The effects of input modality and word difficulty on word recognition accuracy in Croatian - English bilinguals

Stojanović, Dragana

UNIVERSITY OF RIJEKA
FACULTY OF HUMANITIES AND SOCIAL SCIENCES
DEPARTMENT OF ENGLISH LANGUAGE AND LITERATURE

Dragana Stojanović
(JMBAG: 0009065388)

The effects of input modality and word difficulty on word recognition accuracy in Croatian - English bilinguals

Submitted in partial fulfilment of the requirements for the M.A. in English Language and Literature and German Language and Literature at the University of Rijeka

Supervisor: Associate Professor Dr. Anita Memišević, PhD

Rijeka, September 2024

IZJAVA O AUTORSTVU DIPLOMSKOG RADA

Ovim potvrđujem da sam osobno napisala diplomski rad pod naslovom *The effects of input modality and word difficulty on word recognition accuracy in Croatian - English bilinguals* i da sam njegova autorica.

Svi dijelovi rada, podatci ili ideje koje su u radu citirane ili se temelje na drugim izvorima u radu su jasno označeni kao takvi te su navedeni u popisu literature.

Dragana Stojanović

Abstract

The present study investigates how input modality (spoken vs. written) and word difficulty (word prevalence and type) affect word recognition accuracy in highly proficient Croatian - English bilinguals. By using a within – participant design, 20 bilinguals completed a lexical decision task in both written, and spoken modality. The results showed no significant difference in word recognition accuracy between modalities, which suggests that proficient bilinguals are able to process words equally well in both spoken and written forms. However, word prevalence and word type showed a significant impact on accuracy. Words with high prevalence were recognised more accurately than words with low prevalence, and cognates were recognised with the highest accuracy. These findings support the non-selective activation model of bilingual word recognition, in which both languages are activated simultaneously. This research provides insights into bilingual word processing and recognition, which may offer implications for future language learning.

Key words: bilingual word processing, word recognition, input modality, word prevalence, Croatian-English bilinguals

Table of contents

1 Introduction

Humans have a unique ability that no other biological species on this planet has – language. There is not one single definition as to what language is. However, one of the recognized definitions of language comes from the American anthropologist-linguist Edward Sapir, who defines language as "… a purely human and noninstinctive method of communicating ideas, emotions, and desires by means of a system of voluntarily produced symbols. These symbols are, in the first instance, auditory and they are produced by the so-called "organs of speech." " (Sapir, 1921, p. 4). From this citation, it is quite obvious that Edward Sapir put forth speech as a central aspect of language, and indeed, for centuries, humans only communicated through speech, i.e., by talking to one another. However, speech has its limitations. Walter J. Ong emphasises that ''sound exists only when it is going out of existence. It is not simply perishable but essentially evanescent, and it is sensed as evanescent. When I pronounce the word 'permanence', by the time I get to the '-pence', the 'perma-' is gone, and has to be gone.'' (Ong, 1982, p. 69), which means that speech is perishable, and requires an immediate communication.

On the other hand, the invention of writing changed this constraint many centuries ago, and written language overcame these limitations. Written language is not restricted by ''time, distance, and acquaintanceship'' as the famous linguist Steven Pinker (1995) writes (p. 16). A person may convey their ideas without the recipient having to be in close proximity to the sender, i.e., speaker. As Walter J. Ong (1982) writes ''Writing makes 'words' appear similar to things because we think of words as the visible marks signalling words to decoders: we can see and touch such inscribed 'words' in texts and books. Written words are residue. Oral tradition has no such residue or deposit'' (Ong, 1982, p. 11), which supports the idea of permeance of the written word, going beyond time and distance. Without this relatively new modality of language in which language could be expressed, we would not be able to grow and evolve as a species (Wolf et al., 2021).

Even though this research primarily focuses only on spoken and written modality, other modalities will be mentioned as well. Each of these modalities engages different areas of the brain and different cognitive processes, which in turn helps us better understand how the brain functions in respect to the language (Anderson et al. 2022). Understanding these modalities becomes even more important when speaking about the bilingual brain, because individuals who are bilingual must process at least two different language systems at the same time (in this research Croatian and English, respectively). Also, understanding how bilingual individuals

process languages in different modalities has become even more important in today's age of globalization, since globalization increases the prevalence of bilinguals in ''Europe (67%), Canada (55%), India (25%), and the United States (20%)'' (Byers-Heinlein et al., 2019). Although there is existing research that has studied either monolingual, or French – English, and Spanish – English bilinguals, we know of no research so far that has dealt with language modalities and the bilingual brain among the Croatian - English speaking group, despite the extensive research on language processing. There still remains a gap in understanding how different input modalities affect word recognition accuracy in individuals who are bilingual, especially for Croatian - English bilinguals, where it is unknown how spoken and written modalities, and the difficulty of the words that are presented in individuals' L2 affect their ability to recognize words (and non-words). This is why this research aims to address this gap by investigating the effects of input modality and word difficulty on word recognition accuracy in Croatian - English bilinguals.

2 Language modalities

In the introductory part of this study, different language modalities were mentioned. However, a clearer definition of language modality is in order. A language modality refers to various methods through which language is perceived and expressed (Anderson et al., 2022, p. 67). In simpler terms, it means that there are different ways in which humans can receive and send messages, i.e., encode and decode linguistic information. This is an exceptionally important aspect of linguistic studies because it enables our understanding of how language is processed, learned (which is especially important for educators) and used in different contexts (e.g., social context, professional context, the aforementioned educational context etc.). According to Anderson et al. (2022) there are four types of language modalities: spoken, signed, tactile and graphemic.

2.1 Spoken modality

Spoken modality, i.e., spoken language is also called the vocal – auditory language (Anderson et al., 2022). The more common name for this modality is 'speech', i.e., speaking. Speaking occurs when a sound is produced through the vocal tract and perceived by the listener via hearing. It is an essential and most natural human form of communication, often included as a key feature of various definitions of human language that scholars propose.[1] The main aspects of spoken language are phonetics (the study of sounds), prosody (the rhythm and intonation in speech, among other) and syntax (the study of sentence structure). Some common examples of spoken language include e.g. everyday conversations and lectures.

Leonard and Chang (2014) have studied how the brain processes speech. They concluded that different areas of the brain have specific roles that deal with understanding and interpreting the words that we hear, which enables us to better understand the complex processes behind the spoken language and its (very important) role in human communication. This understanding is fundamental for further research on how different modalities of language (spoken, written, signed, etc.) affect word recognition and language processing, which is the basis for this study as well.

[1] Starting with the definition of the notable linguist Edward Sapir who said that "Language is a purely human and noninstinctive method of communicating ideas, emotions, and desires by means of a system of voluntarily produced symbols. These symbols are, in the first instance, auditory and they are produced by the so-called "organs of speech."" (Sapir, 1921, p. 7)

2.2 Signed modality

Signed modality, i.e., signed language is also called the manual – visual language (Anderson et al., 2022). It is produced by making different types of movements, usually (and most often) through hands, face and body. The recipient (i.e., the 'listener') perceives it through seeing, i.e., eyes. This is a modality primarily used by the deaf and the heard-of-hearing communities. It is interesting to note that there is not one single (unified) signed language, but rather many signed languages across the globe. For example, there is the American Sign Language (ASL) in the United States, or the Hrvatski znakovni jezik (HZJ) in Croatia, among others.

The key features of every signed language are iconicity, spatial grammar and facial expressions (Anderson et al., 2022). Iconicity basically means that the signs often visually resemble their meaning. For example, the sign for 'to drink' in ASL involves a gesture, where the hand is shaped like it is holding a glass and approaches the mouth, which (visually) resembles the act of drinking. Or for example, the sign for the word 'knjiga' in HZJ involves gestures where the palms of a person signing open and close as if opening and closing a book. Spatial grammar is the second feature of the signed language. It uses space around the individual to indicate location and distance. For example, when a person who signs wants to portray a conversation between two people, he or she would turn their body to the left when 'speaking' as one character and to the right when 'speaking' as the other. The third and final feature of the signed language are facial expressions. Facial expressions are important because they express emotions and grammatical features. In ASL, when a person asks/signs a question, the eyebrows are raised and the head is tilted slightly forward, which indicates that a question is being asked. These main features (iconicity, spatial grammar and facial expressions) show just how complex signed languages are. Together, they make sign languages a complete and an efficient way of communication.

2.3 Tactile modality

Tactile modality, i.e., tactile sign language (or tactile signing and haptic communication) is also called the manual-tactile language (Anderson et al., 2022). It is produced and perceived through touch by individuals who are either blind, or deaf and blind. Tactile signing involves adaptations of sign language that are felt rather than seen, which enables communication through touch. Braille, which is a system that uses raised dots which

represent letters and numbers, is one of the great examples of a tactile language. A person moves their fingers over the dots to read the text. Other forms, i.e., methods of communicating are called tracking and tactile fingerspelling. Tracking is a way of communicating in which an individual puts their hand on the wrist of the person they are communicating with to communicate their message. Tactile fingerspelling is a method of communicating in which an individual traces letters and signs onto the palm of the individual with whom they are communicating (*Tactile Sign Language | What Is Tactile Sign Language?,* 2022). A case study conducted by Suzana Obretenova (2010) has shown that individuals who are deafblind use occipital cortical regions (together with posterior superior temporal and inferior frontal language areas), which is interesting, because this area is used for seeing in people who are not deafblind, and the interpreter, who could see and hear, and who also participated in this study (along with the deafblind individual) did not have these areas of the brain activated. This proves that the brain is an adaptable organ and that it can find new ways to maintain the function of language even when some senses are not present.

2.4 Graphemic modality

Graphemic modality, i.e., graphemic language is a language that is produced by making markings on a certain surface (paper, stone, etc.) and is perceived through seeing. Another name for this modality is written language, i.e., writing. The afore mentioned modalities are called 'primary modalities'. However, writing is considered to be a 'secondary modality' because it is a representation of another primary modality (speaking) (Anderson et. al, 2022). It is also different from the other primary modalities because it requires conscious *learning*, unlike e.g. spoken and signed modality which are *acquired*. (Anderson et. al, 2022). As previously mentioned in the introductory part of this thesis, writing differs from other modalities in that it transcends time and space, but also in the fact that written systems have been independently invented only a few times in history (Wright, 2014), and often borrowed from other language communities, like e.g. Arabic numbers that have been borrowed by other language users, which would imply that writing is a process that is very rare and complex and evolves a lot slower (and not spontaneously) than e.g. speaking (Anderson et al. 2022). Because of these differences, as well as the fact that, since it is not subconsciously acquired and thus considered not to be a part of the Universal Grammar, scholars, i.e., linguists do not treat written language the same as other modalities (Doner, n.d.). However, writing, i.e. graphemic language should not be considered an 'inferior' language, since there are numerous forms of graphemic

language that are not directly linked to the spoken language but do display some complex properties (Henner and Robinson, 2023).

It is important to note that in everyday communication, humans usually use more than just one modality at the same time during communication, for example, when they use hand gestures during speaking. This would mean that language is in its essence ''multimodal'' (Henner and Robinson, 2023, p. 14). Seldomly will an individual only use one modality when communicating.

For the purpose of this thesis, the primary focus will be on the spoken and written modalities because these are the most common modalities that occur in communication and education in most bilingual settings, especially in Croatian-English bilingual settings (schools, social media, everyday conversations, etc.). Also, in bilingual education, the emphasis should be on developing proficiency in skills such as reading and writing (which falls under the scope of graphemic, i.e., written modality) as well as listening and speaking (which falls under the scope of spoken modality). Proficiency in reading, writing, listening and speaking is considered to be an ultimate goal in reaching an overall language competence (Hymes, 1972). This would suggest that the findings of this research, i.e., thesis would have a significant impact on practices and policies in education and could provide further practical applications in e.g. designing and improving the curriculum for the English language in Croatian schools.

3 Processing input in written and spoken modalities

As mentioned in the introductory part, the primary focus of this thesis are the written and spoken modalities. However, the brain's ability to process linguistic input through these modalities is a complex process that includes different cognitive and neurological processes, which are the scope of research for psycholinguistics and neurolinguistics (Gazzaniga et al., 2015). The two language sciences work together in explaining how these complex processes are vital if one wants to achieve an effective communication in everyday life, i.e. if one wants to successfully send, but also receive a message. Even though the message that is conveyed is the same (identical) across both written and spoken modality, the brain must first process the input through modality-specific processing systems (Wolf et al., 2021).

Findings of neurolinguistics support the existence of some shared brain regions when processing both written and spoken languages, such as the superior temporal sulcus (STS), which is responsible for decoding and comprehending input in spoken and written modalities (Wilson et al., 2018).

Psycholinguistics, additionally, tries to provide an explanation for the cognitive processes that try to explain how language is represented in the mind. One such process is the notion of the 'mental lexicon', which will be further discussed in this chapter. All of these processes help scholars gain better understanding of how the brain transforms symbols (in written language), as well as sounds (in spoken language) into meaningful language.

3.1 Defining word recognition

In order to explain which system the brain uses to derive meaning from spoken and written modalities, one must first start from defining a term crucial for this thesis: word recognition. Word recognition is a process by which the brain identifies and interprets individual words in both spoken and written form, converting them into linguistic representations (Moreno & van Orden, 2002). It involves several steps, i.e., stages, which are the perception of auditory or visual word forms, then matching these forms to stored representations in the mental lexicon, and, subsequently, the retrieval of meaning. This is a crucial process for listening and reading, because it allows individuals to quickly and accurately understand the provided input. Understanding word recognition is essential for studying the

deeper processing mechanisms of the brain. However, it is only the first step in explaining the more advanced cognitive processes that are involved, which will be explained in more detail below.

The process of word recognition begins with the brain's ability to decode linguistic input, a process called perceptual analysis (Gazzaniga et al, 2015). When encountering an input in the *spoken* modality, the listener first has to ''decode the acoustic input'' (Gazzaniga et al., 2015, p. 480). The input is then converted into the phonological format because ''that is how the lexical representations of auditory word forms are stored in the mental lexicon'' (Gazzaniga et al., 2015, p. 480). A similar process occurs with the *written* modality, however, due to the different input (written instead of spoken), there is a difference in the earlier processing steps. Here the reader first must extract and decode the orthographic units of the word from the visual input. According to Gazzaniga et al. (2015), these orthographic units can be 'mapped' onto, i.e., linked to *orthographic* word forms in the lexicon (p. 481), or converted into phonological units, which then activate the *phonological* word forms in the lexicon, similar to the process that is used in spoken word comprehension. For example, in this initial stage, an individual may encounter the word 'cat' in either spoken or written modality. In this initial stage the brain processes raw sensory, i.e., perceptual input, in this case either sounds, i.e., phonemes that constitute the word 'cat' in spoken modality, or graphemes in the written modality.

However, according to Wolf et al. (2021), the processing system our brain uses to decode words in these modalities is 'abstract' rather than 'perceptual' in nature, which would suggest that higher-level cognitive operations involve accessing and integrating various types of representational information (such as phonological, semantic, orthographic and syntactic) that are stored in the mental lexicon, rather than relying solely on sensory input (what one sees and hears). After the brain receives the input from the senses such as the hearing and seeing (perceptual), it moves to a 'deeper' level of processing (abstract). This change from the initial perceptual to the abstract processing is crucial for the brain to be able to recognise and interpret words. At this deeper level of processing, even before the input is fully acknowledged, the brain activates potential word candidates and converts the initial input into activated lexical representations, which is a process called *lexical activation* (Taft, 2001). In the example with the word 'cat', during this process, words that are similar, such as 'cap', 'cab', etc., are also retrieved (this is called the lexical neighbourhood). The brain then decides which word in its

lexicon best matches the received input, i.e., which is the most appropriate word based on the input received.[2] This process is called *lexical competition* (Harley, 2005), which would mean that these words compete until the word that matches the input (the word 'cat') is found, i.e., activated. Once the conversion of the input is done, the corresponding, i.e., the fully matching lexical representations in the mental lexicon can be accessed, a process called *lexical access* (Harley, 2005). All of these information are connected in the brain and help with fully understanding the word 'cat' efficiently. This account is supported by influential theoretical models of word recognition such as the Cohort model (Marslen-Wilson, 1987) and the Interactive Activation Model (IAC) (McClelland & Rumelhart, 1981).

3.2 Understanding the mental lexicon

The deeper, i.e., abstract processing is closely tied to the concept of the mental lexicon, a term first hinted at by the linguist Ann Treisman (1961) in her doctoral thesis (as cited in Coltheart et al., 2001). Treisman then used the term 'dictionary', instead of 'lexicon':

> "There is a single channel for recognising words, presumably comprising the matching of signals with some kind of 'dictionary' ... some of whose units have their thresholds for activation permanently or temporarily lowered" (p. 210, as cited in Coltheart et al., 2001)

Another famous linguist R.C. Oldfield (1966) then described the mental lexicon, again, as a mental *dictionary* ''in which information about word meaning is retrieved'' (Sanches et al., 2017), however, many scholars today argue against defining the lexicon as a mere *dictionary*, because they deem it too narrow of an explanation, and compare it more to a World Wide Web where information can be updated and changed (Brown, 2006; Farahian, 2011; Libben, 2008). Gregory W. Yelland (1994) proposes a definition of the mental lexicon as ''a memory system dedicated to the storage of all that we know about the words in our vocabulary'' (p. 1). Wolf et al. (2021) state that the mental lexicon

> ''… contains an entry for every word an individual knows. Each lexical entry within this lexicon consists of different types of representational information. These types of representations encompass semantic

[2] In monolingual word recognition, many possible words initially become active when encountered with a word in either modality, and the individual is usually not aware of them; only the word that is (eventually) recognised comes to awarness (supported by the Cohort model (spoken word) and the IAC model (written word)) (Dijkstra, 2005).

(meaning), phonological (sound), orthographic (spelling) and syntactic (grammatical) information" (p. 12)

An average native American-English speaker may possess in their vocabulary around 42000 words (Brysbaert et al., 2016) and can recognise, as well as produce, three words per second (Gazzaniga et al., 2015). Given the size of the vocabulary of the individual, as well as the speed of processing, there arises a fundamental question as to how these words are stored in the mental lexicon. Due to these reasons, the mental lexicon has to be efficiently organised, and the most efficient way of organisation is in the form of the smallest meaningful unit in a language, called morpheme. It is also the representational unit in the mental lexicon. An example would be 'happy, 'unhappy', 'unhappily'. The root of these words is 'happy', which forms one morpheme, the prefix (and the morpheme in itself) 'un' changes the meaning of the root morpheme and becomes a new word, and 'unhappily' is another word that consists of three morphemes. This is the first organisational principle of the mental lexicon, the second being that the words that are used more frequently are retrieved faster than the words that are used less frequently (an example will be provided in the following chapter of this thesis titled "Lexical quality hypothesis"). The third principle is the previously mentioned lexical neighbourhood, which includes words that differ from a given word by only one phoneme or letter (an example provided in the section "Defining word recognition" in this thesis). A phoneme is the smallest unit of sound that can change the meaning of a word. According to Gazzaniga et al. (2015), research has shown that words with more lexical neighbours are "identified more slowly during language comprehension than words with few neighbours" (p. 476) (for example, 'cat' has many neighbours, but 'jeans' has only a few), as already mentioned in previous chapter titled "Understanding the mental lexicon". The fourth, and last, organisational principle, would be that the mental lexicon is organised based on the semantic relationship between words, which was supported through semantic priming studies that use lexical decision tasks to show how words are recognised, such as the one by Holderbaum & de Salles (2011), who demonstrated that both children and adults responded faster to semantically related word pairs. This would suggest that related words are stored closely together in the mental lexicon.

3.3 Lexical quality hypothesis

According to Perfetti (2007, as cited in Wolf et al., 2021), there is a crucial hypothesis, called 'lexical quality hypothesis', within the framework of the mental lexicon, which states that these different types of representations (phonological, semantical, syntactical, and orthographical representations) in a single lexical entry, i.e., word, can vary in their completeness and precision. This means that different aspect of the same word can differ in their 'lexical quality'. For example, an individual may have a clear phonological representation of the word 'cat' in their brain but may have a less precise or even a non-existent orthographic one, if the individual, e.g., cannot read. Another crucial hypothesis according to Perfetti (2007, as cited in Wolf et al., 2021) is that the processing of higher quality lexical representations (i.e., those that are more complete and more precisely defined) occurs more efficiently than the lower quality lexical representations. To return to the example mentioned above: the word 'cat'. We could compare it with the word 'metamorphosis'. The word 'cat' is a word that would be processed faster by the majority of people than the word 'metamorphosis' would, because of its high lexical representation. An individual would be much more certain about the word 'cat' in terms of its meaning, spelling, sound and usage in a sentence, in comparison to the word 'metamorphosis'. These hypotheses bring about the following questions: if the quality of the modality-specific representations (spoken vs. written) can differ, could it be that processing a word that is presented in one modality is more effiective than processing the same word in another modality?

There has been some research that suggests that familiar words are more accurately and faster, i.e., more efficiently, recognised when they are presented in written, rather than in spoken modality (Connine, Mullennix, Shernoff, & Yelen, 1990; Turner, Valentine, & Ellis, 1998). However, the research on influence of the modality in which a word is presented on word recognition is still scarce in terms of bilingual brain, and mostly focuses on L2 word recognition in only one modality (Cornut, 2021), especially in the context of Croatian-English bilinguals. Here, it becomes important to investigate whether bilinguals, or in this case, Croatian-English bilinguals exhibit the same or at least similar patterns of modality-specific word recognition efficiency. Since the bilinguals possess and process two language systems simultaneously, the modality in which a word is presented may significantly influence the accuracy of their word recognition.

4 Defining and classifying bilingualism

Defining the term ‘bilingualism’ remains to be a complex challenge in linguistics (Bialystok, 2010). It is still unclear, as to what degree must an individual be bilingual in order to fall under this term, i.e., category. Must an individual be equally proficient in both languages, or is it enough for an individual to understand a language perfectly but not be able to produce it, to be considered bilingual? There are many factors, such as linguistic, cultural, social and psychological, that must be taken into account when defining bilingualism (Hamers & Blanc, 1983; Perregaux, 1994, as cited in Cornut, 2021). There are two main positions that discuss this complexity: the *maximalist* and the *minimalist position*. The first position, called the *maximalist position*, is introduced by Christophersen (1948) and states that a bilingual person is ’’a person who knows two languages with approximately the same degree of perfection as unilingual speakers of those languages’’ (p. 4). The second position, the *minimalist position*, as introduced by Macnamara (1967, as cited in Cornut, 2021) argues that an individual is ’’bilingual if he/she has a minimum of L2 proficiency in at least one of the four language skills (i.e., speaking, spelling, reading and comprehension)’’ (p. 8). There are, however, certain definitions along the spectrum or categories in between these two positions called the bilingual *continuum* (Rothman et al., n.d.), a noteworthy one being by Francois Grosjean (1994) who defined bilinguals as ’’those people who use two (or more) languages (or dialects) in their everyday lives’’ (p. 164) regardless of the way, i.e., mode and age of acquisition of that language. Nevertheless, to expect a perfect mastery of both languages equally, is not something that is easily achievable, and may be considered ‘utopian’ (Cornut, 2021). Therefore, it is vital to accurately classify bilinguals, i.e., bilingualism, in order to properly define the characteristics of the population being studied.

There are several classifications proposed in the literature, depending on the parameters used to distinguish bilingual individuals. This thesis will focus on classifications that are most relevant to the research objectives of this thesis, such as the age of acquisition and the context in which the language is used. The first classification worth mentioning takes into account the age of acquisition of both languages as the key factor. Scholars differentiate between *simultaneous bilingualism*, which is the acquisition of both languages simultaneously before the age of three (McLaughlin, 1995), as well as *successive bilingualism*, which is the acquisition of the second language (L2) after the age of three or older (McLaughlin, 1995), as

well as *late bilingualism,* which refers to the acquisition of the second language either in adolescence or adulthood (Adler, 1977, Moradi, 2014, as cited in Cornut, 2021). The second classification, relevant in general, but also for the purpose of this research, refers to the context in which L2 was learned. Here, one differentiates between the *compound* and *coordinate bilingualism.* Compound bilingualism refers to ''an individual who learns two languages in the same environment so that he/she acquires one notion with two verbal expressions'' (D'Acierno, 1990, p. 1). This could refer to a child who acquires two languages in the same household. Coordinate bilingualism refers to ''an individual who acquires the two languages in different contexts (e.g., home and school), so the words of the two languages belong to separate and independent systems.'' (D'Acierno, 1990, p. 1). This is the most common form of bilingualism in Croatian students, according to a research by Erk and Ručević (2021), who included 147 young English learners, out of which all of the participants started learning English either in kindergarten, in school, or attended courses in foreign language schools. It is important to mention that the age when the participants started to learn English ranged from 2 years and 11 months to 7 years and 3 months (p. 149). Regardless of the definitions and classifications, it is widely acknowledged that both written and oral communication skills are essential for becoming skilled in a language, and these skills rely heavily on effective language processing, which will be covered in the following chapter.

4.1 Bilingual mental lexicon and word recognition

There has been an ongoing debate whether bilingual individuals possess one lexicon for both languages, or separate mental lexicons. Some research discuss the existence of separate lexicons for both languages and believe that information from one language can be transferred to the other through translation. However, there is research that suggests that languages are stored in one single mental lexicon, and that linguistic information is stored in one semantic system. Some research, however, proposes a more nuanced theory, where there are separate lexicons. However, there is some evidence of shared storage (Bastkowski, 20023; Singleton, 1999, as cited in Masrai, 2005). Singleton (1999) also proposes that, although stored separately, the L1 and L2 lexicons are connected with each other and do indeed 'communicate' with one another. This would mean that ''while phonological and morphosyntactic forms differ across languages, meanings and/or concepts are largely, if not completely, shared (d. Costa, 2005; Kroll & Stewart, 1994)'' (Pavlenko, 2009, p. 125).

Building on this foundation, there arises a question on how different monolingual and bilingual lexicons operate, whether similarly or differently and whether bilingual lexical access operates similarly to monolingual, whether access is restricted to only one language at the time (language-selective access), or whether both languages are activated at the same time, i.e., simultaneously (language non-selective access) (Dijkstra, 2005; de Groot et al, 2000). Although the basic mechanisms function the same in both mono- and bilinguals (e.g., phonological decoding), empirical studies suggest that during reading and listening, i.e., encountering language input in different modalities, bilinguals activate both languages (language non-selective lexical access) (Dijkstra, 2005). This is supported by several models that have been developed, which try to explain how bilinguals manage and process lexical access when there is more than one language involved, the most important ones being the Bilingual Interactive Activation (BIA) model, and its successor, the BIA + model (Dijkstra & van Heuven, 1998; Dijkstra & van Heuven, 2002), which suggest that both languages are activated simultaneously, compete and interact with one another, and are shaped by various factors such as context, individual's language proficiency and task demands.

4.2 Bilingual Interactive Activation model (BIA) and Bilingual Interactive Activation model plus (BIA+)

The ongoing debate on the bilingual mental lexicon has led to the development of several models on word recognition in bilinguals, the most notable ones (and the most relevant for this research) being the Bilingual Interactive Activation model (the BIA model), and its successor, the BIA+ model (Dijkstra & van Heuven, 1998).

The BIA model was a bilingual response to McClelland and Rumelhart's Interactive Activation and Competition model (1981) . It assumes a shared lexicon for both languages, and was designed to explain how bilinguals recognize written words by stating that both languages are activated simultaneously, i.e., at the same time, during word recognition. Such evidence is found, among others, through the use of stimuli such as *cognates*, which are ''words that have an orthography and a meaning that are similar or identical across languages, such as tomato in English and tomaat in Dutch'' (Dijkstra, 2005, p. 179). For example, in the context of a Croatian-English bilingual who encounters an English word 'problem' while reading, this is what happens: since the BIA model assumes multiple levels of processing, it begins by

activating letter features such as lines and curves that make a letter (the round shape of the letter ’p’, for example). This then triggers specific *letter nodes* for each position of a letter in a word. The word ‘problem’ activates the letters ‘p’, ‘r’, ‘o’, ‘b’, ‘l’, ‘e’, and ‘m’ (in a correct position in a word). These letter nodes then trigger *word nodes,* which activate potentially both the English word ‘problem’, as well as the Croatian word ‘problem’ (which, in this case, is a cognate word and shares the same meaning and spelling in both languages). Initially, since words from both languages are activated simultaneously, this leads to a competition between them. This is where *language nodes* come into play, which act as filters that help the brain decode the correct language in use based on contextual cues. In the above-mentioned case of the Croatian-English bilingual, the fact that they know that the text they are reading is in English, the English language node will act as a suppressant for the words in the other language not in use at the time, which is the Croatian ‘problem’. The outcome of the competition is based on the frequency of use, meaning that highly frequent words will be recognised faster and more easily (Cornut, 2021).

The BIA model focused purely on orthographic level of representation, however, as Dijkstra and van Heuven (2002) mention: ’’we assume that the distinction between word retrieval and task/decision system will also be valid for bilingual auditory word recognition and bilingual word production (cf. Green, 1998)’’ (p. 193). The BIA+ model expands on the BIA further by also including phonological and semantic levels of representation within a single integrated, i.e., shared lexicon for both languages. Even though BIA+ is an updated version of the BIA, which would suggest that BIA+ is still quite similar to the BIA, there are some key differences: the first difference was already mentioned, and it is the addition of phonological and semantic levels of representation to the orthographic one, however, these new representations ’’constitute a specific subsystem the activation of which is *delayed* compared to L1 representations’’ (Cornut, 2021, p. 15). The second difference would be differentiating between linguistic and non-linguistic context effects. An example of linguistic context would be that when a bilingual reads in a particular language, their brain is ‘attuned’ to that language and its words (along with its grammar and syntax, among others), and in the BIA+ model, the linguistic context (factors that are related to language) affects how words are recognised based on the language in use, the frequency of certain words from that language, etc. The non-linguistic context (factors that are external, i.e., not related to the language), according to Dijkstra & van Heuven (2002), influences how decisions are made about word recognition,

however, it does not directly affect the activation levels of words during the recognition process itself:

> ”A second option is that non-linguistic context information does not affect the activity in the identification system itself, but leads to an adaptation of decision criteria only… The BIA+ model proposes that (in contrast to linguistic effects) non-linguistic context effects in reading can be accounted for only by the second option.” (p. 15)

This last difference would be replacing the top-down inhibition mechanism from language nodes by word nodes (Cornut, 2021). As with any model, there is some debate over some of the assumptions that the BIA+ model proposes. However, it has been supported by various experiments and studies, which demonstrated that this model explains well how bilinguals process words in different languages (Dijkstra et al., 1999; Duyck, 2005, as cited in Cornut, 2021).

5 Previous research on L2 word recognition

The majority of studies on L2 word recognition primarily focused on researching written word recognition. However the need to include spoken word recognition has been an interest of linguistic research only in relatively recent years. When it comes to empirical data regarding word recognition in L2, a study by Van Heuven et al. (1998) highlights that words are recognised faster than pseudowords. They found that L1 neighbourhood effects can inhibit word recognition in L2, while L2 neighbourhood effects can facilitate it. Additionally, a study by Oganian et al. (2016) found that participants responded faster to words than pseudowords, with fewer errors. Additionally, Dijkstra et al. (2010), found that cognates are recognised faster and with fewer errors than non-cognates, with identical cognates (e.g., the cognate 'hotel' in Croatian and English) being recognised the fastest. Additional research on this topic was conducted by Frances et al. (2021) whose findings suggest that both orthographic and phonological similarity play a key role in word recognition. Words that are similar in both their spelling and sound across languages (e.g., the word 'chocolate' in English and Spanish) are recognised more easily than those with low similarity (e.g., 'onion' in English and 'cebolla' in Spanish), where both the spelling and pronunciation are different. However, when orthographic and phonological similarities did not align (if a word looked similar but sounded different), it hindered word recognition. When it comes to word prevalence, Brysbaert et al. (2016) introduced the concept of word prevalence (how widely a word is known within a population) and found that word prevalence strongly influences lexical decision times, independent of word frequency. Words that are known by a larger portion of the population are recognized faster, highlighting the importance of considering word prevalence as a key factor in word processing.

However, research that takes modality into account when it comes to word recognition in bilinguals (or multilinguals) remains scarce. Veivo (2015) conducted an experiment with a group of Finnish learners of French, who were bilingual in Finnish and English, where participants showed higher accuracy in recognising written words compared to the spoken ones. Findings from Cornut (2021) suggest that the way bilinguals process and recognise words in their L2 is not entirely independent of the modality in which the words are presented. The study also found that cognate-facilitation effect is modality dependent, i.e., it aids word recognition in written modality, but hinders recognition in spoken modality. Additionally, Frances et al. (2021) emphasised that word recognition is easier when both the modality and the type of similarity align. For example, words with orthographic similarity are recognised

more accurately when presented in written modality, and words with phonological similarity are recognised more accurately when heard. However, when similarity crosses modalities (e.g., when a word sounds similar in both languages but is presented written form), word recognition becomes more difficult.

In summary, while extensive research has been conducted on word recognition in L2, it mostly focused on written modality, which raised the need for including spoken modality in order to gain a more comprehensive understanding. Findings from various studies show that factors such as word type and word frequency all play significant roles in word recognition across different modalities. However,, as some studies show, due to the modality-dependent nature of these effects, there is a need for further research that would try to address and clarify how bilinguals process words in different conditions. Additionally, no known research deals with these factors among the Croatian-English bilingual group, which is why addressing these gaps may prove beneficial for informing practical applications in, e.g., language teaching and assessment.

6 Methodology

6.1 Aim

The present study aims to investigate how input modality and word difficulty affect the accuracy with which Croatian – English bilinguals recognise words.

6.2 Research questions

The study aims to answer the following three research questions:

1. Does modality (written vs. spoken) affect word recognition accuracy in highly proficient Croatian-English bilinguals?
2. Does word prevalence (familiar vs. less familiar) affect word recognition accuracy in highly proficient Croatian-English bilinguals?
3. Does word type (cognates, non-cognates, pseudowords) affect word recognition accuracy in highly proficient Croatian-English bilinguals?

6.3 Research method

The research employed a within-participants design, meaning that each participant was exposed to all conditions of the experiment (both modalities and all word types, to increase statistical power by ensuring that each participant served as their own control) with two primary factors: modality (written vs. spoken) and word difficulty, which was divided into word prevalence (familiar vs. less familiar) and word type (cognates, non-cognates, pseudowords). The study was conducted by means of a lexical-decision task, where 20 Croatian-English bilinguals were divided into Group A and Group B (10 participants in each group) and presented with words and pseudowords, i.e., non-words in different modalities (spoken and written) for each group and had to carry out a binary decision task, i.e., a lexical decision, and decide whether the presented items were real words or pseudowords. The average response time was measured. Given the high proficiency of the Croatian – English bilingual participants in this study, the stimuli, i.e., words with high prevalence and words with low prevalence

(including cognates) were selected from ''Word prevalence measures for 62,000 English words'' by Brysbaert et al. (2019), which derived its norms from English native speakers, but could likely reflect the participants' familiarity with the English vocabulary. The non-words were generated by using UniPseudo, a universal pseudoword generator (Barra et al., 2023). The complete list of stimuli, including word prevalence and type, can be found in the Appendix section of this thesis.

6.3.1 Participants

20 participants took part in this research, 12 male and 8 female, all of them being of legal age (18 – 45 years old, M age = 30 years old, SD = 7.43), all of them having Croatian as their mother tongue. The justification for this sample is to ensure a manageable and detailed analysis within the scope of this thesis, i.e., research. They had normal or corrected-to-normal vision and hearing and gave their written consent prior to the test. Participants were not awarded or paid for their participation in this research. The data collected was anonymous to protect their privacy. The participants were recruited from a variety of backgrounds and age groups to ensure a diverse sample, and inclusion criteria was based on high-proficiency test results that were administered to them prior to this research. They were divided into two groups (Group A and Group B) and were given a general proficiency test in form of a *C-test* (which had to have a score of 75% for a participant to be considered highly proficient), a *questionnaire* that included questions about their language background, frequency of English use, along with contexts in which they use, as well as hear/read English, and a self-assessed proficiency in the four language skills (listening, reading, writing and speaking), as well as the *word recognition test* itself.

Based on their questionnaire responses, there were 9 (45%) participants that attended, or used to attend, grammar schools in Rijeka, and 11 (55%) that attended, or used to attend vocational schools in Rijeka. There were 6 (30%) participants that stated that they started learning English between ages 0-3, 12 (60%) of them that stated that they started learning English between the ages of 4-7, and 2 (10%) that stated that they started learning English between the ages 8-12. There were 17 participants that reported primarily learning English either in their formal education, i.e., their elementary school (70%) or in language courses (15%). The rest (15%) stated that they learned English either through television or video games. The majority of participants, 16 of them, (80%) stated that they use English often in their everyday life, and 4

(20%) participants stated that they use English sometimes. When it came to language use, all of them reported using English in contexts of school, social media, and watching TV. There were two participants that added 'reading books/articles' in English under their 'language use'. When it comes to 'listening habits' and 'reading habits', there were 15 (75%) participants that stated that they often hear English, 4 (20%) that stated that they sometimes hear English and one (5%) participant that stated that they always hear English. All of the participants reported listening to English in school, listening to music, on social media, watching YouTube videos and playing video games. Two participants (10%) added listening to podcasts or radio. When it came to the 'reading habits' section, all of them reported that they often read in English and reported social media posts as the type of materials in English that they read. Additionally, there were three (15%) participants who added reading books in English. Under language proficiency, nearly all of the participants rated themselves highly in terms of their proficiency in listening, speaking, reading and writing skills:

Score	Listening	Reading	Writing	Speaking
1	0	0	0	0
2	0	0	0	0
3	0	0	3	1
4	3	5	14	17
5	17	15	3	2
	4.85	4.75	4	4.05

Table 1: Self-reported weighted average proficiency in listening, reading, writing and speaking skills

The data in Table 1 shows that participants rated themselves highly on all four skills, with listening being the highest rated skill with an average score of 4.85 and speaking with the lowest score among them with 4.05, respectively. The self-reported proficiency segment in the questionnaire suggests that participants possess a high proficiency in the English language, which was one of the main prerequisites for solving the tasks in this research. None of the participants reported taking a standardised English proficiency test.

6.3.2 Test materials and procedure

Questionnaire: The questionnaire was completed two days prior to the main testing phase (along with the C-test) and it contained a total of 15 questions aimed to collect detailed information about the participants' age, language background, and usage habits in respect to the English language. The questionnaire was divided into six sections, each aiming to capture different aspects of participants' encounter and experience with the English language. It entailed: participant information (age, gender, native language and educational background, as well as any vision/hearing impairments), language background (age when they started learning English, context in which English was learned, i.e., school, language course, etc.), language use (frequency and context of English use), language proficiency (both self-reported on a five-point Likert scale, as well as any standardised proficiency test and their score), as well as their listening and reading habits (frequency and context in which they encounter English in spoken/written modality).The questionnaire was designed to be easy yet comprehensive, i.e., containing information relevant for this study, and took approximately 10 minutes to complete. It was administered in a written format, prior to undertaking language proficiency (C-test) assessment, so as not to influence the participants' attitude towards any of the segments required in the questionnaire. A full copy of the questionnaire is included in the Appendix.

C-test: The C-test was administered as a language proficiency test, two days prior to the main testing phase. The test consisted of 5 short passages, each containing a 20-word gap, where the second half of the word was omitted. Participants were given the instruction to complete the missing parts of the words in order to form appropriate words. The participants solved the test in a quiet and supervised environment, and had a maximum of 20 minutes to complete the test. A successful completion (75% correct answers or more) of the C-test was a prerequisite for moving onto the main test, i.e., research. Each correct answer was awarded 1 point, out of total of 100 points. C-test was chosen due to its reliability when measuring language proficiency, especially for L2 learners.

Word recognition test: For this experiment, the 20 participants were divided into two groups (Group A and Group B), each group corresponding to the modality that was tested (visual, i.e., written vs. auditory, i.e., spoken). On each trial of the word recognition test,

participants from group A were asked to respond to 30 individual target words in their L2, i.e., English (10 cognates, 10 non-cognates and 10 pseudowords) that were presented visually and had to decide whether the presented stimulus was a word or non-word, i.e., pseudoword. The participants from group B had the same task, but in spoken instead of written modality. After solving the test for the corresponding groups, the other modality was administered, which meant that after completing the written word recognition task, the group A had to solve the spoken modality as well. The same went for group B. This approach was selected in order to investigate a potential influence of modality on word recognition accuracy, as well as to avoid random variations and individual differences among groups, and to control for potential order effects. The tests were developed using PsychoPy (Pierce et al., 2019) and hosted on Pavlovia Surveys (Open Science Tools, Nottingham, UK). The audio stimuli were recorded using the software programme Audacity© (Audacity - Copyright © 1999-2018) by the researcher, in the standardised American-English accent. All the stimuli, in both modalities, were presented only once and in randomised order (by using PsychoPy's built in randomisation features), to prevent learning effects and to ensure that participants' responses were indeed true word recognition abilities, rather than memorisation.

Procedure: The experiment, i.e., the word recognition test was administered in a conference hall, and completed on digital devices. The total duration of the word recognition test (including instructions and a short break) was 20 minutes. The test was administered in the same location for both groups, but at different sessions.

Prior to the actual test, all participants were required to sign a consent form, confirming their agreement to participate in the study. They were informed of their right to withdraw from the study at any time without any consequences, which was reiterated on the day of the test itself. This ensured that the participants were fully informed and aware of the study's purpose and their right as participants.

After the initial instructions and a brief orientation on how to navigate the tasks, the participants solved a short pilot study with different stimuli than the ones presented in the main task. This ensured they understood the procedure and felt confident with the task. Ten participants (Group A) completed the written word recognition test first, followed by the spoken recognition test, while another ten participants (Group B) completed the spoken test first, followed by the written test. Each group completed the task under almost identical conditions, with consistent environment, instructions and procedures. Both groups were supervised by the researcher, to

ensure consistency. Both groups had to decide whether the presented word in either modality was an actual English word or not. As previously mentioned, the words were randomised for each group, i.e., the stimuli did not follow the exact same order for both groups. They were instructed that some of the presented target words were made up non-words, i.e., pseudowords, and that the order of words and pseudowords was randomised prior to experiments.

Participants accessed the test via links that were sent by the researchers *on site*. There were two different links for each of the modality – one link for the written word recognition test, and the other for the spoken word recognition test. The links were sent separately, so as to avoid confusion (Group A first received the written word recognition test, and after completing it, then received the spoken word recognition test. The same went for Group B, but in the reverse order). Between tests, the participants were allowed a short break during which they were not allowed to discuss the test, which was mentioned prior, while giving instructions on how to solve the task.

The written recognition test featured three components: instructions (that were presented at the beginning of the test), a fixation cross (that appeared before every stimulus) and the stimuli themselves. The stimuli were presented on a grey background, with the target word displayed in white text, centred on the screen. The text was presented in Arial font, ensuring that the words were readable. Before each stimulus, a fixation cross appeared in the centre of the screen for approximately one second to help the participants focus their attention. The stimulus did not have a specific duration, i.e., the target word remained on the screen until the participant gave their response either by pressing the "Y" key (meaning "yes, this is a word") or the "N" key (meaning "no, this is not a word").

The spoken word recognition test featured the same three components: an instruction, a fixation cross before each word and the stimulus which was presented in audio form. The screen featured a grey background (the same as in the visual word recognition) and remained blank during the presentation of the auditory stimuli. A fixation cross appeared for one second before each presented audio stimulus. The stimulus played for a maximum of two seconds, and the screen remained blank until the participant responded by either pressing the "Y" key or the "N" key.

Participants' responses were automatically recorded by PsychoPy (in an .xslx format), which ensured that the data collected was accurate. All collected data was anonymous to protect the privacy of the participants.

6.4 Results

The present section presents the findings of this study. There were three research questions covered in this thesis, which will be covered separately in the following sections.

The first research question aimed to investigate whether modality (spoken vs. written) affects word recognition accuracy in Croatian – English bilinguals. In order to address this research question, the accuracy of word recognition across the written and spoken modalities was compared for both Group A and Group B. The following table shows the summary of accuracy of both groups for both spoken and written modality:

	Correct spoken	**Incorrect spoken**	**Correct written**	**Incorrect written**	**Total**
Group A	257	43	258	42	600
Group B	251	49	248	52	600
Total	508	92	506	94	1200

Table 2: Word recognition accuracy by modality and group

As seen in Table 1, Group A showed a correct response rate of 86% in the written modality and 85.67% in the spoken modality. A chi-square test of independence yielded a statistic of 0.014 (p-value = 0.907), which shows an insignificant statistical relevance and association between modality and accuracy.

Group B, on the other hand, showed a correct response rate of 82.67% in the written modality and 83.67% in the spoken modality. A chi-square statistic was 0.107 (p-value = 0.743), which also indicates no significant relationship.

The data combined from both groups indicated that the chi-square statistic was 0.016 (p-value = 0.899), which further supports the lack of a significant association between modality and word recognition accuracy.

The second research question aimed to investigate whether word prevalence (familiar vs. less familiar) affects word recognition accuracy in highly-proficient Croatian – English bilinguals. To explore this question, the accuracy rates for words with high prevalence and words with low prevalence were analysed across all participants. Out of 400 words with high prevalence, participants correctly identified 384 (96%) and made 16 errors, which resulted in high accuracy rate for this particular category. On the other hand, out of 400 words with low prevalence, participants correctly identified 295 (73.75%) words, and made 105 errors, which indicates a lower accuracy rate.

The following table summarizes the results:

Word prevalence	**Correct responses**	**Incorrect responses**	**Total responses**	**Accuracy Rate (%)**
High	384	16	400	96.00
Low	295	105	400	73.75

Table 3: Accuracy rates for words with high prevalence and words with low prevalence

A chi-square test was conducted to determine if the difference in accuracy between words with high prevalence and words with low prevalence words was statistically significant. The results of the chi-square test were significant (77.129, with p-value < 0.00001), which suggests that word prevalence significantly affects word recognition accuracy.

The third and final research question aimed to investigate whether word type (cognates, non-cognates and pseudowords) affect word recognition accuracy in Croatian-English bilinguals. In order to answer this question, accuracy rates were calculated separately for each word type within both Group A and Group B, and then combined for an overall analysis, which will be shown in the following charts and tables:

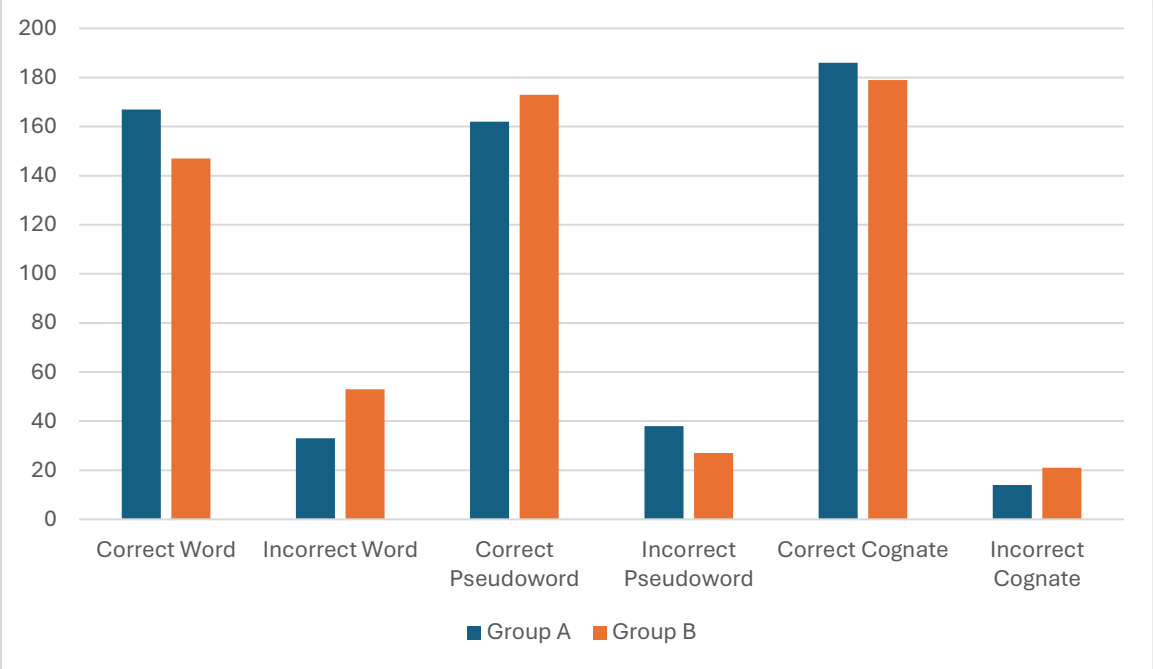

Figure 1: Comparison of word recognition accuracy by word type and group

The data in Figure 1 indicate that cognates had the highest recognition accuracy in both groups, with Group A achieving 186 correct responses, and Group B 179 correct responses. Non-cognates showed a slightly lower accuracy, with Group A recording 167 correct responses and Group B 147 correct responses. Pseudowords had the lowest accuracy rates, though this difference was more pronounced in Group B, where there were 173 correct and 27 incorrect responses, compared to Group A with 162 correct and 38 incorrect responses.

After calculating the chi-square between each category, the following results were obtained for Group A: the comparison between words and pseudowords resulted in chi-square value of 0.428, with a p-value of 0.513 which indicates no significant difference. The comparison between words and cognates resulted in a chi-square value of 8.703 with a p-value of 0.003, indicating a significant difference. The comparison between pseudowords and cognates produced a chi-square value of 12.732 with a p-value of 0.00036, also indicating a significant difference.

For Group B, the comparison between words and pseudowords resulted in a chi-square value of 10.563 and a p-value of 0.00154, which revealed a significant difference. When comparing words to cognates, the chi-square value was 16.979 and a p-value of 0.00004, which revealed a highly significant difference. Lastly, the comparison between pseudowords and

cognates showed no significant difference, with a chi-square value of 0.853 and a p-value of 0.356.

After analysing the word recognition accuracy for each group separately, a combined analysis was performed to evaluate the overall impact of word type on recognition across Group A and Group B. The combined results reveal that, in total, there were, 314 words correctly recognised with 86 incorrectly. For pseudowords, 335 were recognised correctly and 65 incorrectly. Cognates showed the highest accuracy rate, with 365 correctly recognised and 35 incorrectly. The following table presents the number of correct and incorrect recognitions for each word type:

Word type	**Correctly recognised**	**Incorrectly recognised**	**Percentage (%)**
Words	314	86	21.5%
Pseudowords	335	65	16.2%
Cognates	365	35	8.8%

Table 4: Word recognition accuracy and error rates by word type

The comparison between words and pseudowords resulted in a chi-square value of 3.6 with a p-value of 0.058, suggesting a slight significance. The comparison between words and cognates showed a highly significant difference, with a chi-square value of 25.327 and a p-value of < 0.00001. The comparison between pseudowords and cognates yielded a chi-square value of 10.2857 with a p-value of 0.00134, which indicates a significant difference.

Lastly, the response times were calculated additionally, due to the fact that response time provides insights into the efficiency of word recognition, where shorter response times usually indicate an easier and more confident processing, and longer response times may indicate difficulty or hesitance. The analysis revealed that correct responses had an average response time of 1.75 seconds (± 1.13 seconds), and incorrect responses had a longer average response time of 2.13 seconds (± 1.06 seconds), with $t = 4.2152$ and $p < 0.00001$.

7 Discussion

The present study aimed to investigate how input modality (spoken vs. written) and word difficulty (word prevalence and word type) affect word recognition accuracy in Croatian-English bilinguals. The findings from this study may offer insight into how these factors contribute to bilingual word processing, specifically in Croatian – English bilinguals.

7.1 Modality and word recognition accuracy

The first research question focused on whether input modality (written vs. spoken) influences word recognition accuracy. Contrary to previous findings by Veivo (2015) and Cornut (2021), which suggested that modality significantly affects word recognition (where written words are recognised more easily than spoken words), the current study found no significant difference in accuracy between written and spoken modalities. Specifically, Group A showed a correct response rate of 86% in the written modality, and 85.67% in the spoken modality. On the other hand, Group B showed a correct response rate of 82.67% in the written modality and 83.67% in the spoken modality. The chi-square for Group A showed a statistic of 0.014, and 0.107 for Group B. The combined chi-square statistic was 0.016. These results showed that there is no statistically significant difference in word recognition accuracy between written and spoken modalities in highly proficient Croatian-English bilinguals.

These findings can be understood within the framework of the non-selective language activation theory (Dijkstra, 2005), which suggests that bilingual speakers activate both languages simultaneously when processing words. Having that in mind, it is possible that the same principle applies to modalities, i.e., both modalities could be processed with equal strength by highly proficient bilinguals, because their lexical representations are well developed in both modalities.

Although previous studies, such as those by Veivo (2015) and Cornut (2021), have suggested some potential advantages of the written modality in certain contexts, the present study indicates that modality does not significantly impact word recognition accuracy for highly proficient bilinguals. However, their findings may stem from the fact that those earlier studies focused on lower proficiency bilinguals, whereas this study exclusively studied highly proficient bilinguals, which would imply that modality becomes less relevant as proficiency

increases. This is further supported by Frances et al. (2021), whose findings suggest that, in bilinguals with intermediate proficiency, word recognition accuracy changes with respect to the modality in which a word is presented. In contrast, the highly proficient bilinguals in this study did not show such differences between written and spoken tasks, which suggests that higher proficiency helps with overcoming these challenges.

In summary, while previous studies found that lower proficiency bilinguals experience differences between recognising written and spoken words, the present study suggests that higher proficiency reduces these differences, which results in a more balanced word recognition. Additionally, future research should consider different levels of proficiency and determine at what point does modality cease to be a factor in word recognition, i.e., at what point does proficiency help bilinguals process written and spoken words equally well.

7.2 Word prevalence and word recognition accuracy

The second research question focused on whether word prevalence (high vs. low) affects word recognition accuracy in highly proficient Croatian-English bilinguals. The results show that words with high prevalence were recognised with significantly higher accuracy (96%) than words with low prevalence (73.75%). Additionally, the results of the chi-square test were significant (77.129), which suggests that word prevalence significantly affects word recognition accuracy.

The results obtained support the findings of Brysbaert et al. (2016), where words with high prevalence, which are familiar to a larger portion of the population, were processed more efficiently and with fewer errors. This is likely because words with high prevalence are more deeply embedded in the mental lexicon, which makes them easier to retrieve and recognise. In contrast, words with low prevalence, which are known by fewer speakers, were recognised with lower accuracy, since less familiar words require additional cognitive strategies in order to recognise and process them.

These findings highlight how word prevalence, i.e., familiarity, influences bilingual word processing and recognition. Words with high prevalence, being more familiar, allow bilinguals to access them more easily, whereas words with low prevalence may require additional effort, especially if they are nor used on an everyday basis. This suggests that word prevalence plays

an important role in respect to word recognition accuracy, where greater familiarity leads to better performance.

From a practical standpoint, these results have implications for language learning and teaching. Words with high prevalence should be included in early language learning, since they are more likely to be recognised and used by learners, which would facilitate faster language acquisition. Additionally, while words with low prevalence can expand a learner's vocabulary, they may require more practice and repeated exposure in order to achieve the same level of recognition accuracy.

In summary, the present study confirms that word prevalence significantly affects word recognition accuracy in highly proficient Croatian-English bilinguals, with words with high prevalence being recognised more accurately than words with low prevalence. Future research could further explore the interactions between word prevalence and proficiency levels, and examine the ways in which to enhance teaching and expose learners of L2 to less familiar words in order to improve their word recognition.

7.3 Word type and word recognition accuracy

The third research question focused on how word type (cognates, non-cognates, pseudowords) affects word recognition accuracy in highly proficient bilinguals. The results show that cognates were recognised with the highest accuracy, followed by pseudowords, and non-cognates, i.e., words, were recognised with the lowest accuracy. These findings align with the findings from Dijkstra et al. (2010), which suggest that cognates are easier to recognise because they activate both languages simultaneously in the bilingual brain. In this study, cognates had a recognition accuracy of 91.25%, pseudowords 83.75%, and non-cognates 78.5%. These findings also align with the Bilingual Interactive Activation (BIA) and the Bilingual Interactive Activation Plus (BIA+) models. These models suggest that bilinguals activate lexical representations in both languages when processing words, which allows both language systems to contribute to word recognition, especially when there are similarities between the two languages.

Contrary to the previous research, such as the one by Oganian et al. (2016), who found that real words are typically recognised with greater accuracy than pseudowords, this study found that pseudowords were recognised more accurately than non-cognates. This difference

may be due to the specific nature of the stimuli or the individual differences of the participants in the current study. It is possible that pseudowords were more easily identified as "non-words" by participants, which led to fewer errors in their recognition accuracy compared to non-cognates. Additionally, Croatian-English bilinguals may have developed stronger skills when it comes to recognising pseudowords, due to their experience with handling two language systems. Thus, they may have a greater ability to detect non-existent lexical items more effectively.

The statistical analysis supports these claims, showing significant differences in accuracy between cognates and non-cognates. Cognates were significantly more accurately recognised, with a chi-square value of 25.327. The difference between pseudowords and non-cognates was also significant, with a chi-square value of 10.285, which shows that pseudowords were easier to recognise than non-cognates. However, the comparison between pseudowords and cognates showed a slight significance, with a chi-square value of 3.6, which indicates a minor difference in favour of cognates. The lower accuracy for non-cognates suggests that words without the phonological and orthographical overlap between the two languages require additional cognitive effort, i.e., they are harder to recognise. Because non-cognates lack the phonological and orthographical overlap, they do not simultaneously activate both language systems (as cognates do). This makes it harder and requires more effort to recognise them.

These findings highlight how important similarities between languages are when it comes to bilingual word recognition. They also suggest that bilinguals develop strategies to identify pseudowords more easily, possibly because their experience with two languages and language systems aids in recognising and dismissing "non-words" more efficiently.

Additionally, the response times provided further insight into word recognition accuracy. Correct responses had an average response time of 1.75 (± 1.13) seconds, while incorrect responses took longer, 2.13 (± 1.06) seconds on average. This suggests that faster response times may be connected to a more confident or easier word recognition, and the longer response times may be connected to an indecisiveness and increased difficulty with respect to recognition. These results may further support the BIA and BIA+ models, where slower response times may suggest that bilinguals face more cognitive effort, i.e., load when uncertain, which requires additional effort for processing and recognition, when lexical representations are harder to access.

These findings may offer practical implications for language learning, especially with respect to cognates. Learners should be encouraged to recognise and use cognates as a bridge between the two languages, as this can increase their confidence and speed in word recognition. Additionally, the higher accuracy rate for pseudowords compared to non-cognates suggests that using language games or activities that include pseudowords could help learners improve their ability to distinguish between real- and pseudowords.

7.5 Limitations and further research

There are possible limitations to the present study, the most significant limitation being the number of participants (N=20), which impedes drawing any general conclusions about the word recognition accuracy in Croatian-English bilinguals. Other limitations include a possibility of participants guessing the correct answers, due to the fact that the test includes a response time. This may have caused an anxiety in participants to perform more quickly. This may have also led to participants pressing the wrong key resulting in an unwanted answer. Another possible limitation would be taking only highly proficient bilinguals into consideration, which does not reflect the population of bilinguals with different levels of proficiency. Additionally, results may differ in the case of another language pair.

Further research with a larger number of participants is required to validate the findings presented in this research and provide a more comprehensive understanding of how the factors such as word type, modality and prevalence affect word recognition accuracy in this population. Moreover, including L1, in this case Croatian language, may identify the differences in processing mother tongue vs. an L2. Future research could investigate whether modality affects word recognition accuracy in less proficient bilinguals or bilinguals with different language pairs. Additionally, a deeper correlation between response times and the accuracy of participants' answers may add value and provide a different aspect and information on how bilinguals process language.

7 Conclusion

The present study investigated how input modality (spoken vs. written), word prevalence and word type affect word recognition accuracy in highly proficient Croatian-English bilinguals. The results showed there were no significant differences between spoken and written word recognition, which suggests that proficient bilinguals recognise words equally well in both cases.

However, word prevalence had a strong impact, with words with high prevalence being recognised more accurately than words with low prevalence. This highlights the importance of frequent exposure to words for better recognition. Additionally, it was proven that word type also played a significant role in word recognition accuracy with cognates being recognised most accurately, followed by pseudowords and then words. This supports the idea that bilinguals may benefit from similarities between languages when it comes to recognising words.

Although these findings provide valuable insights, the study did have its limitations, primarily due to a small sample size, as well as its focus on only highly proficient bilinguals. Future research should take into account less proficient bilinguals or other language pairs to investigate whether the results would be the same in different contexts and conditions.

In conclusion, the present study contributes to our understanding of how Croatian-English bilinguals process words, especially the influence of word prevalence and type. These findings may be useful in further developing language learning in this particular bilingual group.

Literature

Anderson, C., Bjorkman, B., Denis, D., Doner, J., Grant, M., Sanders, N., & Taniguchi, A. (2022). *Essentials of linguistics* (2nd ed.). eCampusOntario.

Babić, S., Finka, B., & Moguš, M. (1996). *Hrvatski pravopis*. Školska knjiga.

Barra J., Pallier C., & New B. (under press). UniPseudo : A Universal Pseudoword Generator. *Quarterly Journal of Experimental Psychology.*

Bastkowski, M. (2010). *Extending the mental lexicon: The L2 mental lexicon*. Retrieved August 8, 2024, from http://www.Grin.com

Bialystok, E. (2010). *Bilingualism. Wiley Interdisciplinary Reviews: Cognitive Science, 1(4), 559–572.* doi:10.1002/wcs.43

Bizzochi, A. L. (2017). How many phonemes does the English language have? *International Journal on Studies in English Language and Literature, 5*(10), 36–46. https://www.arcjournals.org/pdfs/ijsell/v5-i10/6.pdf

Blumenfeld, H. K., & Marian, V. (2007). Constraints on parallel activation in bilingual spoken language processing: Examining proficiency and lexical status using eye-tracking. *Language and Cognitive Processes, 22*(5), 633-660. https://doi.org/10.1080/01690960601000746

Brysbaert, M., Stevens, M., Mandera, P., & Keuleers, E. (2016). How many words do we know? Practical estimates of vocabulary size dependent on word definition, the degree of language input and the participant's age. *Frontiers in Psychology, 7*. https://doi.org/10.3389/fpsyg.2016.01116

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

Brown, D. (2006). Vocabulary activities to enhance your textbook. Paper presented at the annual conference of the Japanese Association of Language Teaching (JALT), Shizuoka.

Byers-Heinlein, K., Esposito, A. G., Winsler, A., Marian, V., Castro, D. C., & Luk, G. (2019). The case for measuring and reporting bilingualism in developmental research. *Collabra: Psychology, 5*(1), 37. https://doi.org/10.1525/collabra.233

Coltheart, M., Rastle, K., Perry, C., Langdon, R., & Ziegler, J. (2001). DRC: A dual route cascaded model of visual word recognition and reading aloud. *Psychological Review, 108*(1), 204–256. https://doi.org/10.1037/0033-295X.108.1.204

Connine, C. M., Mullennix, J., Shernoff, E., & Yelen, J. (1990). Word familiarity and frequency in visual and auditory word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16*(6), 1084–1096. https://doi.org/10.1037/0278-7393.16.6.1084

Cornut, C. (2021). *L2 word recognition among late bilinguals: Does modality matter? Investigation of the modality effect in English second language among typical and dyslexic-readers*. [Doctoral dissertation, Université de Lille].

Costa, A. (1999). Lexical selection in bilinguals: Do words in the bilingual's two lexicons compete for selection? *Journal of Memory and Language*.

Christophersen, P. (1948). *Bilingualism: An inaugural lecture delivered on Foundation Day, November 17th, 1948*. Methuen & Co. Ltd.

D'Acierno, M. R. (1990). *Three types of bilingualism* [Paper presentation]. Annual Meeting of the International Association of Teachers of English as a Foreign Language, Dublin, Ireland. https://files.eric.ed.gov/fulltext/ED321574.pdf

Diependaele, K., Lemhöfer, K., & Brysbaert, M. (2013). The word frequency effect in first- and second-language word recognition: A lexical entrenchment account. *The Quarterly Journal of Experimental Psychology, 66*, 843-863. https://doi.org/10.1080/17470218.2012.720994

Dijkstra, T., & van Heuven, W. J. B. (1998). The BIA model and bilingual word recognition. In J. Grainger & A. M. Jacobs (Eds.), *Localist connectionist approaches to human cognition* (pp. 37-54). Psychology Press. https://doi.org/10.4324/9780203774533

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. (2005). Bilingual visual word recognition and lexical access. In J. F. Kroll & A. M. B. de Groot (Eds.), *Handbook of bilingualism: Psycholinguistic approaches* (pp. 179-201). Oxford University Press.

Dijkstra, T., Miwa, K., Brummelhuis, B., Sappelli, M., & Baayen, H. (2010). How cross-language similarity and task demands affect cognate recognition. *Journal of Memory and Language*, 62(3), 284-301. https://doi.org/10.1016/j.jml.2009.12.003

Doner, J. (n.d.). 4.1. Language modality. In *Essentials of Linguistics* (2nd ed.). Pressbooks.openedmb.ca. Retrieved July 16, 2024, from https://pressbooks.openedmb.ca/wordandsentencestructures/chapter/language-modality/#return-footnote-1467-1

Dong, Y., Gui, S., & Macwhinney, B. (2005). Shared and separate meanings in the bilingual mental lexicon. *Bilingualism: Language and Cognition, 8*(3), 221–238. https://doi.org/10.1017/s1366728905002270

Erk, M., & Ručević, S. (2021). Early English language acquisition: How early is early enough? *Suvremena Lingvistika, 47*(92). https://doi.org/10.22210/suvlin.2021.092.02

Farahian, M. (2011). Mental lexicon: Its organization and representation. *BRAIN. Broad Research in Artificial Intelligence and Neuroscience, 2*(4), 56-59. https://www.researchgate.net/publication/265421243_Mental_Lexicon_Its_Organization_and_Representation

Frances, C., Navarra-Barindelli, E., & Martin, C. D. (2021). Inhibitory and facilitatory effects of phonological and orthographic similarity on L2 word recognition across modalities in bilinguals. *Scientific Reports, 11*(1). https://doi.org/10.1038/s41598-021-92259-z

Gazzaniga, M. S., Ivry, R. B., & Mangun, G. R. (2015). *Cognitive neuroscience: The biology of the mind*. W.W. Norton & Company, Inc.

Groot, A. M. B. de, Delmaar, P., & Lupker, S. J. (2000). The processing of interlexical homographs in translation recognition and lexical decision: Support for non-selective access to bilingual memory. *The Quarterly Journal of Experimental Psychology: Section A, Human Experimental Psychology, 53*(2), 397-428. https://doi.org/10.1080/713755891

Grosjean, F. (1989). Neurolinguists, beware! The bilingual is not two monolinguals in one person. *Brain and Language*. https://www.francoisgrosjean.ch/bilin_bicult/3%20Grosjean.pdf

Harley, T. A. (2005). *The psychology of language: From data to theory* (2nd ed.). Psychology Press. (Original work published 2001, Taylor & Francis e-Library edition 2005).

Henner, J., & Robinson, O. (2023). Unsettling languages, unruly bodyminds: A Crip Linguistics manifesto. *Journal of Critical Study of Communication and Disability, 1*(1), 7–37. https://criticalstudycommunicationdisability.org/index.php/jcscd/article/view/4

Holderbaum, C. S., & de Salles, J. F. (2011). Semantic Priming Effects in a Lexical Decision Task: Comparing third Graders and College Students in two Different Stimulus Onset Asynchronies. *The Spanish Journal of Psychology*, *14*(2), 589–599. https://doi.org/10.5209/rev_sjop.2011.v14.n2.7

Hui, C. (2022). A critical review and assessment of the Revised Hierarchical Model. *US-China Foreign Language, 20*(11). https://doi.org/10.17265/1539-8080/2022.11.005

Hymes, D. (1972). On communicative competence. https://nimshav.github.io/EthnoComm-Repository/EOC_Library/Hymes%20-%201972%20-%20On%20Communicative%20Competence.pdf

Leonard, M. K., & Chang, E. F. (2014). Dynamic speech representations in the human temporal lobe. *Trends in Cognitive Sciences, 18*(9), 472–479. https://doi.org/10.1016/j.tics.2014.05.001

Libben, G. (2008). Disorders of lexis. In B. Stemmer & H. A. Whitaker (Eds.), *Handbook of the neuroscience of language* (pp. 147-154). Elsevier. https://doi.org/10.1016/B978-0-08-045352-1.00014-8

Lillo-Martin, D., & Meier, R. (2011). On the linguistic status of "agreement" in sign languages. https://www.academia.edu/4675865/On_the_linguistic_status_of_agreement_in_sign_languages

Tactile Sign Language | What is Tactile Sign Language? (2022, July 4). Fable. https://makeitfable.com/glossary-term/tactile-sign-language/

Marslen-Wilson, W. D., & Welsh, A. (1978). Processing interactions and lexical access during word recognition in continuous speech. *Cognitive Psychology, 10*(1), 29-63. https://doi.org/10.1016/0010-0285(78)90018-X

Masrai, A., & Milton, J. (2015). The impact of L1 lexical organisation on L2 vocabulary acquisition. *Language in Focus, 1*(1), 15–34. https://doi.org/10.1515/lifijsal-2015-0002

McLaughlin, B. (1995). *Fostering second language development in young children: Principles and practices*. Educational Practice Report: 14. National Center for Research on Cultural Diversity and Second Language Learning, Office of Educational Research and Improvement (ED). https://eric.ed.gov/?id=ED392273

McClelland, J. L., & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception: Part 1. An account of basic findings. *Psychological Review, 88*(5), 375-407. Retrieved July, 16 from https://legacy.cs.indiana.edu/~port/teach/641/McClellandRumelhart.IAC.model.1981.pdf

Moreno, M. A., & van Orden, G. C. (2001). *Word Recognition, Cognitive Psychology of. International Encyclopedia of the Social & Behavioral Sciences, 16556–16561.* doi:10.1016/b0-08-043076-7/01553-9

Norris, D. (1994). Shortlist: A connectionist model of continuous speech recognition. *Cognition, 52*(3), 189-234.

Obretenova, S. (2010). Neuroplasticity associated with tactile language communication in a deaf-blind subject. *Frontiers in Human Neuroscience, 3*. https://doi.org/10.3389/neuro.09.060.2009

Oganian, Y., Froehlich, E., Schlickeiser, U., Hofmann, M. J., Heekeren, H. R., & Jacobs, A. M. (2016). Slower Perception Followed by Faster Lexical Decision in Longer Words: A Diffusion Model Analysis. *Frontiers in Psychology*, 6, 1958. https://doi.org/10.3389/fpsyg.2015.01958

Oldfield, R. C. (1966). Things, words, and the brain. *Quarterly Journal of Experimental Psychology, 18*, 340-353.

Ong, W. J. (1982). *Orality and literacy: The technologizing of the word*. Routledge. (Original work published 1982, Reissued 2002, Taylor & Francis e-Library edition 2005).

Pavlenko, A. (2009). 6. Conceptual Representation in the Bilingual Lexicon and Second Language Vocabulary Learning. In A. Pavlenko (Ed.), *The Bilingual Mental Lexicon: Interdisciplinary Approaches* (pp. 125-160). Bristol, Blue Ridge Summit: Multilingual

Matters. https://www.anetapavlenko.com/pdf/Chapter_6_Conceptual_Representation_in_the_Bilingual_Lexicon_and_Second_Language_Vocabulary_Learning.pdf

Perfetti, C. (2007). Reading ability: Lexical quality to comprehension. *Scientific Studies of Reading, 11*(4), 357-383. https://doi.org/10.1080/10888430701530730

Pinker, S. (2007). *The language instinct*. HarperCollins.

Ramoo, D. (2021). *Psychology of language*. BCcampus. https://opentextbc.ca/psyclanguage/

Rothman, J., Bayram, F., Deluca, V., González, J., Kubota, M., & Puig-Mayenco, E. (n.d.). *Understanding Language and Cognition through Bilingualism*. Retrieved August 3, 2024, from https://munin.uit.no/bitstream/handle/10037/33089/article.pdf?sequence=2&isAllowed=y

Sapir, E. (1921). *Language: An introduction to the study of speech*. Harcourt, Brace. Reprinted in 2004, Dover Books on Language.

Saussure, F. d. (2013). *Course in general linguistics*. Bloomsbury Publishing.

Singleton, D. M. (1999). *Exploring the second language mental lexicon*. Cambridge University Press.

Taft, M. (2001). Lexical access. In N. J. Smelser & P. B. Baltes (Eds.), *International Encyclopedia of the Social & Behavioral Sciences* (pp. 8602-8607). Pergamon.

Turner, J.E., Valentine, T. & Ellis, A.W. Contrasting effects of age of acquisition and word frequency on auditory and visual lexical decision. *Mem Cogn* 26, 1282–1291 (1998). https://doi.org/10.3758/BF03201200

van Heuven, W. J. B., Dijkstra, T., & Grainger, J. (1998). Orthographic neighborhood effects in bilingual word recognition. *Journal of Memory and Language, 39*(3), 458–483. https://doi.org/10.1006/jmla.1998.2584

Veivo, O. (2015). Orthographic bias in L3 lexical knowledge: Learner-related and lexical factors. *Language, Interaction and Acquisition*. John Benjamins Publishing Company. https://www.academia.edu/121447926/Orthographic_bias_in_L3_lexical_knowledge_Learner_related_and_lexical_factors

Wilson, S. M., Bautista, A., & McCarron, A. (2018). Convergence of spoken and written language processing in the superior temporal sulcus. *NeuroImage, 171*, 62–74. https://doi.org/10.1016/j.neuroimage.2017.12.068

Wolf, C., Meyer, A. S., Rowland, C. F., & Hintz, F. (2021). The effects of input modality, word difficulty and individual differences in reading experience on word recognition accuracy. *Collabra: Psychology, 7(1), Article 24919*. https://doi.org/10.1525/collabra.24919

Wright, J. (2014). *The Evolution of Writing | Denise Schmandt-Besserat*. Utexas.edu. https://sites.utexas.edu/dsb/tokens/the-evolution-of-writing/

Yelland, G. (1994). *Word Recognition and Lexical Access*. *4*. https://www.inf.unioeste.br/~jorge/MESTRADOS/LETRAS%20-%20MECANISMOS%20DO%20FUNCIONAMENTO%20DA%20LINGUAGEM%20-%20PROCESSAMENTO%20DA%20LINGUAGEM%20NATURAL/ARTIGOS%20INTERESSANTES/Word%20Recognition%20&%20Lexical%20Access.pdf

Appendix

1. Questionnaire

Questionnaire for Participants

Section 1: Participant Information

- **Age:** [ ] years
- **Gender:** [ ] Male [ ] Female [ ] Other
- **Native Language:** [ ] Croatian [ ] Other (please specify) _____
- **Education level:** _____
- **Do you have any visual or hearing impairments that may affect your ability to participate in this study?**

  [ ] Yes

  [ ] No

  If yes, please specify:_____

Section 2: Language Background

1. **At what age did you start learning English?**
   - 0-3 years
   - 4-7 years
   - 8-12 years
   - 13 years or older

2. **Where did you primarily learn English?**
   - Home
   - School
   - Language courses
   - Other (please specify)

     _____

Section 3: Language Use

1. **How often do you use English in your daily life?**
   - Never
   - Rarely
   - Sometimes

- Often
- Always

2. **In which contexts do you use English? (Check all that apply)**
    - Home
    - School
    - Work
    - Social media
    - Watching TV/movies
    - Reading books/articles
    - Other (please specify)

_____

Section 4: Language Proficiency

1. **How would you rate your proficiency in the following skills? (1 = Very Poor, 5 = Native-like)**
    - **Listening:** [1] [2] [3] [4] [5]
    - **Speaking:** [1] [2] [3] [4] [5]
    - **Reading:** [1] [2] [3] [4] [5]
    - **Writing:** [1] [2] [3] [4] [5]

2. **Have you taken any standardized English proficiency tests (e.g., Cambridge English, Oxford Placement Test…)?**
    - Yes (please specify your score and the test)
      _____
    - No

Section 5: Listening Habits

1. **How often do you hear English?**
    - Never
    - Rarely
    - Sometimes
    - Often
    - Always

2. **In what contexts do you primarily listen to English?** (Check all that apply)
    - Conversations with friends or family

- In school (e.g., during English lessons)
- Watching TV shows/movies
- Listening to music
- Podcasts or radio
- Watching YouTube videos or online tutorials
- Video games
- Other (please specify): _____

Section 6: Reading Habits

1. **How often do you read in English?**
    - Never
    - Rarely
    - Sometimes
    - Often
    - Always

2. **What types of materials do you read in English? (Check all that apply)**
    - Books
    - Newspapers
    - Magazines
    - Online articles
    - Social media posts
    - Academic papers
    - Other (please specify)

_____

2. List of stimuli:

word	word_type	word_prevalence
drama	cognate	2.429
hotel	cognate	2.331
leader	cognate	2.338
moment	cognate	2.576
yoghurt	cognate	0.63
ampoule	cognate	0.037
patina	cognate	0.544
ethno	cognate	0.29
recidivism	cognate	0.292
amortization	cognate	0.469
brambly	word	0.47
disserve	word	0.426
includable	word	0.328
shylock	word	0.328
aforetime	word	0.33
insurgent	word	1.773
profuse	word	1.774
docile	word	1.778
impeachment	word	1.934
luminous	word	2.094
interingly	pseudoword	NA
lassen	pseudoword	NA
cleast	pseudoword	NA
bastacle	pseudoword	NA
tift	pseudoword	NA
clow	pseudoword	NA
bure	pseudoword	NA
eneel	pseudoword	NA
clary	pseudoword	NA
inveriative	pseudoword	NA