger than the number of object clefts with low-frequency verbs. Subject clefts are easier to process, as shown by the Hopp and Tily studies and the results of the EI task. Thus, it might be that, when combined with high frequency words, subject clefts turned out to be easier for the L2 group to comprehend and repeat. In contrast, the more complex object clefts are more difficult, and low frequency verbs increase the difficulty, causing the most severe disruptions.

Although this result might hint at a partial interaction between cleft type and frequency, results point to a clear frequency-driven pattern. For L2 speakers, frequency was the main factor in the EI task and all other effects emerged because of frequency. In fact, high frequency was crucial in helping the L2 speakers to the point that the more complex object clefts with a high-frequency verb were produced more correctly than the easier subject cleft with low-frequency verbs. In the L2 group, every choice made by the processing system was first determined by frequency, and then the effects were reflected in the syntactic structure. Support to this interpretation comes from the result that the number of correct clefts with high-frequency verb was significantly larger than the number of clefts with low-frequency verbs, regardless of the syntactic structure.

One reason for the difference between the results of the L2 group found in this study and those of Hopp (2016) and Tily et al. (2010), might have been caused by the difference between the type of task. Of course, the Self-paced Paced Reading reflect finer-grained measurements, where the processing stages are measured word by word. On the contrary, the EI does not achieve the same degree of precision. However, the different results might also be linked to the characteristics of the EI task in a different way. In the case of Hopp (2016), to take on the SPR task, the L2 group needed only a comprehension stage. In the EI task, L2 speakers needed first to comprehend the stimulus sentence and then to access the correct features to reactivate the words. In terms of lexical access, the difference between the two tasks mirrors the difference between recalling and recognising a word (Schmitt, 2010). In a SPR task lexical activation is chiefly determined by word-from features and the right form-meaning connection tends to be reached very quickly (Schmitt, 2019). In the EI task, this stage, being based on aural input, it is more difficult (Milton, 2009). Furthermore, it is followed by a recall stage where features must be linked to a specific phonological form. Although the recall stage is primed by the stimulus sentences, it adds cognitive cost to the processing system of L2 speakers. For this reason, a task that needs both recognition and retrieval stages like the EI might be extremely sensitive to frequency effects, thus explaining why L2 speakers showed the largest ones with low frequency words. As noted by Gollan et al. (2011), production tasks, entailing a meaning-based retrieval stage, a structural stage, and an articulatory stage (Levelt & Meyer, 1999) can produce more pronounced frequency effects.

In contrast, it seemed that L1 speakers were completely unaffected by low-frequency verbs in that they could access them as much as they could with low-frequency verbs. The lack of effects is at odds with Hopp (2016) and Tily et al. (2010) studies. Hopp (2016) found that L1 speakers showed an interaction between the type of cleft sentence and low frequency in both the verb and the postverb region. Tily et al. (2010) also found effects of frequency and cleft in both regions. One explanation might be that verbs whose frequency is beyond the first two most frequent bands of the NVDB, might still have been easy to access for L1 speakers. As Brysbaert et al. (2018) noted, people with high vocabulary knowledge are affected by extremely low-frequency words only. This might be the case for L1 speakers in the present experiment. An alternative explanation is that L1 speakers might have taken advantage by the priming that occurred during the comprehension stage of the EI task. Priming and frequency effects interacted, so that listening to the stimulus sentence increased the level of activation of the target low-frequency verbs eliminating the effects of frequency. In this respect, while the EI task captures two processing stages at the same time, it seems to be not fine-grained enough to really distinguish between the two stages. Therefore, it misses crucial information to capture more subtle effects, such as those of L1 speakers.

An additional aim of this study was to evaluate the impact of frequency on feature availability and sentence processing. The prediction was that participants with high vocabulary knowledge can access more lexical features and use them to produce more correct cleft sentences. On the contrary, participants with a low vocabulary knowledge cannot access lexical features as readily and cannot used them fast enough to produce correct cleft sentences. Considering the results of the statistical analysis, only the L2 speakers showed this pattern: their less robust lexical representations made them more sensitive to low frequency verbs, as predicted by the Lexical Entrenchment Hypothesis (Cop et al., 2015; Diependaele et al., 2013). Additionally, in this study, L2 speakers showed that vocabulary knowledge and frequency are a reliable predictor of syntactic processing difficulties, although their effect was additive.

The notion of feature availability assumes that lexical features in low-frequency words are especially hard to access for L2 speakers. The causes are the lower degree of organization in their lexical networks (Meara, 2006, 2009), non-selectivity (Dijkstra & van Heuven, 2002; Kroll et al., 2008) and their higher sensitivity to the low availability of low frequency words. In this sense, the results found in the present study are linked to the I-Lex results described in Chapter 3. In that experiment L2 speakers showed that the poor organization of lexical representation is reflected in the difficulty of providing low-frequency words. That effect is reflected, in the present study, in the largest disadvantage that L2 speakers showed when processing cleft sentences with low-frequency verbs.

The last point to examine is how to frame the results into the feature-based lexical network model described in Chapter 5. The model is based on two straightforward assumptions. The first is that words are described by complex feature structures modelled according to HPSG and that the type of link that is activated between two or more words is determined by the type of features. In the following paragraphs, I will adapt the model to the result of the study, presenting first the feature-based lexical network model for subject and object cleft with high frequency verbs. Given that the present study focused on the relationship between frequency, feature availability and syntactic processing, I setup a model wherein the links within the lexical network are syntactic links.

The lexical networks for the subject and object clefts with high frequency verbs are given in Figure 6.8 and Figure 6.9. The networks describe subject and object clefts given in Examples **6.16** and **6.17**.

6.16 Amanda che spaventava Filippo
Amanda who scared Filippo

6.17 Amanda che Filippo spaventava
Amanda who Filippo scared

Figure 6.8

Feature-based Lexical Networks for a Subject Cleft with High Frequency Verbs

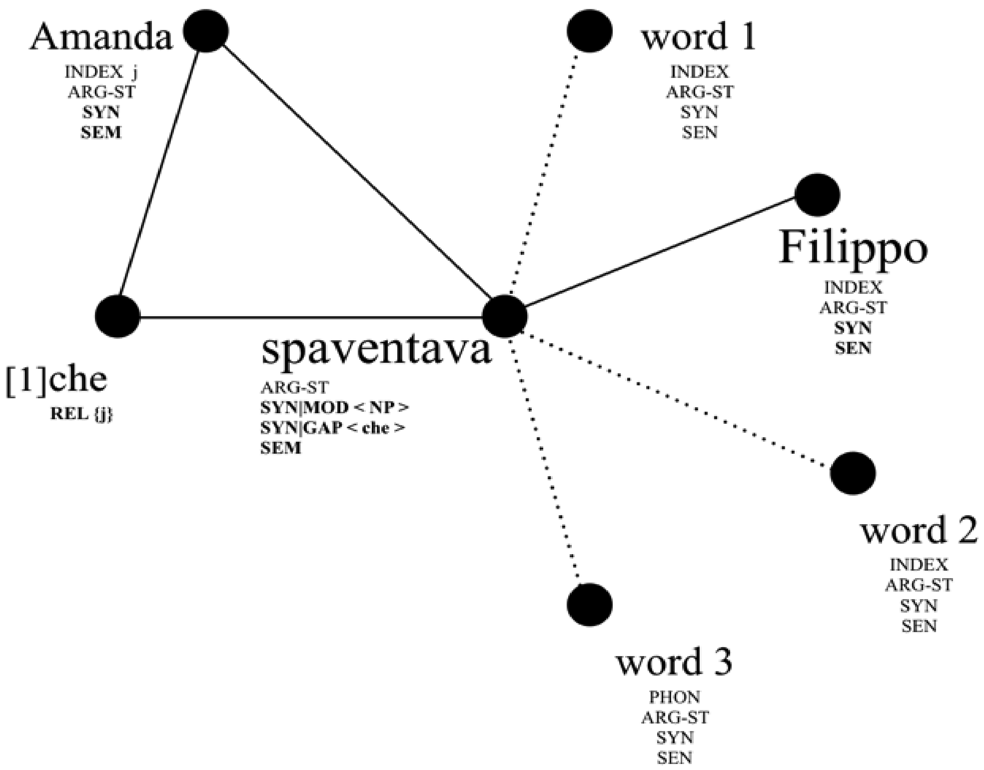
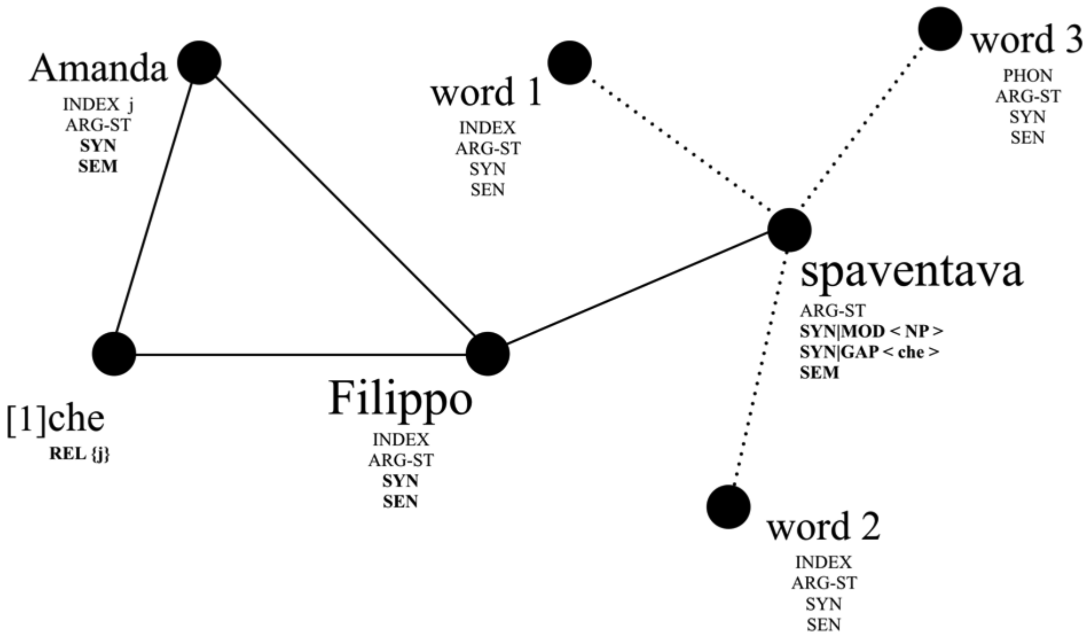


Figure 6.9

Feature-based Lexical Networks for an Object Cleft with High Frequency Verbs



The lexical networks contain only the primary features SYN and SEM to signpost the syntactic and semantic information. As said in Chapter 4, the feature SYN contains all syntactic information, while the feature SEM contains all semantic information. In contrast, the more specific feature GAP and MOD appear on the verb *spaventava* (*scared*) to properly highlight the grammar of the relative clause, along with the feature INDEX on the modified NP *Amanda* and the feature REL encoded in the relative pronoun *che*. The features that activate the links in the relative clause are in bold and the activated syntactic links are described by a straight line, while the links that are not activated are described by a dotted line. Once the correct features become available, one of the many links that connect words receives the amount of activation needed to form a syntactic dependency. An activated syntactic link is expected to be stronger than an existing link due to the amount of activation added by syntactic features. In effect, structural links produce a meaningful segment formed according to the rules of the grammar, which is the strongest type of connection between two words. One aspect that I have introduced in both networks is that activation can potentially occur to all the links that have already been formed and incorporated into the network. The dotted links to *word1*, *word2,* and *word3* describe connections between words in the lexical network that are part of the organization of mental representations. However, as they are not used in the sentence, they are not activated.

One technical aspect of the network that need to be highlighted is that the models in Figure 6.8 and Figure 6.9 show that both syntactic and semantic features are in bold, i.e., they both form links in the network. As I have explained in Chapter 5, when syntactic features determine activation, meaning features are also activated. In fact, HPSG is a constraint-based framework where multiple features are accessed and deployed without going through ordered stages. Therefore, some features are prevalent in activating some links, but the processing system uses the entire set of available features by default. The claim is clear, for example, in the two-way mapping that occurs via the feature ARG-ST between meaning and syntactic features (see Chapter 4, Section 4.4). I will return on this point in the discussion (Chapter 8), showing how to model constraint-based activation more accurately.

The feature based lexical networks for clefts sentences with low frequency verbs are given in Figure 6.10 and Figure 6.11. The networks describe subject and object clefts given in example **6.18** and **6.19**.

6.18 Amanda che intimoriva Filippo
Amanda who frightened Filippo

6.19 Amanda che Filippo intimoriva
Amanda who Filippo frightened

Figure 6.10

Feature-based lexical networks for Subject Cleft with low frequency verbs

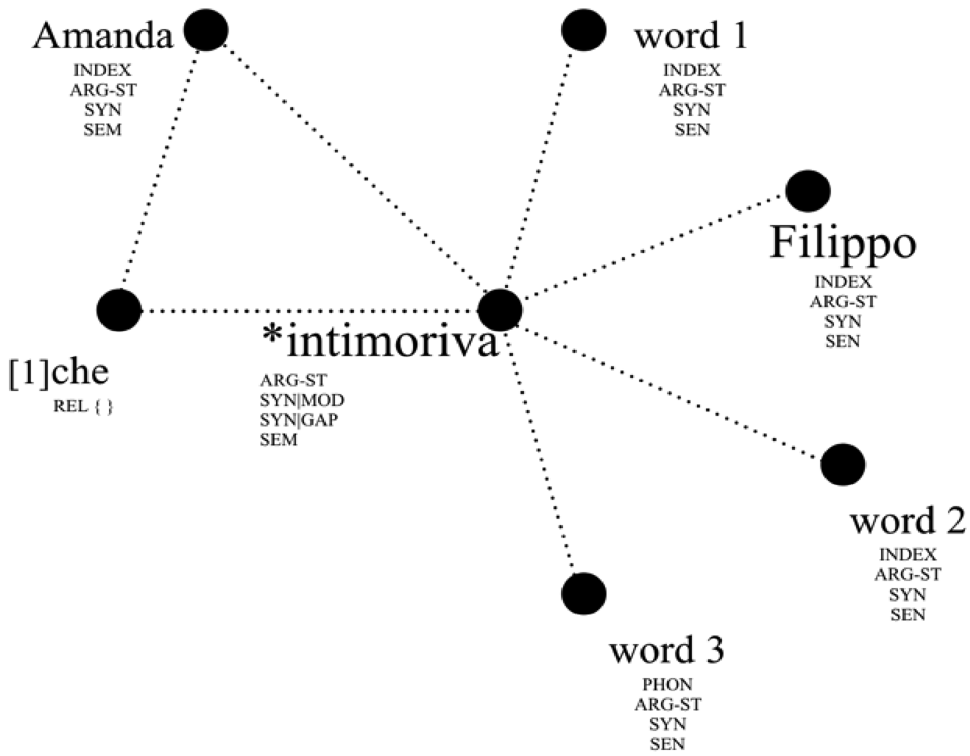
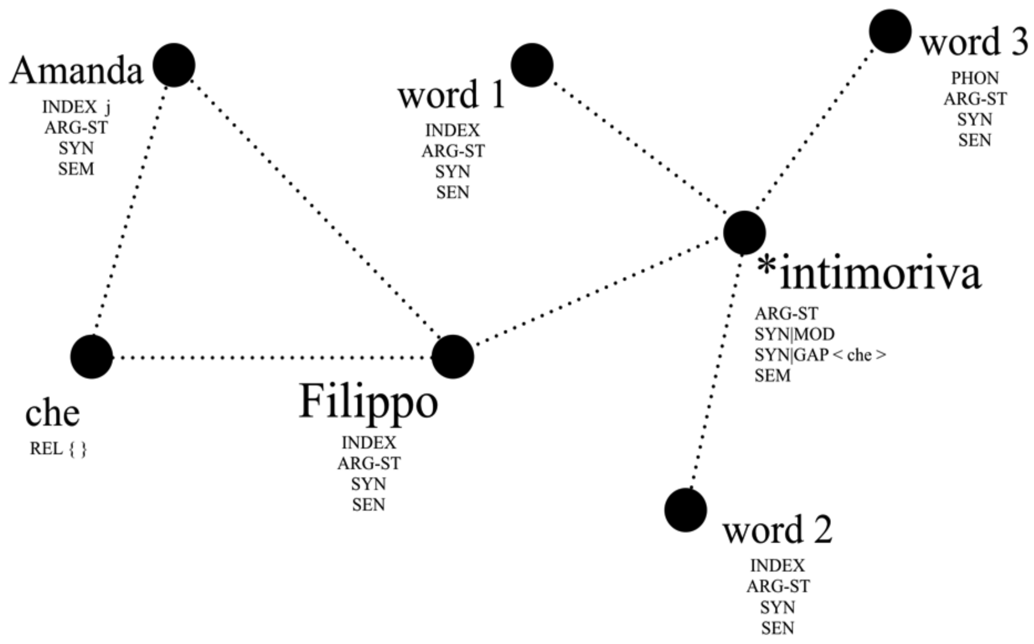


Figure 6.11

Feature-based lexical networks for Object Cleft with low frequency verbs



The two feature-based lexical network models describe clefts with low-frequency verbs that, in his study, were problematic only for L2 speakers. According to the notion of feature availability, with low frequency verbs lexical features were more difficult to access and the stimulus cleft sentences could not be correctly repeated. When listening to low frequency verbs, the processing system could not readily access phonological features (following Mellow (2004) this is shown by the asterisk before the verb). As a result, the other lexical features were also not accessed, so the syntactic links in the network could not be activated. In fact, unlike in the previous two models, here the lexical feature SYN is not in bold, and all the links are described by a dotted line. This reflects that the NP *Filippo* cannot be associated to the verb as its complement, not even via any default process. Finally, the relative pronoun *che* cannot be coindexed with the NP *Amanda* and with the feature GAP. In both subject and object clefts the feature GAP does not encode any information, so that the processing system cannot relate it neither to the feature SPR nor to the feature COMPS.

The feature-based descriptions and the lexical networks that I have proposed for both types of relative clauses show how to unify syntactic and lexical processing in one model. Feature-based lexical representations, under the precise HPSG formalism, account for lexical and syntactic processing. Depending on lexical features a link is formed in the lexical network so that a new word can be added to it based on meaning, syntactic or word-form features, although there is evidence that meaning links are prevalent (De Deyne & Storms, 2008a; Fitzpatrick, 2006, 2007, 2009; Steyvers & Tenenbaum, 2005). Although every time a meaningful and grammatical segment needs to be produced or interpreted syntactic features need to be activated the process does not need to occur in a specific order. The feature-based lexical network is a constraint-based model that does not rely on a first stage based on the use of syntactic features.

The model I am proposing is well suited to explain the processing disruptions experienced by the L2 speakers. First, as suggested by I-Lex scores, their lexical networks have fewer links and are less dense, especially when it comes to links from and to infrequent words. The correlation between I-Lex and EI scores in low-frequency verb cleft sentences supports this assumption. The L2 lexical networks are missing links, and the activation process cannot occur with low-frequency words. In contrast, that L2 were less affected when the EI stimulus sentence had a high frequency verb, means that those verbs were linked to other words in their lexical network. With high-frequency verbs, the difference between object and subject cleft sentences was not significant, indicating that L2 speakers could access the correct syntactic features and activate the correct structural links. However, the significant effects of language show that L2 speakers could not access all the lexical features, even those encoded in high frequency verbs. This part is not modelled in Figure 6.8 and Figure 6.9, but it is easy to assume that some of the features were harder to access for the L2 group. Given the nature of the task phonological features might have caused major problems, followed by the features RESTR and SPR and COMPS. The models I propose here are more focused on syntactic links, since the study aimed at the interaction between frequency and syntactic structure. However, as I have explained in Chapter 5, lexical features show a more complex pattern of interaction. Thus, a more accurate model to account for partial access to lexical features is described in the discussion in Chapter 8.

6.5.11 *Limitations*

The results of the statistical analysis showed an expected effect of frequency and vocabulary knowledge in the L2 speakers' group. The results are in line with the claims that L2 have less entrenched lexical representations and this affects their lexical and syntactic processing (see, e.g., Brysbaert et al., 2017; Diependaele et al., 2013; Hopp, 2016). However, the study is limited in its power, given the small number of participants in the L2 group and especially in the L1. For example, a sufficiently powered experiment with only a simple comparison of two groups would require around 50 participants (Brysbaert, 2019). However, the pattern of results shows an encouraging trend that converges with the hypothesis, supported by several studies, that vocabulary knowledge and frequency are crucial in syntactic processing (David et al., 2009; Hopp, 2016; Kawaguchi, 2016; Rogers et al., 2018; Tily et al., 2010).

Another problem is the small number of items. Although all participants read all target items, there is still the risk that the effects found in the study might be due to lexical properties of individual items rather than a general trend. However, the results of the Cronbach alpha test showed that no items appeared to be too easy or too difficult, suggesting that the effects captured in the experiment are unlikely to be related to specific properties of the target items. Furthermore, the use of mixed-effects models should be able to control for the potential effects caused by the properties of the items. As noted by Baayen et al. (2008) mixed-effects models offer the possibility to weigh how properties of individual items affect the variance in the statistical analysis. In this respect, having set items as random effects in all the models should partially offset the variance caused by idiosyncratic properties of the experimental items (Baayen et al., 2008; Cunnings, 2012). One key factor that adds confidence to the results obtained in the study is the absence of convergence problems. The notion of convergence refers to the possibility of having, in the experiment, observations that do not fit well with the mixed-effects model (Barr et al., 2013). This amounts to say that the effect of the independent variables gives rise to patterns of observations clustered together which cannot be captured by the random effects, namely items and participants. In the EI study, the clustering in the observations was determined by the participants' individual ability to deal with the two independent variables. For example, some participants were more affected by vocabulary knowledge in some of the items, and by frequency in others. Thus, adding subject and items as random intercepts to the statistical model helps estimate the effects of the clustering on the overall model. If the estimates in the model can be clearly defined, the model it is said to converge. As noted by Barr et al. (2009):

> "the likelihood that a model will converge depends on two factors: (1) the extent to which random effects in the model are large, and (2) the extent to which there are sufficient observations to estimate the random effects." (p. 276).

Since the statistical model used in the study did converge, one can draw the conclusion that, despite the small number of participants, the full range of random-effects was large enough to capture the clustering in the data. Of course, this does not rule out the fact that a larger sample size would have provided more solid results. In this respect, the small number of items and participants is still a limitation in the present study.

Another potential limitation is that I could not include a measure of working memory in the study. Excluding it was primarily due to practical reasons. The experimental task had to be carried out in the limited amount of time available to L2 speakers during the language course in which they were enrolled. To avoid the potential risk of losing participants in the face of longer experimental sessions, I opted for only two tasks. Working memory has been shown to be a reliable predictor for the L2 speakers' ability to process syntactic structure although the evidence is mixed (see Juffs & Rodrigues, 2015 for review). The use of a working memory measure would have improved the design of the study and helped to gauge the degree of cognitive control of the L2 group (Bialystok et al., 2008).

An aspect that has been described in Chapter 2 and Chapter 4 is that the lexicon of the L2 speakers is shared between the two languages. This suggests a potential role for the L1 that could be captured, for example, using words like homographs or cognates. However, in the experiment, the role of the L1 could not be examined for two reasons. First, the number of L1s of the L2 speakers was extremely varied and without a numerically dominant group to be used as a benchmark (see Table 6.2). Such a variety would have been unlikely to produce a significant effect of the native language. Furthermore, as it emerged from the questionnaires, all participants were also advanced English learners, so it would have been hardly feasible to separate the effects of the native languages and L2 English.

Although the EI task was able to capture processing difficulties in L2 speakers, it failed to do the same with L1 speakers. One hypothesis might be that the type of task is better-suited for L2 speakers or that the low frequency verbs used in the cleft sentences was not low enough to affect the L1 group. Using a finer-grained measurement task that can measure the time course of language processing might be a way to separate the two effects. Hence, in the following chapter I will incorporate a self-paced reading task into the experiment. The type of task, offering precise measures of reading time in a word-by-word manner, is expected to uncover frequency and vocabulary effects in both L1 speakers and L2 speakers.

Finally, improvements could be made in the way the EI task is delivered and scored. As a measure of productive knowledge its results tend to be interpreted categorically, in terms of correct or incorrect repetitions. However, like other tasks that involve oral production, there are typical traits of spoken language that could be considered and added to the scores. The design of the task could be improved by examining where pauses or false starts occur during the attempted repetitions of the stimulus sentences. Furthermore, measuring the time it takes to repeat the segments of interest might improve the overall picture. This adjustment would need a trial phase to determine which measurements are better suited to capture the effects of the structure under scrutiny. This type of analysis was not possible in the present experiment due to the technical limitations of the tools used. The main software used to treat the EI responses, GarageBand, does not have the degree of precision necessary to perform such fine-grained measurements.

6.5.12 *Conclusions*

The study was based on the same item design as Hopp's (2016) study, but a productive online task was used instead of the self-paced reading task. The EI task replicated the frequency effects on L2 speakers but not the interaction between frequency and cleft type. Contrary to previous research, the L1 speakers did not show any effect of frequency. The most important result is that, in the L2 group, an interaction between vocabulary knowledge and syntactic processing arose. A higher lexical knowledge led to an increased number of correct cleft sentences repeated in the EI task with low frequency verbs, showing that as their vocabulary increases, L2 speakers are less sensitive to the effects of low-frequency words.

Frequency effects and the interaction between vocabulary and syntactic processing show that the difference between L1 and L2 speakers might be entirely due to their lexical representations. This contrast the claim of that the syntactic processing is inherently different in L2 speakers (Clahsen & Felser, 2006, 2018). What emerges from this study is that lexical effects, in the form of lexical frequency and vocabulary knowledge, are widespread and underlie every other aspect of language processing. In the next chapter, I will focus on a different type of filler-gap dependency to evaluate the range of frequency effects and feature availability on syntactic processing. Given the limitations of the EI in capturing fine-grained processing stages, I will use a self-paced reading task.

Chapter 7 Study 2

7.1 Introduction

The results of the EI study have shown that frequency plays a key role in the processing of complex structures, such as filler-gap dependencies. In L2 speakers, the accessibility of lexical syntactic features like SPR, COMPS, and GAP is hindered in the case of low-frequency verbs, so that in the relative segment of cleft sentences, the processing system cannot build the right structure. The results also showed that vocabulary is a good predictor of the ability to repeat correct cleft sentences, but only in L2 speakers. The L1 group was not affected by the frequency effects, as they could easily access the lexical features needed to form the sentence and there was no interaction between vocabulary and frequency.

In the study presented in Chapter 6, the impact of frequency on lexical features was localised to the relative verb. Since both the modified NP and the subject NP of relative clauses were proper names, they were highly accessible discourse referents (Gordon et al., 2001) that did not add significant lexical features. The interaction between frequency and vocabulary knowledge could arise only from the relative verb and was determined by the availability of the lexical features encoded in it. The information that the participant used to form a correct cleft came entirely from the meaning relations within the subcategorization frame of the verb.

The present experiment has two interrelated objectives. First, it aims to extend the results of Hopp (2016) and Tily et al. (2010) on a different type of filler-gap dependency. Second, using a different task for syntactic processing from the one used in the study presented in Chapter 6, the aim is to obtain a finer-grained measurement.

Concerning the first aim, in the present study, I set out to investigate the role of frequency and vocabulary knowledge when the processing system handles information not encoded in the subcategorization frame of the verb in the relative clause. To have a comparable type of filler gap dependency, the cleft sentences used in Hopp (2016) and Tily et al. (2010) were the starting point. Cleft sentences have a focused part followed by a relative clause. Importantly, the modified NP in the cleft sentence is the predicative complement of the verb *to be*.

A similar type of filler-gap dependency are object-modifying relative clauses. Like in other types of relative clauses, the relative pronoun in object-modifying relative clauses can take different grammatical roles. Example **7.1** shows an object modifying subject relative clause (henceforth only subject relative clauses)

7.1	Alice	ha baciato	lo studente	che	prepara	la torta
			INDEX j	REL j		
	Alice	has kissed	the student	who	prepares	the cake
			INDEX j	REL j	SPR $< NP_j >$	

Alice has kissed the student who is preparing the cake

The modified NP *lo studente* (*the student*) is the complement of the main sentence *Alice ha baciato* (*Alice has kissed*) and is modified by the following relative clause. The relative pronoun *che* (*who*) is coindexed with the NP *lo studente* (*the student*), and it is the subject of the relative sentence. Although the role of the modified NP is different from that of the cleft sentences, it is still a complement introduced by the main verb, although not focused. Therefore, the lexical information that the processing system needs to use is still encoded in the verb of the relative sentence. What is needed is a type of relative clause that has the same structure as Example **7.1** but requires the processing system to also use the broader set of lexical relation that are structurally encoded in the entire sentence. A type of relative sentences that fulfils this requirement are object-modifying oblique relative sentences, that is, relative clauses where the relative pronoun and the modified NP have the syntactic role of adjuncts (Kroeger, 2004). An example of an object-modifying oblique relative clause is given in **7.2**.

7.2	Alice	ha baciato	lo studente	con cui	prepara	la torta
	INDEX k		INDEX j	REL j	SPR < NP_k >	
	Alice	has kissed	the student	with whom	prepares	the cake
	INDEX k		INDEX j	REL j		

Alice has kissed the student with whom she is preparing the cake

I have already described relative clauses that have an adjunct filler in Chapter 4 (Sections 4.11 and 4.11.1), showing how HPSG handles them using the same features as the ones employed in relative clauses. Hence, I will omit the formal description and focus on the peculiar type of lexical relations that such relative clauses generate. The relative element, both in English and in Italian, is formed by the preposition and the pronoun, respectively, *with whom* and *con cui*. Together with the modified NP *lo studente* (*the student*), it has the role of the adjunct in the relative sentence. The modified NP is also the complement of the main sentence *Alice ha baciato* (*Alice has kissed*). The verb in the relative clause has its subcategorization frame already fulfilled by the subject *Alice* and the object *la torta* (*the cake*). In HPSG, this is shown by the co-indexing between the NP *Alice* and the index of the NP in the SPR list of the verb *prepara* (*prepares*). The wh element and the adjunct NP *lo studente* (*the student*) are co-indexed, as shown by the same value for the features REL and INDEX[3].

[3] In grammar frameworks that, unlike HPSG, make use of empty elements, the verb of the relative sentence would have a PRO subject. The English gloss would be:

$Alice_{\mathbf{k}}$ has seen the student with whom $PRO_{\mathbf{k}}$ prepares the cake

Given the type of syntactic relations outlined above, in an object-modifying oblique relative clause (henceforth only oblique relative clauses) the processing system needs to integrate two different sources of lexical information. The first one is encoded in the adjunct, and the second one is encoded in the verb of the relative sentence. The numbers and the types of meaning expressed by normal adjuncts, including relative adjuncts, are potentially unrestricted (Kroeger, 2004). However, the number and type of lexical items that can be part of the subcategorization frame are limited by the verb meaning (Bouma et al., 2001; Kroeger, 2004). In Example **7.2**, while there are certainly many things that a person can prepare, the number of entities and objects with which something can be prepared is much higher. Therefore, oblique relative clauses are expected to need the activation of more meaning information than that related to the verb, like, for instance, the means, the manners, and the time (Halliday, 2004; Kroeger, 2004). To sum up, oblique relative clauses have a structure that is similar to that of cleft sentences since the relative pronoun modifies a complement of the main sentence. Contrary to cleft sentences, the modified complement is not part of the verb subcategorization frame, in fact it is an adjunct. For this reason, the potential links between the adjunct and the relative verb are based on a high number of meaning relations that are not foreshadowed by the meaning of the relative verb. In this sense, the use of oblique relative clauses offers the opportunity to extend the results of Hopp (2016) and Tily et al. (2010), by examining a filler-gap dependency where lexical access is directly affected not only in terms of frequency but also in terms of meaning features.

The second aim of the present study, as said, is to adopt a different task to obtain a finer-grained measurement of syntactic processing. Indeed, one of the limitations of the study presented in Chapter 6 was that the EI can only give an overall view of syntactic processing, while the word-by-word stages are not captured. Therefore, while the EI was useful in extending Hopp (2016)'s results to production, it appeared not accurate enough to highlight the interaction between frequency and structure in L1 speakers. Thus, in the presented study, I have opted for a self-paced reading task (henceforth SPR). An SPR task offers a finer-grained measurement of how frequency and feature availability impact syntactic processing as the processing system operates in an incremental manner. In both object-modifying relative clauses and cleft sentences, the key constituents of the filler-gap dependencies become available incrementally. Therefore, a word-by-word measurement can track down how lexical features become available as soon as they are encountered. To measure this lexical interaction, I am going to use the I-Lex task. As shown in the partial correlation of vocabulary scores with syntactic processing in the experiment described in Chapter 6, I-Lex is able to capture the links between vocabulary and syntactic processing.

The chapter starts with a review of the few studies that have looked at the processing of relative clauses having adjuncts or optional indirect objects. Some of these studies have also examined the effect of semantic information encoded in the subcategorization frame. The main section of the chapter analyses the design and results of a self-paced reading study on oblique and subject relative clauses. The study aims to capture the relationship between the type of relative clause, frequency, and lexical knowledge. After discussing the results, like I did in Chapter 6 with the EI study, I will model them using the feature-based lexical network developed in Chapter 5.

7.2 Processing of Relative Clauses with Adjuncts or Optional Indirect Objects

Thus far, no studies have focused on Italian object-modifying relative sentences, and, in the field of SLA, research has chiefly focused on the difference between subject and object relative clauses (see Chapter 6, Sections 6.2 and 6.3). Some studies have examined filler gap dependencies where the filler had the role of adjunct in the relative clause, finding a relationship between plausibility and syntactic processing (Cunnings et al., 2010; Felser et al., 2012; Omaki & Schulz, 2011; Pickering & Traxler, 2001; Traxler & Pickering, 1996; Williams, 2006). Plausibility shows whether the meaning of the filler fits the meaning of the verb that encodes the gap. For example, in sentence **7.3** the NP *which girl* is a plausible filler for the verb *push*, while in sentence **7.4** the NP *which river* is not a plausible filler for the verb *push*.

7.3 Which girl did the man push the bike into late last night?

7.4 Which river did the man push the bike into late last night?

Several studies have shown that the processing system finds it more difficult to handle implausible information than plausible information. This is reflected in the processing slowdowns after the implausible segment. Plausibility effects suggest that the processing system activates some rapid semantic processing early on in the interpretations of the sentence (Traxler & Pickering, 1996). The fact that the meaning features are used to interpret filler-gap dependencies is in line with the assumption proposed in Chapter 5, that all lexically encoded information plays a key role in syntactic processing.

Other studies have focused on the structure examined in the present experiment, that is, relative clauses that modify the complement of the main clause (Felser & Roberts, 2007; Miller, 2014b, 2014a; Rastelli, 2013; Roberts et al., 2007). Felser & Roberts (2007) examined object-modifying relative clauses in L1 and L2 speakers. An example of the items used in the study is shown in Example **7.5**.

7.5 Jane loved the tiger to which the black beetle offered the sweet [TIGER/PAINTBRUSH] strawberry cake [TIGER/PAINTBRUSH] __ at the party last week.

The relative clause has the wh-phrase *to which* that is the indirect object of the subcategorization frame of the verb *to offer*, and it is also coindexed to the object NP of the main sentence, *the tiger*. The study was carried out using the cross-modal priming technique. It consists of presenting images while participants read the sentence and asking them to classify the images based on their semantic features. The images used in the study depicted a referent that was related (the picture of a tiger) or unrelated (the paintbrush) to the modified NP. Participants were asked whether the animal or object in the images was alive or not. The images were presented at different points, slightly before the gap and in the gap position after the object NP *the sweet strawberry.* The images presented in the experiment and the position are shown in square brackets. Reaction times to perform the semantic decision task on the images were recorded. Felser & Roberts (2007) expected an interaction between participants performance on the target images, their relationship with the referent, and the structural position in the sentence. Their reasoning was that, if participants could reactivate the filler and the modified NP (i.e., *the tiger to which*) at the right structural position, the unrelated image (i.e., paintbrush), when compared to the related image (i.e., tiger), would have a confounding effect, causing reaction times to be slower. At the pre-gap position, no structurally driven activation of the filler occurs, so no difference between the related and unrelated picture is expected to appear.

The results showed that the L2 speakers responded to the related images more quickly, without showing a significant difference in the position where the images were presented. Interestingly, the L1 speaker group showed the same facilitation effect for the related pictures, but the effect was significantly larger at the gap position. Felser & Roberts (2007) concluded that L1 speakers were making use of merely structural information, so the same referent sped up their performance in the gap position only. In contrast, L2 speakers were making use of semantic information, keeping the filler active throughout the sentence. However, Miller (2014a), using the same cross-modal priming task with cleft sentences, obtained a different pattern. She found that the L2 participants showed response patterns that reflected the activation of the filler at the correct structural position, as in the L1 group. She interpreted the finding as evidence that L2 learners can process real-time input like L1 speakers.

However, a non-structural interpretation of the results is also possible. As explained in Chapter 2 and Chapter 5, and as shown in the results of the I-Lex study in Chapter 3, lexical representations in L2 speakers are, overall, more susceptible to interference from vocabulary knowledge and frequency effects. In the case of Example **7.5**, L2 speakers must handle meaning and syntactic features at the same time for a long segment of the sentence. In addition, as described in Chapter 5, meaning and word-form features interact in L2 speakers. As a result, it could be that L2 speakers are computing the correct syntactic structure, but are less precise in using the lexical information to join the filler to the gap. This explanation echoes what Cunnings (2017) described as the role of interference in retrieving the correct lexical cues during the resolution of filler-gap dependencies.

If the hypothesis holds, then facilitating lexical access should diminish the effects on syntactic processing. Support to this claim comes from Miller (2014), who investigated the differences in the processing of relative clauses with indirect object in English learners of L2 French. She used cognate words to weigh the role of lexical access. Example **7.6** shows an item used in the study, which has the same structure as those used in Felser & Roberts (2007). The key segment in the target items was the modified NP, *le papillon*. In one version they were non-cognate words, like *papillon* and *butterfly* in Example **7.6**, and in the other version there were cognate words like, for instance *eléphant* and *elephant*.

7.6 Charlotte critique le papillon à qui Robert a offert
Charlotte criticizes the butterfy to whom Robert offered

le nouveau [BUTTERFLY/CALCULATOR] ballon [BUTTERFLY/CALCULATOR]
the new [BUTTERFLY/CALCULATOR] ball [BUTTERFLY/CALCULATOR]

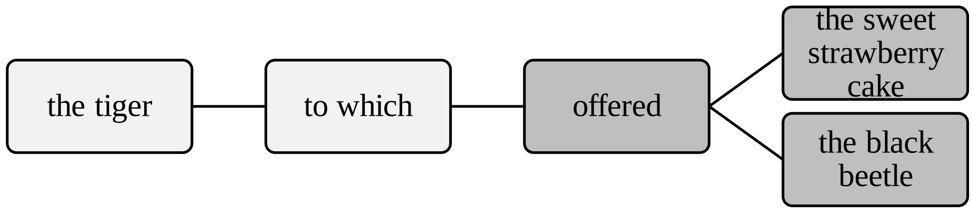
dans le jardin lundi soir.
in the garden Monday evening

Like in Felser & Roberts (2007) participants had to judge related (i.e. butterfly) and unrelated (i.e. calculator) images presented at a pre-gap and gap position, respectively, before and after the word *ballon* (the content of the images is described by the words in square brackets). The result showed that, with cognate words, the expected slower reaction times for the unrelated image (e.g., calculator) occurred at the gap position for both L1 and L2 speakers. Miller (2014) argued that faster lexical processing of cognates led the processing system to have more available resources to work out the right syntactic structure.

Taken together, the studies described so far suggest that in object-modifying relative clauses, the great distance between the filler and the gap can delay lexical access, causing disruptions to syntactic processing. Another factor that might have played a role in the results is the lexical relation between the filler and the relative clauses used in both studies. To resolve the filler gap dependency, the processing system must integrate the lexical features of the core arguments, object, and subject, with the lexical features of the indirect object, which is not a core argument (Kroeger, 2004). In this respect, indirect objects resemble oblique complements, as they have a broader relationship with the meaning of the verb. This is shown in Figure 7.1 which describes Example **7.5** used by Felser & Roberts (2007).

Figure 7.1

A Flow Chart for the Lexical Relations in Example 7.5



The meaning links activated by the lexical information encoded in the indirect object are on the left-hand side in a light-grey background. The meaning links activated by the lexical information encoded in the verb are on the right-hand side in a dark-grey background.

7.3 The Study

The studies reviewed above showed that when the processing system handles lexical relations not entirely predictable from the subcategorization frame of the verb, integrating meaning features might play a key role in syntactic processing. The present experiment builds on the results of the studies described above and aims to investigate to what extent lexical integration can affect syntactic processing. It tests this in two ways. First, by disentangling the information encoded in the verb from the information encoded in the filler. To do so, as I have explained in the Introduction, I will use object-modifying oblique relative clauses where the modified NP has the role of adjunct within the relative clause (Example **7.2**). This type of relative clause requires a more complex process of lexical integration based on a large set of meaning features. I will compare them to object-modifying subject relative clauses like Example **7.1**, where no complex lexical integration occurs, as the modified NP is the subject of the relative verb. Second, I will test whether vocabulary knowledge, measured with I-Lex, and frequency show a relationship with the lexical integration process needed to interpret object-modifying oblique relative clauses. The design of the experiment draws on the role of frequency found in Hopp (2016) and Tily et al. (2010) and on the interaction between frequency and vocabulary for L2 speakers shown in the EI experiment but it uses a more complex process of lexical integration.

7.3.1 *Research Questions*

Research question (1): to what extent will frequency interact with syntactic structure?

In line with the findings in Hopp (2016) and Tily et al. (2010), low-frequency verbs are slower to access and are expected to give rise to slower reading times. The interaction between frequency and the availability of lexical features is more likely to affect L2 speakers especially with low-frequency verbs (Brysbaert et al., 2018). In effect this was the pattern that emerged from the EI study presented in Chapter 6. Although the L1 speakers did not show any sensitivity to frequency, the L2 group did. As noted in Chapter 4 (see Section 4.11.1) subject relative clauses are expected to be easier to interpret because the processing system has to handle the feature SPR as it will in any other normal sentence. On the contrary, in the case of oblique relative clauses, the processing system must interpret the meaning of the relative clause handling the information encoded in the two features COMPS and GAP. Dealing with information encoded in two distinct features reflects the processing cost. In addition to this, in the present experiment, the wh-element of oblique relative clauses is distant from the post-verbal region where it is associated with the gap. As shown in Example **7.2**, in oblique relative sentences, before associating the relative pronoun and the gap, the processing system needs to interpret the VP of the relative clause. Therefore, based on evidence that keeping filler active for longer requires more cognitive resources (Gibson, 2000; Gordon et al., 2002; Parker et al., 2017) oblique relative sentences should slow down reading times. In contrast, with subject relative clauses, the filler can be immediately positioned in the subject slot, and there is no need to keep it active in the memory. Frequency effects are expected to worsen this structural difference between oblique and subject relative clauses, causing slowdowns with low-frequency verbs for L2 speakers. The effects are expected to be less prominent in L1 speakers, showing a pattern similar to that found in Hopp (2016).

Research question (2): to what extent will productive vocabulary knowledge will facilitate reading times?

Oblique relative clauses build on a broad scope meaning relationship between the verb, its complements and other adjuncts with which it may occur. In the Introduction I have explained that adjuncts add further pieces of information to the meaning encoded in the verb subcategorization frame. They can be about, for instance, the goals, the way, and the time associated to event described in the verb meaning. The meaning relations between adjuncts and verbs are not predictable from the verb's meaning although they must be semantically plausible in relation to it (Bouma et al., 2001; Kroeger, 2004). Considering example 7.2 at page 242, repeated here for convenience, the processing system must access the meaning feature encoded in the verb *prepara* (*prepares*) to link it to the subject NP *Alice*. This process is based on the subcategorization frame of the relative verb: the subject of the verb *prepara* is likely to be an agent and the NP *Alice* fits well with this prediction.

7.2	Alice	ha baciato	lo studente	con cui	prepara	la torta
	INDEX k		INDEX j	REL j	SPR $< NP_k >$	
	Alice	has kissed	the student	with whom	prepares	the cake
	INDEX k		INDEX j	REL j		

Alice has kissed the student with whom she is preparing the cake

In addition, the processing system must link the modified NP *lo studente* (*the student*) and the relative pronoun *con cui* (*with whom*) to the relative verb. In this case the process is only loosely based on the meaning of the relative verb. First, the number of entities and objects with which something can be prepared is extremely large and potentially unrestricted.

Second, the entities and objects with which something can be prepared are not associated to any specific thematic role encoded into the verb subcategorization frame. Hence, at the verb segment, the processing system must access the lexical network and activate additional meaning and structural links between the verb and the adjunct (i.e., between the NP *lo studente* and the verb *prepara* in example 7.2 above). This has a cognitive cost for the processing system since those additional links to the verb meaning are not entirely predictable. On the contrary, the cognitive cost should be less pronounced with subject relative sentences. The reason is that the expectations about which word can be the subject of the relative clause are encoded in the verb meaning and quickly fulfilled as soon as the processing system arrives at the verb segment. If so, a larger vocabulary knowledge should be able to reduce reading times of oblique relative sentences to a higher degree than subject relative sentences. Furthermore, the EI study in Chapter 6 has shown that frequency and vocabulary knowledge interact in the L2 speakers. A similar pattern is expected in the present study, with the effects of vocabulary knowledge being more prominent with low-frequency verbs. Furthermore, the use of the SPR is expected to show a similar relationship between reading times and L1 speakers.

7.3.2 *The Design of the Study*

The study builds on a self-paced reading task (henceforth SPR) and the I-Lex task. The I-Lex task is the same as the one used in the validation test and in the EI study in the previous experiment in Chapter 6. As already explained, I-Lex is a word association task, with 30 stimulus words to which participants can associate up to four words. Every word that is not among the first 1000 most frequent Italian words is scored one point. The type of knowledge that the I-Lex capture has been extensively examined in Chapter 3. In the present study, it is expected to capture the vocabulary knowledge of the participants and to interact with their reading times on the SPR. In the EI experiment in Chapter 6 the task has shown to be able to capture the relationship between vocabulary knowledge of L2 speakers and how they process syntax. Given that the SPR offers a finer-grained measurement of sentence processing in terms of reading times, the relationship with I-Lex should be more apparent.

The second task is a self-paced reading task, an online reading task in which participants read sentences broken into words or phrases and can control their reading pace by pressing a key. The time that participants need to read a segment or a word is captured by measuring the time between each key press: the whole mechanism allows researchers to determine with great precision when processing costs or facilitation arises (Jiang, 2013; Keating & Jegerski, 2015). The task has been extensively used in L1 and L2 research and is considered to tap into automatic processing and implicit knowledge (Marsden et al., 2018). Two main assumptions underlie the task. First, the processing effects emerge as longer reading times for the experimental items relative to a baseline condition. Second, L2 speakers, because of their lower degree of automaticity, are expected to show slower reading times than L1 speakers (Keating & Jegerski, 2015; Marsden et al., 2018).

Furthermore, L2 speakers could rely on their L1 reading skills or be influenced by their explicit knowledge of the second language (Jiang, 2013; Marsden et al., 2018). Therefore, it is not always easy to differentiate slowdown effects caused by a lack of automaticity from slowdown effects caused by a lack of sensitiveness to the structure examined in the task. However, the assumption that processing disruptions cause slower reading times is tenable, overall. Although reading times might be influenced by several factors, nevertheless, when the reading pace slows down, it is a clear sign that some problems are straining cognitive processes. As noted by Marsden et al. (2018), the fact that the readers decided their reading pace by pressing a key, "leaves control over exposure time to the participant (as in natural reading) and, as such, can concurrently measure processing time, thus reflecting online cognitive mechanisms" (p. 2).

Therefore, the SPR is well suited to capture the processing difficulties in oblique relative sentences. The target items aim to measure the lexical relations in a long segment of the sentence, from the relative pronoun to the word after the complement of the relative verb (see Section 7.3.3. Since it measures how the sentence is processed word by word, the SPR task can offer a deeper insight into syntactic processing. Additionally, since the SPR ensures precise reading times for every word of interest, the comparison to the I-Lex scores is expected to be more precise.

7.3.3 *Materials*

The SPR task has 16 target sentences and 68 filler sentences (see Appendices 6 and 7 for the list of the target sentences and the fillers). The target items consist of a doublet of two sentences, one having an oblique relative clause and one having a subject relative clause. There are 8 items with low-frequency verbs and 8 items with high-frequency verbs, the same number of items used in the EI experiment in Chapter 6. To avoid constructing too contrived target sentences, high- and low-frequency verbs had different meanings. Unfortunately, the problems concerning the small number of participants and items explained for the EI study applied to the present study as well. Since the data collection took place in during the same course, students could be invited but not compelled to take part into the study. For the same reasons as those I laid out in Chapter 6 (see Section 6.5.3) I opted for a similar design with a fairly small number of items. The choice, in this case, was also backed up by the results of the EI study, where the number of items was large enough to make the statistical model converge. In addition, to avoid the target items from being recognised, the number of filler sentences was kept high.

An example of a target item with a high-frequency verb is given below in Example **7.7**. The first part of the is given only once, followed by the two types of relative clauses.

7.7 Il negozio ha iniziato una promozione
The shop$_k$ has started a promotion
The shop has started a promotion

7.7a) che offre gli sconti in base agli acquisti
which offers the reductions based on the purchases
which offers reductions based on the purchased

7.7b) con cui offre gli sconti in base agli acquisti
with which PRO$_k$ offers the reductions based on the purchases
through which it offers reductions based on the purchased

All target sentences started with a transitive clause (*The shop has started a promotion; il negozio ha iniziato un promozione*) and the object modifying wh-element. Oblique relative clauses have the wh-element within a PP *with which* (*con cui*) while subject relative clauses have the relative pronoun *che* (*which*). After it, the verb of the relative clause appears (*offre*), followed by the complement (*gli sconti*) and the final adjunct (*in base agli acquisti*). There are two important points related to the target items. The first point is that, in Italian, the relative pronouns *che* and *cui* have the same form for animate and inanimate referents. Therefore, contrary to what the English equivalents *which* might imply, the Italian forms are both bare relatives. The second point is that, as described in Section 7.1, in oblique relative sentences, the subject of the main clause is also the subject of the relative clause. In subject relative sentences, the modified NP is the subject of the relative clause. Thus, subject relative clauses do not give rise to the same lexical relations as oblique relative clauses and are expected to work as control sentences.

To select high-frequency verbs, I followed the same procedure as in the EI experiment, using verbs that were in the first 2000 most frequent words of the NVDB (Chiari & Mauro, 2010, 2015). The assumption was that L2 speakers with a high proficiency level would be familiar with the first 2000 most frequent items. In contrast to the EI study, the low-frequency verbs were taken from the second frequency band of the NVDB list, which contains the second 2000 most frequent Italian words. The choice was different from the EI experiment in Chapter 6, because the accuracy of the SPR task as a measurement task could easily capture a less sizable frequency gap. The verbs used to construct the target relative clauses are shown in Table 7.1. Figures for the itTenTen corpus are also given for comparison (Jakubícek et al., 2010; Jakubíček et al., 2013). As noted in Chapter 6 there are some discrepancies with the corpus frequency, due to the different ways in which the corpora that underlie the NVDB list and the itTenTen corpus were assembled.

Table 7.1

List of the Verb used in the Relative Clauses

High-Frequency Verbs		Low-frequency verbs	
Target item	Freq. per million	Target item	Freq. per million
lavorare (work)	243.74	assegnare (assign)	61.97
cambiare (change)	217.77	mescolare (blend)	35.69
offrire (offer)	344.47	procurare (provide)	17.77
pagare (pay)	146.42	archiviare (file)	11.9
Average frequency	238.1	Average frequency	31.83

7.3.4 ***Participants***

The L2 speakers that took part into the study are a subset (N = 20) of the 82 advanced L2 speakers' group. The same factors noted in Chapter 3 and Chapter 6 about the group apply to the present group:

- Participants have at least a B2 level of proficiency established through language certification.
- Participants attend a C1 level Italian language course for 4 hours every morning, Monday to Friday
- Participants attend every afternoon a two-hour conference on topics related to Italian culture from Monday to Friday

The L1 speaker group (N=12) is the same group that took part into both the I-Lex study and the EI study. All participants completed a consent form and a questionnaire with questions about their age, their education, and their L2 (see Appendix 2). Data collected from the questionnaire are provided in Table 7.2.

Table 7.2

Data of L1 and L2 Speakers Collected in the Questionnaire

	Age	Gender	Education	Proficiency	First language
L1 (N=12)	M = 28.2 SD = 9.4	F = 3 M = 9	BA = 7 PhD = 5	NATIVE	Italian (12)
L2 (N=21)	M = 25.5 SD = 4.8	F = 13 M = 8	BA = 12 MA = 4 PhD = 4	B2-C1	Spanish (4), Russian (2), English (2), Serbian (2), Hungarian (2), Arabic (2) Chinese, Portuguese, Polish, Finnish, French, German (1)

7.3.5 *Procedure*

As in the previous two experiments, before performing SPR and I-Lex, the L2 and L1 speakers received the same consent form, the same brief on the purpose and method of the study, and the same questionnaire (Appendix 2, see Appendix 1 for Ethics Assessment Status). They were asked to read them carefully and fill in all the parts, asking questions if anything would not be clear. The I-Lex task was carried out in one session by L2 speakers, in pencil and paper form, and lasted approximately 15 minutes. As described in Chapter 3 the L1 speakers completed the task directly on a spreadsheet in one session of the same duration as the L2 speakers. The SPR task was carried out by L2 speakers individually in four consecutive days after they had attended the afternoon conferences. Only the experimenter and one participant at the time were present during each session. Each SPR session took approximately 30 minutes. Regarding the L1 group, each participant carried out the SPR task individually after completing the EI. The procedure followed for the SPR task was the same in both groups. Each participant was sitting in front of a computer screen, a Macbook Pro with a 13-inch retina screen, in a quiet room. The experimenter explained to the participants that they were about to carry out a reading comprehension task that contained various grammatical structures. He carefully avoided making any reference to the structure under scrutiny. Participants first read instructions about the task written in English by the experimenter, who also explained the procedure orally to make sure everything was clear. Before starting the actual experiment, the participants performed 8 examples to familiarise themselves with the task. After each example, the experimenter took care to ask if they had any questions and the task was sufficiently clear. After the training stage the experimenter launched the SPR task. The stimulus sentences were presented in a segment-by-segment fashion, in white letters in the centre of the screen on a black background. All words were in 20-point Mono font. The task was administered via OpenSesame software (version 3.1.6).

Each participant read a randomized version of the SPR task. At the beginning of each sentence, participants were asked to focus on a fixation cross in the centre of the screen, in the same position as the first word of the sentence. To move from one word to the next, participants used the space bar. After the final word of each target item, to move to the following stimulus sentence, participants needed to press the space bar: a new fixation cross in the centre of the screen marked the beginning of the next sentence.

After each target item, a comprehension question was presented to ensure that the participants paid close attention to the meaning of the sentences. As for the fillers, since they aimed to have participants engage with the task, following Keating & Jegerski (2015), we included one question for every three or four stimuli. This led to 25 comprehension questions for the set of fillers. The questions for the target items had all the same format: they reframed in a declarative form the event described in the relative clause, and the yes or no answer was about the meaning of the relative clause. Participants could answer *yes* by pressing the space bar or *no* by pressing the N key. An example (**7.8**) is given below based on example **7.7** – recall that Italian does not have an auxiliary for questions (see Apeendix 6 and 7).

Sentence:

7.8 Il negozio ha iniziato una promozione che offre gli sconti
The shop has started a promotion which offers the reductions

The shop has started a promotion which offers the reductions

in base agli acquisti
based on the purchases

based on the purchases

Question:

Il negozio offre sconti in base agli acquisti?
The shop offers reductions based on the purchases?

Does the shop offer reductions based on the purchase?

7.3.6 *Results*

The first step in the analysis of the data concerns the number of correct responses to the comprehension questions produced by the participants during the SPR (section 7.3.7. The second step is the analysis of frequency and relative type effects on the reading times (section 7.3.8. The third step is the analysis of the interaction between the I-Lex vocabulary scores and reading times (Section 7.3.11.

7.3.7 *Analysis of the Comprehension Questions*

Before carrying out the statistical analysis, the accuracy scores were averaged between items, obtaining a score for each type of relative clause in each frequency condition. Overall, the accuracy of the responses to the target items was high, showing that participants were engaged in the task and showed a satisfactory level of comprehension. The results are given in Table 7.3.

Table 7.3

Absolute Number of Comprehension Question Accuracy

	L1 Speakers		L2 Speakers	
	High-frequency M (*SD*)	Low-Frequency M (*SD*)	High-frequency M (*SD*)	Low-Frequency M (*SD*)
Subject Relatives (out of 4)	3.6 (0.6)	3.5 (0.7)	3.8 (0.4)	3.9 (0.3)
Oblique Relatives (out of 4)	2.9 (0.2)	3.2 (0.8)	2.4 (0.6)	3.4 (0.6)

Since the results were averaged between items and the analysis aimed to find an interaction between frequency, type of relative, and group, there was no need to set up a linear mixed-effects model with random effects. A Shapiro-Wilk normality test revealed that the distribution was not normal in both groups (L1: $W = 0.75$, $p < 0.005$; L2: $W = 0.73$, $p < 0.005$). Therefore, the analysis was carried out using a Kruskal-Wallis rank sum test. The analysis of both groups revealed a significant main effect of frequency ($H(1) = 9.09$, $p < 0.005$), a non-significant effect of language ($H(1) = 0.26$, $p = 0.61$), and a significant effect of relative type ($H(3) = 52.69$, $p < 0.0005$).

The analysis of the L1 group did not show a significant main effect of frequency ($H(1) = 0.97$, $p = 0.32$) but a significant main effect of relative type ($H(3) = 11.10$, $p = 0.01$).

Table 7.4

Comprehension Questions: Post hoc Analysis, L1 Speakers

	High-Freq Subj	High-Freq Obl	Low-Freq Subj	Low-Freq Obl
High-Freq Subj	–			
High-Freq Obl	0.01	–		
Low-Freq Subj	0.84	0.01	–	
Low-Freq Obl	0.25	0.22	0.28	–

Note. Only p-values are shown.

The Dunn post hoc analysis (see Table 7.4), showed a significant difference in response accuracy between subject ($M = 3.6$, $SD = 0.6$) and oblique ($M = 2.9$, $SD = 0.2$) relative clauses with high-frequency verb ($z = -2.95$, p = 0.01). Accuracy was lower ($z = -2.76$, p = 0.01) for oblique relative clauses with high-frequency verbs ($M = 2.9$, $SD = 0.2$) compared to subject relative clauses with low-frequency verbs ($M = 3.5$, $SD = 0.7$).

The analysis of the L2 group showed a significant main effect of frequency ($H (1) = 8.79$, $p < 0.004$) and a significant main effect of relative type ($H (3) = 42.41$, $p < 0.0005$). The Dunn post hoc analysis (Table 7.5) showed a marked effect of type of relative clause. With low-frequency relative verbs, L2 speakers gave more correct responses to subject relative clauses (*M* = 3.9, *SD* = 0.3) than to object relative clauses (*M* = 3.4, *SD* = 0.6) and the difference was significant (*z* = - 2.20, *p* = 0.04). Relative clauses with high-frequency verbs showed the same pattern, with more correct responses to subject relative clause (*M* = 3.8, *SD* = 0.4) than to object relative clauses (*M* = 2.4, *SD* = 0.6) and the difference was significant (*z* = - 2.20, *p* = 0.04). In terms of frequency effects, responses to oblique relative clauses with high frequency verbs (*M* = 2.4, *SD* = 0.6) were lower than responses to oblique relative clauses with low-frequency verbs (*M* = 3.4, *SD* = 0.6) and the difference was significant (*z* = -3.67, *p* < 0.0005). A significant difference (*z* = -5.88, *p* < 0.0005) was also found between oblique relative clauses with high frequency verbs (*M* = 2.4, *SD* = 0.6) and subject relative clauses with low frequency verbs (*M* = 3.9, *SD* = 0.3).

Table 7.5

Comprehension Questions: Post hoc Analysis, L2 Speakers

	High-Freq Subj	High-Freq Obl	Low-Freq Subj	Low-Freq Obl
High-Freq Subj	–			
High-Freq Obl	< 0.0005	–		
Low-Freq Subj	0.6	< .0005	–	
Low-Freq Obl	0.11	< .0005	0.04	–

Note. Only p-values are shown.

The pattern that arises from the response's accuracy shows that for both the L1 and the L2 groups oblique relative sentences were harder to comprehend. The post hoc analysis performed on each group showed a more blurred pattern. The L1 group produced fewer correct responses to oblique relative clauses with high-frequency verbs than to their subject counterpart. No such effect arose with relative clauses with low-frequency verbs. In the L2 group, oblique relative clauses were harder to comprehend in both high- and low-frequency verbs. This suggests that oblique relative clauses were the chief source of inaccuracy, and, in both groups, high-frequency verbs did not seem to help the correct interpretation.

7.3.8 *Analysis of the SPR Reading Times*

In the analysis of SPR reaction times, it is customary practice to remove target items if participants gave the wrong answer to the following comprehension questions (Jiang, 2013; Keating & Jegerski, 2015). In contrast, in this study, the analysis of all target items was carried out, including those whose answers were not correct. The reason for the choice is that the SPR seeks to investigate how participants, both L1 and L2 speakers, read and process sentences rather than how they read and process sentences to subsequently give a correct answer about their content. The difference is not as trivial as it may seem. Answering correctly is not a mere result of syntactic processing, as it also draws on more general strategies and preferences in interpreting sentences (Marsden et al., 2018). The claim that the wrong response originates from the inability to build the right syntactic structure is too strong and unwarranted (Hopp, 2006). Therefore, the present study has a very similar design to the studies of Hopp (2016) and Tily et al. (2010) and it aims to compare its results with those of the two studies. In their studies, they both included all the items in their analysis, regardless of the comprehension questions. Both studies had a level of accuracy far lower than in the present study. In Hopp's (2016) study, the accuracy ranged from 57.5% to 90.3%, while in Tily et al. (2010), it ranged from 74.7% to 86.8%. The accuracy in the present study is higher than that of Hopp (2016), ranging from 60% to 97%, and similar to that of Tily et al. (2010).

Another customary practice is the use of residuals reading times. Residual times are drawn from submitting the raw reading times of each participant to a regression equation. Word length is the predictor, and reaction times yielded by each word are the outcome (Ferreira & Clifton, 1986). The result is a regression line based on the interaction between word length and reading times. Using the model residuals, normalises the differences in reading times of each participant (Keating & Jegerski, 2015; Marsden et al., 2018). In contrast, in this study, I will examine raw reading times. Since the experiment investigates frequency, a factor related to the lexical representations of individuals (Ellis, 2011) converted residual reading times can conflate any frequency effects, overshadowing individual differences.

Following the methodology of the EI experiment presented in Chapter 6, the data were analysed with a linear mixed effect model The analysis was carried out using the lmer package available on R (Baayen et al., 2008; Cunnings, 2012; R Core Team 2013). The same model was applied to the four regions set up for statistical analysis. The regions that were analysed are shown in Table 7.6.

Table 7.6

Regions Used in the Analysis of Reading Times

Region 1 Relative	Region 2 Verb	Region 3 Complement	Region 4 Post-complement
che/con cui	offre	gli sconti	in base
which/with which	offers	the sales	based on

Note. The example given shows a high frequency verb

The first region, the relative region, was included because of the structural difference between the two wh-phrases. The verb region includes the frequency effect, while the complement region is crucial to determine slowdowns in reading times at the point where the process of attaching the oblique filler is likely to start. The post-complement region should single out, if present, spillover effects.

The model had reading times, relative type (subject vs. oblique), frequency (low vs. high) and language (L1 vs. L2) as fixed effects, and subject and items as random intercepts. Before entering the data into the statistical model, all reaction times that were greater than +/- 2SDs from the mean were excluded. Like in the statistical analysis carried out in Chapter 6, I have adopted the assumption of Schielzeth et al. (2020) on the robustness of mixed-effects models when the normality of distribution of random effects and residuals is violated. Random slopes could not be added to the model because the model with random slopes gave a large eigenvalue ratio, indicating a low correlation between predictors. In cases with a large eigenvalue ratio, assessing the importance of each predictor may not be reliable enough (A Field et al., 2012). For each model in each region, residuals that were greater than +/- 2SDs from the mean were excluded, and the final analysis was carried out on the trimmed model. The distributions of reading times and of the residuals of the four models are in Appendix 8 (Figures 1, 2, and 3).

Research question (1): to what extent will frequency interact with syntactic structure?

The reading times for the relative region are shown in Table 7.7. It appears that the L2 speakers were overall marginally slower than the L1 speakers. Both groups show similar reading times on both oblique and subject relative sentences.

Table 7.7

Reading Times in the Relative Region

	Relative Region			
	L1		L2	
	High-frequency *M* (*SD*)	Low-frequency *M* (*SD*)	High-frequency *M* (*SD*)	Low-frequency *M* (*SD*)
Subject clauses	578 (203)	558 (215)	651 (269)	602 (205)
Oblique clauses	529 (199)	514 (152)	609 (232)	627 (258)

In the segment of the relative pronoun, no effect was found between sentences with low- and high-frequency verbs (*ß* = -27.97, *SE* = 36.34, *t* = - 0.75, *p* = 0.78) or between oblique and subject relative clauses (*ß* = -45.29, *SE* = 36.24, *t* = - 1.25, *p* = 0.22). Language did not affect reading times (*ß* = -55.90, *SE* = 56.59, *t* = 0.98, *p* = 0.32) and did not interact with frequency (*ß* = 4.71, *SE* = 36.22, *t* = 0.13, *p* = 0.89) nor with the type of relative clause (*ß* = 29.08, *SE* = 35.93, *t* = 0.80, *p* = 0.41). The interaction between type and frequency was not significant (*ß* = 37.62, *SE* = 51.25, *t* = 0.73, *p* = 0.44) as it was the three-way interaction between frequency, language, and type of relative clause (*ß* = 1.43, *SE* = 50.64, *t* = 0.02, *p* = 0.97). No significant effects emerged in the post hoc analysis in both groups.

The data for the verb region (Table 7.8) show that the L2 speakers had overall slower reading times, especially in low-frequency verbs. The difference between the two types of relative clause is again fairly small, suggesting that oblique relative clauses were not particularly hard to read in the verb region. No effects of relative clause type (*ß* = -8.23, *SE* = 46.17, *t* = - 0.17, *p* = 0.8) or frequency (*ß* = 0.91, *SE* = 45.61, *t* = 0.02, *p* = 0.9) were observed.

Table 7.8

Reading Times in the Verb Region

	Verb Region			
	L1		L2	
	High-frequency *M* (*SD*)	Low-frequency *M* (*SD*)	High-frequency *M* (*SD*)	Low-frequency *M* (*SD*)
Subject clauses	558 (205)	563 (204)	629 (253)	742 (332)
Oblique clauses	574 (262)	580 (223)	630 (250)	761 (298)

Although the effect of language was not significant overall (*ß* = 90.15, *SE* = 68.29, *t* = 1.32, *p* = 0.19) there was a significant interaction between frequency and language in that L2 speakers were significantly slower than L1 speakers in reading oblique and subject relative sentences with low frequency verbs (*ß* = 95.44, *SE* = 41.01, *t* = 2.32, *p* < 0.5). The interaction between relative type and frequency was not significant (*ß* = 22.97, *SE* = 65.16, *t* = 0.35, *p* = 0.72) nor was the overall three-way interaction between frequency, language, and relative clause type (*ß* = 4.04, *SE* = 58.50, *t* = 0.06, *p* = 0.94). In the verb region, L2 speakers were slower with low-frequency verbs, while L1 speakers did not experience any frequency effect in their reading times.

The reading times in the complement region are shown in Table 7.9. In this region, oblique relative clauses did not seem to be more difficult to read for both L1 and L2 speakers. Overall, L2 speakers show slower reading times, although the difference between low- and high-frequency relative clauses is smaller than in the verb region. There was no effect of frequency (*ß* = -4.52, *SE* = 65.32, *t* = -0.06, *p* = 0.94) and no effect of relative clause type (*ß* = 63.48, *SE* = 65.68, *t* = 0.96, *p* = 0.34). Language turned out to be a significant predictor, with L2 speakers generally slower than L1 speakers (*ß* = 298.4, *SE* = 86.768, *t* = 3.43, $p < 0.005$).

Table 7.9

Reading Times in the Complement Region

	Complement Region			
	L1		L2	
	High-frequency *M* (*SD*)	Low-frequency *M* (*SD*)	High-frequency *M* (*SD*)	Low-frequency *M* (*SD*)
Subject clauses	658 (319)	640 (300)	951 (395)	919 (418)
Oblique clauses	712 (283)	670 (302)	941 (371)	885 (397)

There was no significant interaction between relative type and frequency (*ß* = -21.51, *SE* = 92.6, *t* = -0.23, *p* = 0.81) or between language and frequency (*ß* = -36.68, *SE* = 60.15, *t* = -0.61, *p* = 0.81) and between language and relative clause type (*ß* = -56.59, *SE* = 60.1, *t* = -0.94, *p* = 0.34). As in the previous regions, the main interaction between frequency, language and relative clause type was not significant (*ß* = 2.04, *SE* = 84.51, *t* = 0.02, *p* = 0.98). The fact that L2 speakers are slower than L1 speakers is an expected effect. However, in this region, the shortcomings in L2 processing are not restricted to relative clauses with low-frequency verbs. This might suggest that at the point where the filler-gap dependency is resolved, L2 speakers are slower than L1 speakers, regardless of the type of structure and the frequency of the relative verb.

Table 7.10 shows the reading time in the post-complement region. The difference in reading times between L1 and L2 speakers is smaller compared to the previous regions.

Again, oblique relative clauses do not seem to put additional strain on the processing system. Relative sentences with low- and high-frequency verbs did not differ (*ß* = -42.8, *SE* = 39.25, *t* = -1.09, *p* = 0.28), there was no difference between relative clause type (*ß* = 12.59, *SE* = 39.50, *t* = 0.30, *p* = 0.76) nor between L1 and L2 speakers (*ß* = 77.41, *SE* = 59.41, *t* = 1.03, *p* = 0.2).

Table 7.10

Reading Times in the post-Complement Region

	Post-complement Region			
	L1		L2	
	High-frequency *M* (*SD*)	Low-frequency *M* (*SD*)	High-frequency *M* (*SD*)	Low-frequency *M* (*SD*)
Subject clauses	603 (189)	555 (141)	694 (251)	722 (267)
Oblique clauses	643 (257)	608 (210)	678 (249)	705 (263)

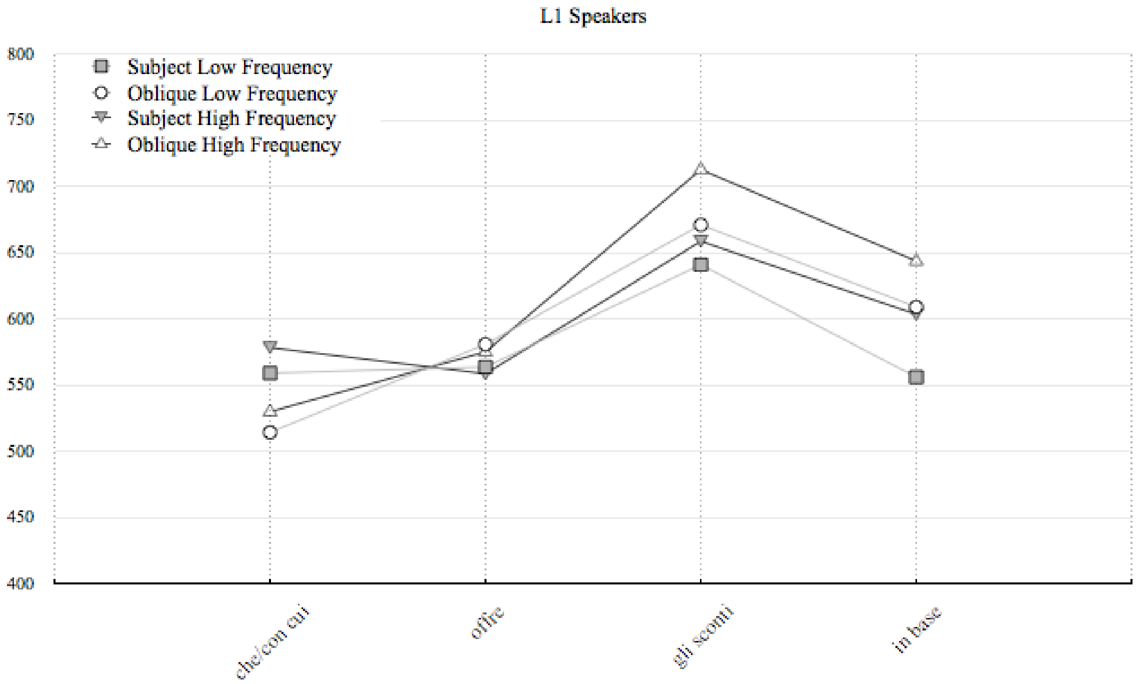
The three predictors did not interact: frequency and relative clause type (*ß* = 16.15, *SE* = 55.76, *t* = 0.29, *p* = 0.77), frequency and language (*ß* = 51.33, *SE* = 37.25, *t* = 1.27, *p* = 0.16), relative clause type and language (*ß* = -26.39, *SE* = 36.97, *t* = - 0.71, *p* = 0.47) and frequency, language and relative clause type (*ß* = -6.62, *SE* = 52.65, *t* = - 0.12, *p* = 0.9). No significant effects emerged in the post hoc analysis in both groups. The lack of effects is not totally surprising, since, at the post-complement region, the filler-gap dependency might already be resolved so that the interpretation of the relative clause occurs without further disruptions. The region did not capture any spill-over effects.

7.3.9 *Analysis by Group*

As the only two regions that showed a significant effect of language were the verb and the complement region, an analysis of each group was carried out separately in each of the two regions. Drawing on the robustness of mixed-effects models against skewed distributions (Schielzeth et al., 2020), the model for both the L1 and the L2 groups had reading times, relative type (subject vs. oblique) and frequency (low vs. high) as fixed effects, with subject and items as random intercepts. Again, in both models, random slopes could not be added because of a large eigenvalue ratio. Histograms showing the distributions of reading times and of the residuals of the two models are given in Appendix 8 (Figures 4 and 5). All reaction times that were greater than +/- 2SDs from the mean were excluded. Once both models were run, the model's residuals that were greater than +/- 2SDs from the mean were excluded. The final analysis was carried out on the models with trimmed residuals. Figure 7.2 and Figure 7.3 show the word-by-word graph of the two groups of participants in the four regions of interest. Given the lack of significance in the relative region and in the post-complement region, the subsequent analysis will focus on the verb and the complement region.

Figure 7.2

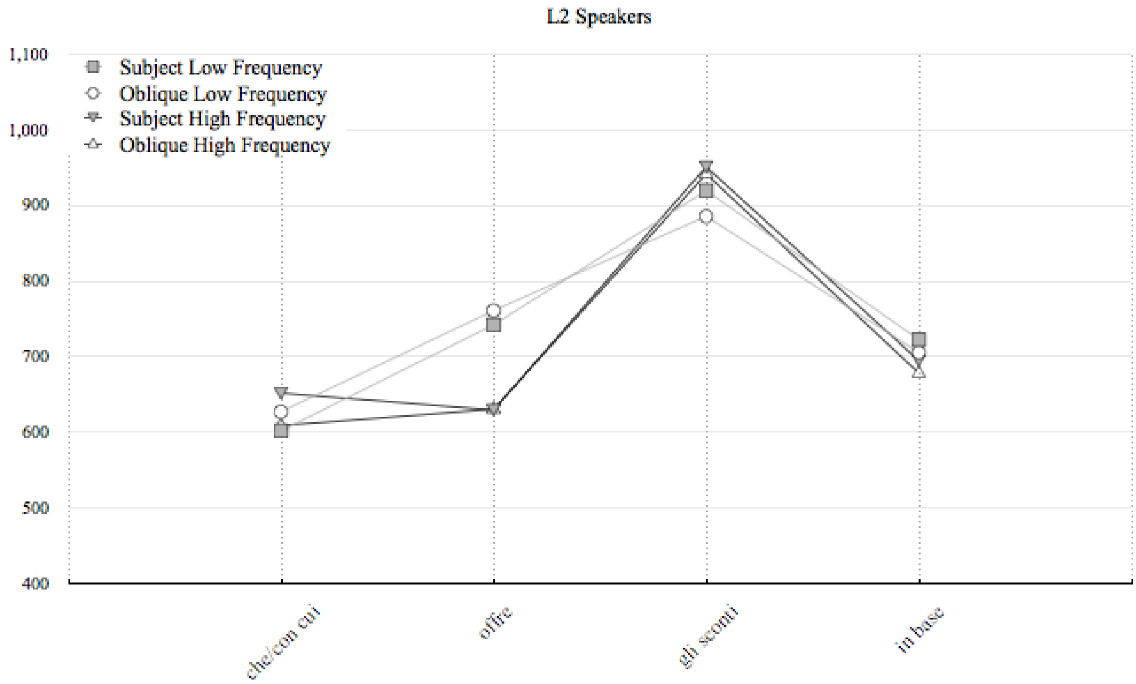
Reading Times of L1 Speakers in all the Four Regions



In the verb region, in the L1 speakers group, there was no significant effect of relative clause type (*ß* = -11.24, *SE* = 47.92, *t* = -0.23, *p* = 0.81) and frequency (*ß* = -5.35, *SE* = 47.63, *t* = -0.11, *p* = 0.91) and the interaction between relative clause type and frequency was also not significant (*ß* = 4.37, *SE* = 67.78, *t* = 0.06, *p* = 0.95). No significant effects emerged in the post hoc analysis. L2 speakers did not show a significant effect of relative clause type (*ß* = -27.61, *SE* = 34.72, *t* = -0.79, *p* = 0.44) but a significant effect of frequency (*ß* = -5.35, *SE* = 47.63, *t* = -0.11, *p* = 0.91). The interaction between the type and frequency of the relative clause was not significant (*ß* = 37.62, *SE* = 49.25, *t* = 0.76, *p* = 0.45). As shown in Table 7.11, a post hoc Tukey test showed that L2 speakers had slower reading times in oblique relative sentences with low-frequency verbs compared to those with high-frequency verbs (*ß* = -128.94, *SE* = 34.71, *t* = -3.71, *p* = 0.01).

Figure 7.3

Reading Times of L2 Speakers in all the Four Regions



There was a significant difference between subject relative clauses with low-frequency verbs and oblique relative clauses with high-frequency verbs (*ß* = 118.93, *SE* = 34.87, *t* = 3.41, *p* = 0.02) and between subject relative clause with high-frequency verb and oblique relative clause with low-frequency verbs (*ß* = -101.33, *SE* = 34.81, *t* = -2.91, *p* = 0.05). The results shown by the L2 speakers in the verb region appeared to be driven by a primary frequency effect, while frequency did not cause a slowdown in reading times in the L1 group.

Table 7.11

Table of the Post Hoc Analysis in the Verb Region: L2 Speakers

	High-Freq Subj	High-Freq Obl	Low-Freq Subj	Low-Freq Obl
High-Freq Subj	–			
High-Freq Obl	0.85	–		
Low-Freq Subj	0.09	0.02	–	
Low-Freq Obl	0.05	0.01	0.99	–

Note. Only p-values are shown.

In the complement region in the L1 speakers' group, there was a significant main effect of the relative clause type (*ß* = 89.31, *SE* = 35.23, *t* = 2.53, *p* = 0.02) but no significant effects of frequency (*ß* = 9.28, *SE* = 35.23, *t* = 0.26, *p* = 0.76) or a significant interaction between relative clause type and frequency (*ß* = -44.65, *SE* = 49.58, *t* = - 0.96, *p* = 0.35). Despite the main effect of relative clause type, no significant effects emerged in the post hoc analysis. L2 speakers did not show a significant effect on any of the predictors of interest. Relative clause type (*ß* = 24.62, *SE* = 86.11, *t* = -0.49, *p* = 0.61), frequency (*ß* = -42.64, *SE* = 86.67, *t* = -0.49, *p* = 0.78) and interaction between relative clause type and frequency (*ß* = -50.76, *SE* = 122.09, *t* = -0.41, *p* = 0.68) were not significant. No significant effects emerged in the post hoc analysis. The results in the complement region show that the L1 speakers had slower reading times with the more complex of the two relative types. Although no effect occurred in the L2 speakers, one must bear in mind that the overall model showed that in this region they were slower than the L1 group in both relative types, experiencing a wider structural effect. In general, L2 speakers were more sensitive to the effects of low-frequency verbs than L1 speakers only in the verb region. In contrast, frequency did not appear to be a problem for L1 speakers in either of the two regions, reflecting the findings of the EI study reported in Chapter 6. Interestingly, the expected effect of complexity for the oblique relative clauses arose only in L1 speakers who showed slower reading times in the complement region.

7.3.10 *Analysis of the Verb + Complement Region*

That significant effects showed only in the verb and the complement region raises the question whether the interaction between frequency and lexical access occurred across the entire complement-verb region in an incremental way. If building a relative clause entails the cognitive cost of attaching the filler to the proper gap, then such a cost might appear in the entire relative clause region. It would start at the verb and continue at the next region where the filler-gap dependency is resolved, and the adjunct filler placed after the complement.

To find whether such effect was present, the reading times at the verb and complement regions were combined and analysed in a single model, separately for each group (see Hopp, 2016 for a similar analysis). The model had reading times, relative type (subject vs. oblique), frequency (low vs. high), and region (verb region and complement region) as fixed effects and subject and items as random intercepts. The distributions for the two regions and the residual are given in Appendix 8 (Figures 4 and 5). Despite the skewness, the analysis is again based on the assumptions of Schielzeth et al. (2020) on the robustness of mixed-effect models with residuals not normally distributed. Before the analysis was performed, the residuals of the model that were greater than +/- 2SDs from the mean were excluded from the model. The data given in Table 7.12 along with the patterns in Figure 7.2 and Figure 7.3 show that L2 speakers were slower in the verb complement region and even slower in relative clauses with low-frequency verbs.

Table 7.12

Total Reading Times for the Verb + Complement Region

	Verb + Complement Region			
	L1		L2	
	High-frequency Mean (*SD*)	Low-Frequency Mean (*SD*)	High-frequency Mean (*SD*)	Low-Frequency Mean (*SD*)
Subject clauses	608 (271)	601 (258)	764 (335)	809 (361)
Oblique clauses	643 (280)	626 (268)	777 (338)	802 (326)

Looking at the results of the L1 speakers group (Figure 7.2) oblique relative clauses were read more slowly (*ß* = 95.20, *SE* = 45.94, *t* = 2.07, *p* = 0.05). The interaction between relative clause type and region was significant, showing that reading times for oblique relative clauses were faster in the verb region but slowed in the complement region (*ß* = -70.94, *SE* = 40.23, *t* = -1.76, *p* = 0.07). A Tukey post hoc analysis revealed that the interaction was driven by a significant slowdown in oblique relative clauses with high-frequency verbs in the complement region (*ß* = 110.21, *SE* = 28.30, *t* = 3.89, *p* = 0.003), while the difference between regions was not significant for oblique relative clauses with low-frequency verbs (*ß* = 63.30, *SE* = 28.96, *t* = 2.18, *p* = 0.36). The results suggest that L1 speakers experienced slowdowns when reading oblique relative clauses, but the effect did not interact with frequency, since it occurred with both high- and low-frequency verbs.

Taking into account the data from the L2 group (Figure 7.3) there were no significant effects for the relative clauses (*ß* = 32.43, *SE* = 45.99, *t* = 0.70, *p* = 0.84). There was a significant interaction between frequency and region, with slower reading times for relative clauses with low-frequency verbs in the verb region (*ß* = 137.09, *SE* = 53.31, *t* = 2.57, *p* =

0.01). A significant effect of region also appeared, in that the reading times in the verb region were significantly faster ($ß$ = -239.48, SE = 37.83, t = -6.33, $p < 0.005$). A Tukey post hoc analysis revealed that the effect was entirely caused by both types of relative clauses with high-frequency verbs (subject clause: $ß$ = 239.47, SE = 37.83, t = 6.32, $p < 0.005$; object clauses: $ß$ = 270.86, SE = 36.37, t = 7.44, $p < 0.005$). Although the relative clauses with low-frequency verbs showed a similar trend, the slowdown in the complement region was not significant (subject clause: $ß$ = 102.39, SE = 37.58, t = 2.72, p = 0.11; object clauses: $ß$ = 83.79, SE = 36.47, t = 2.29, p = 0.29).

7.3.11 *Vocabulary and Sentences Processing Analysis*

To answer the second research question about the interaction between vocabulary and syntactic processing, the I-Lex scores and the reading times in the SPR were analysed. Like in the experiment described in Chapter 6, the scores based on the I-Lex1000 frequency list were used because, as the results of the I-Lex study showed, it is better suited to capture differences in proficiency than the I-Lex2000 list. The I-Lex scores, both absolute and percentage scores, are given in Table 7.13 and Figure 7.4.

Table 7.13

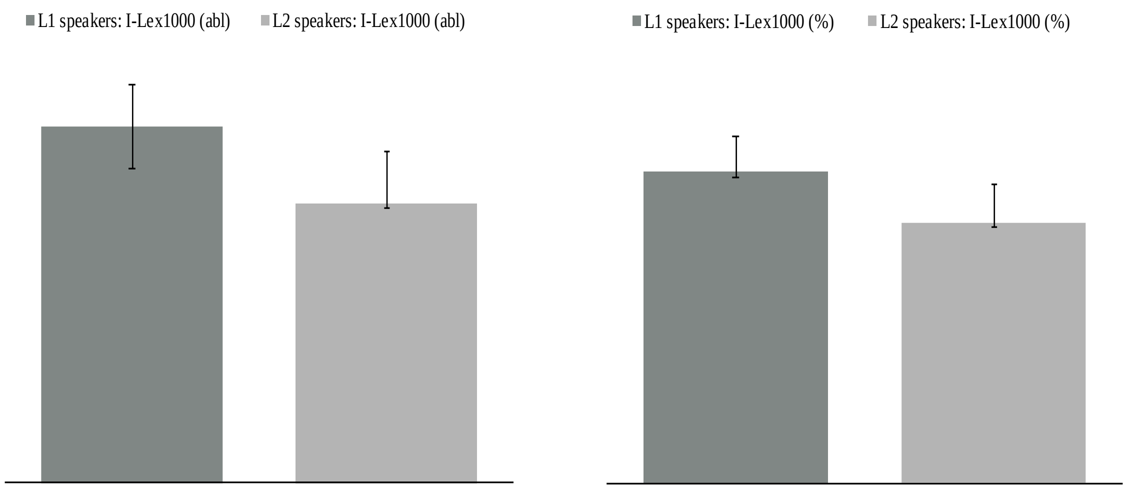
I-Lex Scores, Absolute (abl) and Percentage (%)

	I-Lex1000 (abl)	I-Lex1000 (%)
L1 Speakers (N = 12)	*M* = 75 (*SD* = 8)	*M* = 66 (*SD* = 7)
L2 Speakers (N = 20)	*M* = 60 (*SD* = 12)	*M* = 56 (*SD* = 8)

Given that the percentage and absolute scores are extremely close, the analysis was carried out only on absolute scores. The results show that the L1 speakers achieved higher scores than the L2 speakers. Before looking at statistical significance, a Shapiro-Wilk normality test was performed, showing that the scores of the two groups were normally distributed (Absolute scores, L1: $W = 0.87$, $p = 0.09$; L2: $W = 0.95$, $p = 0.48$; Percentage scores, L1: $W = 0.91$, $p = 27$; L2: $W = 0.92$, $p = 0.11$). The analysis, carried out with a Welch two sample t-test, showed that the difference between the L1 group ($M = 75$, $SD = 8$) and the L2 group ($M = 60$, $SD = 12$) was significant (t (20) = 4.38, $p < 0.0005$).

Figure 7.4

I-Lex Scores, Absolute (abl) and Percentage (%)



Note. Bars represent *SD*

Research question (2): to what extent will productive vocabulary knowledge will facilitate reading times?

The comparison between frequency, vocabulary knowledge measured with the I-Lex task, and relative type is based on two interrelated assumptions, one linked to the statistical model and one linked to the preparation of the data for the statistical analysis.

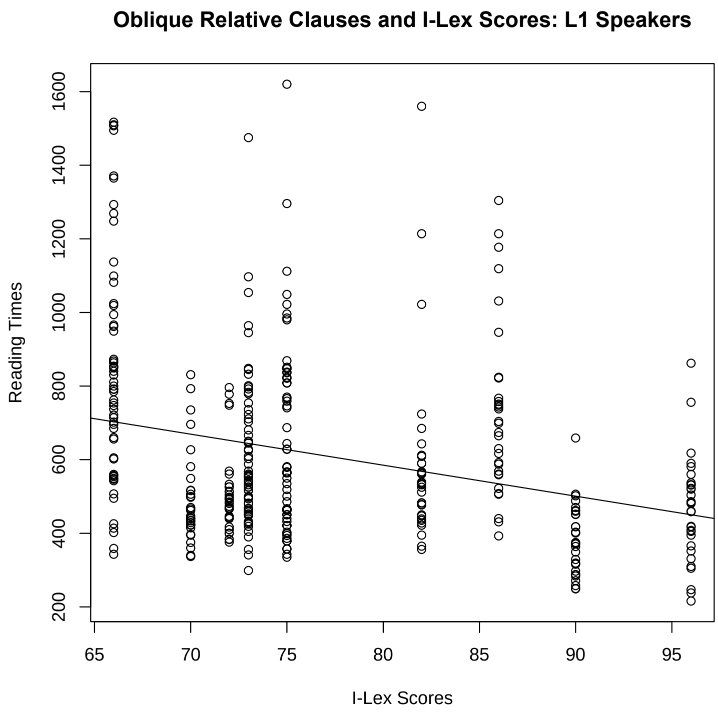
First, studies that have investigated the relationship between frequency and vocabulary knowledge have looked at the effect on the overall fit of the model when vocabulary scores were included (Brysbaert et al., 2017; Diependaele et al., 2013). The standard procedure consists of comparing a statistical model without I-Lex scores as a predictor with a model with I-Lex scores. The key point is to determine whether the model with the I-Lex scores can significantly explain more of the variance in the data (Cunnings, 2012). Based on these assumptions, two mixed-effect models with and without I-Lex scores were set separately for L1 and L2 speakers. The first model had I-Lex scores, relative type (subject vs. oblique), frequency (low vs. high) as fixed effects, and subject and items as random intercepts. The second model had the same fixed and random effects apart from the I-Lex scores. The distributions of the data and the residuals of the model are given in Appendix 8 (Figure 6). The analysis was based on the assumptions of Schielzeth et al. (2020) for robustness of the mixed-effects models with skewed distributions.

Second, for the vocabulary and sentence processing analysis, the SPR reading times were arranged differently from the previous analysis on frequency and syntactic structure. The reason is that oblique relative clauses require the processing system to integrate two distinct sources of lexical information. The first source of information is accessed in the segment where the relative pronoun and the coindexed adjunct NP are read. The second source of information is accessed in the verb and in the following segments when the filler-gap dependency is resolved (see Section 7.1). Therefore, the objective of the analysis was to capture how vocabulary knowledge affects the integration of lexical information from the onset to the end of the relative clause. Therefore, the reading times in the pronoun, verb, complement and post-complement region were averaged for each participant, and the I-Lex scores were compared to the reading times of the entire region (henceforth relative clause region).

Regarding L1 speakers, the model (Figure 7.5) with the I-Lex scores had a significantly better fit (x^2 = 10.49, *p* = 0.03).

Figure 7.5

Plot of Oblique Relative Clauses and I-Lex Scores: L1 Speakers

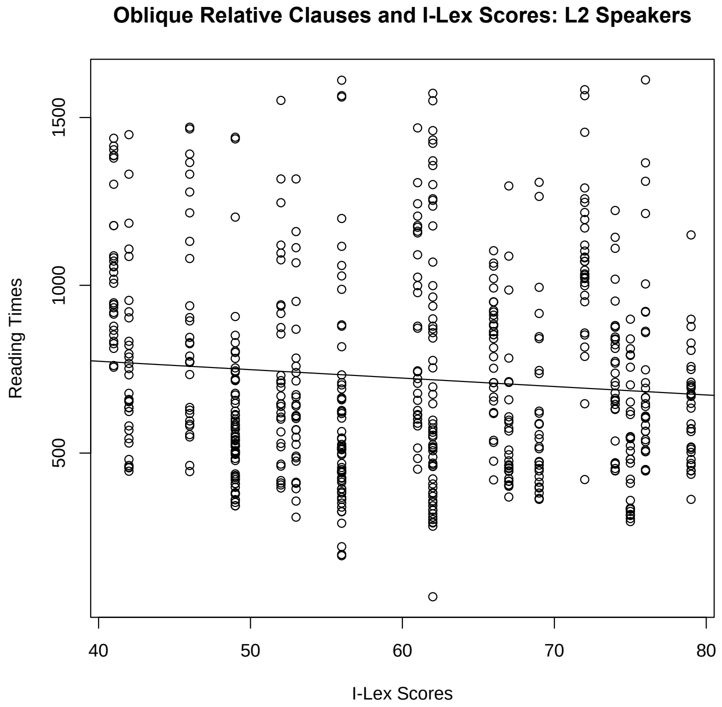


Vocabulary scores were not significant in predicting reading times in the relative clause region (*ß* = -4.91, *SE* = 3.21, *t* = -1.52, *p* = 0.15). The interaction between vocabulary scores and the type of relative clause was significant, producing a slight drop in the reading times of oblique relative clauses (*ß* = -3.39, *SE* = 1.44, *t* = - 2.34, *p* = 0.01). Adding vocabulary scores as a predictor in the model showed significantly slower reading times for oblique relative clauses (*ß* = 294.63, *SE* = 118.12, *t* = 2.49, *p* = 0.01). The general pattern for the L1 group suggests that vocabulary knowledge can reduce the strain on the processing system caused by more complex oblique relative sentences.

Turning to L2 speakers the model (Figure 7.6) with the I-Lex scores did not show a significantly better fit (x^2 = 6.63, p = 0.15). Although vocabulary scores did not improve the model, an analysis was carried out on the model to see what the underlying pattern was. The results mirrored the pattern of the L1 group.

Figure 7.6

Plot of Oblique Relative Clauses and I-Lex Scores: L2 Speakers



I-Lex scores were not a significant predictor of reading times in the relative clause region (*ß* = 1.32, *SE* = 5.23, *t* = 0.25, *p* = 0.8). The interaction between vocabulary scores and the type of relative clause was significant (*ß* = -2.83, *SE* = 1.42, *t* = -1.99, *p* = 0.04). Adding vocabulary scores as predictors in the model showed an almost significant trend towards slower reading times for oblique relative clauses (*ß* = 165.47, *SE* = 92.39, *t* = 1.79, *p* = 0.07). The interaction suggests that, like in the L1 groups, vocabulary knowledge produced a slight advantage in the reading times of oblique relative clauses. The results for the L2 group suggest that vocabulary knowledge can facilitate the lexical integration that occurs in oblique relative sentences. However, the model on which the analysis was carried out was not the model with the best fit, and the results need to be taken with caution. As noted by Cunnings (2012), a model with a lower fit explains less variance in the data.

7.3.12 *Discussion*

Looking at the by-region analysis in the overall model, the L2 speakers showed frequency effects only in the verb region. When low-frequency verbs hindered lexical access, reading times in the verb region were slow. However, the analysis carried out in the verb+complement region showed that the reading times of the L2 group increased in the complement region with high-frequency verbs. This effect shows a pattern similar to that found for the L2 group in Hopp's (2016) study. In the region that included the cleft verb, he only found a significant effect of frequency, while in the region after the verb, he found a significant effect of syntactic structure but no effect of frequency. Although the results of the present study replicated the presence of frequency effects limited to the verb region, no structure effects appeared in the following region. In the post hoc analysis Hopp (2016) showed a significant effect of syntactic structure for low frequency effects, which in the present study did not appear. In fact, the only effect appeared in relative clauses with high-frequency verbs. The effect was not driven by syntactic structure or, as expected by L2 speakers, by low-frequency verbs, but was caused by a general difficulty in integrating the filler. The results of the L2 speakers align only partially to the objective of extending Hopp's (2016) results to another syntactic structure.

The analysis of the L1 group reading times was different. In the verb region, the L1 group did not show any significant effect of frequency. In the complement region, the frequency was not significant, but a significant effect of syntactic structure emerged, with slower reading times for oblique relative sentences. The lack of frequency effects for L1 speakers in both the verb and complement regions closely resembles the pattern found in the EI study. In contrast to the results of the EI study, the L1 group showed a main effect of relative clause type in the complement region, where reading times for oblique relative clauses were slower. The results of the L1 speakers are in line with the hypothesis made in the first research question, wherein the oblique relative clauses increase processing cost because of the greater distance between the filler and the gap (Gibson, 2000; Gordon et al., 2002; Parker et al., 2017). The lack of any effect of interaction between frequency and relative type in every region suggests that L1 speaker reading times were exclusively affected by the type of syntactic structure. As in the EI study described in Chapter 6, L1 speakers could easily access the lexical features of high- and low-frequency verbs. Therefore, the only effect was caused by the integration cost of integrating the filler with the gap in the complement region.

This result resonates with the findings of the EI study, where L1 speakers did not show frequency effects and produced fewer correct object clefts. More importantly, the results show a pattern that resembles that found by Hopp (2016). In his study, L1 speakers, both in the region with the cleft verb and in the post-cleft region, showed a significant effect of structure but no effect of frequency and no interaction between the two. While in this study no effect of structure appeared in the relative verb, the same lack of interaction between frequency and structure appeared in the region after the verb. Therefore, in the present study, although frequency effects were absent, the L1 group showed a structural pattern similar to that found in Hopp (2016) with a different syntactic structure. However, Hopp's (2016) found also an effect of low-frequency verbs on syntactic structure in the post-verb region that is absent in the present study (see also Tily et al. (2010) for a similar effect).

One explanation for the differences between the present study and Hopp's (2016) might be due to the fact that oblique relative clauses, unlike object clefts, require a process of lexical integration (see Introduction) that might have overshadowed frequency effects in both L1 and L2 speakers. Overall, the similarity between the present study and Hopp's (2016) one shows that the SPR is better suited to capture fine-grained effects. As it allows for more precision in singling out the effects as the processing system moves across the sentence, it can reveal a more nuanced pattern than the EI task used in the previous experiment.

The second research question aimed to find the effects of vocabulary knowledge, measured with the I-Lex task, on the processing of oblique relative clauses. The hypothesis draws on the assumption that, with oblique relative clauses, the processing system needs to integrate the lexical information encoded in the relative verb with that encoded in the adjunct. In subject relative clauses, the processing system needs to integrate only the lexical information encoded in the relative verb. Overall, L2 speakers showed a trend towards the expected interaction between vocabulary knowledge and structural effects. When their vocabulary knowledge was controlled for, their reading times were slower with oblique relative clauses that required them to integrate lexical information encoded in the verb and in the adjunct. However, lexical knowledge helped the processing system handle all the lexical features that needed to be deployed to interpret oblique relative clauses. The analysis of the L1 group showed the same pattern. When their vocabulary knowledge was controlled, vocabulary scores significantly reduced reading times for oblique relative clauses. Like L2 speakers, vocabulary knowledge was crucial for L1 speakers to integrate lexical information encoded in the verb and in the adjunct.

The results match the predictions made in the second research question. With oblique relative sentences, the processing system must rely on a broad and well-organized lexical network to map meaning into syntactic relations. In oblique relative sentences, the adjunct filler is encountered before the verb. The lexical information encoded in it can be modelled as the lexical network made up of the potential verbs that can be modified by the adjunct. When the processing system moves from it to the verb region, the network must be updated, adding the links between the verb and all its potential complements. This process of lexical integration is easier when supported by vocabulary knowledge. Higher I-Lex scores mean that the processing system can rely on an enhanced network with more links and more words.

One caveat is necessary to interpret these results. In the case of L2 speakers, the model that had I-Lex scores as predictor was the model with the worst fit that explained less variance in the data. Therefore, while the link between vocabulary knowledge and the processing of oblique relative sentences holds for L1 speakers, it must be regarded more as a trend in the case of L2 speakers.

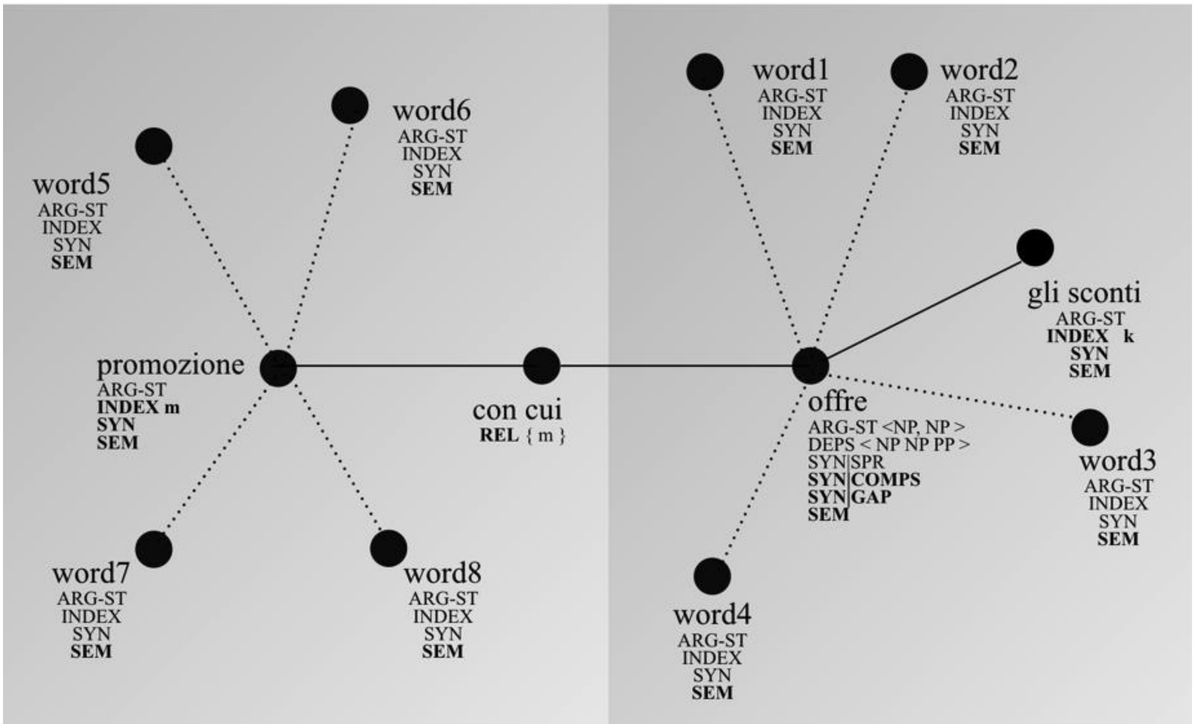
The last part of the section is devoted to weighing the precision of the feature-based lexical model in explaining the relationship between vocabulary knowledge and syntactic processing. I have already shown in Chapter 6, modelling the EI task results, how low-frequency verbs reduce feature availability, preventing the processing system from activating syntactic links. In this section, I will focus on how to represent the integration process that occurs when the processing system interprets oblique relative sentences. Regarding oblique relative sentences, the small-scale lexical network in Figure 7.7 shows the two sections of the lexical network, for Example **7.9**, which contain the modified adjunct and the oblique relative clause.

7.9	Una promozione	con cui	offre	gli sconti
	A promotion	with which	offers	the reductions

The first section of the lexical network, based on the information encoded in the modified adjunct *promozione* (*promotion*), is shown on the left-hand side of Figure 7.7. The second, based on the information encoded in the relative verb *offre* (*offers*), is shown on the right-hand side. The representation is extremely simplified. In a real lexical, network, links would be more numerous. Therefore, the difference between the sections related to the adjunct and the verb would be captured by the different values of the measurement that I have described in Chapter 2. They would be, for instance, the number of links, the average distance between the words in the network, and the clustering coefficient. However, the simple visualisation in Figure 7.7 helps to understand how the processing system integrates two distinct sources of lexical information. The light grey colour is used as the background of the adjunct lexical networks, and the darker grey background for the subcategorizations frame lexical networks.

Figure 7.7

Feature-based Network for an Oblique Relative Clause



To keep the model simple to interpret, all links are considered meaning-based links, as the bold feature SEM show. The labels word5, word6, word7, and word8 are hypothetical words related to the meaning encoded into the modified NP *promozione* (*promotion*). On the contrary, the labels word1, word2, word3, and word4 are hypothetical words related to the meaning encoded in the verb *offre* (*offers*). The activation of the links within the network is feature-driven in the same way as shown for the clefts in the EI task. When the processing system arrives at the wh-phrase, the link with the modified NP is activated, and, as a result, the feature REL and the feature INDEX are co-indexed. At this point, the processing system preactivates a section of the lexical network that encodes all the links between the NP and the words with which it can combine. In the verb region, the processing system activates the link between the wh-phrase and the verb and the section of the lexical network encoded in the subcategorization frame of the relative verb. At this stage, the processing system needs to integrate the lexical information that lies in the two sections of the lexical networks. High vocabulary knowledge can support and speed up the process, while low vocabulary knowledge can slow it down.

7.3.13 *Limitations*

There are several limitations to this study. Limitations concerning working memory, the number of participants, and the number of items overlap entirely with what I have observed for the EI task in Chapter 6. However, unlike the results of the previous experiment, in this study the main pattern found for L2 and L1 speakers was more similar to that found in (Hopp, 2016) using a comparable syntactic structure. Although this does not exclude, *per se*, that the variance was related to the small number of items, it gives support to the claim that an effect really emerged from the results. As already noted in Chapter 6, the use of mixed-effects models should be able to control for the small number and how properties of individual items affect the variance in the statistical analysis.

Looking at the results of the SPR tasks, although they manage to replicate the overall pattern of (Hopp, 2016) the picture that emerges is somewhat blurred. Frequency effects did not emerge in L1 speakers, and the result of the study could not answer to the question raised at the end of Chapter 6. A more precise measurement task than the EI did not capture frequency effects in the L1 speakers at all. This might again be because the frequency of the low-frequency relative verbs was not low enough. A different explanation could be that the type of relative used in the study was able to capture vocabulary knowledge but was not complex enough to reflect frequency effects. In effect, even the L2 speakers had frequency effects limited to the verb region which disappeared in the complement region. The main factor that the analysis of the experiment revealed was the effect of lexical integration in both groups.

As described in the previous EI study, the role of the L1 could not be investigated because the L2 group was extremely varied and without a numerically dominant group that could be used as a reference. Half of the L2 participants in the experiment spoke a different language. Furthermore, as revealed in the questionnaires, all participants were also advanced English learners. Since relative clauses and cleft sentences have the same structural properties in Italian and English, it would be hardly feasible to separate the effects of the L1s and the L2 English.

7.3.14 *Conclusions*

The study was designed to investigate the relationship between vocabulary knowledge, frequency, and the ability to process two types of filler-gap dependency. The interaction between frequency and vocabulary knowledge did not arise in the statistical analysis. However, frequency affected the reading times of L2 speakers in the relative verb region, showing their sensitivity to low-frequency words. The interaction between syntactic structure and vocabulary knowledge measured through the I-Lex task is an important result in that it shows, for the first time, a clear facilitation effect of lexical knowledge on the ability to process a complex syntactic structure. As in Chapter 6, the feature-based lexical network model was used to model the effects of the study, that is, the interaction between vocabulary knowledge and reading times with oblique relative sentences. The model was shown to be flexible enough, but there are still technical aspects about the way it works that need to be formalised more precisely. This is one of the objectives that I will pursue through the discussion in Chapter 8.

Chapter 8 Discussion

In this chapter, I will describe the key features of a unified model of lexical and sentence processing, the feature-based lexical network model. The theoretical foundations are based on the feature-based HPSG grammar described in Chapter 4 and the assumptions outlined in Chapter 5 about the relationship between vocabulary knowledge and syntactic processing. The functions and the properties of the model are based on the experimental results in Chapter 3, Chapter 6, and Chapter 7, where the feature-based lexical network model has been used to explain the results of the two experiments on syntactic processing. I will then evaluate the limits of the model and the implications for future research in second language acquisition.

8.1 Frequency Effects and Vocabulary Knowledge in Language Processing

Both research on lexical and sentence processing have found effects of frequency on language processing. In lexical processing research, effects of frequency have arisen in various tasks based on semantic or word-form features or an interaction of both, like lexical decision tasks or semantic classification tasks (e.g., Brysbaert et al., 2017, 2018; Diependaele et al., 2013; Duyck et al., 2008; Gollan et al., 2008, 2011). In sentence processing research, the constraint-based models were the first to recognise that frequency impacts lexical and syntactic representations to the same extent (MacDonald et al., 1994; McRae & Matsuki, 2013). Usage-based models have proceeded on a similar path finding effects of frequency in the way the processing system uses information encoded in the verb subcategorization frame (e.g., Ellis et al., 2014; Ellis & Ferreira-Junior, 2009).

Furthermore, in lexical research, several studies have provided evidence that frequency and vocabulary knowledge interact (Brysbaert et al., 2017, 2018; Cop et al., 2015; Diependaele et al., 2013; Lemhöfer et al., 2008). However, this interaction has not been thoroughly framed in language processing models that included frequency as a key determinant. Usage-based approaches have shown how, in the learning process, the frequency of a verb in a construction interacts with semantic factors that make some exemplars of the construction easier to learn (Ellis, 2011). For example, the verb object-object construction *give the kid a new toy* or the verb object-location construction *put the toy on the table* are learned via the most prototypical verbs *give* and *put* (Ellis, 2011; Ellis et al., 2014). However, prototypicality has not been linked to vocabulary knowledge. It refers to the fit of an individual verb into the network based on semantic similarity between all the verbs that are used in a specific construction. Prototypicality has been regarded more as a general property of constructions, whereby "one particular verb accounts for the lion's share of instances of each particular argument frame; this pathbreaking verb also is the one with the prototypical meaning from which the construction is derived" (N. Ellis, 2011, p.199). Constraint-based models have incorporated the frequency of semantic, syntactic, and morphological features in their representations (MacDonald et al., 1994; McRae et al., 1998; McRae & Matsuki, 2013). However, they have not fully worked out the potential link between the frequency of lexically encoded features and vocabulary knowledge.

To my knowledge, only three studies have examined the relationship between vocabulary knowledge and syntactic processing, without using online processing tasks (David et al., 2009; Kawaguchi, 2016; Rogers et al., 2018). They used a receptive vocabulary measurement task and found a correlation between vocabulary size and the ability to produce correct target syntactic structures. Studies that have focused on frequency effects on sentence processing have interpreted frequency as a measure of lexical access, but have not examined whether vocabulary knowledge and frequency interacted (Hopp, 2016; Tily et al., 2010).

As explained in Chapter 2, recent work on lexical processing, based on online lexical tasks such as word naming or lexical decision, has clearly shown that vocabulary knowledge interacts with frequency (Brysbaert et al., 2017; Cop et al., 2015; Diependaele et al., 2013). Even when lexical representations are examined in terms of properties of lexical networks, frequency effects have arisen. For example, De Deyne et al. (2013), De Deyne & Storms (2008), Steyvers & Tenenbaum (2005), and De Deyne et al. (2019) examined the correlations between frequency and the number of links and other various measures of density in lexical networks, which reflect lexical relations within the networks. The correlations were significant and relatively high. Importantly, the statistical analysis showed that the two measures were intertwined, although they captured a separate portion of lexical information (De Deyne et al., 2019). Thus, considering the frequency-vocabulary interaction as a major factor in lexical processing has well-grounded theoretical and data-driven motivations.

The present research is one of the first to integrate frequency and productive vocabulary knowledge as predictors of sentence processing in two online tasks, an oral production and an online reading task. The aim is to describe the interaction between vocabulary knowledge and frequency with a unified model of sentence processing that uses the same set of tools to describe lexical and structural effects. Therefore, I have proposed a model (see Chapter 5) to describe the relationship between frequency and vocabulary knowledge in terms of feature availability. The key assumption of the model is that frequency may be used as a reliable tool to gauge the availability of lexical features. As the frequency decreases, lexical features tend to become less available to the processing system. Frequency is the main factor in reducing the availability of lexical features, but it also interacts with other traits of L2 speakers, and it can potentially be linked to other variables such as non-selectivity and underspecified lexical representations.

The first piece of evidence that frequency is a reliable measure of lexical knowledge came from the results of I-Lex, the new productive vocabulary measure of Italian based on a word association format. In the study presented in Chapter 3, I-Lex showed that there is a significant relationship between proficiency and vocabulary scores, measured in terms of the L2 speaker's ability to provide infrequent words. The results showed that it is harder for L2 speakers to produce infrequent words compared to L1 speakers because their lexical networks are not as organised and have fewer links (Meara, 2006, 2009). As a result, the L1 speaker group outscored the L2 advanced group which, in turn, outscored the L2 pre-intermediate group. Furthermore, when the lexical networks drawn from the responses given by each group were compared, the difference in the scores between the L1 and L2 speakers was still significant. These results tie in with the interaction between frequency and vocabulary knowledge described by the Lexical Entrenchment Hypothesis, whereby less entrenched lexical representations lead to larger frequency effects (Diependaele et al., 2013; see also Gollan et al., 2008). The I-Lex results show that the words that are the hardest for the L2 speakers to add to the lexical network (De Deyne & Storms, 2008a; Steyvers & Tenenbaum, 2005) are infrequent words. As a result, L2 speakers are more sensitive to the effects of low-frequency words in online processing tasks.

To empirically test this assumption, I used both frequency and vocabulary in two sentence-processing studies showing a relationship between the two factors in both studies. In the EI study presented in Chapter 6, the L2 group showed frequency effects and a significant interaction between vocabulary knowledge and the number of cleft sentences with low-frequency verbs correctly repeated. In the SPR study presented in Chapter 7, frequency effects appeared again in the L2 group. However, the SPR experiment also showed that lexical knowledge of the L2 group helps the processing system handle the more complex syntactic structure of the object-modifying oblique relative clauses. Crucially, in the same study, the pattern matched that of the L1 speakers. The effects of frequency and vocabulary knowledge were additive. Frequency effects appeared in the L2 group in one region only, while vocabulary knowledge effects emerged in both groups and affected the reading of the entire object-modifying oblique relative clause. These results are the first to show that frequency and vocabulary knowledge can facilitate or slow down syntactic processing in L2 Italian.

The results of the two experiments provided support for the theoretical assumptions outlined in Chapter 4 and Chapter 5. Lexical and syntactic processing can be regarded as one process fed by the same feature-based information that can be described by a unified model. This is the focus of the next section.

8.2 Feature-Based Networks

8.2.1 *Introduction*

One major factor that the present project takes on is the creation of a simple formalism that can unify lexical and syntactic representations at all levels. As shown in Chapter 4, HPSG incorporates the information on which constraint-based models have been working, adding a formal set of principles and rules (Pollard & Sag, 1994; Sag et al., 2003). Unlike constraint-based models, HPSG rules and feature descriptions can be generalised to all sorts of syntactic and lexical phenomena (Sag & Wasow, 2011, 2014; Wasow, 2019). Thanks to the hierarchical organisation of the lexicon and the on-line type construction mechanism (Davis & Koenig, 2019; Koenig & Jurafsky, 1995) lexical and syntactic representations can be part of a unified theoretical approach.

Having the same type of feature-based description for every level of representations offers a theoretical framework for describing lexical relations as the main component of the processing system. I have described some general traits of the feature-based lexical network model in Chapter 5 and then applied the model to the results of Chapter 6 and Chapter 7. However, the precise mechanisms that underlie how the feature-based lexical network works have not been fully formalised. In this section, I will fully describe how such a network can work using the HPSG formalism along with some basic principles that underpin the architecture of lexical networks. This is the first model based on HPSG that has been applied to describe, in a precise theoretical framework, the results of two experiments carried out on L2 speakers of Italian.

The architecture of a feature-based lexical network encompasses two distinct processes that interact but work at a different level of granularity. The first is the process of activating links between words and is dependent on the organisation of the lexical network, namely, the number of links and the density. The second process is a lexical process that describes how lexical features are accessed. This section describes how the activation process works. The following section describes in which way the model captures how lexical features are accessed depending on their availability.

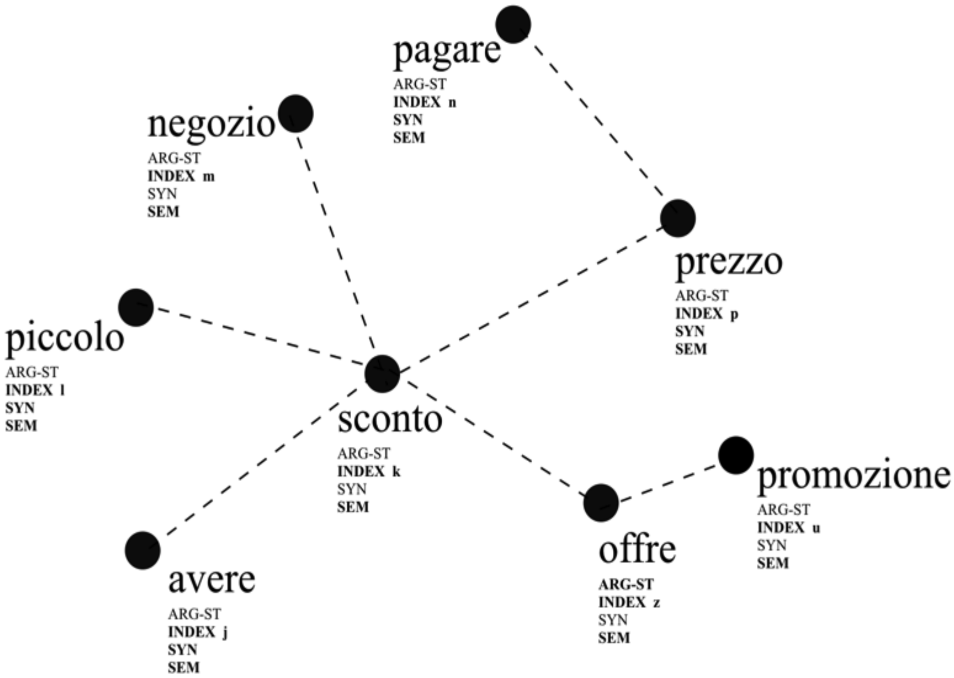
8.2.2 *Type of Links within the Lexical Network*

An effective setup for a lexical network is to have two simple conditions where links can be either activated or inactivated (Meara, 2006). The activation process occurs when two or more words are connected to build or interpret a sentence or any other type of meaningful lexical sequence for communicative purposes. The number and setup of the inactivated links are determined by the vocabulary size of individuals: the full set of inactivated links corresponds to the structure of the mental lexicon. In short, a lexical network has a certain setup of links that can be activated to form a specific linguistic structure and express meaning.

Consider the Italian word *sconto* (*sale/reduction*), which was used in one of the SPR items (see Appendices 6 and 7). In a lexical network, the word might have links to many other words. The resulting network is shown in Figure 8.1.

Figure 8.1

An Example of Lexical Network



Some connections might be with words having a similar meaning like *negozio* (*shop*) or *prezzo* (*price*), or with words that have more specific meaning relations based on one individual experience with language. Collocational or structural patterns might also determine some connections: in Italian, one collocational link is with the verb *avere* (*to have*) in the collocation *avere lo sconto* (*to have a sale/reduction*). A similar pattern with less collocational strength might be with, for example, verbs or adjectives that tend to cooccur with the word, like *offrire* (*offer*) or *piccolo* (*small*). Of course, neighbour words might also be connected to one another without having a connection with the word *sconto* (*sale/reduction*) themselves. So, the verb *offrire* (*offer*) might have a link with the noun *promozione* (promotion), and the word *prezzo* (*price*) might be linked to *pagare* (*pay*).

The model shows that more than one feature can underlie a link between two words. Meaning links between words are described by the feature SEM, while structural links are described by the feature SYN. The features that determine the links are in bold. Notice that every word has its own value for the feature INDEX as it refers to a different discourse referent. The feature INDEX is in bold because meaning links might be determined by relations based on sharing the same discourse context (Halliday, 2004). In HPSG this amounts to saying that some words in the network may be elements of the same semantic frame and refer to a shared set of discourse referents (Baker et al., 2003; Fillmore & Baker, 2009). In fact the words in the network share the meaning information associated to the frames Commercial_transaction, Commerce_pay and Giving (Baker et al., 2003).

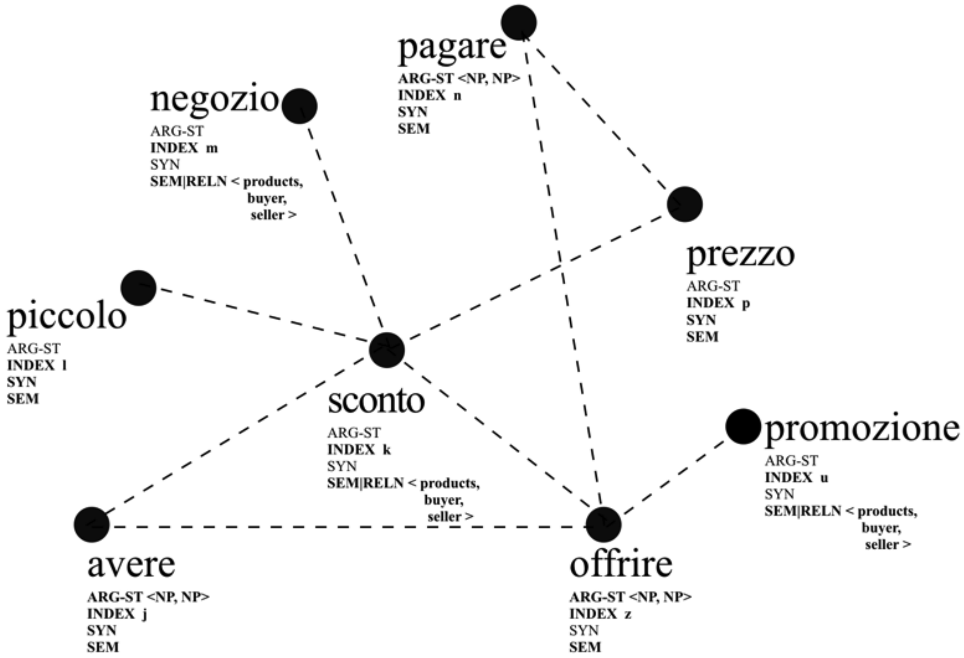
Such an approach might be a potential problem in setting up a reliable model in that the processing system can use multiple features at the same time, drawing on a complex and multilayered system. However, the proliferation of features is not a problem if it can be constrained with the mechanism of lexical hierarchies adopted by HPSG (Davis & Koenig, 2019; Sag et al., 2003). In lexical hierarchies, most of the features are encoded in lexical entries by default. As a result, lexical entries of the same type share the same features, and many features are shared by several types of lexical entries (Davis & Koenig, 2019). Therefore, with a hierarchical organization of the lexicon, there is no need to list all the features that determine a link. For example, in one of the hierarchies shown in Chapter 4 (Section 4.9), all transitive verbs inherit, by default, the feature ARG-ST with two elements in its list, with the subject and object, ARG-ST < NP, NP >. So, all transitive verbs encode, by default, the information to form a combinatorial link with another NP.

There is one further aspect of feature-based lexical networks that needs to be analysed, as it is pivotal in forming meaning links between words. In Chapter 4, I have briefly discussed that the meaning of each lexical entry is encoded in the feature RESTR and that its value is a type of feature structure called *predication*. While the feature RESTR and the notion of predication can easily describe verbs and their meaning links, defining a set of restrictions for the lexical meaning of all the words in the networks needs a specific meaning feature. The feature RELN described in Chapter 4 (Section 4.4) encodes the type of meaning relations that a lexical entry encodes and can be described, in turn, as a set of features (Koenig & Davis, 2006). As it is embedded in the feature structure *predication*, it can do the work in a straightforward way.

The feature RELN can be deployed to work out the links between nouns within the network, like that between the word *sconto* (*sale/reduction*) and *negozio* (*shop*) in that the type of meaning relations that the two words encode have a lot in common. Like any other type of lexical relation, the feature RELN can be modelled by using semantic hierarchies where, like in syntactic hierarchies, many meaning features are simply inherited by default. Hence, the feature RELN may list shared meaning features that the two nouns like *sconto* (*sale/reduction*) and *negozio* (*shop*) inherit by default. Using an approach based on Frame Semantics (Baker et al., 1998, 2003), the feature representations might be something along the lines of RELN < products, buyer, seller >, and could be easily extended to word *prezzo* (*price*) as well. An updated network with the syntactic and semantic information encoded in the hierarchical relations is shown in Figure 8.2.

Figure 8.2

An Example of Lexical Network with Inherited Features RELN and ARG-ST



Although the example shown for nouns can be easily extended to verbs as well, the meaning links between the three verbs in the lexical network of Figure 8.2 are less likely to rely on shared features. For example, the verb *offrire* (*offer*) and the verb *avere* (*have*) have two different values for the feature RELN. The meaning of the first verb *offrire* (*offer*) might be described as RELN < present > while the meaning of the second verb *avere* (*have*) is described by RELN < possess >. It is more likely that the default inheritance mechanism might give rise to a combinatorial link between the three transitive verbs in the network *offrire* (*offer*), *avere* (have) and *pagare* (*pay*), as they inherit, as explained before, the feature structure ARG-ST < NP, NP >.

Adding feature-based representations to the lexical entries in lexical networks allows one to have more precision in describing word meaning and, therefore, the type of links that exist between words. In addition, having lexical entries arranged in a hierarchy ensures that a great deal of lexical features is encoded by default and shared between words that have equivalent properties. Two verbs might be linked if their feature ARG-ST inherits the same default value or if their feature RESTR inherits some similar meaning relations. Two nouns can be linked if their feature RELN shares some meaning properties, if their features PHON or AGR share a part of their phonological form, etc.

8.2.3 *Activating Links within the Lexical Network*

The feature-based network in Figure 8.2 raises the question of how to incorporate phrasal constructions into it to differentiate existing links from links that are activated to form a phrasal construction, such as the *Head Filler Construction.* Consider again the target items used in the SPR study:

8.1 Il negozio ha iniziato una promozione
The shop has started a promotion
The shop$_j$ has started a promotion$_k$

Subject:

8.1a) che		offre	gli sconti	in base	agli acquisti
which	PRO$_k$	offers	the reductions	based on	the purchases

which offers reductions based on the purchased

Oblique:

8.1b) con cui	offre	gli sconti	in base	agli acquisti
with which	offers	the reductions	based on	the purchases

through which it$_j$ offers reductions based on the purchased

In the SPR study, I explained how a feature-based network describes the activation of the links between *sconto*, *offrire,* and *promozione* to build the relative clause *con cui offre uno sconto* and link it to the modified NP *una promozione* to form the *Head Filler*

Construction. However, the example was based on a small-scale network and did not examine in detail the role of key factors related to networks such as clustering, density, and the presence of hubs.

A potential problem is that syntactic relations are chiefly local or can be broken down into smaller segments based on local relations (Sag et al., 2003). However, not all words with local connections in a sentence have local connections in the lexical network. The properties of the network are crucial here: with a low average path length, a high clustering, and the presence of hubs, all lexical networks guarantee a high degree of communication between all the words (Solé et al., 2010). Although not all links within the network might connect a verb directly to one of its complements or a noun to an adjectival modifier, the distance between the two, in terms of links, would be minimal.

For example, the word association network examined by De Deyne & Storms (2008) had a clustering of 0.6 and an average path length of 3.2 while the syntactic networks examined by Cancho & Solé (2001) had an average path length of 2.6 and a clustering between 0.4 and 0.6. Thus, setting up a syntactic relation between two words needs the activation of only about 3 links. Furthermore, the high clustering guarantees that two words can be linked through more than one path. Words that are in a syntactic relation, like being the complement or the subject of a verb, even when not directly connected in the network, are remarkably close and clustered. As a result, activating a syntactic link will not put any additional strain on the processing system.

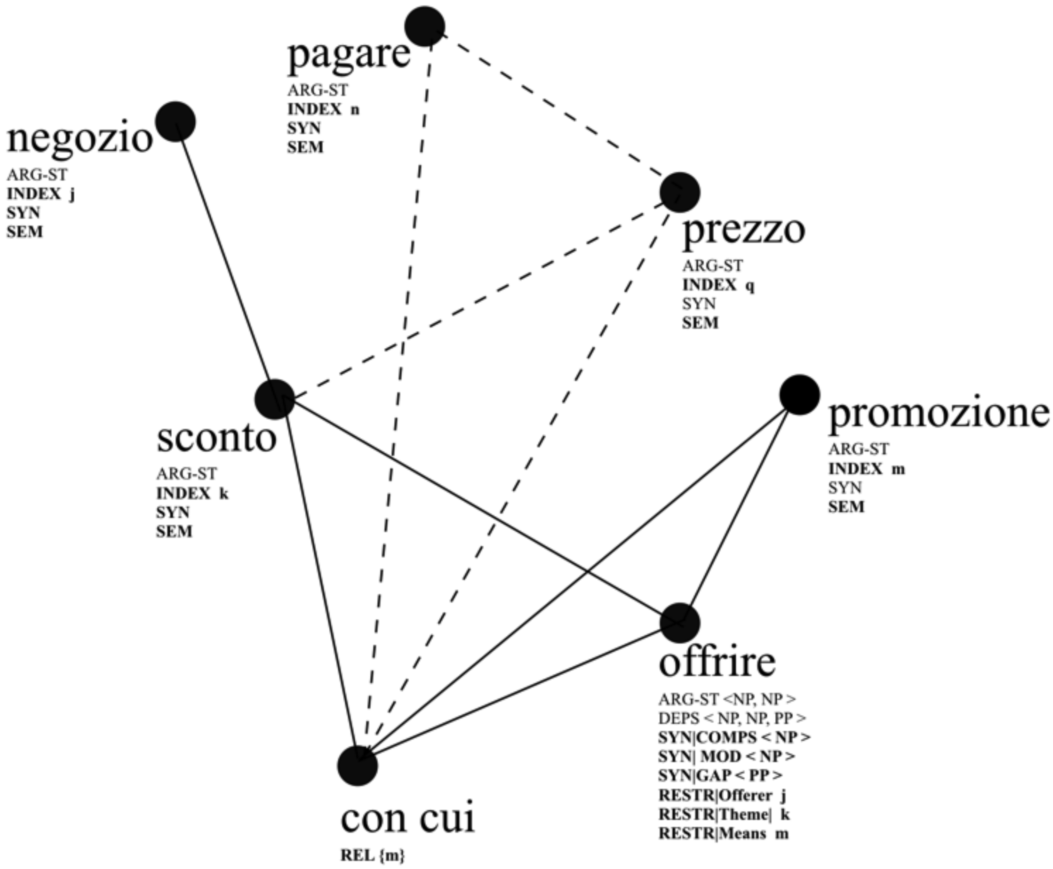
More importantly, networks do not activate one link at the time but make use of spreading activation (e.g., Collins & Loftus, 1975; De Deyne & Storms, 2015; Hudson, 2010). When a word in the network is activated, the neighbour words connected with it are also activated, so that searching and retrieving other words is facilitated (De Deyne & Storms, 2015). Every word is connected to another word via meaning, form, and syntactic features, so activation spreads throughout the network based on the information encoded into lexical features (De Deyne & Storms, 2015; de Groot & van Hell, 1998b; Hudson, 2010). Driven by similarity between the type of information the array of activated lexical features converges, at the end, on the intended word and the link is activated (Hudson, 2010). Spreading activation is also a key factor in bilingual lexical processing models: when a word is activated, its similarity with the input word, the activation spreads based on the similarity of phonological and semantic representations (Dijkstra et al., 2019; Dijkstra & van Heuven, 2002).

In the case of syntactic features, spreading activation is somehow simplified by the fact that, in HPSG, a limited set of features underlie the activation of syntactic links, the features SPR, COMPS, MOD, and GAP. Crucially, in line with the way spreading activation works, syntactic links are not activated in isolation in HPSG. In fact, its constraint-based setup assumes that syntactic features are mapped into the semantic information encoded in the feature RESTR via the feature ARG-ST. Finally, many of the functions words that are used to form syntactic structures happen to be hubs (Solé et al., 2010). This is true for prepositions, auxiliary forms, and other function words. Therefore, relative pronouns are also hubs in lexical networks. As explained in Chapter 2 (Section 2.3), hubs are words that have a very large number of links with almost all the other words in the network (Cancho, 2005; Solé et al., 2010). Thus, in the relative clause employed in the SPR study, the relative pronoun has links towards any other word.

As a result, it does not make much difference if local syntactic relations are activated between words that are separated by more than one link in the network. The potential problem caused by the further links to travel is offset by spreading activation, the concurrent activation of meaning features, and the presence of hubs, as shown in Figure 8.3. To facilitate reading, some of the meaning links that were in the network depicted in Figure 8.3 have been omitted in the new network.

Figure 8.3

An Example of Lexical Network with a Relative Pronoun as a Hub



The relative pronoun *cui* is linked to every other word in the network, given the status of function word and network hub. All the features of the relative clause are present and encoded in the relative verb *offrire* (*offer*). The feature GAP contains the prepositional wh-phrase. The feature ARG-ST contains the NP for subject *negozio* (*shop*) and the NP for the object *sconti* (sales/reductions), while the feature DEPS has the modified adjunct and all the other dependents. The feature RESTR encodes the information about the thematic roles in the subcategorization frame and the adjunct and creates meaning links among other words in the network. As a result, the nouns *prezzo* (*price*) and *promozione* (*promotion*) are both a fitting theme for the verb *offrire* (*offer*) and can be co-indexed with the feature RESTR|theme.

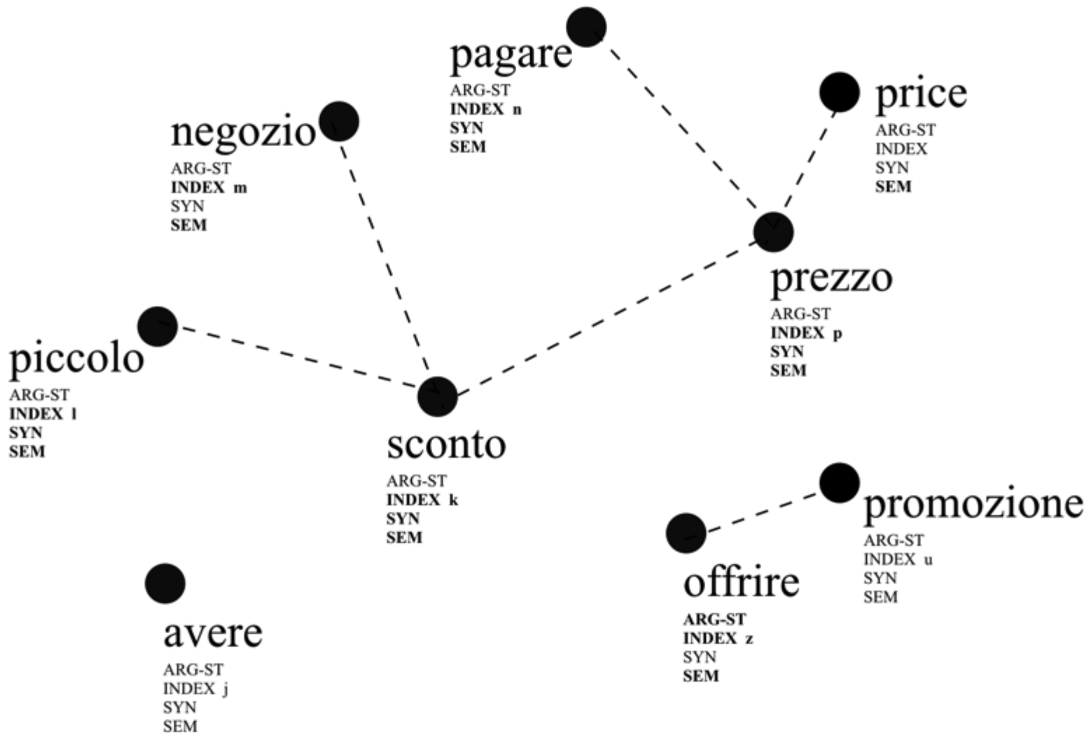
Under the spreading activation mechanism, nouns that match the same semantic role may compete to be co-indexed. Once the level of activation for *prezzo* (*price*) is outweighed by that of *promozione* (*promotion*), the relevant syntactic features of the target word are accessed and activated so that the relative clause *una promozione con cui offre gli sconti* is built. The activated syntactic links between the words that form the oblique relative clause **8.1** are described by solid lines, whereas the links that are inactivated are described by dashed lines. As explained before, despite being two links away from the verb, the subject of the main clause *negozio* (*shop*) can be easily linked to verb via spreading activation.

However, although travelling through the network based on the information encoded in the lexical features is usually very effective, things may still go wrong. The findings of the EI and SPR experiments showed that low frequency can make activation extremely difficult for L2 speakers, whose lexical networks do not have as many links as L1 ones. When lexical networks have fewer links, the activation process might not occur because a link between two words has not yet been formed. The I-Lex study in Chapter 3 showed that this might be the case with infrequent words, which L2 speakers are not able to associate to stimulus words to the same extent as L1 speakers. Furthermore, the EI study showed that low-frequency verbs affected the ability to correctly repeat cleft sentences, suggesting that the features of low-frequency words are harder to access and connect to one another. Having fewer links and struggling to access lexical features, explain the results of the SPR study. Since the lexical networks of L2 speakers have fewer infrequent words, when reading low-frequency relative verbs, they appeared slow in accessing the lexical features within the filler gap dependency. Furthermore, in the SPR task, the L2 speakers also showed structural effects in the complement region that were not caused by frequency. This result, rather than being caused by activation levels problems, is related to the weakness in the overall organisation of their network. Having networks less organised can affect the ability to recall and provide infrequent words (Meara, 2006; Wilks & Meara, 2002).

Considering all the effects found in the three studies, a simple example of how a L2 feature-based lexical network might look like is given in Figure 8.4. The network is, of course, a hypothetical example that aims to make clear that L2 lexical networks have a different quality relative to L1 speakers. The main assumption about the network is a simple one: some links are not present as expected from the less organised L2 lexical networks (Meara, 2009). The collocational link between *sconto* (*sale*/*reduction*) and the verb *avere* (*have*) and the meaning link to the verb *offrire* (*offer*) are missing. In addition, because of non-selectivity, the L2 networks also contain the translation equivalent word *price* that, depending on the task, might cause inhibition (Kroll et al., 2008). The English word overlaps with the Italian equivalent in terms of meaning features and in terms of interlexical neighbourhood.

Figure 8.4

An L2 Lexical Network with Missing Links

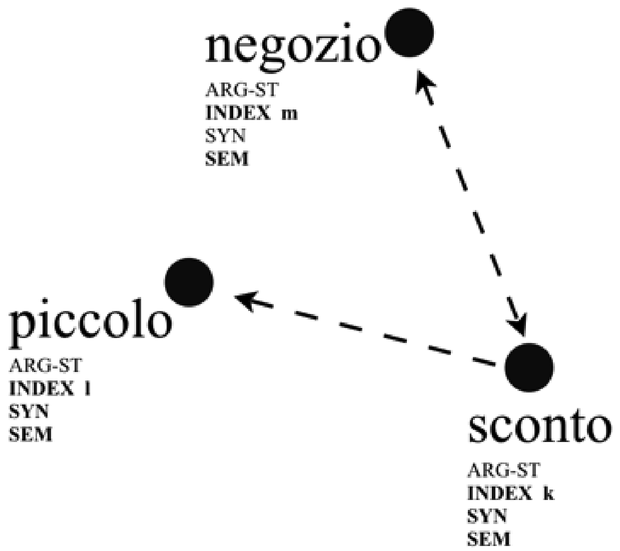


The fact that some links are missing affects the L2 processing system: when interpreting the sentence, it does not have a link to activate, but a link that must be formed from scratch. While in the receptive SPR task the process is cued by the sentence being read, in the productive EI task the link might not be formed or activated at all.

The idea that no activation is caused by the lack of a lexical link is speculative. An alternative explanation may be that the link is present, but there are not enough available features to activate it. As I explained in Chapter 5, in feature-based networks, links can have strength or weight depending on the number of features that underlie the activation process. If the available features are too few, the strength of the connection may not be enough to activate a link. A different hypothesis might be that some links in the network have a direction; that is, one word has a link towards and receives a link from another word. Therefore, a word that has links that run in both directions can be activated by any other word in the network. In contrast, words with links that only run in one direction can be activated by an external stimulus but not by a link coming from another word in the network (Fitzpatrick & Thwaites, 2020; Meara, 2009). For example, Figure 8.5 describes a section of the network in Figure 8.4.

Figure 8.5

A Fragment of Lexical Network with Directional Links



The link between *negozio* (*shop*) and *sconto* (*sale/reduction*) has an arrow at both ends. This indicates that the link between the two words may be activated in both directions. The word *sconto* (*sale/reduction*) can activate a link to the word *negozio* (*shop*) and the word *negozio* (*shop*) can activate a link to the word *sconto* (*sale/reduction*). In contrast, the link between *sconto* (*sale/reduction*) and *piccolo* (*small*) has an arrow at one end only. This means that *sconto* (*sale/reduction*) can activate a link to *piccolo* (*small*), but *piccolo* (*small*) cannot activate a link to *sconto* (*sale*/*reduction*). Hence, the word *piccolo* (*small*) can be accessed only when another word is already active in the discourse.

The most workable approach is to hold to both the notion of strength and the assumption that in L2 networks many links are missing. Missing links can occasionally be formed, but, to remain in the network, they need enough strength, which is determined by the ability of L2 speakers to access the right lexical features. The process of accessing lexical features to form and strengthen links is explained in the next section.

8.3 Feature-Based Lexical Access

The notion of availability assumes that the number of features that L2 speakers can access is limited by several factors. Feature availability is affected by the sensitiveness of L2 speakers to frequency effects, the organization of their lexical representations, and non-selectivity. Frequency can reduce the availability of lexical features and, as the I-Lex results in Chapter 3 showed, can also curtail the number of links within the networks. The finding of Chapter 3 has been replicated in both studies on syntactic processing in Chapter 6 and Chapter 7, where low-frequency verbs were difficult to process for the L2 group. Frequency affects the speed and the ease at which words are accessed and combined, especially in L2 speakers, whose lexical representations are already less robust and less organised (de Groot & van Hell, 1998b; Dijkstra & van Heuven, 2002; Kroll et al., 2008; Meara, 2006; Wilks & Meara, 2002).

In terms of the feature-based lexical network model, language representations of L2 speakers may lack some lexical features or some values for lexical features. When features or the information they encode (i.e., their value) are not available, this affects the ability to successfully access word (de Groot & van Hell, 1998a; Dijkstra & van Heuven, 2002;

Grosjean & Byers-Heinlein, 2018; Levelt & Meyer, 1999). This process can be straightforwardly framed into the HPSG formalism described in Chapter 4.

In HPSG having the right type of features encoding the right type of information, i.e., the right value, is the condition to have well-formed lexical structures (Pollard & Sag, 1994; Sag et al., 2003). The feature grammar used in HPSG does not focus extensively on lexical access; therefore, I will adapt the Vocabulary Test Capture Model developed by Fitzpatrick & Clenton (2017), which can incorporate the precision of HPSG feature-based descriptions while preserving more general claims about L2 lexical access. The Vocabulary Test Capture Model describes how the level of vocabulary knowledge underlies the production and use of a word and the quantitative aspect of lexical representations (Fitzpatrick & Clenton, 2017). The model was originally used to compare the lexical knowledge needed to carry out Lex30 with other vocabulary tasks. However, its flexibility is such that it can be easily used as a general model for lexical access. The model is shown in Figure 8.6. The column *Quality of learner's word knowledge* describes the number of distinct lexical features that an individual can use to produce a word. Knowledge is based on a scale that starts with the knowledge of the form and ends with the knowledge of semantic appropriateness and grammatical accuracy. The light grey shape on the right-hand side describes the type and the breath of vocabulary knowledge that Lex30 can capture.

Figure 8.6

The Vocabulary Test Capture Model

Quality of learner's word knowledge	Learner's overall lexical resource (number of words available for production)
1. can produce these word forms	
2. can use these words for appropriate referents or L1 words	
3. can use these words with semantic appropriateness in context	Lex30
4. can use these words with semantic appropriateness and grammatical accuracy in context	
	test task activation events

Note. Adapted from Fitzpatrick & Clenton (2017), p.860

Fitzpatrick & Clenton (2017) highlight that the model "can be taken as representative of all the items a learner has the capacity to produce at each level, across various functions, contexts, and topics" (p.861). An individual that produces associates from levels 4 and 3 is expected to know the meaning of the word, its contexts of use, and its grammatical properties. It is plausible that a thorough and sizable vocabulary knowledge can produce these types of lexical items. In contrast, to produce an associate at level 1, an individual might know only the form of the word and be hardly aware of any other of its lexical features. This kind of knowledge might not need to be backed up by a large and well-organized lexicon.

The version of the model shown in Figure 8.6 can be easily adapted to the I-Lex scores. The adaptation involves a small step, as the Vocabulary Test Capture Model was originally used to gauge the degree of lexical knowledge measured by Lex30. Given the similarity of the format in the two tasks, what holds for Lex30 can be suitably employed for I-Lex. The degree of feature availability can be easily framed by combining the four levels of learner's knowledge with the lexical features of HPSG. First, the labels of the four levels need to be preliminarily rephrased in the following form:

1. To be able to produce word forms, an individual must access the word-form features.
2. To be able to use a word for an appropriate referent or for an L1 word, an individual must access semantic features.
3. To be able to use words with semantic appropriateness in context, an individual must access semantic and syntactic features.
4. To be able to use words with semantic appropriateness and grammatical accuracy in context, an individual must access semantic, syntactic, and morphological features.

The next step is to adjust the new labels to the Vocabulary Test Capture Model and add the appropriate HPSG feature. This is described in the Model for Lexical Access and Feature Availability in Figure 8.7. The first thing to notice about the model shown in Figure 8.7 is that the relationship between word knowledge and lexical features is straightforward using the HPSG formalism.

Figure 8.7

Model for Lexical Access and Feature Availability

Level of lexical access and usage	**Feature availability**
1. can produce word forms accessing the right phonological features	PHON, HEAD
2. can use a word for an appropriate referent or for an L1 word accessing the right meaning features	PHON, INDEX, RESTR, RELN
3. can use words with semantic appropriateness in context accessing the right meaning and combinatorial features	RESTR, ARG-ST
4. can use words with semantic appropriateness and grammatical accuracy accessing the right syntactic features	COMPS, SPR, GAP, PER, NUM, GEND

Note. Based on Fitzpatrick & Clenton (2017), p.860

The feature PHON encodes the orthographic or phonological form of a word; when the form of word is activated, it is likely that the part of speech to which it belongs is accessed. In level 2, the meaning and the form of word are accessed to. The feature INDEX works out the link between the word and the entity to which it refers, while the features RESTR and RELN encode information about the main meaning properties. At level 3, the feature RESTR maps to the feature ARG-ST so that they encode the information about what a word combines with based on its meaning. At level 4 the syntactic features COMPS, MOD and SPR (SYN features) and the morphological features PER, NUM, GEND encode the information needed to build correct structural relations with other words in context. The feature GAP encodes, instead, the information that helps to form a relation in more extended contexts.

An important feature of the Model for Lexical Access and Feature Availability is that it describes lexical access of single words and lexical access of words in context. The first two levels describe how a word can be accessed through phonological or orthographic features only or through phonological or orthographic features and meaning features only. The last two levels of the model describe how the processing system can access a word drawing on information about the structural and meaning relations that it encodes. Of course, the difference between levels is not clear in language usage. Word association tasks, for example, show that associates to the stimulus words are provided based on the lexical features of the four levels (e.g., De Deyne & Storms, 2008; Fitzpatrick, 2009).

The Model for Lexical Access and Feature Availability describes the process that occurs when the processing system accesses lexical features encoded in words in lexical networks. It explains how feature availability can be shaped as a staged process in which some types of lexical feature can be used to form connections between words. The framing of this notion into a series of simple principles is a step forward in streamlining a feature-based model of lexical and sentence processing. Therefore, in the next section, I will show how it can be used to describe several examples taken from the lexical items used in the experiments.

The simplest case of lexical access is described at level 1 of the Model for Lexical Access and Feature Availability. The feature description of the activation of the phonological form of a word is given in Figure 8.8 using the verb *spaventava* (*scared*) from one of the items used in the EI study in Chapter 6. To simplify the structure the features SYN and SEM stand for the features SPR, COMP and RESTR. When the form of a word is known, through the default inheritance mechanism, some features are automatically activated together. The processing system may recognise the part of speech the word belongs to; hence, the feature HEAD becomes available, too. Knowing the feature HEAD ensures that the features SPR, COMP, and RESTR are activated as a default and that the list of the feature ARG-ST contains the first element, either the subject or the determiner. The question marks highlight that, at this stage of lexical knowledge, all lexical features are available by default, but they do not encode any information that can be put into use.

Figure 8.8

Activation of a Word Form

$$\begin{bmatrix} \begin{bmatrix} \text{PHON} & <\text{spaventava}> \\ \text{HEAD} & <\text{verb}> \end{bmatrix} \\ \begin{bmatrix} \text{ARG-ST} & <\text{NP}_{???}> \\ \text{SYN} & <???> \\ \text{SEM} & <???> \end{bmatrix} \end{bmatrix}$$

Example **8.2** and the feature structure shown in Figure 8.9 describe the opposite case, where the correct phonological features do not become fully available. In Chapter 5 I have explained that in bilingual models of lexical processing, phonological access is based on the similarity between input and words already stored in the mental lexicon. This might well be the case for Example **8.2**. The processing system has not accessed all the correct phonological features, but has accessed a similar, but incorrect, lexical item. As the participant listened to the rest of the stimulus sentence, the wrong form *guidava* (*drove*) turned out to have a meaning too different, not compatible with the complement of the target form *liquidava* (*refunded*) and the final adjunct.

8.2 Target: Nella società era Filippo che liquidava gli artigiani alla fine del lavoro
In the firm was Filippo who refunded the artisans at the end of the job
In the firm it was Filippo who refunded the artisans at the end of the job

P1: era qualcuno che guidava
was somebody who drove
It was somebody who drove

Figure 8.9

Activation of a Wrong Verb

$$
\begin{bmatrix}
\begin{bmatrix}
\text{PHON} & <\text{*guidava}> \\
\text{HEAD} & <\text{verb}> \\
\text{ARG-ST} & <\text{NP}_{???}>
\end{bmatrix} \\
\begin{bmatrix}
\text{SYN|SPR} & <\text{*qualcuno}> \\
\text{SYN|COMPS} & <???> \\
\text{SEM|RESTR} & <???>
\end{bmatrix}
\end{bmatrix}
$$

Syntactic information is available again by default inheritance. However, since the meaning information does not match the meaning of the other words in the sentence, the mapping between the meaning and the subcategorization frame does not occur. Therefore, the list of the feature ARG-ST contains only the default first member, the subject. As a result, no useful meaning information can be encoded by the features RESTR and COMPS as shown by the question marks. Thus, based on the assumption that the incorrectly retrieved form is transitive and can have a subject, the L2 participant produced the generic indefinite subject pronoun *qualcuno* (*somebody*).

An example for level 3 is provided in sentence **8.3**. Here the L2 participant could access the right words with semantic appropriateness and could access all the correct meaning and combinatorial features. However, the relative pronoun *che* (*who*) and the lexical feature GAP are missing. Therefore, the processing system cannot support the link between the relative pronoun and the relative verb with the necessary grammatical accuracy.

8.3 Target: Al lavoro era Amanda che spaventava Filippo con i richiami

P9: al lavoro era Amanda spaventava Filippo con i richiami

In the feature description shown in Figure 8.10 all the features of the relative verb have been correctly accessed. Although the form of the verb and the feature ARG-ST shows a correct mapping with the semantic features, the mapping with the syntactic features is incorrect. Missing the relative pronoun has prevented the processing system from forming the correct subject relative clause, so that the content of the feature GAP is empty. Furthermore, in the case of subject clefts, the content of the feature SPR should also be empty as the subject and the relative pronoun filler coincide. The meaning conveyed by the sentence is still correct, with the focused information lying at the front of the sentence, but the cleft sentence is not fully grammatical and sounds unusual for Italian speakers.

Figure 8.10

An example of Cleft Lacking Grammatical Accuracy

$$\begin{bmatrix} \begin{bmatrix} <\text{spaventava}> \\ \text{ARG-ST} < \text{NP}_j , \text{NP}_k > \end{bmatrix} \\ \begin{bmatrix} \text{SYN|SPR} & < \text{NP}_j > \\ \text{SYN|COMPS} & < \text{NP}_k > \\ \text{SYN|GAP} & < ??? > \end{bmatrix} \\ \begin{bmatrix} \text{SEM|Stimulus} & j \\ \text{SEM|Experiencer} & k \end{bmatrix} \end{bmatrix}$$

As can be seen from the cases described so far, the feature-based lexical network model can achieve a great degree of precision. Furthermore, it is flexible enough to account for complex interactions between different types of linguistic information. In Chapter 9, after giving an overview of the research project, I will discuss to what extent the model can be further improved and tested. Then, I will highlight potential new applications and its relevance for second language acquisition studies.

Chapter 9 Conclusions

In the present research project, I set out to reconcile, under one formal model, two different strands of research: lexical processing and sentence processing. Research on lexical processing has shown that lexical representations, frequency, and vocabulary size affect the way L2 speakers access lexical features. Research on sentence processing started to shift towards frameworks that encode grammatical information into lexical items. A specific framework, Head-Driven Phrase Structure Grammar (Abeillé & Borsley, 2019; Pollard & Sag, 1994; Sag et al., 2003) has converged on constraint-based models that associate information relevant to language processing with the lexicon.

In Chapter 2 I have described how frequency interacts with vocabulary size in lexical processing and how it underlies the process by which words are integrated into the mental lexicon. Since their lexical representations are weaker and less organised, it is more difficult for L2 speakers to access lexical features compared to L1 speakers. Reduced feature availability hinders the processing system from working efficiently. To measure feature availability, in Chapter 3, I developed and validated a productive vocabulary measurement task for L2 Italian called I-Lex. The aim was to have the first productive vocabulary task to be compared with online tasks that measure syntactic processing in L2 Italian.

In Chapter 4 I described how, in the HPSG theoretical approach, lexical entries encode phonological, semantic, and syntactic information and make the grammar work with the addition of few phrasal constraints. In Chapter 5 I showed that a feature-based model of lexical representations is best-suited to be implemented into constrained-based models. Both share the view that all lexically encoded information can be used simultaneously to process language. Building on the research about the properties and the setup of lexical networks, I have incorporated the HPSG feature system into a network-based model of lexical representations. The result was the creation of a feature-based lexical network model that could describe lexical access and the activation of structural links between words with the same set of lexical features.

The model was tested in two different studies described in Chapter 6 and Chapter 7. In Chapter 6 an Oral Elicited Imitation task investigated frequency and vocabulary effects on cleft sentences, while, in Chapter 7, a Self-paced Reading Task investigated frequency and vocabulary effects on relative clauses. To evaluate the effects of lexical knowledge, I added the results of the I-Lex task to the analysis of the experimental data. The results revealed that both frequency and vocabulary size interacted with the ability of L2 speakers to process both cleft and relative clauses. The interaction between vocabulary size and sentence processing provides evidence that accessing lexical features is a crucial stage for processing syntactic structures. Next, the feature-based lexical network set up in Chapter 5 was used to model the interaction between lexical knowledge and sentence processing.

Finally, in the discussion chapter, I have presented a complete description of the feature-based lexical network, drawing on examples from the Self-Paced Reading and the Elicited Oral Imitation tasks. I have also discussed the limitations of the model and proposed future directions to implement it. I have highlighted the necessity to test the model with a larger spectrum of linguistic phenomena to weigh its flexibility and to refine its setup by adding weighed links, based on natural language data.

9.1 Limitations

A limitation of the present research project is that both experiments have investigated a filler-gap dependency that does not have any syntactic ambiguity. All the items in the EI and SPR task had one interpretation. However, most of the research in L2 processing has been carried out based on the syntax-first models, thus using ambiguous structures (e.g., Cunnings et al., 2010; Felser et al., 2012). With the present results, it is not clear how the feature-based network could be adapted to explain, for instance, relative attachment or the effects of the subject – verb – object bias (McRae & Matsuki, 2013).

Another limitation regards the tasks adopted in the experiments described in Chapter 6 and 7. Since feature-based representations can achieve a great level of complexity, more precise measurements, such as eye tracking, could weigh how lexical features interact during lexical processing. Eye tracking design can distinguish, with great accuracy, what type of information encoded in lexical items is used in a particular task. Being able to add more granularity to measurement tools could lead to modeling more precisely how lexical features are accessed and links are activated.

The feature-based lexical network model has been implemented based on theoretical assumptions and experimental evidence. The notions that underpin the model were flexible enough to model the experimental results of Chapter 6 and Chapter 7. However, the model in its current version is not flexible enough to distinguish between unavailable features and lexical features that are unknown to L2 speakers. There are several techniques that can be exploited to ensure that participants knew the experimental items, such as questionnaires and translation tasks that can be administered after the experiment (e.g., Jiang, 2013). They were not used in the present research project, but it would be useful to compare the predictions of the model with the post-experimental tasks carried out by the students.

Finally, as noted in explaining the limitations of the SPR and EI experiment, investigating the effects of vocabulary knowledge in terms of lexical features requires a large number of participants. Frequency information and vocabulary size are factors related to individual language representations and are susceptible to a high degree of variation. In this type of research, large-scale studies are more likely to produce properly powered results that incorporate individual variations.

9.2 Implications for Future Research

Using the model proposed in the present research project could shed light on types of dependency that have not normally been analysed in terms of lexical effects. Agreement patterns, for example, is another good candidate. In HPSG, as described in Chapter 4, agreement is achieved through lexical features. It would be advisable to test whether agreement can be affected by frequency and vocabulary knowledge to the same extent as the filler gap dependency examined in Chapter 6 and Chapter 7.

It would be advisable to examine how the feature-based lexical network can be tested on the effects of non-selectivity on language processing. Taking into account the basic architecture of the model, incorporating lexical features and links from two languages would require a specific implementation. Although partial cross-linguistic overlap is likely to add complexity to model, drawing on more recent models (e.g., Dijkstra et al., 2019), the inclusion of non-selectivity seems utterly feasible.

An aspect that needs further research is how frequency information could be formally integrated into feature-based descriptions. In HPSG the feature values are always categorical in that they either have or do not have a specific value. However, as the results of the experiment have shown, features might be able to encode complex sets of information that can be accessed simultaneously. In a recent paper Wasow argued that it is entirely feasible that "some HPSG feature structures might allow multiple values for the same feature, but with probabilities (adding up to 1) attached to each value." (Wasow, 2019, p. 25). I would be extremely interested in developing a variant of HPSG where features can encode gradient information and test it in specific experiments.

9.3 Implications for SLA

I have proposed a feature-based lexical network model that describes how links between words are activated to combine them into meaningful grammatical strings. This opens the possibility of having a single account for phenomena that research in second language acquisition has considered part of distinct language representations.

Furthermore, the model can be implemented in a principled way using an existing theoretical framework. HPSG provides a fully fledged formalism for many languages in addition to English. In this respect, the model can be fully formalised and adapted to the study of different L2s and different phenomena. This is a novel trait of the model, as models currently used in second-language research are not fully formalised in terms of lexical and syntactic representations. Therefore, the model, along with the results of the experiments presented in Chapter 6 and Chapter 7, offers a novel approach on L2 Italian that can be adopted and expanded in future research.

References

Abdel Rahman, R., & Sommer, W. (2003). Does phonological encoding in speech production always follow the retrieval of semantic knowledge?: Electrophysiological evidence for parallel processing. *Brain Research. Cognitive Brain Research*, *16*(3), 372–382. https://doi.org/10.1016/S0926-6410(02)00305-1

Abdel Rahman, R., van Turennout, M., & Levelt, W. J. M. (2003). Phonological encoding is not contingent on semantic feature retrieval: An electrophysiological study on object naming. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*(5), 850–860. https://doi.org/10.1037/0278-7393.29.5.850

Abeillé, A., & Borsley, R. D. (2019). Basic properties and elements. In S. Mű̈ller, A. Abeillé, R. D. Borsley, & J. Koenig (Eds.), *Head-Driven Phrase Structure Grammar: The handbook* (pp. 1–43). Language Science Press.

Adelman, J. S., Brown, G. D. A., & Quesada, J. F. (2006). Contextual diversity, not word frequency, determines word-naming and lexical decision times. *Psychological Science*, *17*(9), 814–823. https://doi.org/10.1111/j.1467-9280.2006.01787.x

Adger, D. (2003). *Core syntax : a minimalist approach / David Adger*. Oxford : Oxford University Press.

Adger, D. (2010). A Minimalist theory of feature structure. *Features: Perspectives on a Key Notion in Linguistics*, *May*, 1–43. https://doi.org/10.1093/acprof:oso/9780199577743.003.0008

Adger, D., & Svenonius, P. (2011). Features in Minimalist Syntax. *The Oxford Handbook of Linguistic Minimalism*, 1–28. https://doi.org/10.1093/oxfordhb/9780199549368.013.0002

Aitchison, J. (2003). *Words in the mind : an introduction to the mental lexicon / Jean Aitchison*. (3rd ed.). Malden, MA : Blackwell Pub.

Altmann, G., & Steedman, M. (1988). Interaction with context during human sentence processing. *Cognition*, *30*(3), 191–238. https://doi.org/10.1016/0010-0277(88)90020-0

Amenta, S., Badan, L., & Brysbaert, M. (2021). LexITA: A Quick and Reliable Assessment Tool for Italian L2 Receptive Vocabulary Size. *Applied Linguistics*, *42*(2), 292–314. https://doi.org/10.1093/applin/amaa020

Andringa, S., Olsthoorn, N., van Beuningen, C., Schoonen, R., & Hulstijn, J. (2012). Determinants of Success in Native and Non-Native Listening Comprehension: An Individual Differences Approach. *Language Learning*, *62*(SUPPL. 2), 49–78. https://doi.org/10.1111/j.1467-9922.2012.00706.x

Anthony, L. (2020). *AntConc (Version 3.5.9)*. Tokyo, Japan: Waseda University. https://www.laurenceanthony.net/software

Arnold, D., & Godard, D. (2019). Relative clauses in HPSG. In *Head-Driven Phrase Structure Grammar: The handbook. Prepublished version.*

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*(4), 390–412. https://doi.org/10.1016/j.jml.2007.12.005

Baker, C. F., Fillmore, C. J., & Cronin, B. (2003). *The Structure of the FrameNet Database*. *16*.

Baker, C. F., Fillmore, C. J., & Lowe, J. B. (1998). The Berkeley FrameNet project. *COLING-ACL '98: Proceedings of the Conference*.

Barabási, A.-L. (2002). Linked: The New Science of Networks. In *Foreign Affairs* (Vol. 81, Issue 5). Perseus Publishing. https://doi.org/10.2307/20033300

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language, 68*(3), 255–278. https://doi.org/10.1016/j.jml.2012.11.001

Beglar, D., & Nation, P. (2013). Assessing Vocabulary. In *The Companion to Language Assessment* (pp. 172–184). John Wiley & Sons, Inc. https://doi.org/10.1002/9781118411360.wbcla053

Bender, E. M., & Emerson, G. (2021). Computational linguistics and grammar engineering. In *Head-Driven Phrase Structure Grammar: The handbook*.

Bernolet, S., Hartsuiker, R. J., & Pickering, M. J. (2007). Shared Syntactic Representations in Bilinguals: Evidence for the Role of Word-Order Repetition. *Journal of Experimental Psychology: Learning Memory and Cognition*, *33*(5), 931–949. https://doi.org/10.1037/0278-7393.33.5.931

Bernolet, S., Hartsuiker, R. J., & Pickering, M. J. (2013). From language-specific to shared syntactic representations: The influence of second language proficiency on syntactic sharing in bilinguals. *Cognition*, *127*(3), 287–306. https://doi.org/10.1016/j.cognition.2013.02.005

Bialystok, E., Craik, F. I. M., & Luk, G. (2008). Lexical access in bilinguals: Effects of vocabulary size and executive control. *Journal of Neurolinguistics*, *21*(6), 522–538. https://doi.org/10.1016/j.jneuroling.2007.07.001

Boas, H. C., & Sag, I. A. (2010). Sign-Based Construction Grammar Editors : *Distribution*.

Bod, R., Hay, J., & Jannedy, S. (2003). Probabilistic Linguistics. In R. Bod, J. Hay, & S. Jannedy (Eds.), *Journal of Chemical Information and Modeling*. The MIT Press.

Borro, I., & Luoni, G. (2015). L’Elicited Oral Imitation Test come strumento di misurazione delle conoscenze implicite e delle abilità di processing. *Italiano LinguaDue*, *7*(2), 71.

Bouma, G., Malouf, R., & Sag, I. (2001). Satisfying constraints on extraction and adjunction. *Natural Language and Linguistic Theory*, *19*(1), 1–65. https://doi.org/10.1023/A:1006473306778

Branigan, H. P., & Pickering, M. J. (2017). An experimental approach to linguistic representation. *Behavioral and Brain Sciences*, *40*(2017). https://doi.org/10.1017/S0140525X16002028

Bresnan, J. (2016). *Lexical-functional syntax* (Second edi). Wiley Blackwell.

Briscoe, E. J., Copestake, A., & De Paiva, V. (1993). *Inheritance, defaults, and the lexicon / edited by Ted Briscoe, Ann Copestake, Valeria de Paiva.* Cambridge England .

Brysbaert, M. (2019). How Many Participants Do We Have to Include in Properly Powered Experiments? A Tutorial of Power Analysis with Reference Tables. *Journal of Cognition*, *2*(1), 1–38. https://doi.org/10.5334/joc.72

Brysbaert, M., Lagrou, E., & Stevens, M. (2017). Visual word recognition in a second language: A test of the lexical entrenchment hypothesis with lexical decision times. *Bilingualism*, *20*(3), 530–548. https://doi.org/10.1017/S1366728916000353

Brysbaert, M., Mandera, P., & Keuleers, E. (2018). The Word Frequency Effect in Word Processing: An Updated Review. *Current Directions in Psychological Science*, *27*(1), 45–50. https://doi.org/10.1177/0963721417727521

Brysbaert, M., Mandera, P., McCormick, S. F., & Keuleers, E. (2019). Word prevalence norms for 62,000 English lemmas. *Behavior Research Methods*, *51*(2), 467–479. https://doi.org/10.3758/s13428-018-1077-9

Brysbaert, M., Stevens, M., Mandera, P., & Keuleers, E. (2016). The impact of word prevalence on lexical decision times: Evidence from the Dutch lexicon project 2. *Journal of Experimental Psychology: Human Perception and Performance*, *42*(3), 441–458. https://doi.org/10.1037/xhp0000159

Bybee, J. (2006). From usage to grammar: The mind's response to repetition. *Language*, *82*(4), 711–733. https://doi.org/10.1353/lan.2006.0186

Bybee, J. (2008). Usage-based Grammar and Second Language Acquisition. In N. C. Ellis & P. Robinson (Eds.), *Handbook of cognitive linguistics and second language acquisition* (pp. 216–236). Taylor & Francis.

Cai, Z. G., Pickering, M. J., Yan, H., & Branigan, H. P. (2011). Lexical and syntactic representations in closely related languages: Evidence from Cantonese–Mandarin bilinguals. *Journal of Memory and Language*, *65*(4), 431–445. https://doi.org/10.1016/j.jml.2011.05.003

Cancho, R. F. I. (2005). The Structure of Syntactic Dependency Networks: Insights from Recent Advances in Network Theory. *Problems of Quantitative Linguistics*, 60–75. https://doi.org/10.1007/s13398-014-0173-7.2

Cancho, R. F. I., & Solé, R. V. (2001). The small world of human language. *Proceedings of the Royal Society B: Biological Sciences*, *268*(1482), 2261–2265. https://doi.org/10.1098/rspb.2001.1800

Cancho, R. F. I., Solé, R. V., & Köhler, R. (2004). Patterns in syntactic dependency networks. *Physical Review E - Statistical Physics, Plasmas, Fluids, and Related Interdisciplinary Topics*, *69*(5), 8. https://doi.org/10.1103/PhysRevE.69.051915

Catala, R. J., & Moreno Espinosa, S. (2005). Using Lex30 to measure the L2 productive vocabulary of spanish primary learners of EFL. In *Vigo international journal of applied linguistics Vial.* (Issue May, pp. 27–44).

Chambers, C. G., & Cooke, H. (2009). Lexical Competition During Second-Language Listening: Sentence Context, But Not Proficiency, Constrains Interference From the Native Lexicon. *Journal of Experimental Psychology: Learning Memory and Cognition*, *35*(4), 1029–1040. https://doi.org/10.1037/a0015901

Chaves, R. P. (2012). On the grammar of extraction and coordination. *Natural Language and Linguistic Theory*, *30*(2), 465–512. https://doi.org/10.1007/s11049-011-9164-y

Chaves, R. P. (2013). An expectation-based account of subject islands and parasitism. In *Journal of Linguistics* (Vol. 49, Issue 2). https://doi.org/10.1017/S0022226712000357

Chiari, I., & Mauro, T. De. (2010). the New Basic Vocabulary of Italian: Problems and Methods. *Italian Journal of Applied Statistics*, *22*(1), 23–36. http://sa-ijas.stat.unipd.it/sites/sa-ijas.stat.unipd.it/files/22(1)_ChiariDeMauro.pdf

Chiari, I., & Mauro, T. De. (2015). The New Basic Vocabulary of Italian as a Linguistic Resource. In R. Basili, A. Lenci, & B. Magnigni (Eds.), *Proceedings of the First Italian Conference on Computational Linguistics CLiC-it 2014* (pp. 113–116). Accademia University Press.

Chomsky, N. (1981). *Lectures on government and binding* (7th ed.). Mouton de Gruyter.

Chomsky, N. (1995). *The minimalist program / Noam Chomsky.* Cambridge, Mass. : The MIT Press.

Chui, A. S. Y. (2006). A study of the English vocabulary knowledge of university students in Hong Kong. *Asian Journal of English Language Teaching*, *16*, 1–23.

Clahsen, H., & Felser, C. (2006). Grammatical processing in language learners. *Applied Psycholinguistics*, *27*(1), 3–42. https://doi.org/10.1017/S0142716406060024

Clahsen, H., & Felser, C. (2018). Some notes on the Shallow Structure Hypothesis. *Studies in Second Language Acquisition*, *40*(3), 693–706. https://doi.org/10.1017/S0272263117000250

Clenton, J., de Jong, N. H., Clingwall, D., & Fraser, S. (2020). Investigating the extent to which vocabulary knowledge and skills can predict aspects of fluency for a small group of pre-intermediate Japanese L1 users of English (L2). In *Vocabulary and the Four Skills* (Issue November, pp. 126–145). Routledge. https://doi.org/10.4324/9780429285400-15

Clifton, C., & Frazier, L. (1989). Comprehending Sentences with Long-Distance Dependencies. In G. N. Carlson & M. K. Tanenhaus (Eds.), *Linguistic structure in language processing* (pp. 273–317). https://doi.org/10.1007/978-94-009-2729-2_8

Collins, A. M., & Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological Review*, *82*(6), 407–428. https://doi.org/10.1037/0033-295X.82.6.407

Cong, J., & Liu, H. (2014). Approaching human language with complex networks. *Physics of Life Reviews*, *11*(4), 598–618. https://doi.org/10.1016/j.plrev.2014.04.004

Cop, U., Keuleers, E., Drieghe, D., & Duyck, W. (2015). Frequency effects in monolingual and bilingual natural reading. *Psychonomic Bulletin and Review*, *22*(5), 1216–1234. https://doi.org/10.3758/s13423-015-0819-2

Copestake, A. (2002). *Implementing Typed Feature Structure Grammars (CSLI Lec- ture Notes 110)*. CSLI Publications.

Corver, N. (2003). *Perfect projections* (R. van Hout, A. Hulk, F. Kuiken, & R. Towell (eds.); pp. 45–68). John Benjamins Publishing Company. https://doi.org/10.1075/lald.30.04cor

Crossley, S. A., Salsbury, T., & Mcnamara, D. S. (2014). Assessing Lexical Proficiency Using Analytic Ratings: A Case for Collocation Accuracy. *Applied Linguistics*, *36*(5), amt056. https://doi.org/10.1093/applin/amt056

Cuetos, F. (1988). Cross-linguistic differences in parsing: Restrictions on the use of the Late Closure strategy in Spanish. *Cognition*, *30*(1), 73–105. https://doi.org/10.1016/0010-0277(88)90004-2

Cunnings, I. (2012). An overview of mixed-effects statistical models for second language researchers. *Second Language Research*, *28*(3), 369–382. https://doi.org/10.1177/0267658312443651

Cunnings, I. (2017). Parsing and Working memory in bilingual sentence Processing. *Bilingualism*, *20*(4), 659–678. https://doi.org/10.1017/S1366728916000675

Cunnings, I., Batterham, C., Felser, C., & Clahsen, H. (2010). *Constraints on L2 learners' processing of wh-dependencies*. 87–110. https://doi.org/10.1075/lald.53.04cun

David, A., Myles, F., Rogers, V., & Rule, S. (2009). Lexical development in instructed L2 learners of French: is there a relationship with morphosyntactic development. In B. Richards, M. H. Daller, D. D. Malvern, P. Meara, J. Milton, & J. Treffers-Daller (Eds.), *Vocabulary Studies in First and Second Language Acquisition* (pp. 147–164). Palgrave Macmillan UK. https://doi.org/10.1057/9780230242258

Davis, A., & Koenig, J.-P. (2019). The nature and role of the lexicon in HPSG. In *Head-Driven Phrase Structure Grammar: The handbook. Prepublished version.* (Vol. 14).

De Deyne, S., Navarro, D. J., Perfors, A., Brysbaert, M., & Storms, G. (2019). The "Small World of Words" English word association norms for over 12,000 cue words. *Behavior Research Methods*, *51*(3), 987–1006. https://doi.org/10.3758/s13428-018-1115-7

De Deyne, S., Navarro, D. J., & Storms, G. (2013). Better explanations of lexical and semantic cognition using networks derived from continued rather than single-word associations. *Behavior Research Methods, 45*(2), 480–498. https://doi.org/10.3758/s13428-012-0260-7

De Deyne, S., & Storms, G. (2008a). Word associations: Network and semantic properties. *Behavior Research Methods*, *40*(1), 213–231. https://doi.org/10.3758/BRM.40.1.213

De Deyne, S., & Storms, G. (2008b). Word associations: Norms for 1,424 Dutch words in a continuous task. *Behavior Research Methods, 40*(1), 198–205. https://doi.org/10.3758/BRM.40.1.198

De Deyne, S., & Storms, G. (2015). Word Associations. In J. R. Taylor (Ed.), *Science* (Vol. 153, Issue 3742). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199641604.013.018

de Groot, A. M., & Brink, R. C. L. van den. (2010). Chapter 15. Foreign language vocabulary learning. In M. Kail & M. Hickmann (Eds.), *Language Acquisition across Linguistic and Cognitive Systems*.

de Groot, A. M., & van Hell, J. (1998a). Disentangling Context Availability and Concreteness in Lexical Decision and W ord Translation. *The Quarterly Journal of Experimental Psychology A*, *51*(1), 41–63. https://doi.org/10.1080/027249898391756

de Groot, A. M., & van Hell, J. (1998b). Conceptual representation in bilingual memory: Effects of concreteness and cognate status in word association. *Bilingualism: Language and Cognition*, *1*(3), 193–211. https://doi.org/10.1017/S1366728998000352

De Mauro, T. (2016). *Il Nuovo vocabolario di base della lingua italiana*. Il Nuovo Vocabolario Di Base Della Lingua Italiana. https://www.internazionale.it/opinione/tullio-de-mauro/2016/12/23/il-nuovo-vocabolario-di-base-della-lingua-italiana

De Vincenzi, M. (1991). *Syntactic parsing strategies in Italian:th e minimal chain principle*. Springer Science+Business Media Dordrecht.

Denhovska, N., Serratrice, L., & Payne, J. (2018). Frequency and working memory effects in incidental learning of a complex agreement pattern. *Lingua*, *207*, 49–70. https://doi.org/10.1016/j.lingua.2018.02.009

Diependaele, K., Lemhöfer, K., & Brysbaert, M. (2013). The word frequency effect in first- and second-language word recognition: A lexical entrenchment account. *Quarterly Journal of Experimental Psychology*, *66*(5), 843–863. https://doi.org/10.1080/17470218.2012.720994

Dijkstra, T. (2003). *Lexical storage and retrieval in bilinguals* (pp. 129–150).

Dijkstra, T., & van Heuven, W. J. B. (2002). The architecture of the bilingual word recognition system: From identification to decision. *Bilingualism: Language and Cognition*, *5*(3), 175–197. https://doi.org/10.1017/S1366728902003012

Dijkstra, T., Wahl, A., Buytenhuijs, F., Van Halem, N., Al-Jibouri, Z., De Korte, M., & Rekké, S. (2019). Multilink: A computational model for bilingual word recognition and word translation. *Bilingualism*, *22*(4), 657–679. https://doi.org/10.1017/S1366728918000287

Dunn, L. M., Dunn, L. M., Bulheller, S., & Hacker, H. (1965). *Peabody picture vocabulary test*. Circle Pines, MN: American Guidance Service.

Dussias, P. E., & Cramer Scaltz, T. R. (2008). Spanish-English L2 speakers' use of subcategorization bias information in the resolution of temporary ambiguity during second language reading. *Acta Psychologica*, *128*(3), 501–513. https://doi.org/10.1016/j.actpsy.2007.09.004

Duyck, W., Van Assche, E., Drieghe, D., & Hartsuiker, R. J. (2007). Visual Word Recognition by Bilinguals in a Sentence Context: Evidence for Nonselective Lexical Access. *Journal of Experimental Psychology: Learning Memory and Cognition*, *33*(4), 663–679. https://doi.org/10.1037/0278-7393.33.4.663

Duyck, W., Vanderelst, D., Desmet, T., & Hartsuiker, R. J. (2008). The frequency effect in second-language visual word recognition. *Psychonomic Bulletin and Review*, *15*(4), 850–855. https://doi.org/10.3758/PBR.15.4.850

Eckman, F. R., Bell, L., & Nelson, D. (1988). On the Generalization of Relative Clause Instruction in the Acquisition of English as a Second Language. *Applied Linguistics*, *9*(1), 1–20. https://doi.org/10.1093/applin/9.1.1

Ellis, N. C. (1996). Sequencing in SLA: Phonological memory, chunking, and points of order. *Studies in Second Language Acquisition*, *18*(1), 91–126. https://doi.org/10.1017/S0272263100014698

Ellis, N. C. (1997). Vocabulary acquisition: Word structure, collocation, word-class, and meaning. In N. Schmitt & M. McCarthy (Eds.), *Vocabulary : description, acquisition and pedagogy / edited by Norbert Schmitt and Michael McCarthy.* (pp. 122–139). Cambridge.

Ellis, N. C. (2002). FREQUENCY EFFECTS IN LANGUAGE PROCESSING. *Studies in Second Language Acquisition*, *24*(2), 143–188. https://doi.org/10.1017/S0272263102002024

Ellis, N. C. (2005). At the interface: Dynamic interactions of explicit and implicit language knowledge. *Studies in Second Language Acquisition*, *27*(2), 305–352. https://doi.org/10.1017/S027226310505014X

Ellis, N. C. (2011). Frequency-based accounts of SLA. *Handbook of Second Language Acquisition*, 193–201.

Ellis, N. C. (2013). The Routledge Handbook of Second Language Acquisition. In S. M. Gass & A. Mackey (Eds.), *The Routledge Handbook of Second Language Acquisition* (pp. 193–210). Routledge. https://doi.org/10.4324/9780203808184

Ellis, N. C., & Ferreira-Junior, F. (2009). Construction learning as a function of frequency, frequency distribution, and function. *Modern Language Journal*, *93*(3), 370–385. https://doi.org/10.1111/j.1540-4781.2009.00896.x

Ellis, N. C., O'Donnell, M. B., & Römer, U. (2014). Second language verb-argument constructions are sensitive to form, function, frequency, contingency, and prototypicality. *Linguistic Approaches to Bilingualism*, *4*(4), 405–431. https://doi.org/10.1075/lab.4.4.01ell

Elston-Gúttler, K. E., & Friederici, A. D. (2005). Native and L2 processing of homonyms in sentential context. *Journal of Memory and Language*, *52*(2), 256–283. https://doi.org/10.1016/j.jml.2004.11.002

Erlam, R. (2006). Elicited Imitation as a Measure of L2 Implicit Knowledge: An Empirical Validation Study. *Applied Linguistics*, *27*(3), 464–491. https://doi.org/10.1093/applin/aml001

Evans, N., & Levinson, S. (2012). Word Processing. In M. J. Traxler (Ed.), *Introduction to Psycholinguistics: Understanding Language Science* (pp. 79–140). Wiley-Blackwell.

Felser, C., Cunnings, I., Batterham, C., & Clahsen, H. (2012). The timing of island effects in nonnative sentence processing. *Studies in Second Language Acquisition*, *34*(1), 67–98. https://doi.org/10.1017/S0272263111000507

Felser, C., & Roberts, L. (2007). Processing wh-dependencies in a second language: A cross-modal priming study. *Second Language Research*, *23*(1), 9–36. https://doi.org/10.1177/0267658307071600

Felser, C., Roberts, L., Marinis, T., & Gross, R. (2003a). The processing of ambiguous sentences by first and second language learners of English. In *Applied Psycholinguistics* (Vol. 24, Issue 3). https://doi.org/10.1017/S0142716403000237

Felser, C., Roberts, L., Marinis, T., & Gross, R. (2003b). The processing of ambiguous sentences by first and second language learners of English. *Applied Psycholinguistics*, *24*(3), 453–489. https://doi.org/10.1017/S0142716403000237

Ferreira, F., & Clifton, C. (1986). The independence of syntactic processing. *Journal of Memory and Language*, *25*(3), 348–368. https://doi.org/10.1016/0749-596X(86)90006-9

Field, A, Miles, J., & Field, Z. (2012). *Discovering Statistics Using R*. SAGE Publications. https://books.google.co.uk/books?id=8n6jEkIj4kIC

Field, Andy, Miles, J., & Field, Z. (2012). *Discovering Statistics Using R*.

Fillmore, C. J. (1982). Frame semantics. In *Linguistics in the Morning Calm*. Hanshin Publishing Co.

Fillmore, C. J., & Baker, C. F. (2009). A Frame Approaching to Semantics. In *The Oxford Handbook of Linguistic Analysis* (pp. 313–340).

Finkbeiner, M., Forster, K., Nicol, J., & Nakamura, K. (2004). The role of polysemy in masked semantic and translation priming. *Journal of Memory and Language*, *51*(1), 1–22. https://doi.org/10.1016/j.jml.2004.01.004

Fitzpatrick, T. (2006). Habits and rabbits. *EUROSLA Yearbook*, *6*, 121–145. https://doi.org/10.1075/eurosla.6.09fit

Fitzpatrick, T. (2007). Word association patterns: Unpacking the assumptions. *International Journal of Applied Linguistics*, *17*(3), 319–331. https://doi.org/10.1111/j.1473-4192.2007.00172.x

Fitzpatrick, T. (2009). Word Association Profiles in a First and Second Language: Puzzles and Problems. In *Lexical Processing in Second Language Learners* (Vol. 21, Issue 1, pp. 38–52). Multilingual Matters. https://doi.org/10.21832/9781847691538-006

Fitzpatrick, T. (2010). Productive vocabulary tests and the search for concurrent validity. In H. Daller, J. Milton, & J. Treffers-Daller (Eds.), *Modelling and Assessing Vocabulary Knowledge* (pp. 116–132). Cambridge University Press. https://doi.org/10.1017/CBO9780511667268.009

Fitzpatrick, T., & Clenton, J. (2010). The challenge of validation: Assessing the performance of a test of productive vocabulary. *Language Testing*, *27*(4), 537–554. https://doi.org/10.1177/0265532209354771

Fitzpatrick, T., & Clenton, J. (2017). Making Sense of Learner Performance on Tests of Productive Vocabulary Knowledge. *TESOL Quarterly*, *51*(4), 844–867. https://doi.org/10.1002/tesq.356

Fitzpatrick, T., & Izura, C. (2011). Word association in L1 and L2: An exploratory study of response types, response times, and interlingual mediation. *Studies in Second Language Acquisition*, *33*(3), 373–398. https://doi.org/10.1017/S0272263111000027

Fitzpatrick, T., & Meara, P. (2004). Exploring the validity of a test of productive vocabulary. *VIAL, Vigo International Journal of Applied Linguistics*, *1*, 55–74.

Fitzpatrick, T., & Thwaites, P. (2020). Word association research and the L2 lexicon. *Language Teaching*, *53*(3), 237–274. https://doi.org/10.1017/S0261444820000105

Flickinger, D., Pollard, C., & Wasow, T. (1983). *The evolution of HPSG*. *i*, 47–83.

Flickinger, D., Pollard, C., & Wasow, T. (1985). Structure-sharing in lexical representation. *Proceedings of the 23rd Annual Meeting on Association for Computational Linguistics* -, 262–267. https://doi.org/10.3115/981210.981242

Frazier, L. (2013). Syntax in sentence processing. In R. P. G. Van Gompel (Ed.), *Sentence processing* (pp. 21–51). New York : Psychology Press.

Frazier, L., & Clifton, C. (1997). Construal: Overview, Motivation, and Some New Evidence. *Journal of Psycholinguistic Research*, *26*(3), 277–295.

Gass, S. M., & Ard, J. (1984). Second laguage acquisition and the ontology of laguage universals. In W. E. Rutherford (Ed.), *Typological universals and second language acquisition* (pp. 33–68). John Benjamins Publishing Company. https://doi.org/10.1075/tsl.5.05gas

Gibson, E. (1998). Linguistic complexity: Locality of syntactic dependencies. *Cognition*, *68*(1), 1–76. https://doi.org/10.1016/S0010-0277(98)00034-1

Gibson, E. (2000). The dependency locality theory: A distance-based theory of lingustic complexity. In Y. Miyashita, A. Marantz, & W. O'Neil (Eds.), *Image, Language, Brain* (pp. 95–126). MIT Press.

Gibson, E., Desmet, T., Grodner, D., Watson, D., & Ko, K. (2005). Reading relative clauses in English. *Cognitive Linguistics*, *16*(2), 313–353. https://doi.org/10.1515/cogl.2005.16.2.313

Gibson, E., & Warren, T. (2004). Reading-Time Evidence for Intermediate Linguistic Structure in Long-Distance Dependencies. *Syntax*, *7*(1), 55–78. https://doi.org/10.1111/j.1368-0005.2004.00065.x

Ginzburg, J. (2012a). *The Interactive Stance: Meaning for Conversation*. Oxford University Press UK.

Ginzburg, J. (2012b). Interaction, Grammar, and the Behavioural Sciences. In *The Interactive Stance* (pp. 1–11). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199697922.003.0001

Gollan, T. H., Montoya, R. I., Cera, C., & Sandoval, T. C. (2008). More use almost always means a smaller frequency effect: Aging, bilingualism, and the weaker links hypothesis. *Journal of Memory and Language*, *58*(3), 787–814. https://doi.org/10.1016/j.jml.2007.07.001

Gollan, T. H., Slattery, T. J., Goldenberg, D., Van Assche, E., Duyck, W., & Rayner, K. (2011). Frequency Drives Lexical Access in Reading but Not in Speaking: The Frequency-Lag Hypothesis. *Journal of Experimental Psychology: General*, *140*(2), 186–209. https://doi.org/10.1037/a0022256

González, R. A., & Píriz, A. M. P. (2016). Measuring the productive vocabulary of secondary school CLIL students: Is Lex30 a valid test for low-level school learners? *Vigo International Journal of Applied Linguistics*, *13*, 31–53.

Gordon, P. C., Hendrick, R., & Johnson, M. (2001). Memory interference during language processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*(6), 1411–1423. https://doi.org/10.1037/0278-7393.27.6.1411

Gordon, P. C., Hendrick, R., & Johnson, M. (2004). Effects of noun phrase type on sentence complexity. *Journal of Memory and Language*, *51*(1), 97–114. https://doi.org/10.1016/j.jml.2004.02.003

Gordon, P. C., Hendrick, R., & Levine, W. H. (2002). Memory-Load Interference in Syntactic Processing. *Psychological Science*, *13*(5), 425–430. http://www.jstor.org/stable/40063874

Gordon, P. C., & Lowder, M. W. (2012). Complex Sentence Processing: A Review of Theoretical Perspectives on the Comprehension of Relative Clauses. *Linguistics and Language Compass*, *6*(7), 403–415. https://doi.org/10.1002/lnc3.347

Green, G. M. (2011). Modeling Grammar Growth: Universal Grammar Without Innate Principles or Parameters. *Non-Transformational Syntax: Formal and Explicit Models of Grammar*, 378–403. https://doi.org/10.1002/9781444395037.ch11

Grosjean, F. (2008). *Studying bilinguals*. Oxford University Press.

Grosjean, F., & Byers-Heinlein, K. (2018). Speech Perception and Comprehension. In *The Listening Bilingual*. Wiley. https://doi.org/10.1002/9781118835722.ch2

Grosjean, F., & Li, P. (2013). *The psycholinguistics of bilingualism /* (F. Grosjean & P. Li (eds.)). Hoboken : John Wiley & Sons Inc.

Guo, T., & Peng, D. (2007). Speaking words in the second language: From semantics to phonology in 170ms. *Neuroscience Research*, *57*(3), 387–392. https://doi.org/10.1016/j.neures.2006.11.010

Halliday, M. A. K. (Michael A. K. (2004). *An introduction to functional grammar* (C. M. I. M. Matthiessen (ed.); 3rd ed. /). London : Arnold.

Hamrick, P., & Pandža, N. B. (2019). Contributions of Semantic and Contextual Diversity to the Word Frequency Effect in L2 Lexical Access. *Canadian Journal of Experimental Psychology*, *74*(1), 25–34. https://doi.org/10.1037/cep0000189

Hare, M., McRae, K., & Elman, J. L. (2003). Sense and structure: Meaning as a determinant of verb subcategorization preferences. *Journal of Memory and Language*, *48*(2), 281–303. https://doi.org/10.1016/S0749-596X(02)00516-8

Hartsuiker, R. J., & Bernolet, S. (2017). The development of shared syntax in second language learning. *Bilingualism*, *20*(2), 219–234. https://doi.org/10.1017/S1366728915000164

Hartsuiker, R. J., Pickering, M. J., & Veltkamp, E. (2004). Is syntax separate or shared between languages? Cross-linguistic syntactic priming in spanish-english bilinguals. *Psychological Science*, *15*(6), 409–414. https://doi.org/10.1111/j.0956-7976.2004.00693.x

Hawkins, R., & Liszka, S. A. (2003). Locating the source of defective past tense marking in advanced L2 English speakers. In R. van Hout, A. Hulk, F. Kuiken, & R. Towell (Eds.), *The Lexicon–Syntax Interface in Second Language Acquisition* (pp. 21–44). John Benjamins Publishing Company. https://doi.org/10.1075/lald.30.03haw

Hopp, H. (2006). Syntactic features and reanalysis in near-native processing. *Second Language Research*, *22*(3), 369–397. https://doi.org/10.1191/0267658306sr272oa

Hopp, H. (2014). Working Memory Effects in the L2 Processing of Ambiguous Relative Clauses. *Language Acquisition*, *21*(3), 250–278. https://doi.org/10.1080/10489223.2014.892943

Hopp, H. (2016). The timing of lexical and syntactic processes in second language sentence comprehension. *Applied Psycholinguistics*, *37*(5), 1253–1280. https://doi.org/10.1017/S0142716415000569

Hopp, H. (2017). Cross-linguistic lexical and syntactic co-activation in L2 sentence processing. *Linguistic Approaches to Bilingualism*, *7*(1), 96–130. https://doi.org/10.1075/lab.14027.hop

Hudson, R. A. (2010). *An introduction to word grammar / Richard Hudson.* Cambridge .

Izumi, S. (2003). Processing difficulty in comprehension and production of relative clauses by learners of english as a second language. *Language Learning*, *53*(2), 285–323. https://doi.org/10.1111/1467-9922.00218

Jakubíček, M., Kilgarriff, A., Kovář, V., Rychlý, P., & Suchomel, V. (2013). The TenTen Corpus Family. *7th International Corpus Linguistics Conference*, 125–127. https://www.sketchengine.co.uk/wp-content/uploads/The_TenTen_Corpus_2013.pdf%5Cnhttp://ucrel.lancs.ac.uk/cl2013/doc/CL2013-ABSTRACT-BOOK.pdf

Jakubícek, M., Kilgarriff, A., McCarthy, D., & Rychlý, P. (2010). Fast Syntactic Searching in Very Large Corpora for Many Languages. *Proceedings of the 24th Pacific Asia Conference on Language, Information and Computation*, 741–747. https://www.aclweb.org/anthology/Y10-1086

Jessop, L., Suzuki, W., & Tomita, Y. (2007). Elicited imitation in second language acquisition research. *Canadian Modern Language Review*, *64*(1), 215–220. https://doi.org/10.3138/cmlr.64.1.215

Jiang, N. (2013). Conducting reaction time research in second language studies. In *Conducting Reaction Time Research in Second Language Studies*. https://doi.org/10.4324/9780203146255

Johns, B. T., Sheppard, C. L., Jones, M. N., & Taler, V. (2016). The role of semantic diversity in word recognition across aging and bilingualism. *Frontiers in Psychology*, *7*(MAY), 1–11. https://doi.org/10.3389/fpsyg.2016.00703

Jones, M. N., Johns, B. T., & Recchia, G. (2012). The role of semantic diversity in lexical organization. *Canadian Journal of Experimental Psychology/Revue Canadienne de Psychologie Expérimentale*, *66*(2), 115–124. https://doi.org/10.1037/a0026727

Juffs, A., & Harrington, M. (1995). Parsing effects in second language sentence processing: Subject and object asymmetries in wh-extraction. *Studies in Second Language Acquisition*, *17*, 483–516.

Juffs, A., & Rodrigues, G. A. (2015). *Second language sentence processing* (G. A. Rodríguez (ed.)). New York, NY : Routledge.

Kaan, E. (2001). Effects of NP Type on the Resolution of Word-Order Ambiguities. *Journal of Psycholinguistic Research*, *30*(5), 529–547. https://doi.org/10.1023/A:1010417614058

Kaan, E. (2014). Predictive sentence processing in L2 and L1. *Linguistic Approaches to Bilingualism*, *4*(2), 257–282. https://doi.org/10.1075/lab.4.2.05kaa

Kawaguchi, S. (2016). Question constructions, argument mapping, and vocabulary development in English L2 by Japanese speakers. In J.-U. Keßler, A. Lenzing, & M. Liebner (Eds.), *Developing, Modelling, and Assessing Second Languages* (pp. 35–63). John Benjamins Publishing Company.

Keating, G. D., & Jegerski, J. (2015). Experimental designs in sentence processing research. *Studies in Second Language Acquisition*, *37*(1), 1–32. https://doi.org/10.1017/S0272263114000187

Keenan, E. L., & Comrie, B. (1977). Noun Phrase Accessibility and Universal Grammar. *Linguistic Inquiry*, *8*(1), 63–99.

Keuleers, E., Stevens, M., Mandera, P., & Brysbaert, M. (2015). Word knowledge in the crowd: Measuring vocabulary size and word prevalence in a massive online experiment. *Quarterly Journal of Experimental Psychology*, *68*(8), 1665–1692. https://doi.org/10.1080/17470218.2015.1022560

Kim, E., Baek, S., & Tremblay, A. (2015). The Role of Island Constraints in Second Language Sentence Processing. *Language Acquisition*, *22*(4), 384–416. https://doi.org/10.1080/10489223.2015.1028630

Koenig, J.-P., & Davis, A. R. (2006). The KEY to lexical semantic representations. *Journal of Linguistics*, *42*(1), 71–108. https://doi.org/10.1017/S0022226705003695

Koenig, J.-P., & Jurafsky, D. (1995). Type Underspecification and On-line Type Construction in the Lexicon. *West Coast Conference on Formal Linguistics (WCCFL13)*, 270–285.

Kroeger, P. R. (2004). *Analyzing Syntax*. Cambridge University Press. https://doi.org/10.1017/CBO9780511801693

Kroll, J. F., Bobb, S. C., Misra, M., & Guo, T. (2008). Language selection in bilingual speech: Evidence for inhibitory processes. *Acta Psychologica*, *128*(3), 416–430. https://doi.org/10.1016/j.actpsy.2008.02.001

Kroll, J. F., & Stewart, E. (1994). Category Interference in Translation and Picture Naming: Evidence for Asymmetric Connections Between Bilingual Memory Representations. *Journal of Memory and Language*, *33*(2), 149–174. https://doi.org/10.1006/jmla.1994.1008

Kroll, J. F., Van Hell, J. G., Tokowicz, N., & Green, D. W. (2010). The Revised Hierarchical Model: A critical review and assessment. *Bilingualism*, *13*(3), 373–381. https://doi.org/10.1017/S136672891000009X

Lau, E., & Tanaka, N. (2021). The subject advantage in relative clauses: A review. *Glossa: A Journal of General Linguistics*, *6*(1), 34. https://doi.org/10.5334/gjgl.1343

Laufer, B. (1994). The Lexical Profile of Second Language Writing: Does It Change Over Time? *RELC Journal*, *25*(2), 21–33. https://doi.org/10.1177/003368829402500202

Laufer, B., & Nation, P. (1995). Vocabulary Size and Use: Lexical Richness in L2 Written Production. *Applied Linguistics*, *16*(3), 307–322. https://doi.org/10.1093/applin/16.3.307

Laufer, B., & Nation, P. (1999). A vocabulary-size test of controlled productive ability. *Language Testing*, *16*(1), 33–51. https://doi.org/10.1177/026553229901600103

Laufer, B., & Ravenhorst-Kalovski, G. C. (2010). Lexical Threshold Revisited: Lexical Text Coverage, Learners' Vocabulary Size and Reading Comprehension. *Reading in a Foreign Language*, *22*(1), 15.

Lemhöfer, K., & Broersma, M. (2012). Introducing LexTALE: A quick and valid Lexical Test for Advanced Learners of English. *Behavior Research Methods*, *44*(2), 325–343. https://doi.org/10.3758/s13428-011-0146-0

Lemhöfer, K., & Dijkstra, T. (2004). Recognizing cognates and interlingual homographs: Effects of code similarity in language-specific and generalized lexical decision. *Memory and Cognition*, *32*(4), 533–550. https://doi.org/10.3758/BF03195845

Lemhöfer, K., Dijkstra, T., Schriefers, H., Baayen, R. H., Grainger, J., & Zwitserlood, P. (2008). Native Language Influences on Word Recognition in a Second Language: A Megastudy. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *34*(1), 12–31. https://doi.org/10.1037/0278-7393.34.1.12

Levelt, W. J. M., & Meyer, A. S. (1999). *Levelt_A_Theory_of_Lexical_Access_BBS_1999.pdf*. 1–75.

Levshina, N. (2015). *How to do linguistics with R : data exploration and statistical analysis*. Amsterdam.

Lewis, R. L., Vasishth, S., & Van Dyke, J. A. (2006). Computational principles of working memory in sentence comprehension. *Trends in Cognitive Sciences*, *10*(10), 447–454. https://doi.org/10.1016/j.tics.2006.08.007

Libben, M. R., & Titone, D. A. (2009). Bilingual Lexical Access in Context: Evidence From Eye Movements During Reading. *Journal of Experimental Psychology: Learning Memory and Cognition*, *35*(2), 381–390. https://doi.org/10.1037/a0014875

Lücking, A., Ginzburg, J., & Cooper, R. (2019). Grammar in dialogue. In *Head-Driven Phrase Structure Grammar: The handbook. Prepublished version.*

MacDonald, M. C., Pearlmutter, N. J., & Seidenberg, M. S. (1994). The lexical nature of syntactic ambiguity resolution. *Psychological Review*, *101*(4), 676–703. https://doi.org/10.1037/0033-295x.101.4.676

Mainz, N., Shao, Z., Brysbaert, M., & Meyer, A. S. (2017). Vocabulary knowledge predicts lexical processing: Evidence from a group of participants with diverse educational backgrounds. *Frontiers in Psychology*, *8*(JUL), 1–14. https://doi.org/10.3389/fpsyg.2017.01164

Mak, W. M., Vonk, W., & Schriefers, H. (2008). Discourse structure and relative clause processing. *Memory and Cognition*, *36*(1), 170–181. https://doi.org/10.3758/MC.36.1.170

Marinis, T., Roberts, L., Felser, C., & Clahsen, H. (2005). Gaps in second language sentence processing. *Studies in Second Language Acquisition*, *27*(1), 53–78. https://doi.org/10.1017/S0272263105050035

Marsden, E., Thompson, S., & Plonsky, L. (2018). A methodological synthesis of self-paced reading in second language research. *Applied Psycholinguistics*, *39*(5), 861–904. https://doi.org/10.1017/S0142716418000036

McDonald, S. A., & Shillcock, R. C. (2001). Rethinking the word frequency effect: The neglected role of distributional information in lexical processing. *Language and Speech*, *44*(3), 295–323. https://doi.org/10.1177/00238309010440030101

McManus, K., & Liu, Y. (2020). Using elicited imitation to measure global oral proficiency in SLA research: A close replication study. *Language Teaching*, 1–20. https://doi.org/10.1017/S026144482000021X

McRae, K., & Matsuki, K. (2013). Constraint-based models of sentence processing. In R. Van Gompel (Ed.), *Sentence processing* (pp. 51–77). Psychology Press.

McRae, K., Spivey-Knowlton, M. J., & Tanenhaus, M. K. (1998). Modeling the Influence of Thematic Fit (and Other Constraints) in On-line Sentence Comprehension. *Journal of Memory and Language*, *38*(3), 283–312. https://doi.org/10.1006/jmla.1997.2543

Meara, P. (2006). Emergent properties of multilingual lexicons. *Applied Linguistics*, *27*(4), 620–644. https://doi.org/10.1093/applin/aml030

Meara, P. (2009). *Connected words. Word associations and second language vocabulary acquisition*. John Benjamins Publishing Company.

Meara, P., & Fitzpatrick, T. (2000). Lex30: An improved method of assessing productive vocabulary in an L2. *System*, *28*(1), 19–30. https://doi.org/10.1016/S0346-251X(99)00058-5

Meara, P., & Fitzpatrick, T. (2015). *Lex30: a test of productive vocabulary*. Lex30: A Test of Productive Vocabulary. https://www.lognostics.co.uk/tools/Lex30/index.htm

Mellow, J. D. (2004). Connectionism, HPSG signs and SLA representations: Specifying principles of mapping between form and function. *Second Language Research*, *20*(2), 131–165. https://doi.org/10.1191/0267658304sr234oa

Miller, A. K. (2014a). Accessing and maintaining referents in L2 processing of wh - dependencies . *Linguistic Approaches to Bilingualism*, *4*(2), 167–191. https://doi.org/10.1075/lab.4.2.02mil

Miller, A. K. (2014b). Intermediate Traces and Intermediate Learners. *Studies in Second Language Acquisition*, *37*(3), 487–516. https://doi.org/10.1017/S0272263114000588

Milton, J. (2009). Measuring second language vocabulary acquisition. In *Measuring Second Language Vocabulary Acquisition*. https://doi.org/10.1080/09571736.2012.677623

Milton, J. (2010). The development of vocabulary breadth across the CEFR levels. *Communicative Proficiency and Linguistic Development: Intersections between SLA and Language Testing Research. Eurosla Monograph*, 211–232. http://eurosla.org/monographs/EM01/211-232Milton.pdf

Milton, J. (2013). Measuring the contribution of vocabulary knowledge to proficiency in the four skills. *L2 Vocabulary Acquisition, Knowledge and Use: New Perspectives on Assessment and Corpus Analysis*, *1983*, 57–78.

Milton, J., Wade, J., & Hopkins, N. (2010). Chapter 6. Aural Word Recognition and Oral Competence in English as a Foreign Language. In *Insights into Non-native Vocabulary Teaching and Learning* (pp. 83–98). Multilingual Matters. https://doi.org/10.21832/9781847692900-007

Moreno Espinosa, S. (2009). Young learners' L2 word association responses in two different learning contexts. In R. J. Catala & Y. Ruis de Zarobe (Eds.), *Content and Language Integrated Learning: Evidence from Research in Europe* (pp. 241–256). Multilingual Matters.

Müller, S. (2015). The CoreGram project: theoretical linguistics, theory development and verification. *Journal of Language Modelling*, *3*(1). https://doi.org/10.15398/jlm.v3i1.91

Müller, S. (2016). *Grammatical Theory*. Language Science Press.

Müller, S., & Wechsler, S. (2014). Lexical approaches to argument structure. *Theoretical Linguistics*, *40*(1–2), 1–76. https://doi.org/10.1515/tl-2014-0001

Nation, I. S. P. (2006). How Large a Vocabulary is Needed For Reading and Listening? *Canadian Modern Language Review*, *63*(1), 59–82. https://doi.org/10.3138/cmlr.63.1.59

Nation, I. S. P. (2013). *Learning vocabulary in another language / I.S.P. Nation.* (Second edi). Cambridge : Cambridge University Press.

Omaki, A., & Schulz, B. (2011). Filler-gap dependencies and island constraints in second-language sentence processing. *Studies in Second Language Acquisition*, *33*(4), 563–588. https://doi.org/10.1017/S0272263111000313

Pan, H. Y., Schimke, S., & Felser, C. (2015). Referential context effects in non-native relative clause ambiguity resolution. *International Journal of Bilingualism*, *19*(3), 298–313. https://doi.org/10.1177/1367006913515769

Papadopoulou, D., & Clahsen, H. (2003). Parsing strategies in l1 and l2 sentence processing. *Studies in Second Language Acquisition*, *25*(4), 501–528. https://doi.org/10.1017/S0272263103000214

Parker, D., Shvartsman, M., & Van Dyke, J. A. (2017). The Cue-based Retrieval Theory of Sentence Comprehension: New Findings and New Challenges. In L. Escobar, V. Torrens, & T. Parodi (Eds.), *Language Processing and Disorders* (pp. 121–144). Cambridge Scholars Publishing.

Pickering, M. J., & Gompel, R. P. G. van. (2006). Chapter 12 – Syntactic Parsing. In *Handbook of Psycholinguistics*. https://doi.org/10.1016/B978-012369374-7/50013-4

Pickering, M. J., & Traxler, M. J. (2001). Strategies for Processing Unbounded Dependencies: Lexical Information and Verb-Argument Assignment. *Journal of Experimental Psychology: Learning Memory and Cognition*, *27*(6), 1401–1410. https://doi.org/10.1037/0278-7393.27.6.1401

Pickering, M. J., & Traxler, M. J. (2006). *Parsing and Incremental Understanding During Reading* (pp. 238–258).

Pienemann, M. (1998). *Language Processing and Second Language Development* (Vol. 15). John Benjamins Publishing Company. https://doi.org/10.1075/sibil.15

Pliatsikas, C., & Marinis, T. (2013). Processing empty categories in a second language: When naturalistic exposure fills the (intermediate) gap. *Bilingualism*, *16*(1), 167–182. https://doi.org/10.1017/S136672891200017X

Pollard, C., & Sag, I. (1994). *Head-Driven Phrase Structure Grammar*. The University of Chicago Press.

Poort, E. D., & Rodd, J. M. (2019). Towards a distributed connectionist account of cognates and interlingual homographs: evidence from semantic relatedness tasks. *PeerJ*, *7*, e6725. https://doi.org/10.7717/peerj.6725

Prior, A., Goldina, A., Shany, M., Geva, E., & Katzir, T. (2014). Lexical inference in L2: predictive roles of vocabulary knowledge and reading skill beyond reading comprehension. *Reading and Writing*, *27*(8), 1467–1484. https://doi.org/10.1007/s11145-014-9501-8

Qian, D. D., & Lin, L. H. F. (2019). The Relationship Between Vocabulary Knowledge and Language Proficiency. In *The Routledge Handbook of Vocabulary Studies* (Issue 1998, pp. 66–80). Routledge. https://doi.org/10.4324/9780429291586-5

Qian, D. D., & Schedl, M. (2004). Evaluation of an in-depth vocabulary knowledge measure for assessing reading performance. *Language Testing*, *21*(1), 28–52. https://doi.org/10.1191/0265532204lt273oa

Rastelli, S. (2013). *Il processing nella seconda lingua*. Carocci.

Rastelli, S. (2018). Neurolinguistics and second language teaching: A view from the crossroads. *Second Language Research*, *34*(1), 103–123. https://doi.org/10.1177/0267658316681377

Read, J. (2019). Key Issues in Measuring Vocabulary Knowledge. In *The Routledge Handbook of Vocabulary Studies* (pp. 545–560). Routledge. https://doi.org/10.4324/9780429291586-34

Reali, F., & Christiansen, M. H. (2007). Processing of relative clauses is made easier by frequency of occurrence. *Journal of Memory and Language*, *57*(1), 1–23. https://doi.org/10.1016/j.jml.2006.08.014

Roberts, L., Marinis, T., Felser, C., & Clahsen, H. (2007). Antecedent Priming at Trace Positions in Children's Sentence Processing. *Journal of Psycholinguistic Research*, *36*(2), 175–188. https://doi.org/10.1007/s10936-006-9038-3

Rogers, V., Playfoot, D., & Milton, J. (2018). 7. The Second Language Lexicon. In C. Wright, T. Piske, & M. Young-Scholten (Eds.), *Mind Matters in SLA* (pp. 117–136). Multilingual Matters. https://doi.org/10.21832/9781788921626-009

Roland, D., & Jurafsky, D. (2002). Verb sense and verb subcategorization probabilities. In P. Merlo & S. Stevenson (Eds.), *The lexical basis of sentence processing: Formal, computational, and experimental issues* (pp. 325–346). John Benjamins Publishing.

Roland, D., Mauner, G., O'Meara, C., & Yun, H. (2012). Discourse expectations and relative clause processing. *Journal of Memory and Language*, *66*(3), 479–508. https://doi.org/10.1016/j.jml.2011.12.004

Rossini Favretti, R. (2000). Progettazione e costruzione di un corpus di italiano scritto: CORIS/CODIS. In R. Rossini Favretti (Ed.), *Linguistica e informatica. Multimedialità, corpora e percorsi di apprendimento* (pp. 39–56). Bulzoni. http://corpora.ficlit.unibo.it/CORISPubs/Rossini2000_ProgettazioneeCostruzione.pdf

Sag, I. (1997). English relative clause constructions. *Journal of Linguistics*, *33*(2), 431–483. https://doi.org/10.1017/S002222679700652X

Sag, I. (2010). English Filler-Gap Constructions. *Language*, *86*(43), 486–545.

Sag, I. (2012). Sign-Based Construction Grammar: An Informal Synopsis. *Sign-Based Construction Grammar*, 39–170.

Sag, I., & Wasow, T. (2011). Performance-Compatible Competence Grammar. *Non-Transformational Syntax: Formal and Explicit Models of Grammar*, *1992*, 359–377. https://doi.org/10.1002/9781444395037.ch10

Sag, I., & Wasow, T. (2014). Flexible Processing and the Design of Grammar. *Journal of Psycholinguistic Research*, *44*(1), 47–63. https://doi.org/10.1007/s10936-014-9332-4

Sag, I., Wasow, T., & Bender, E. M. (2003). *Syntactic Theory: A Formal Introduction*. Centre For The Study Of Language And Information.

Saito, K., Webb, S., Trofimovich, P., & Isaacs, T. (2016). Lexical Profiles of Comprehensible Second Language Speech. *Studies in Second Language Acquisition*, *38*(4), 677–701. https://doi.org/10.1017/S0272263115000297

Santi, A., Grillo, N., Molimpakis, E., & Wagner, M. (2019). Processing relative clauses across comprehension and production: similarities and differences. *Language, Cognition and Neuroscience*, *34*(2), 170–189. https://doi.org/10.1080/23273798.2018.1513539

Schielzeth, H., Dingemanse, N. J., Nakagawa, S., Westneat, D. F., Allegue, H., Teplitsky, C., Réale, D., Dochtermann, N. A., Garamszegi, L. Z., & Araya-Ajoy, Y. G. (2020). Robustness of linear mixed-effects models to violations of distributional assumptions. *Methods in Ecology and Evolution*, *11*(9), 1141–1152. https://doi.org/10.1111/2041-210X.13434

Schmitt, N. (2010). *Researching vocabulary : a vocabulary research manual / Norbert Schmitt.* Houndmills, Basingstoke, Hampshire.

Schmitt, N. (2014). Size and Depth of Vocabulary Knowledge: What the Research Shows. *Language Learning*, *64*(4), 913–951. https://doi.org/10.1111/lang.12077

Schmitt, N. (2019). Understanding vocabulary acquisition, instruction, and assessment: A research agenda. *Language Teaching*, *52*(02), 261–274. https://doi.org/10.1017/S0261444819000053

Schoonbaert, S., Hartsuiker, R. J., & Pickering, M. J. (2007). The representation of lexical and syntactic information in bilinguals: Evidence from syntactic priming. *Journal of Memory and Language*, *56*(2), 153–171. https://doi.org/10.1016/j.jml.2006.10.002

Schwartz, A. I., & Kroll, J. F. (2006a). Bilingual lexical activation in sentence context. *Journal of Memory and Language*, *55*(2), 197–212. https://doi.org/10.1016/j.jml.2006.03.004

Schwartz, A. I., & Kroll, J. F. (2006b). Language Processing in Bilingual Speakers. In *Handbook of Psycholinguistics*. https://doi.org/10.1016/B978-012369374-7/50026-2

Schwarze, C. (2009). *Grammatica della lingua italiana* (A. Colombo (ed.)). Carocci.

Shantz, K., & Tanner, D. (2017). Talking out of order: task order and retrieval of grammatical gender and phonology in lexical access. *Language, Cognition and Neuroscience*, *32*(1), 82–101. https://doi.org/10.1080/23273798.2016.1221510

Shantz, K., & Tanner, D. (2020). Electrophysiology finds no inherent delay for grammatical gender retrieval in non-native production. *Bilingualism*, *23*(1), 98–118. https://doi.org/10.1017/S1366728918001232

Shi, L., & Qian, D. (2012). How Does Vocabulary Knowledge Affect Chinese EFL Learners' Writing Quality in Web-based Settings? — Evaluating the Relationships Among Three Dimensions of Vocabulary Knowledge and Writing Quality. *Chinese Journal of Applied Linguistics*, *35*(1). https://doi.org/10.1515/cjal-2012-0009

Shieber, S. M. (1986). *An Introduction to Unification-based Approaches to Grammar*.

Shook, A., & Marian, V. (2013). The Bilingual Language Interaction Network for Comprehension of Speech. *Bilingualism (Cambridge, England)*, *16*(2), 304–324. https://doi.org/10.1017/S1366728912000466

Solé, R. V., Corominas-Murtra, B., Valverde, S., & Steels, L. (2010). Language Networks: Their Structure, Function, and Evolution. *Complexity*, *15*(6), 20–26. https://doi.org/10.1002/cplx.20305

Spinelli, G., Forti, L., & Jared, D. (2020). Learning to assign stress in a second language: The role of second-language vocabulary size and transfer from the native language in second-language readers of Italian. *Bilingualism*, *May*. https://doi.org/10.1017/S1366728920000243

Spivey-Knowlton, M. J., & Tanenhaus, M. K. (1998). Syntactic ambiguity resolution in discourse: Modeling the effect of referential context and lexical frequency within an integration-competition framework. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*(6), 1521–1543.

Spivey-Knowlton, M., & Sedivy, J. C. (1995). Resolving attachment ambiguities with multiple constraints. *Cognition*, *55*(3), 227–267. https://doi.org/10.1016/0010-0277(94)00647-4

Spivey, M. J., & Marian, V. (1999). Cross Talk Between Native and Second Languages: Partial Activation of an Irrelevant Lexicon. *Psychological Science*, *10*(3), 281–284. https://doi.org/10.1111/1467-9280.00151

Stæhr, L. S. (2008). Vocabulary size and the skills of listening, reading and writing. *The Language Learning Journal*, *36*(2), 139–152. https://doi.org/10.1080/09571730802389975

Stæhr, L. S. (2009). vocabulary knowledge and advanced listening comprehension in english as a foreign language. *Studies in Second Language Acquisition*, *31*(4), 577–607. https://doi.org/DOI: 10.1017/S0272263109990039

Starreveld, P. A., De Groot, A. M. B., Rossmark, B. M. M., & Van Hell, J. G. (2014). Parallel language activation during word processing in bilinguals: Evidence from word production in sentence context. *Bilingualism*, *17*(2), 258–276. https://doi.org/10.1017/S1366728913000308

Staub, A. (2010). Eye movements and processing difficulty in object relative clauses. *Cognition*, *116*(1), 71–86. https://doi.org/10.1016/j.cognition.2010.04.002

Stella, G., Pizzoli, C. E., & Tressoldi, P. . (2000). *Peabody: Test di vocabolario recettivo. [Peabody: Receptive vocabulary test]* (Omega Ediz).

Steyvers, M., & Tenenbaum, J. B. (2005). The Large-Scale Structure of Semantic Networks: Statistical Analyses and a Model of Semantic Growth. *Cognitive Science*, *29*(1), 41–78. https://doi.org/10.1207/s15516709cog2901_3

Sullivan, M. D., Poarch, G. J., & Bialystok, E. (2018). Why is lexical retrieval slower for bilinguals? Evidence from picture naming. *Bilingualism*, *21*(3), 479–488. https://doi.org/10.1017/S1366728917000694

Tily, H., Fedorenko, E., & Gibson, E. (2010). The time-course of lexical and structural processes in sentence comprehension. *Quarterly Journal of Experimental Psychology*, *63*(5), 910–927. https://doi.org/10.1080/17470210903114866

Tokowicz, N. (2014). *Lexical Processing and Second Language Acquisition*. Routledge. https://doi.org/10.4324/9780203551387

Tracy-Ventura, N., McManus, K., Norris, J. M., & Ortega, L. (2014). ‘Repeat as Much as You Can’: Elicited Imitation as a Measure of Oral Proficiency in L2 French. In *Measuring L2 Proficiency* (pp. 143–166). Multilingual Matters. https://doi.org/10.21832/9781783092291-011

Traxler, M. J., Morris, R. K., & Seely, R. E. (2002). Processing Subject and Object Relative Clauses: Evidence from Eye Movements. *Journal of Memory and Language*, *47*(1), 69–90. https://doi.org/10.1006/jmla.2001.2836

Traxler, M. J., & Pickering, M. J. (1996). Plausibility and the processing of unbounded dependencies: An eye-tracking study. *Journal of Memory and Language*, *35*(3), 454–475. https://doi.org/10.1006/jmla.1996.0025

Traxler, M. J., Williams, R. S., Blozis, S. A., & Morris, R. K. (2005). Working memory, animacy, and verb class in the processing of relative clauses. *Journal of Memory and Language*, *53*(2), 204–224. https://doi.org/10.1016/j.jml.2005.02.010

Trueswell, J. C. (1996). The role of lexical frequency in syntactic ambiguity resolution. *Journal of Memory and Language*, *35*(4), 566–585. https://doi.org/10.1006/jmla.1996.0030

Trueswell, J. C., Tanenhaus, M., & Garnsey, S. M. (1994). Trueswell_Tanenhuas_Garnsey_1994.pdf. In *Journal of Memory and Language* (Vol. 33, Issue 3, pp. 285–318).

Trueswell, J. C., & Tanenhaus, M. K. (1994). Toward a Lexicalist Framework for Contraint-Based Syntactic Ambiguity Resolution. In L. Frazier & K. Rayner (Eds.), *Perspectives on Sentence Processing* (pp. 155–179). Lawrence Erlbaum Associated.

Trueswell, J. C., Tanenhaus, M. K., & Kello, C. (1993). Verb-Specific Constraints in Sentence Processing: Separating Effects of Lexical Preference From Garden-Paths. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *19*(3), 528–553. https://doi.org/10.1037/0278-7393.19.3.528

Uchihara, T., & Saito, K. (2019). Exploring the relationship between productive vocabulary knowledge and second language oral ability. *Language Learning Journal*, *47*(1), 64–75. https://doi.org/10.1080/09571736.2016.1191527

Uchihara, T., Saito, K., & Clenton, J. (2020). Re-examining the relationship between productive vocabulary and second language oral ability. *Vocabulary and the Four Skills*, *November*, 146–165. https://doi.org/10.4324/9780429285400-16

Ullman, M. T., & Lovelett, J. T. (2018). Implications of the declarative/procedural model for improving second language learning: The role of memory enhancement techniques. *Second Language Research*, *34*(1), 39–65. https://doi.org/10.1177/0267658316675195

Van Assche, E., Drieghe, D., Duyck, W., Welvaert, M., & Hartsuiker, R. J. (2011). The influence of semantic constraints on bilingual word recognition during sentence reading. *Journal of Memory and Language*, *64*(1), 88–107. https://doi.org/10.1016/j.jml.2010.08.006

Van Assche, E., Duyck, W., & Brysbaert, M. (2013). VERB PROCESSING BY BILINGUALS IN SENTENCE CONTEXTS. *Studies in Second Language Acquisition*, *35*(2), 237–259. https://doi.org/10.1017/S0272263112000873

Van Dyke, J. A., & Johns, C. L. (2012). Memory Interference as a Determinant of Language Comprehension. *Language and Linguistics Compass*, *6*(4), 193–211. https://doi.org/10.1002/lnc3.330

Van Dyke, J. A., & Lewis, R. L. (2003). Distinguishing effects of structure and decay on attachment and repair: A cue-based parsing account of recovery from misanalyzed ambiguities. *Journal of Memory and Language*, *49*(3), 285–316. https://doi.org/10.1016/S0749-596X(03)00081-0

Van Dyke, J. A., & McElree, B. (2011). Cue-dependent interference in comprehension. *Journal of Memory and Language*, *65*(3), 247–263. https://doi.org/10.1016/j.jml.2011.05.002

Van Gompel, R. P. G. (2013). *Sentence processing*. New York : Psychology Press.

van Hout, R., Hulk, A., Kuiken, F., & Towell, R. (2003). *The Lexicon–Syntax Interface in Second Language Aquisition* (R. van Hout, A. Hulk, F. Kuiken, & R. Towell (eds.)). John Benjamins Publishing Company.

Vanpatten, B. (2015). Foundations of processing instruction. *International Review of Applied Linguistics in Language Teaching*, *53*, 109–191.

VanPatten, B. (2004). *Processing instruction theory, research, and commentary / edited by Bill VanPatten.* Mahwah, N.J. : L. Erlbaum Associates.

Vitevitch, M. S. (2008). What Can Graph Theory Tell Us About Word Learning and Lexical Retrieval? *Journal of Speech, Language, and Hearing Research*, *51*(2), 408–422. https://doi.org/10.1044/1092-4388(2008/030)

Vosse, T., & Kempen, G. (2000). Syntactic structure assembly in human parsing: a computational model based on competitive inhibition and a lexicalist grammar. *Cognition*, *75*(2), 105–143. https://doi.org/10.1016/S0010-0277(00)00063-9

Walker, D. A. (2003). JMASM9: Converting Kendall's Tau For Correlational Or Meta-Analytic Analyses. *Journal of Modern Applied Statistical Methods*, *2*(2), 525–530. https://doi.org/10.22237/jmasm/1067646360

Walters, J. (2012). Aspects of validity of a test of productive vocabulary: Lex30. *Language Assessment Quarterly*, *9*(2), 172–185. https://doi.org/10.1080/15434303.2011.625579

Wang, Y., & Treffers-Daller, J. (2017). Explaining listening comprehension among L2 learners of English: The contribution of general language proficiency, vocabulary knowledge and metacognitive awareness. *System*, *65*, 139–150. https://doi.org/10.1016/j.system.2016.12.013

Warren, T., & Gibson, E. (2002). The influence of referential processing on sentence complexity. *Cognition*, *85*(1), 79–112. https://doi.org/10.1016/S0010-0277(02)00087-2

Wasow, T. (2019). Processing. In *Head-Driven Phrase Structure Grammar: The handbook. Prepublished version.*

Webb, S. (2005). RECEPTIVE AND PRODUCTIVE VOCABULARY LEARNING: The Effects of Reading and Writing on Word Knowledge. *Studies in Second Language Acquisition*, *27*(01). https://doi.org/10.1017/S0272263105050023

Wechsler, S. (2020). The Role of the Lexicon in the Syntax–Semantics Interface. *Annual Review of Linguistics*, *6*(1), 67–87. https://doi.org/10.1146/annurev-linguistics-011619-030349

Weissman, B., & Tanner, D. (2018). A strong wink between verbal and emoji-based irony: How the brain processes ironic emojis during language comprehension. *PLOS ONE*, *13*(8), e0201727. https://doi.org/10.1371/journal.pone.0201727

Whitford, V., & Titone, D. (2012). Second-language experience modulates first- and second-language word frequency effects: Evidence from eye movement measures of natural paragraph reading. *Psychonomic Bulletin & Review*, *19*(1), 73–80. https://doi.org/10.3758/s13423-011-0179-5

Wilks, C., & Meara, P. (2002). Untangling word webs: graph theory and the notion of density in second language word association networks. *Second Language Research*, *18*(4), 303–324. https://doi.org/10.1191/0267658302sr203oa

Williams, john n. (2006). Incremental interpretation in second language sentence processing. *Bilingualism: Language and Cognition*, *9*(1), 71–88. https://doi.org/10.1017/S1366728905002385

Witzel, J., Witzel, N., & Nicol, J. (2012). Deeper than shallow: Evidence for structure-based parsing biases in second-language sentence processing. *Applied Psycholinguistics*, *33*(2), 419–456. https://doi.org/10.1017/S0142716411000427

Wu, S.-L., & Ortega, L. (2013). Measuring global oral proficiency in SLA research: A new elicited imitation test of L2 Chinese. *Foreign Language Annals*, *46*(4), 680–704. https://doi.org/10.1111/flan.12063

Yap, M. J., Balota, D. A., Sibley, D. E., & Ratcliff, R. (2012). Individual differences in visual word recognition: Insights from the English Lexicon Project. *Journal of Experimental Psychology: Human Perception and Performance*, *38*(1), 53–79. https://doi.org/10.1037/a0024177

Yap, M. J., Tse, C.-S., & Balota, D. A. (2009). Individual differences in the joint effects of semantic priming and word frequency revealed by RT distributional analyses: The role of lexical integrity. *Journal of Memory and Language*, *61*(3), 303–325. https://doi.org/10.1016/j.jml.2009.07.001

Appendix 1

This is an automated confirmation email for the following project. The Ethics Assessment status of this project is: APPROVED

Applicant Name: Gabriele Luoni
Project Title: The Relationship between Vocabulary Knowledge, Lexical Frequency and Syntactic Processing in Second Language Learners of Italian.
Project Start Date: 01/04/2016
Project Duration: four years
Approval No: SU-Ethics-Student-051020/3106

Appendix 2

Content:

Consent Form

Brief for the I-Lex and the EI tasks

Brief for the I-Lex and the SPR tasks

Brief and Consent Form for the online I-Lex

Instructions for the I-Lex and the EI tasks

Instructions for the I-Lex and the SPR tasks

Questionnaire

Debrief for the I-Lex and the EI tasks

Debrief for the I-Lex and the SPR tasks

Consent form

Swansea University
Prifysgol Abertawe

Title of study:
The Relationship between Lexicon and Syntactic Development

This form is for you to state whether or not you agree to take part in the study. Please read and answer every question. If there is anything you do not understand, or if you want more information, please ask the researcher.

Have you read and understood the information leaflet about the study?	Yes	No
Have you had an opportunity to ask questions about the study?	Yes	No
Do you understand that the information you provide will be held in confidence by the research team?	Yes	No
Do you understand that you may withdraw from the study for any reason, before the end of the data collection? (To withdraw, simply do not hand in your answer sheet.)	Yes	No
Do you understand that the information you provide may be used in future research?	Yes	No
Do you agree to take part in the study?	Yes	No

All data is held by the Department of English Language and Literature in accordance with the Data Protection Act.

Your name (in BLOCK letters): ____

Your signature: ____

Researcher's name: Gabriele Luoni (884477@swansea.ac.uk)

Date: ____

Title of the study:

The relationship between vocabulary knowledge and syntactic development in second language acquisition

The study you are invited to take part is about the relationship between vocabulary and syntax. Before you decide whether to take part it is important for you to understand why the research is being done and what it will involve. Please take the time to carefully read the following information. If you have any doubt you are invited to ask the researcher.

Research Coordinator:

Gabriele Luoni, PhD researcher, Department of English Language & Literature, Swansea University

What is the research about?

The research investigates the relationship between vocabulary and syntax during the acquisition of Italian as a second and foreign language.

Who is carrying out the research?

Gabriele Luoni, Swansea University

What does the study involves?

You will be requested to take part in 2 different tests, which will be delivered during the language course:

1. Vocabulary knowledge task on paper
2. Oral production test (repeating sentences)

The Vocabulary task will be delivered during one of the daily lecturers and will last 15 minutes. The Oral production test will be delivered to each student individually and will last 10 minutes. Before you begin each task, you will be given a brief explanation and a consent form to be signed saying that you agree to participate.

What are the possible risks of taking part?

There is no foreseeable risk to taking part

Are there any benefits to participating?

The befits are that you may be part of a piece of research that is taking his first steps and has revealed interesting and hugely promising results. You may help the research to move further to deeply understand the relationship between vocabulary and other complex phenomena of second language acquisition.

What will happen to the data I provide?

Your data will carefully examined alongside data of other participants, in order to assess how word knowledge take part into syntactic development. Data will be stored securely at Swansea University.

What about confidentiality?

Your identity will be kept strictly confidential. Names with not be collected with answer sheet and all subsequent publications will only report anonymised data.

Will I know the results?

Individual results will not be made available, but a summary of the results of the whole project will be provided at the end of the course.

Title of the study:

The relationship between vocabulary knowledge and syntactic development in second language acquisition

The study you are invited to take part of is about the relationship between vocabulary and syntax. Before you decide whether to participate into it, it is important for you to understand why the research is being done and what it will involve. Please take the time to carefully read the following information. If you have any doubt you are invited to ask the researcher.

Research Coordinator:

Gabriele Luoni, PhD researcher, Department of English Language & Literature, Swansea University

What is the research about?

The research investigates the relationship between vocabulary and syntax during the acquisition of Italian as a second and foreign language.

Who is carrying out the research?

Gabriele Luoni, Swansea University

What does the study involve?

You will be requested to take part into 2 different tests, which will be delivered during the language course:

1. Self Paced Reading
2. A vocabulary test on paper

The tasks will be completed over the entire course, in two different occasions, each lasting 15 minutes. The Self Paced Reading test will be delivered to each student individually.
Before you begin each task you will be given a brief explanation and a consent form to be signed stating that you agree to participate.

What are the possible risks of taking part?

There are no foreseeable risk to taking part

Are there any benefits to participating?

The benefits are that you may be part of a piece of research that is taking his first steps and has so far revealed interesting and hugely promising results. You may help the research to move further in order to understand the relationship between vocabulary knowledge and other complex phenomena of second language acquisition. Several studies showed that such a relationship exists and further data may be useful in order to better understand how exactly word's features operate in the acquisition of language. This could hugely improve the work of both linguists and language teachers.

What will happen to the data I provide?

Your data will be carefully examined alongside data of other participants, in order to assess how word knowledge foster syntactic development. Data will be stored securely at Swansea University.

What about confidentiality?

Your identity will be kept strictly confidential. Names with not be collected with answer sheet and all subsequent publications will only report anonymised data.

Will I know the results?

Individual results will not be made available, but a summary of the results of the whole project will be provided at the end of the course.

Brief

As part of Gabriele Luoni's PhD work, we are developing an Italian vocabulary test and would like to invite you to take part. You do not have to take part. All of your information and answers will be anonymous and kept in line with Swansea University's ethical guidelines.

There are two parts to the test:
In the first part you will see an Italian word and you should write four responses that you think of immediately when you see the word. You can leave some blanks but write at least two words for each response.

<u>Example:</u>

casa

porta, arrivare, tetto, lontana

The second part is a brief background questionnaire about you and your language experience.

Consent form

Please tick the box below to agree to take part in this study.
You agree that your data will be used anonymously for the purposes of developing this test and presenting the results in academic contexts (e.g. at a conference, in the PhD thesis, in a research article).
You may withdraw at any time. Please email the researchers if you no longer wish to take part.
Gabriele Luoni: [redacted] OR Dr Vivienne Rogers: [redacted]

○ I agree to take part in this study.

Instructions

In test 1 you will see a list of words. For every word you have to write up to 4 words that come to mind and write the in the four blanks on the right columns.

In test 2 you will first hear a recorded sentence. You will wait 3 seconds, then you have to repeat the sentence just heard as correctly as possible, seeking not to change it.

Instructions

In test 1 you will see a list of words. For every word you have to write up to 4 words that come to mind and write the in the four blanks on the right columns.

In test 2 you will read sentences on a computer screen in a word by word manner. When the symbol + will pop up on the screen press the space key on the keyboard to make the first world of the sentence appear. The press the space key to replace the previous word with a new one. Only one word is visible at any given time. When, after pressing the space key, you see a black screen, it means that the sentence is finished. Press the space key so that you can see again the symbol + and start reading the next sentence. Some sentences are followed by comprehension questions about the meaning of the sentence just read. Press the space key to answer YES and the N key to answer NO.

Personal Details questionnaire

1. Name …………………………………………………..

2. Year of birth ………………………………………..

3. What is your native language? (“Native language” = the language that you learned first, as a baby. If you have more than one native language, please list both/all, and please indicate which language you consider to be your dominant native language.)

………

………

4. Please list all the other languages that you have studied in the table and choose the descriptions that best match each language. (Choose as many as apply.)

Language	Began learning in primary school	Began learning in secondary school	Began learning at university	Continued studying at university	Other (please detail)

5. What is your highest educational qualification that you have already received?

……….

6. Finally, if there is anything else about your background that has anything to do with language and might be relevant, please explain:

………

………

What do my results measure?

Thank you for completing the tests which were part of the research project “The Relationship between Lexicon and Syntactic Development”. It is important that you understand that all these measures rely only on some facets of your overall proficiency and are not to be taken as a test of general language competence.

Vocabulary test:

I-Lex measures an aspect of your productive vocabulary knowledge

Oral Elicited Imitation Test:

The score are going to be assigned on the basis of different criteria: target structure correctly repeated, overall meaning of the sentence correctly repeated, minor errors that don't affect the meaning and major errors that affect the general meaning

What do my results measure?

Thank you for completing the tests which were part of the research project “The Relationship between Lexicon and Syntactic Development”. It is important that you understand that all these measure rely only on some facets of your overall proficiency and are not to be taken as a test of general language competence.

Vocabulary test:

I-Lex measures an aspect of your productive vocabulary knowledge

Self Paced Reading

Reaction times will be measured for four target segments of each sentence: the relative pronoun, the verb after the relative pronoun, the word after the verb and the following word.

Appendix 3

Chapter 3 – I-Lex

Per ogni parola scrivi fino a 4 altre parole che ti vengono in mente
(For each word you have to supply 4 other words that it makes you think of)

abito				
abitudine				
attaccare				
calcio				
cartolina				
chiudere				
commercio				
cuocere				
dente				
esperienza				
finestra				
frutto				
malattia				
mobile				
obbligo				
patata				
posto				
reale				
riposare				
riso				
scienza				
scrivere				
sostanza				
speranza				
sporco				
stupido				
tavola				
televisione				
tenere				
togliere				

Appendix 4

Content:

Elicited Oral Imitation Task.

Target Items: a) subject cleft with high frequency verb, b) object cleft with high frequency verb, c) subject cleft with low frequency verb, d) object cleft with low frequency verb

a) Al lavoro era Amanda che spaventava Filippo con i richiami
   At work it was Amanda who scared Filippo with the rebukes
b) Al lavoro era Amanda che Filippo spaventava con i richiami
   At work it was Amanda who Filippo scared with the rebukes
c) Al lavoro era Amanda che intimoriva Filippo con i richiami
   At work it was Amanda who frightened Filippo with the rebukes
d) Al lavoro era Amanda che Filippo intimoriva con i richiami
   At work it was Amanda who Filippo frightened with the rebukes

a) Nel negozio era Andrea che spingeva Maria a rubare il vino
   In the shop it was Andrea who forced Maria to steal the wine
b) Nel negozio era Andrea che Maria spingeva a rubare il vino
   In the shop it was Andrea who Maria forced to steal the wine
c) Nel negozio era Andrea che istigava Maria a rubare il vino
   In the shop it was Andrea who compelled Maria to steal the wine
d) Nel negozio era Andrea che Maria istigava a rubare il vino
   In the shop it was Andrea who Maria compelled to steal the wine

a) Nella società di costruzioni era Filippo che pagava gli artigiani alla fine dei lavori
In the construction company it was Filippo who paid the artisans at the end of the work

b) Nella società di costruzioni era Filippo che gli artigiani pagavano alla fine dei lavori
In the construction company it was Filippo who the artisans paid at the end of the work

c) Nella società di costruzioni era Filippo che liquidava gli artigiani alla fine dei lavori
In the construction company it was Filippo who refunded the artisans at the end of the work

d) Nella società di costruzioni era Filippo che gli artigiani liquidavano alla fine dei lavori
In the construction company it was Filippo who the artisans refunded at the end of the work

a) Nelle competizioni sportive era Andrea che impressionava Maria per i risultati
In the sport contests it was Andrea who impressed Maria for the results

b) Nelle competizioni sportive era Andrea che Maria impressionava per i risultati
In the sport contests it was Andrea who Maria impressed for the results

c) Nelle competizioni sportive era Andrea che sconcertava Maria per i risultati
In the sport contests it was Andre who astonished Maria for the results

d) Nelle competizioni sportive era Andrea che Maria sconcertava per i risultati
In the sport contests it was Andre who Maria astonished for the results

Appendix 5

Content:

Elicited Oral Imitation Task: Filler Items.

1. Nelle università italiane la ricerca deve assolutamente produrre nuovi risultati
   In the Italian universities the research must absolutely produce new results
2. E’ una bellissima canzone di cui ricordo ancora la melodia
   It is a wonderful song of which I still remember the melody
3. Il primo anno di fisica all’università ha capito l’importanza della matematica
   During the first year of physics at the university (he/she) realised the importance of math
4. Tutte le strutture dell’impianto sportivo sono state riparate nell’ultimo periodo
   All the facilities of the sport plant have been fixed in the last period
5. Da molto la crisi in cui si trova l’industria alimentare provoca problemi alle banche
   For a long time the crisis in which lies the food industry creates problems to the banks
6. Il piccolo paese è diventato un importante centro industriale
   The small village has become an important industrial hub
7. Un'importante azienda italiana le ha offerto un contratto a tempo indeterminato
   An important Italian firm has offered her an indefinite contract
8. Dopo la festa la gente ha parcheggiato l’auto vicino alla piazza
   After the fair the people has parked the car near to the square
9. Durante l’assemblea ha fatto un lungo discorso con cui ha oltraggiato tutti i cittadini onesti
   During the meeting (he/she) gave a long talk with which he/she insulted all the honest citizens
10. Siamo arrivati in un fantastico paesino vicino al mare
    We have arrived in a wonderful little village next to the sea
11. All’hotel ci hanno dato dei buoni consigli sui musei da vedere
    At the hotel they gave us good suggestions about the galleries to visit
12. Il manuale di storia da cui il docente ha preso le domande è complesso per gli studenti
    The history textbook from which the teacher drew the questions is complex for the students
13. Un blog è un ottimo modo per mostrare al pubblico i risultati del nostro lavoro
    A blog is the best way to show to the public the results of our work
14. L’appartamento si trova al pianoterra e ha anche un piccolo giardino indipendente
    The flat sits at the first floor and has also a small independent garden

15. Raccontiamo la storia di un'azienda che è famosa nel mondo
    We tell the story of a company which is famous worldwide
16. Una lunga vacanza offre l'occasione giusta per dimenticare i problemi quotidiani
    A long holiday offers the proper opportunity to forget daily problems
17. Dopo la telefonata l’hotel ha inviato una mail per il pagamento
    After the phone call the hotel has sent an email for the payment
18. I passeggeri dell’autobus erano in viaggio verso l’aeroporto
    The passengers of the bus were travelling to the airport
19. Ho telefonato all’impiegato con cui hai parlato rigurado il prezzo
    I called the employee with whom you have talked about the price
20. Per partecipare ai corsi è necessario un certificato sportivo del proprio medico
    To partake into the courses it is necessary a sport certificate from your doctor
21. Nella guida dello studente si trovano i consigli per affrontare gli esami universitari
    In the student guide there are the suggestions to take on university exams
22. Ho incontrato il bambino con cui hai giocato sulla spiaggia
    I have met the kid with whom you played on the beach
23. Lunedì l'agenzia ci ha venduto i biglietti del treno per il viaggio
    On Monday the agency sold us the train tickets for the journey
24. Non hanno capito il motivo per cui ha manipolato i risultati
    They did not understand the reason why he/she rigged the results

Appendix 6

Content:

Self-Paced Reading.

Target Items. Items 1, 7, 11, 15: subject relative clauses with high frequency verbs. Items 2, 8, 12, 16: oblique relative clauses with high frequency verbs. Items 3, 5, 9, 13: subject relative clauses with low frequency verbs. Items 4, 6, 10, 14: oblique relative clauses with low frequency verbs.

Target Items: Comprehension Questions

Target Items

1. La società ha creato un ufficio che paga le ricevute in base alle scadenze
   The firm has created an office which pays the receipts based on the deadlines
2. La società ha creato un ufficio con cui paga le ricevute in base alle scadenze
   The firm has created an office with which it pays the receipts based on the deadlines
3. La compagnia ha aperto una sede che archivia le fatture in base al costo
   The company has opened headquarters which files the invoices based on the cost
4. La compagnia ha aperto una sede con cui archivia le fatture in base al costo
   The company has opened headquarters with which it files the invoices based on the cost
5. Il comune ha aperto un ufficio che assegna i fondi in base alle spese
   The council has opened an office which assigns the funds based on the expenses
6. Il comune ha aperto un ufficio con cui assegna i fondi in base alle spese
   The council has opened an office with which it assigns the funds based on the expenses
7. Il negozio ha iniziato una promozione che offre gli sconti in base agli acquisti
   The shop has started a promotion which offers the reductions based on the purchases
8. Il negozio ha iniziato una promozione con cui offre gli sconti in base agli acquisti
   The shop has started a promotion with which it offers the reductions based on the purchases

9. Il ristorante ha comprato un apparecchio che mescola gli ingredienti in base alle ricette
   The restaurant has bought an appliance which mixes the ingredients based on the recipes
10. Il ristorante ha comprato un apparecchio con cui mescola gli ingredienti in base alle ricette
   The restaurant has bought an appliance with which it mixes the ingredients based on the recipes
11. Il laboratorio ha realizzato un programma che cambia i codici in base ai prodotti
   The lab has realised a program which changes the codes based on the products
12. Il laboratorio ha realizzato un programma con cui cambia i codici in base ai prodotti
   The lab has realised a program with which it changes the codes based on the products
13. L'azienda ha aperto un negozio che procura i prodotti in base alle ordinazioni
   The company has opened a shop which procures the products based on the orders
14. L'azienda ha aperto un negozio con cui procura i prodotti in base alle ordinazioni
   The company has opened a shop with which it procures the products based on the orders
15. La fabbrica ha costruito un attrezzo che lavora gli ingredienti in base alle ricette
   The factory has made a tool which works the ingredients based on the recipies
16. La fabbrica ha costruito un attrezzo con cui lavora gli ingredienti in base alle ricette
   The factory has made a tool with which it works the ingredients based on the recipies

Comprehension Questions

The numbers correspond to the Target items.

1. L'ufficio paga le ricevute in base alle scadenze? Y
   Does the office pay the receipts based on the deadlines?
2. L'ufficio paga le riveute in base alle tasse? N
   Does the office pay the receipts based on the taxes?
3. La sede archivia le fatture in base al costo? Y
   Do the headquarters file the invoices based on the cost?
4. La sede archivia le fatture in base al prezzo? N
   Do the headquarters file the invoices based on the price?
5. L'ufficio assegna i fondi in base alle fatture? N
   Does the office assign the funds based on the invoices?
6. L'ufficio assegna i fondi in base alle tasse? N
   Does the office assign the funds based on the invoices?
7. La promozione offre sconti in base agli acquisti? Y
   Does the promotion offer the reductions based on the purchases?
8. La promozione offre sconti in base ai costi? N
   Does the promotion offer the reductions based on the costs?
9. L'apparecchio mescola gli ingredienti in base alla temperatura? N
   Does the appliance mix the ingredients based on the temperature?
10. L'apparecchio mescola gli ingredienti in base alle ricette? Y
    Does the appliance mix the ingredients based on the receipes?
11. Il programma cambia i codici in base al prezzo? N
    Does the program changes the codes based on the price?
12. Il programma cambia i codici in base ai prodotti? Y
    Does the program changes the codes based on the products?
13. Il negozio procura i prodotti in base alle ordinazioni? Y
    Does the shop procure the products based on the orders?
14. Il negozio procura i prodotti in base ai clienti? N
    Does the shop procure the products based on the customers?
15. L'attrezzo lavora gli ingredienti in base alle ricette? Y
    Does the tool work the ingredients based on the recipies?

16. L'attrezzo lavora gli ingredienti in base alla temperatura? N

    Does the tool work the ingredients based on the temperature?

Appendix 7

Content

Self-Paced Reading Task.

Filler Items.

1. Molti laureati italiani sono pronti al trasferimento in un altro paese
   Many Italians with a bachelor degree are ready to move in another country
2. Nel centro del paese hanno aperto due nuovi negozi di elettronica
   In the centre of the village they have opened two new electronics shops
3. Sono andati nel nuovo teatro vicino al centro per il concerto
   They have gone in the new theatre next to the city centre for the concert
4. L'hotel è elegante e adatto a un soggiorno di studio o di lavoro
   The hotel is elegant and suitable for a study or a work stay
5. Per Andrea vivere e studiare a Bologna è stata una bella esperienza
   For Andrea living and studying in Bologna has been a nice expereince
6. Sua moglie ha lavorato in un negozio di abbigliamento come commessa
   His wife has worked in a clothe shop as a shop assistant
7. Abitavano da molti anni in un fantastico paesino sulle colline
   They lived for many years in a wonderful little village on the hills
8. L'impianto sportivo ha aperto una sede dove si possono fare gli abbonamenti
   The sport centre has opened headquarters where it is possible to make a subscription
9. Un blog è molto utile per mostrare i risultati del nostro lavoro
   A blog is very useful to show the results of our work
10. Ci sono più di 60 milioni gli animali domestici che vivono in Italia
    There are more than 60 milllion pets who live in Italy
11. La chiesa è antica e conserva molte opere rinascimentali
    The church is old and preserves many Renaissance ouvres
12. Ha lavorato per oltre 20 anni per una piccola libreria del centro storico
    He/she worked for over 20 years for a small bookstore in the old town
13. Dopo la passeggiata al parco Maria è rientrata in ufficio
    After the stroll in the park Maria has gone back into the office

14. La colazione è il pasto più importante della giornata
    Breakfast is the most important meal of the day
15. Dopo la laurea in economia Giovanna è rientrata nel campo della pubblicità
    After the degree in ecnomics Giovanna has returned to the advertising sector
16. Una lunga vacanza aiuta a dimenticare i problemi quotidiani
    A long holiday helps to forget daily problems
17. Sono andati con gli amici in un piccolo teatro vicino al centro
    They went with their friends in a small theatre near to the city centre
18. L'azienda ha offerto un contratto a tempo indeterminato per i nuovi impiegati
    The company has offered an indefinite contract to the new employees
19. Nel centro del paese hanno costruito un nuovo stadio per la squadra di calcio
    In the centre of the village they have built a new arena for the football team
20. Ha lavorato per una grande società di edilizia senza un contratto fisso
    He/She has worked for a big building company without a secure contract
21. Ha lavorato per una grande fabbrica di automobili da corsa
    He/She has worked for a big factory of racing cars
22. Nel centro del paese si possono trovare molti nuovi ristoranti etnici
    In the centre of the village it is possible to find many new ethnic restaurants
23. Nella guida si trovano i consigli per superare gli esami del primo anno
    In the booklet there are suggestions to pass the first year exams
24. Nella libreria in centro si trovano molti libri di storia antica
    In the bookstore in the city centre there are many books on ancient history
25. La casa ha un piccolo giardino di fronte alla porta di ingresso
    The house has a small garden in front of the entrance door
26. Il municipio ha aperto un ufficio dove si possono ritirare i moduli delle tasse
    The council has opened an office where it is possible to collect the tax forms
27. L'impianto sportivo è stato riparato grazie a un'impresa privata
    The sport centre has been repaired thanks to a private company
28. Chiunque può avere un impianto telefonico senza collegarsi alla rete telefonica
    Anyone can have a telephone line without connecting to the telephone network
29. I passeggeri erano in viaggio verso l'aeroporto internazionale di Roma
    The passengers were travelling to the international Airport of Rome

30. Come mobilità urbana e trasporti Milano è una città molto moderna
In terms of urban mobility and transports Milan is a very modern city
31. Il gruppo ha pubblicato un disco dopo aver trovato un nuovo chitarrista
The band has issued a record after having found a new guitar player
32. La polizia ha arrestato un ladro dopo una lunga fuga per le vie della città
The police have apprehended a thief after a long chase through the city alleys
33. Nel negozio era Andrea che istigava Maria a rubare il vino
In the shop it was Andrea who compelled Maria to steal the wine
34. Dopo un breve crisi negli anni 80 la compagnia ha aumentato i suoi guadagni
After a brief crisis in the eighties the company has increased its revenues
35. Il gruppo ha iniziato una serie di concerti di beneficenza in piccoli locali
The band has started a series of charity concerts in small venues
36. Al lavoro era Amanda che Filippo spaventava con i richiami
At work it was Amanda who Filippo scared with the rebukes
37. La compagnia telefonica ha sviluppato i collegamenti anche nei piccoli paesi
The telephone company has developed the network even in the small villages
38. Andrea ha una nuova macchina fotografica da usare per le prossime vacanze
Andrea had a new camera to use during next holidays
39. La biologia è una delle discipline scientifiche che hanno avuto maggiori sviluppi
Biology is one of the scientific subjects which has had greater developments
40. Nel negozio era Andrea che Maria spingeva a rubare il vino
In the shop it was Andrea who Maria forced to steal the wine
41. Al lavoro era Amanda che spaventava Filippo con i richiami
At work it was Amanda who frightened Filippo with the rebukes
42. Il negozio ha venduto molti prodotti soprattutto nei paesi del Centro America
The shop has sold numerous products especially in central American countries
43. La compagnia telefonica ha investito molto denaro per ampliare la rete
The telephone company has invested a lot of money to enlarge the network
44. Dopo essere stati al museo abbiamo visitato anche una fabbrica di machine da scrivere
After having been at the museum we also visited a factory of typewriters
45. Una società di software ha assunto centinaia di giovani programmatori
A software firm has hired hundreds of new programmers
46. Nella banca era Andrea che spingeva Maria a rubare i soldi
In the shop it was Andrea who forced Maria to steal the wine

47. C'è un nuovo ufficio che si occupa di preparare i documenti per le tasse
There is a new office which is in charge of preparing tax forms
48. Amanda andrà in Cina il prossimo mese dopo la fine del semestre
Amanda will go to Cina next month after the end of the term
49. I suoi vicini facevano molto rumore durante la festa di compleanno
His/Her neighbours were very loud during the birthday party
50. Nella società di costruzioni c'erano molti problemi con i pagamenti
In the construction company there were many problems with the payments
51. L'azienda dopo i recenti problemi conta oggi un centinaio di nuovi punti vendita
The firm after the recent problems has now a hundred new shops
52. Al lavoro era Amanda che Filippo intimoriva con i richiami
At work it was Amanda who Filippo frightened with the rebukes
53. I buoni programmi di investimento spingono lo sviluppo tecnologico delle aziende
Good investment programmes boost the technical development of the companies
54. Vicino alla stazione si trovano i palazzi storici e il teatro antico
Close to the station there are historical buildings and the old theatre
55. Il negozio ha venduto molti più prodotti dopo aver abbassato i prezzi
The shop has sold many more products after lowering the prices
56. Dopo molti problemi il negozio ha aumentato il numero delle vendite
After many problems the shop has increased the number of sales
57. L'azienda ha aperto una sede dopo che il nuovo progetto è stato approvato
The firm has opened headquarters after approving the new project
58. Uno studio ha dimostrato che usare Facebook ha un effetto benefico sulla memoria
A study has demonstrated that using Facebook has a beneficial effect on memory
59. Il dipartimento ha aperto una sede dove i nuovi impiegati possono fare tirocinio
The department has opened headquarters where new employees can do an internship
60. A scuola era Amanda che intimoriva Filippo con i richiami
At school it was Amanda who frightened Filippo with the rebukes
61. Nel bar era Andrea che Maria istigava a rubare il vino
In the bar it was Andrea who Maria compelled to steal the wine
62. Ha comprato un vestito per il matrimonio dei suoi vicini di casa
He/She has bought a dress for his/her neighbour wedding
63. L'hotel offre numerosi servizi fra cui una terrazza panoramica e un centro congressi
The hotel offers many facilities like a panoramic terrace and a conference centre

64. Ci sono dei nuovi clienti che aspettano in sala di attesa
    There are some new customers who are waiting in the waiting room
65. Per l'iscrizione è necessario pagare le tasse prima che inizi il primo semestre
    For the enrollement it is necessary to pay the fees before the first term starts
66. La compagnia telefonica ha aperto una nuova sede in centro
    The telephone company has opened new headquarters in the city centre
67. Per raggiungere l’aeroporto di Milano Filippo deve prendere l'autobus in centro
    To get to Milan airport Filippo needs to catch the bus in the city centre
68. C'erano molti ristoranti economici dove si poteva mangiare solo vegetariano
    There were many cheap restaurants where it was possible to have vegetarian meals

Fillers: Comprehension Questions

The numbers correspond to the Filler items which have questions

5. Andrea ha avuto una brutta esperienza vivendo a Bologna? N

Did Andrea had an unpleasant experience when living in Bologna?

6. Sua moglie ha lavorato in un ristorante? N

Did his wife work in a restaurant?

7. Abitavano in un piccolo paese vicino al mare? N

Did they live in a small village close to the sea?

10. In Italia vivono 90 milioni di animali domestici? N

In Italy are there 90 million of pets?

11. La chiesa conserva molte opere barocche? N

Does the church preserve many Baroque ouvres?

22. In centro si possono trovare molti ristoranti etinici? Y

Are there many ethnic shops in the city centre?

24, La libreria ha molti libri di cucina? N

Does the bookstore have books about cooking?

26. Nell'ufficio si possono ritirare i moduli per le tasse? Y

Is it possible to take the tax forms in the office?

27. Un'impresa privata ha riparato l'impianto sportivo? Y

Did a private company repair the sport centre?

29. I passeggeri erano in viaggio verso l'aeroporto di Milano? N

Were the passengers traveling to Milan airport?

31. Il gruppo ha trovato un nuovo batterista? N

Did the band found a new drummer?

33. Andrea istigava Maria a rubare? Y

Did Andrea compel Maria to steal?

34. La compagnia ha guadagnato molti soldi? Y

Did the company earn a lot of money?

36. Filippo spaventava Amanda? Y

Did Filippo scare Amanda?

40. Maria spingeva Andrea a rubare? Y

Did Maria forced Andrea to steal?

41. Amanda spaventava Filippo? Y

Did Amanda scare Filippo?

42. Il negozio ha venduto molti prodotti in Australia? N

Did the shop sell many products to Australia?

45. La società ha assunto una nuova segretaria? N

Did the firm hire a new secretary?

46. Andrea spingeva Maria a rubare? Y

Did Andrea force Maria to steal?

47. L'uffico prepara i documenti per le tasse? Y

Does the office prepare the tax forms?

48. Amanda andrà in Portogallo? N

Will Amanda go to Portugal?

52. Amanda intimoriva Filippo? N

Did Amanda frightened Filippo

54. Il teatro antico è vicino al parco? N

Is the old theatre close to the park?

60. Filippo intimoriva Amanda? N

Did Filippo scare Amanda?

61. Andrea istigava Maria a rubare? N

Did Andrea compel Maria to steal?

66. La compagnia telefonica ha chiuso l'ufficio in centro? Y

Did the telephone company close an office in the city centre?

Appendix 8

Content:

Self-paced Reading Task: Histograms with the distributions of the reading times and of the residuals for all the Linear-mixed Effects Models.

Figure 1

Histograms of the Reading Times Distributions in L1 Speakers in each Region

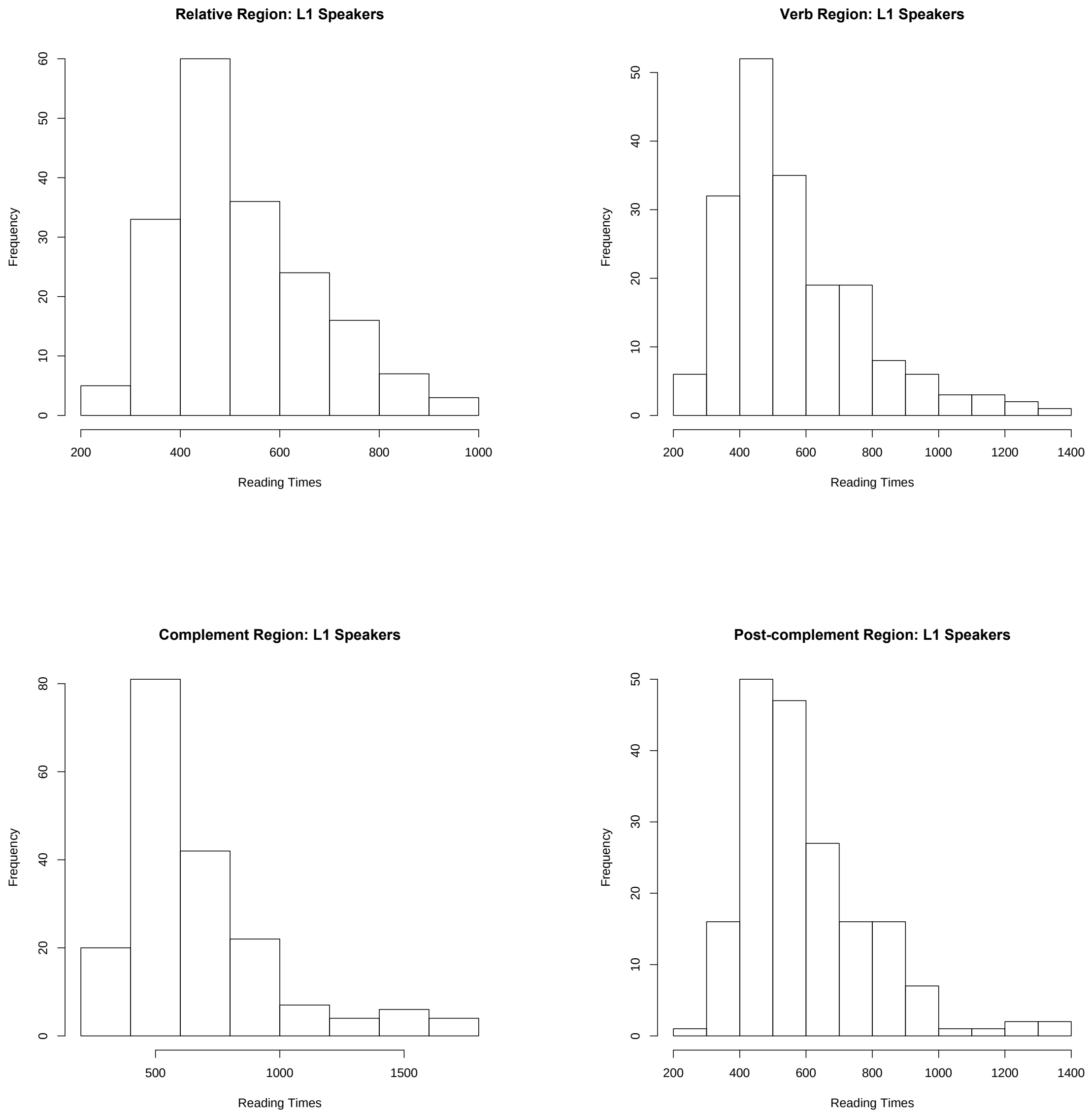


Figure 2

Histograms of the Reading Times Distributions in L2 Speakers in each Region

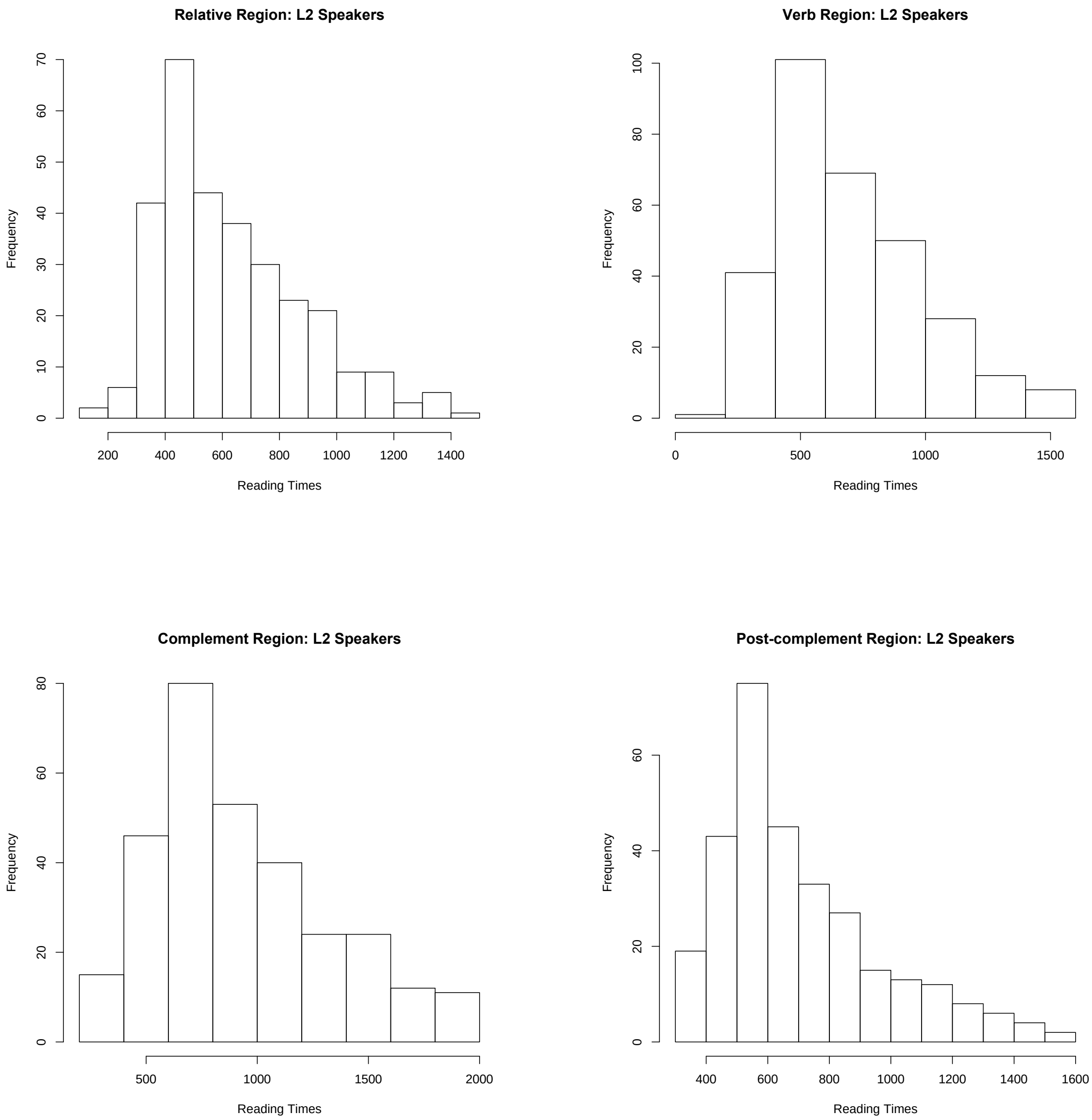


Figure 3

Histograms of the Residuals Distributions in the Statistical Model

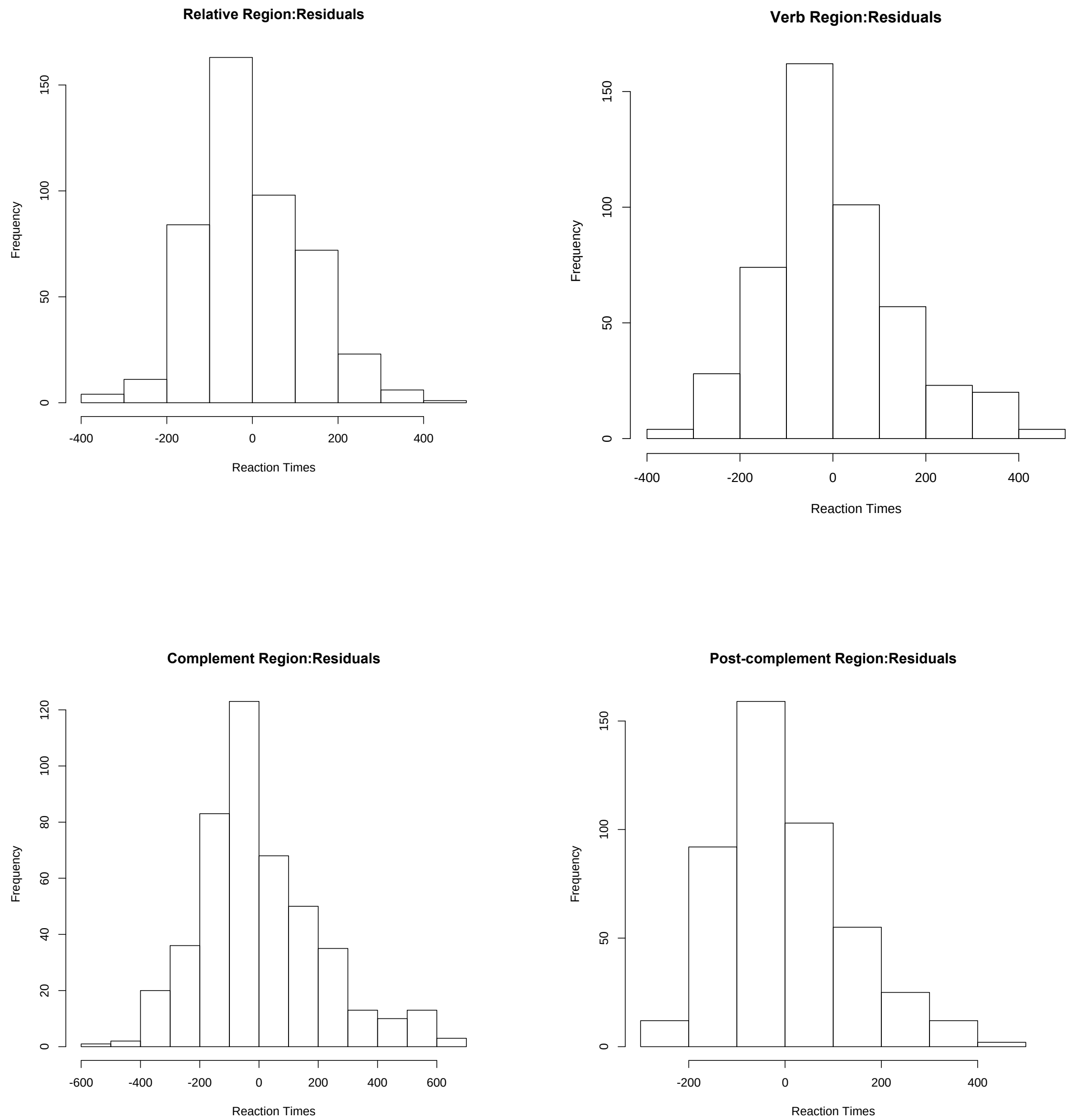


Figure 4

Verb Region: Distributions of Reading Times and Residuals for the L1 and L2 group

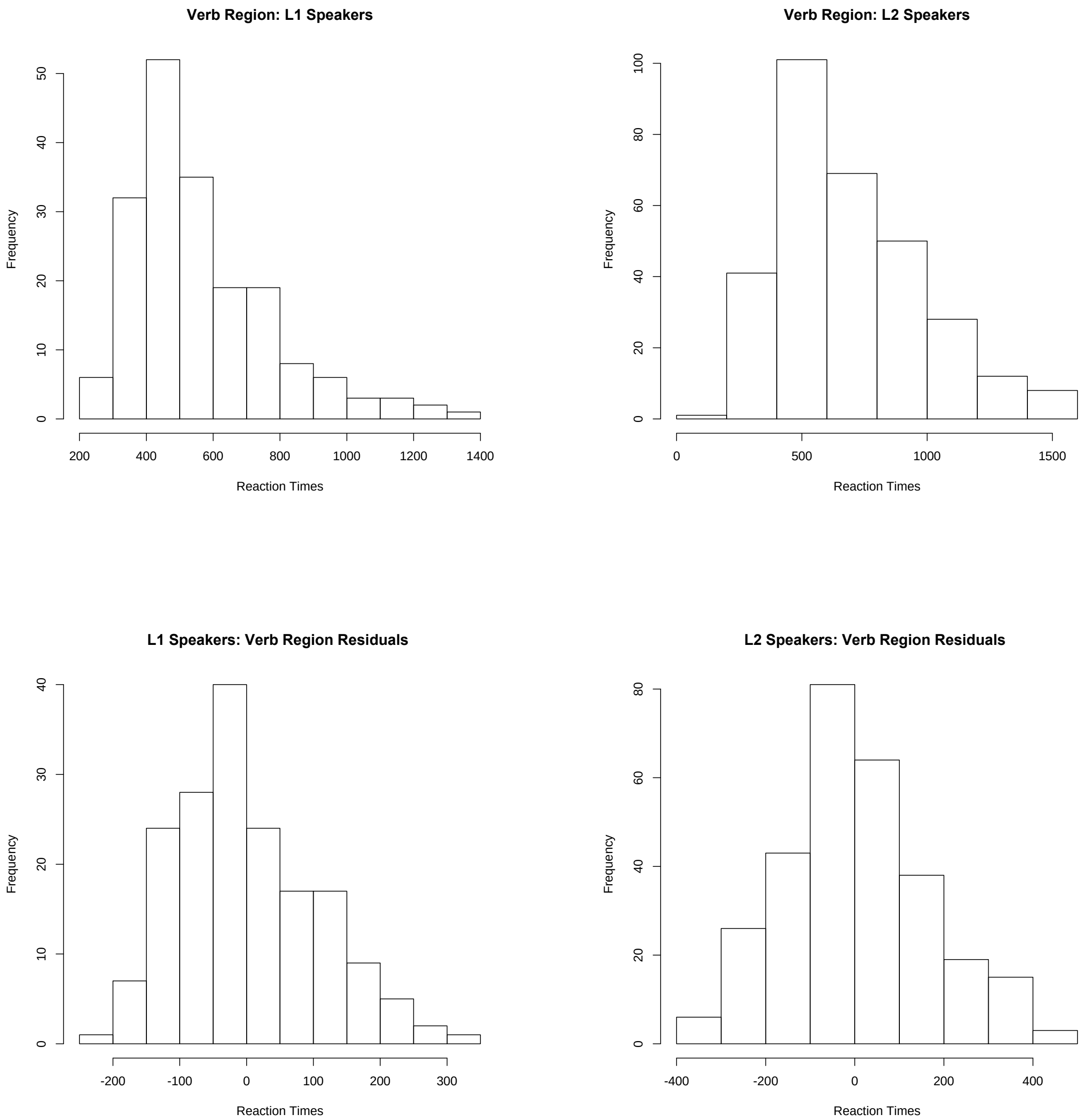


Figure 5

Complement Region: Distributions of Reading Times and Residuals for the L1 and L2 group

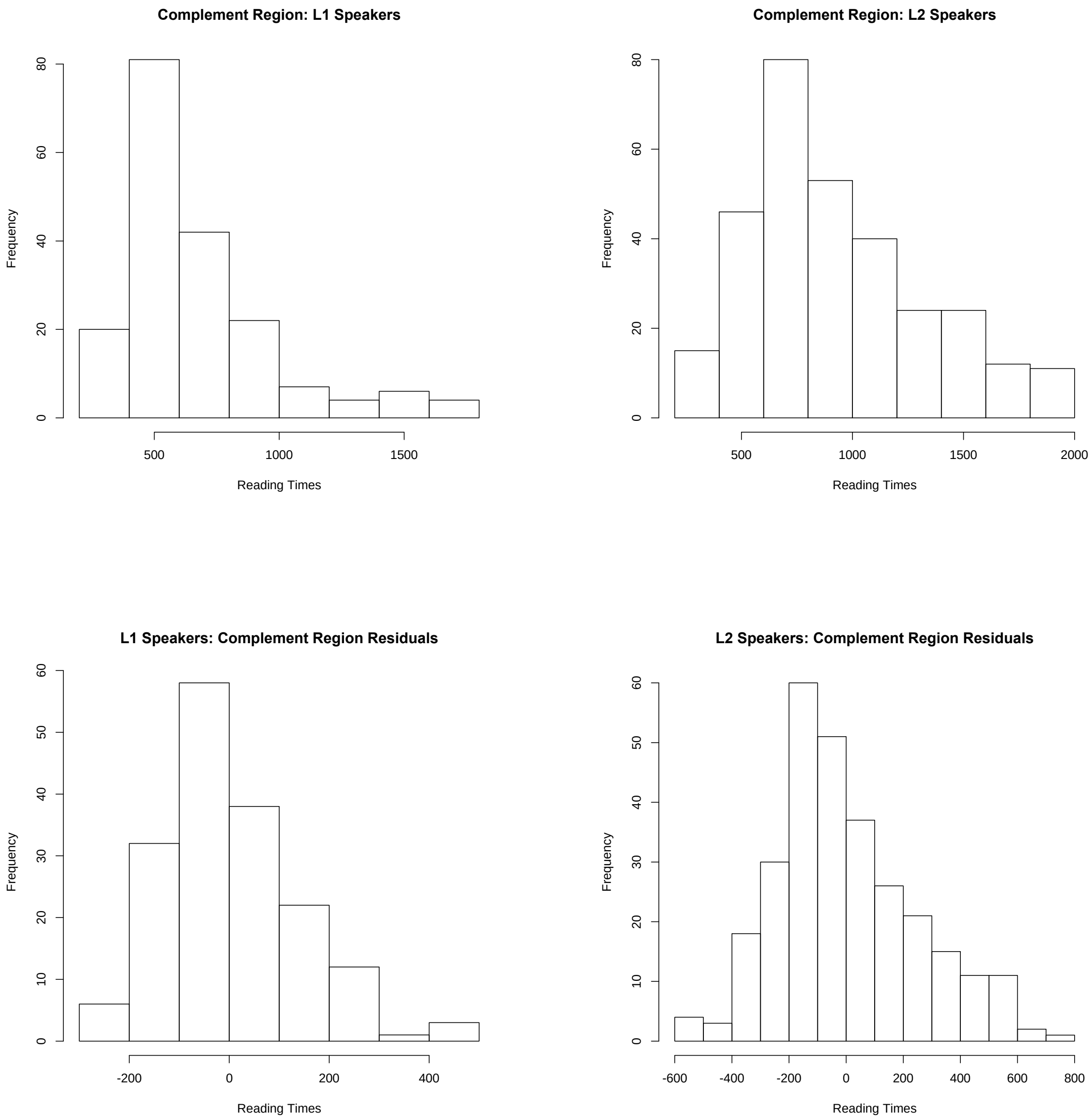


Figure 6

Histograms for the Relative Clause Region (Relative, Verb, Complement, post-Complement): Mean Reading Times for the Region and Model's Residuals