



UNCUYO
UNIVERSIDAD
NACIONAL DE CUYO



FACULTAD DE
EDUCACIÓN

Universidad Nacional de Cuyo

Facultad de Educación

Thesis

Master in Language and Cognition

*Influence of Lexical and Sublexical Typicality on the
Neighbor Generation and Wordlikeness Tasks in English as
a Second Language*

Master's Student: Andro Dalmazzo Oros M.D.

Thesis Director: Alejandro Wainelboim PhD.

Thesis Codirector: Jorge García Carbajo PhD.

Mendoza, Argentina

August 2025

PREFACE

This final work is submitted as part of the requirements for obtaining the academic degree of the Master in Language and Cognition from the Education Faculty at the *Universidad Nacional de Cuyo*. It has not been previously submitted for the attainment of any other degree at this or any other university.

Under the supervision of Dr. Alejandro Javier Wainseboim (independent researcher at INCIHUSA - CONICET), whose research focuses on the neuropsychological bases of emotional, categorical, and linguistic processing in adults, and the co-supervision of Dr. Jorge Carlos García Carbajo, who holds a Master's degree in teaching Spanish as a second language. This work is framed within the fields of psycholinguistics, orthographic processing, and bilingualism.

Author's Signature

December 2, 2024

Universidad Nacional de Cuyo, Mendoza, Argentina

ACKNOWLEDGEMENTS

According to the experience of some, a Master's thesis does not typically include many acknowledgments due to its concise nature. However, that is not the case here, as this thesis could only be completed thanks to the advice, guidance, and support of numerous individuals who helped me navigate a field entirely foreign to what my medical career had accustomed me to.

I must extend my heartfelt thanks to my dear friend, the psychiatrist Martín Farina, as he was the bridge that led me to meet my first mentor and thesis co-director, Dr. Jorge García Carvajo. After completing the coursework phase of the Master's program in Language and Cognition in 2022, like many of my fellow postgraduates, I found myself without a mentor to turn to, someone who could provide the needed guidance and support to move forward with the thesis phase. To complicate matters, my interests did not align with any of the research conducted by professors in the Master's program, leaving me with slim chances of finding a mentor to supervise my work. Nevertheless, I was one of the fortunate few who managed to secure an outstanding mentor in Jorge. His passion for teaching is the driving force behind his unwavering dedication to training researchers, a level of support that often felt undeserved. I am well aware that thesis supervisors do not usually devote much attention to their Master's students, as their primary focus tends to be on doctoral candidates. Even so, with Jorge, I came to understand the necessary norms and requirements to be accepted into the complex world of academia. His diligence compelled me to dedicate countless hours to improving the writing style of my texts, week after week. It must be said that, without Jorge's presence, I am not sure if I would have been able to complete or even continue this postgraduate program, which is why my gratitude to him is boundless.

I must also express my gratitude to both Dr. Adolfo M. García, the original founder of the Master's program, who displayed biblical patience in having me as a student and answering all kinds of questions, and to Dr. Ana Laura Diez Martini, a former disciple of Jorge. Both of them recommended the guidance of Dr. Alejandro Wainseboim, who, in line with his expertise in psycholinguistics and neurophysiology, agreed to supervise my thesis as its director. In doing so, he became my theoretical and methodological mentor, enabling me to refine my rudimentary ideas into well-founded proposals. Alejandro's guidance has been pivotal to this work, and I cannot be more grateful for his formative contributions.

I am deeply grateful to Dr. Gisella Müller, who connected me with Dr. Maria Victoria Magariños and Dr. Marcela Raggio, faculty members from the Facultad de Filosofía y Letras (Faculty of Philosophy and Literature), who were instrumental in granting me access to the participants of my study at the Universidad Nacional de Cuyo (Argentina). My gratitude also extends to the entire English Department, including Dr. Grisel Salmaso, M.A. Verónica Segura, and Rud Mariné Ruiz, who further facilitated the realization of a seminar on the subject. Additionally, I would like to thank the Facultad de Filosofía y Letras as a whole, as this study would not have been possible without its support.

I extend my heartfelt thanks to my friend Mauricio Salcedo for introducing me to the world of graph theory, a foundation that has been highly relevant in conceptualizing the sublexical domain. I am also grateful to Engineer Álvaro Concha, whose mathematical insights were crucial for refining my sublexical token frequency equation. Special thanks go to Ana Repetto and Dr. Alejandra Arroyo, whose recommendations in the statistical realm greatly enriched the results of this work. I am deeply thankful to Dr. Edith Kaan and Dr. Eleonora Rossi from the University of Florida (USA), who welcomed me into their community and work.

I would like to express my gratitude to Dr. Rebecca Treiman, from Washington University in St. Louis (USA), for providing publications that had a decisive influence on my theoretical proposal; to Dr. Jonathan Grainger, from Aix-Marseille University (France), with whom I had the pleasure of clarifying questions related to phonological processing; to Drs. Mike Hammond and Dr. Howard C. Nusbaum, from the University of Arizona and the University of Chicago (USA), respectively, who provided the datasets with which I worked; to Dr. Kenn Apel, from the University of South Carolina (USA), who guided me on mathematical aspects related to the calculation of graphotactic probability; to Dr. David R. Badcock, from the University of Western Australia (Australia), for his patience in instructing me in the use of spatial frequency filtering; and to Dr. Michael Ullman, from Georgetown University (USA), who gave me the opportunity to address specific questions regarding some of the topics covered in this work.

I am also sincerely grateful to those who served as members of my thesis defense committee: Dr. Manuel Perea, from the Universitat de València (Spain), whose scholarly work was fundamental to the development of a large part of my theoretical framework; Dr. Valeria Abusamra, from the University of Buenos Aires (Argentina), who, despite her many commitments, generously took the time to respond to my inquiries; and Dr. Virginia Jaichenco, from the same institution, whom I also had the privilege of having as a professor during my graduate studies.

I would like to express my special thanks to Dr. Antonio Iniesta, from McGill University (Canada), who, although he was unable to serve on my committee due to institutional constraints, showed exemplary generosity and willingness to help, even taking the time to read this work in its entirety.

Finally, I am profoundly grateful to my family, who supported my research endeavors despite their unfamiliarity with the world I ventured into. Above all, I thank one of the most important people that have passed through my life, Paula, who not only introduced me to the existence of this Master's program in the first place but also proved to be an unwavering source of support. She has shared in my joys, obsessions, and "eureka moments," making her an integral part of this journey.

GENERAL INDEX

INDEX OF TABLES AND FIGURES.....	X
LIST OF ABBREVIATIONS.....	XII
ABSTRACT.....	XIII
INTRODUCTION	
<i>Problem Statement</i>	1
<i>Research Objectives and Hypotheses</i>	4
<i>Scope and Limitations</i>	5
<i>Organization of the Thesis</i>	6
FIRST PART. THEORETICAL REVIEW	
CHAPTER 1. Nuances of the Mental Lexicon	
1.1. <i>Incompleteness within the Lexical Representation</i>	8
1.2. <i>Lexical Neighborhood</i>	14
1.3. <i>Confusability Approach of Similarity</i>	21
CHAPTER 2. Appearance of the Sublexicon	
2.1. <i>Memory Systems and the Dual-Route Model</i>	30
2.2. <i>Sublexical Development and Consistency Degree</i>	45
2.3. <i>Gestaltian Approach of Similarity</i>	63
CHAPTER 3. Resolution of the Conceptual Conundrum	
3.1. <i>Similarity, Typicality, and Familiarity</i>	72
3.2. <i>Lexical-Sublexical Metrics</i>	80
3.3. <i>Sublexical Token Frequency</i>	91
SECOND PART. EXPERIMENTAL RESEARCH	
CHAPTER 4. Methodological Design	
4.1. <i>Overall Design and Sampling Strategy</i>	98
4.2. <i>Pre-experimental Tests</i>	98
4.2.1 <i>Vocabulary Size Test (VST)</i>	99
4.2.2 <i>Language History Questionnaire 3 (LHQ3)</i>	99
4.3. <i>Creation of the Experimental Tests</i>	100
4.3.1 <i>Step 1: Filtering of the BLP Database</i>	100
4.3.2 <i>Step 2: SL-TF Calculation and Sublexical Irrelevancy Hypothesis</i>	101

4.3.3	Step 3: Alphabetic and Numerical Sublexical Hierarchy.....	105
4.3.4	Step 4: Stochastic Assembly of Pseudowords.....	106
CHAPTER 5. Analysis and Results		
5.1.	<i>Modified Orthographic Neighbor Generation Task</i>	109
5.1.1.	<i>Description and Hypothesis</i>	109
5.1.2.	<i>Participants</i>	110
5.1.3.	<i>Materials and Methods</i>	111
5.1.4.	<i>Analysis</i>	112
5.1.5.	<i>Results</i>	112
5.1.6.	<i>Preliminary Discussion</i>	115
5.2.	<i>Wordlikeness Tasks for Pseudowords</i>	119
5.2.1.	<i>Description and Hypothesis</i>	119
5.2.2.	<i>Participants</i>	121
5.2.3.	<i>Wordlikeness Judgement</i>	121
5.2.3.1.	<i>Materials and Methods</i>	121
5.2.3.2.	<i>Analysis</i>	122
5.2.3.3.	<i>Results</i>	122
5.2.3.4.	<i>Preliminary Discussion</i>	124
5.2.4.	<i>Wordlikeness Arrangement</i>	126
5.2.4.1.	<i>Materials and Methods</i>	126
5.2.4.2.	<i>Analysis</i>	127
5.2.4.3.	<i>Results</i>	127
5.2.4.4.	<i>Preliminary Discussion</i>	128
CHAPTER 6. General Discussion		
6.1.	<i>Analysis per-Participant</i>	132
6.2.	<i>Analysis per-Word</i>	134
6.3.	<i>Analysis per-Pseudoword</i>	135
CONCLUSIONS.....		139
LIST OF REFERENCES.....		141
APPENDIX.....		182

INDEX OF TABLES AND FIGURES

CHAPTER 1. Nuances of the Mental Lexicon

1.1. *Incompleteness within the Lexical Representation*

Figure 1. Vertical Axis versus Horizontal Axis9

Figure 2. Vertical Axis (from top-to-bottom)..... 10

Figure 3. Horizontal Axis (from left-to-right).....12

Figure 4. Alleged Nature of the Incomplete Orthographic Representation..... 13

1.2. *Lexical Neighborhood*

Figure 5. Substitution-only Restriction (Coltheart's N)..... 14

Figure 6. Substitution & Addition/Omission Restriction (NNB)..... 15

Figure 7. Substitution, Addition/Omission & Morphological Derivatives Restriction (CLEARPOND's N).....16

Figure 8. Substitution, Addition/Omission & Transposition Restriction (N*)..... 17

Figure 9. Substitution & Addition/Omission Restriction (OLD20) vs Substitution, Addition/Omission & Transposition Restriction (OLD20*)... 19

1.3. *Confusability Approach of Similarity*

Figure 10. Perceptual Magnetic Effect..... 23

Figure 11. Distortion due to the Vernacular Phonetic Prototypes..... 25

Figure 12. Lexicalization Error.....26

Figure 13. Lexical Token Frequency and Perceptual Field.....28

CHAPTER 2. Appearance of the Sublexicon

2.1. *Memory Systems and the Dual-Route Model*

Figure 14. Episodic-to-Semantic Memory System..... 31

Figure 15. Cascaded Activation of the Lexical and Semantic Domain.....32

Figure 16. Dual-Route Model.....37

Figure 17. Extrapolated Sublexical Representations from the Lexicon... 38

Figure 18. Alleged Magnocellular versus Parvocellular Spatial Resolution.....41

2.2. *Sublexical Development and Consistency Degree*

Table 1. Alternative bipartite subsyllabic nomenclature..... 46

Figure 19. Transparent versus Opaque Networks..... 47

Figure 20. The Symbolic Pre-Alphabetic Phase.....	51
Figure 21. The Phonemic Segmentation Phases.....	53
Figure 22. Overcoming the Inconsistency Problem.....	54
Figure 23. Lack of Precision in the English Alphabetic Level.....	57
Figure 24. The Sublexical Advantages of the English Rhyme Unit.....	58
Figure 25. Vocalic Syllabic Stage.....	61
2.3. <i>Gestaltian Approach of Similarity</i>	
Figure 26. C-Similarity Inhibition versus G-Similarity Facilitation in Priming.....	65
CHAPTER 3. Resolution of the Conceptual Conundrum	
3.1. <i>Similarity, Typicality, and Familiarity</i>	
Figure 27. The Domains of C-Similarity and G-Similarity.....	74
Table 2. Luce & Luce (2001) Neighbor Generation Task Example.....	76
Table 3. Lexical Typicality and the Neighbor Generation Task.....	77
3.2. <i>Lexical-Sublexical Metrics</i>	
Table 4. Lexical Typicality and the Neighbor Generation Task.....	82
CHAPTER 4. Methodological Design	
4.3. <i>Creation of the Experimental Tests</i>	
Figure 28. Sublexical Token Frequency.....	104
CHAPTER 5. Analysis and Results	
5.1. <i>Modified Orthographic Neighbor Generation Task</i>	
Table 5. Selected Rhymes.....	111
Table 6. Correlations of Words per Rhyme with L-TF and C-Similarity.....	114
Table 7. Correlations of Words per Level/N ^o of Letters with L-TF and C-Similarity.....	115
5.2. <i>Wordlikeness Tasks for Pseudowords</i>	
Table 8. <i>Pseudowords for Wordlikeness Judgement</i>	121
Table 9. <i>Correlation between the Average Wordlikeness Judgement and the Predictor Variables</i>	124
Table 10. <i>Pseudowords for Wordlikeness Arrangement</i>	126
Table 11. <i>Correlation between the Average Wordlikeness Arrangement and the Predictor Variables</i>	128

LIST OF ABBREVIATIONS

L1 : *First Language*

L2 : *Second Language*

Coltheart's N: *Neighborhood Size (only substitutions)*

NNB : *Number of such Neighbors (includes additions and omissions)*

N* : *Neighborhood Size (includes transpositions)*

CLEARPOND : *Cross-Linguistic Easy Access Resource for Phonological and Orthographical Neighborhood Densities*

LD : *Levenshtein Distance*

OLD20 : *Orthographic Levenshtein Distance 20*

OLD20* : *Orthographic Levenshtein Distance 20 (includes transpositions)*

POLD20 : *Phonological Levenshtein Distance 20*

L-TF : *Lexical Token Frequency*

SL-TF : *Sublexical Token Frequency*

BLP : *British Lexicon Project*

BNC : *British National Corpus*

CELEX : *Center for Lexical Information*

C-Similarity : *Confusability Approach of Similarity*

G-Similarity : *Gestaltian Approach of Similarity*

ABSTRACT

Similarity and typicality are core concepts within psycholinguistics. However, when using the metrics that allegedly represent them, the boundaries between them become blurred, which is a result of a lack of a thorough and explicit definition for both. To exemplify this, this work focused on two tasks whose results depend on similarity and typicality. On one hand, the neighbor generation task involves presenting a target stimulus to trigger the recall of all words that structurally resemble it. The total number of retrieved words has been equated with neighborhood size. Nevertheless, a preliminary analysis of the results from the original work shows that retrieval is more related to lexical frequency (L-TF). Thus, only the most frequent words in the neighborhood appear to be the most recalled, regardless of overall neighborhood size. On the other hand, the wordlikeness judgment task rates pseudowords on a scale from 1 to 7, based on how they are perceived as part of a given language. This "English-likeness" depends on the typicality of the sequences that comprise them. The problem here is that the paradigmatic sublexical typicality metric, the tactic probability, does not significantly predict the results of the wordlikeness perception. Moreover, this measure does not even seem to perform better than the conventional lexical similarity measure (neighborhood size) in this regard. This work aims to clarify the concepts of similarity and typicality, evaluate which ones can better predict the results in both tasks, and propose a new sublexical typicality metric. The study is conducted in English due to its monosyllabic nature, allowing for a focus on orthographic processing while controlling other phonological variables that are unrelated to the study (e.g., prosodic features, syllabification). The Spanish-speaking selected population has a high level of competence in the English language, making it comparable to native processing. The present work propose the existence of two types of similarity: one intradomain (lexical) —based on perceptual distance (C-similarity)— and another interdomain (lexical-sublexical) — based on the presence of shared features (G-similarity). The results show that the neighbor generation task was consistently predicted by the L-TF of the recalled neighbors. In contrast, neighborhood size (C-similarity) was only predictive when the target's neighborhood had a higher compounded lexical typicality (Σ L-TF). Regarding the wordlikeness task, it was found that the tactic probability metric was a nonspecific measure of G-similarity, whereas it was proposed a new sublexical typicality measure (SL-TF). The best predictor of the wordlikeness results was the combination of SL-TF —for sublexically relevant sequences— and Σ L-TF —for

sublexically irrelevant sequences. This is framed as the Sublexical Irrelevancy Hypothesis, which posits that sublexical traits are stored as independent units ($SL-TF > 0$) only after surpassing a minimum number of words that contain the pattern (relevancy threshold). If this is not surpassed, the pattern is lexically processed through comparison with other words ($\Sigma L-TF$). The relevancy threshold that yielded the best results was 4 ± 1 .

INTRODUCTION

Problem Statement

The similarity and typicality concepts are essential notions in psycholinguistics, as they underlie the theoretical foundation of the most classical metrics in the field. Examples include the similarity neighborhood (Hammond, 2004), the clustering coefficient (Chan & Vitevitch, 2009), type, and token frequency (Berg, 2014). These metrics are subdivided based on their linguistic domains: lexical —relating to lexemes or word forms—; semantic —relating to meanings—; morphological —relating to morphemes, the smallest linguistic units with meaning—, and sublexical —relating to sublexemes, fragments of lexemes without meaning. Due to this variety of domains, multiple similarities and typicalities exist (e.g., morphological typicality, semantic similarity). However, despite their importance, the concept of similarity has not been well-defined (Luce & Pisoni, 1998; Bybee, 2010), leading to problematic interpretations in studies that rely on it. As will be discussed later, the lack of consensus in the terminology also extends to the alleged typicality metrics, meaning that their definition, which has so far been implicit, also requires revision. This work will focus on this conceptual problem, particularly emphasizing two linguistic domains: the lexical and the sublexical.

The lexical and sublexical domains are foundational to dual-route models alongside the semantic domain (e.g., Diependaele et al., 2010; Hautala et al., 2022). The quicker lexical route recognizes known words using a mental dictionary referred to as the lexicon. By contrast, the slower sublexical route processes unfamiliar words and pseudowords —words without meaning— by extrapolating information from already stored words. This latter route enables plausible spellings and pronunciations of any unknown stimulus.

Based on these domains, the similarity and typicality metrics allow for predicting and interpreting the results from various psycholinguistic studies. The most commonly used metrics are *word frequency* (i.e., lexical typicality) and *lexical neighborhood* (i.e., lexical similarity), which are strongly associated with inhibitory effects occurring during lexical access (see Perea, 2015, for a review). This inhibition has been accounted for as a consequence of the competition between two or more words to reach effective activation (Grainger & Jacobs, 1996).

At the same time, the *tactic probabilities*, defined as the rate of occurrence of a

sequence across different words, are used to predict the outcomes of sublexical tests, such as the wordlikeness judgement task for pseudowords (e.g., Bartolotti & Marian, 2014). This task involves evaluating pseudowords based on their subjectively perceived plausibility as part of a given language. For example, the pseudoword *strooth* is perceived as more "English-like" than *blobrze* because the trigram ⟨brz⟩ has an almost zero tactic probability of occurrence in the English language¹.

The relevance of this metric lies in the fact that neighbors sharing the same vowel and final consonants (e.g., *toon*, *boon*, *moon*) —a sublexical unit known as the *rhyme*— facilitate the processing among themselves (e.g., Ziegler & Perry, 1998; Rapp & Samuel, 2002; Stella, 2018). However, not all sequences produce the same effects. For instance, when multiple neighbors share the same initial consonants (e.g., *clown*, *close*, *clout*) —a sublexical unit known as the *onset*—, the effect between them tends to be inhibitory (e.g., Donnelly & Kidd, 2021)². These effects suggest that there is a logic behind the importance of each sequence, a question that has yet to be resolved by any dual-route model.

Despite its importance, the sublexical domain is the most overlooked area of psycholinguistics. As a result, it is common to find oversimplifications of its mechanisms that fail to capture its more complex aspects. For instance, many authors using the dual-route model refer to the sublexical domain as merely a system of "grapheme-to-phoneme conversion" —or "phoneme-to-grapheme conversion," depending on the source of the input (e.g., Playfoot et al., 2018; Macchi et al., 2019; Choi & Sim, 2019). This oversimplification implies that the sublexical domain is processed solely in an alphabetic manner (i.e., letter-by-letter and phoneme-by-phoneme), disregarding the possibility of a sequence-based analysis.

The dual-route model's terminology sometimes suggests a presumed sensory bias for each route. For example, the lexical route is often referred to as the "orthographic route," whereas the sublexical route is tantamount to a "phonological

1 Italic letters are used in this work to identify a word in the orthographic domain, and angular quotation marks introduce letters or letter sequences. Finally, slashes refer to phonemes or words in the phonological domain, always represented using the alphabet of the *International Phonetic Association* (1996, 1999), based on general American English.

2 Throughout this thesis, both the onset and the rhyme may be referred to using the abbreviated CVC nomenclature (Consonant-Vowel-Consonant) from Treiman et al. (1995).

route" (Macchi et al., 2019; Borghesani et al., 2020; Lee et al., 2022). These are inaccurate, as both routes make use of both sensory domains. Linguistic information — lexical and sublexical— is stored orthographically and phonologically, enabling the meaning to be accessed regardless of the input form. These sensory modalities activate one another through correspondences that link orthographic information to its phonological equivalent.

Thus, the sublexical domain is affected by oversimplifications, issues in the nomenclature, and metrics that may not adequately represent it. The clearest example of the latter lies in the results of the wordlikeness judgement task for pseudowords, a topic of debate in the field. The tactic probability —a paradigmatic metric of sublexical typicality— does not always correlate satisfactorily with some wordlikeness judgement results. The lexical neighborhood, a metric of lexical similarity, has even outperformed it in some cases (e.g., Bailey & Hahn, 2001; Janse & Newman, 2013; Apel et al., 2013). Such findings have led some to consider the tactic probability as nothing more than a more complex similarity metric (e.g., Frauenfelder et al., 1993; Hammond, 2004). Others, taking a more radical stance, have questioned the existence of the entire sublexical domain, viewing it as a mere extension of the lexical domain (Bailey & Hahn, 2001).

However, the core issue is overlooked by solely focusing on the sublexical aspect. Is the tactic probability metric indeed a measure of sublexical typicality, or was it conceived from the start as one of similarity? This need for conceptual clarification revisits the original psycholinguistic problem: How are the concepts of similarity and typicality defined? Only after addressing this question can we ask the next: Which measure of sublexical typicality can reliably predict the wordlikeness of each pseudoword?

The lack of terminological clarification between these concepts hinders the creation of metrics and distorts the interpretation of results. This misinterpreting is evident in the neighbor generation task, which involves presenting a target and asking participants to think of and list all known words that share a structural similarity (Luce & Large, 2001). According to the authors, the number of recalled words (entropy) is tantamount to the lexical neighborhood of the target. Based on this conclusion, the following question arises: Does entropy reflect the similarity neighborhood size of the target stimulus or the lexical frequency of the recalled words?

In conclusion, this thesis is framed within the field of lexical and sublexical processing of the English language. The monosyllabic nature of English allows for

working with whole words and pseudowords without the need to account for syllabic and prosodic processing. Although the Argentine population predominantly speaks Spanish as their first language (L1), the study was conducted with individuals with advanced knowledge of English as a second language (L2), enabling the simulation of native-like results in the language. Consequently, this thesis also falls within the scope of bilingualism.

Research Objectives and Hypotheses

The general objectives of this work are as follows:

1. To clarify the concepts of similarity and typicality, along with their associated lexical and sublexical metrics.
2. To determine which lexical or sublexical metric correlates best with the results of visual wordlikeness tasks for pseudowords and orthographic neighbor generation tasks in a population of non-native Argentine English speakers.

The specific objectives herein proposed are as follows:

1. To describe the theoretical framework of both the lexical and sublexical domains in the English language, along with their relevant nuances, compared to Spanish.
2. To elucidate the cause of contradictions associated with the lexical neighborhood and the tactic probability metrics through a conceptual clarification of similarity and typicality.
3. To create an alternative sublexical typicality metric for assembling pseudowords that may predict the outcomes of sublexical processes in English.
4. To compare the differential impact between lexical typicality and lexical similarity in the orthographic neighbor generation task results in a population of non-native Argentinian English speakers.
5. To examine the predictive power of the new sublexical typicality metric, compared to conventional lexical and sublexical metrics, in visual wordlikeness tasks for pseudowords in a population of non-native Argentinian English speakers.

Based on the previous objectives, the research hypotheses are as follows:

1. The similarity described for the lexical domain is based on the idea that two

words can be confused when they are perceptually close; in contrast, the similarity resulting from the interaction of the lexical and sublexical domains is based on the existence of a shared and relevant segment between the two words.

2. The lexical neighborhood and the tactic probability metrics are nonspecific, as they are similarity measures that can only partially reflect sublexical typicality.

3. Sublexical token frequency better reflects sublexical typicality in visual wordlikeness tasks for pseudowords, as it considers both lexical typicality and similarity based on shared features.

4. The number of words recalled in the orthographic neighbor generation task shows a stronger correlation with lexical typicality than lexical similarity in a population of non-native Argentinian English speakers.

5. The numerical hierarchy —organized based on sublexical typicality (for relevant sequences) and lexical typicality (for irrelevant sequences)— is the best predictor of the results of the visual wordlikeness task for pseudowords in a population of non-native Argentinian English speakers.

Scope and Limitations

This work will not delve into the semantic, morphological, contextual, and syntactic processes, although these may be referenced throughout the text. While both the tasks and the theoretical review focus on the orthographic domain, there will be numerous citations of studies with a phonological emphasis. One example is Kuhl's (1994) perceptual magnetic effect, which is used as a didactic tool to explain one similarity type, although it will be applied here to the orthographic aspect.

Although the dominant linguistic environment in Argentina is Spanish, the selection of English is deliberate. English is a grapho-phonologically inconsistent language, with words having an average length of 5.1 (Bochkarev et al., 2015). Such an average leads most of its words to be monosyllabic, allowing for the exclusion of other phonological variables, such as syllabification and prosodic features (e.g., tone, stress, emphasis). Additionally, English is the most studied language in psycholinguistics, offering more information for the theoretical review and its conclusions. However, since the study was conducted with non-native English speakers, the results may not fully represent those of native speakers. Therefore, this work is necessarily framed within the field of bilingualism.

Beyond this thesis's results, this work is expected to contribute to four key

aspects. First, it should emphasize the sublexical domain as a complex and independent aspect, not as a mere part of the lexical domain. Second, conceptual clarification should resolve the conflict between similarity and typicality. Third, the new sublexical typicality metric should represent a step forward in sublexical understanding while also providing a better tool for assembling pseudowords with controlled wordlikeness. Fourth, this is the first study that combines English L2 speakers with English sublexical processing, marking a novelty in bilingualism. Given that the participants' L1 is Spanish, this study offers a window into the differential influences between the sublexical consistencies and inconsistencies of two of the most widely spoken languages in the world.

Organization of the Thesis

The thesis is structured into six chapters. The first three focus on the topic's theoretical review, while the remaining three address the main experiment itself.

In Chapter 1, the concepts of the mental lexicon, lexical typicality, the nature of lexical representations, and similarity based on confusability are tackled. Chapter 2 describes the sublexical domain according to its biological foundations, the dual-route model, and its development in the literacy process, all to understand its importance and functions. It also reviews some consistency differences with Spanish, the notion of similarity based on shared features, sublexical typicality, and tactic regularities. Chapter 3 delimits the concepts of similarity and typicality, categorizes some classic metrics as lexical-sublexical, and proposes a new sublexical typicality measure: the sublexical token frequency. Chapter 4 presents the methodological design of the study, participant sampling, and the steps required to assemble the data collection instrument. Chapter 5 is dedicated to the analysis of the collected data and discusses the results in relation to the proposed hypotheses, while Chapter 6 recaps all the study results.

FIRST PART:
THEORETICAL REVIEW

CHAPTER 1

Nuances of the Mental Lexicon

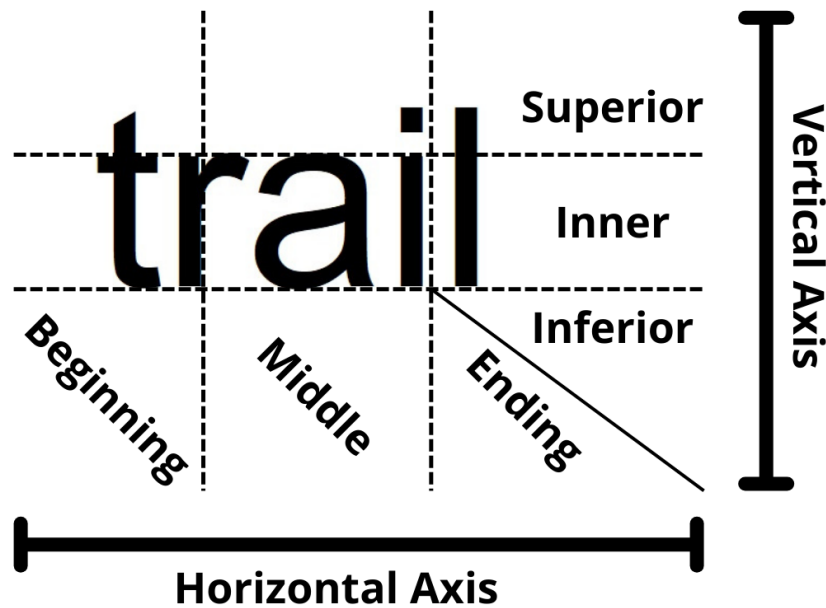
1.1. *Incompleteness within the Lexical Representation*

For most adults, learning a new visual lexeme (i.e., a printed word) occurs more through reading than writing. After repeated encounters with the stimulus, the word is stored as a new lexical entry in the lexicon, i.e., the mental dictionary that contains the knowledge of all known words. This information is later used for recognition in future situations (Jabeen & Shahzad, 2023). However, the recognition process does not require a meticulous and precise analysis of the stimulus; thus, memory only stores the most important features for identification (Burt & Tate, 2002), while it neglects others. This bias results in an incomplete lexical representation, whose "gaps of information" are dispensable for recognizing the stimulus. As long as the re-reading of stored words continues, the representation is updated and refined, improving future processing (Wong et al., 2011). To compensate, these gaps of information can be "filled" through tasks that require a fine-grained analysis of all the word's features, such as spelling or writing (Burt & Tate, 2002). Moreover, handwriting combines the motor-kinesthetic practice with the visualization of the word, promoting an effective internalization of the lexical representation (Shahar-Yames & Share, 2008; Ouellette, 2010), possibly through the use of embodied memory (Mangen & Balsvik, 2016).

The gaps of information do not appear to be random. Due to encoding interference of spatial information, perceptual uncertainty seems to exist in letter positioning (e.g., Beech & Mayall, 2005; Gómez et al., 2008). Certain areas of the word appear to be more important or salient than others, leading to a word identification process that relies on some specific traits. By contrast, the gap area is overlooked and more prone to errors. This particularity in the visual perception of the word can be explained by dividing the stimulus into two axes (Figure 1): a vertical axis (from top-to-bottom) and a horizontal axis (from left-to-right).

Figure 1

Vertical Axis versus Horizontal Axis



The *vertical axis* divides words into superior, inner, and inferior layers, a division associated with letter-shape. Lowercase letters —the most familiar format (Perea & Rosa, 2002)— can be categorized into three shape types: 1) ascending letters (b, d, f, h, k, l, t); 2) descending letters (g, p, q, y); and 3) neutral letters (a, c, e, m, n, o, r, s, u, v, w, x, z) (cf., Perea, 2015). Except for ⟨i⟩ and ⟨y⟩, all vowels are classified as neutral (i.e., they are entirely contained in the inner layer), while only the ⟨j⟩ has both ascending and descending features (i.e., it dwells in all layers). The salient features of the superior and inferior layers —those of the ascending and descending letters— were coined as *external features* (Wilkins et al., 2009). In contrast, elements of the inner layer are referred to as *internal features*, encompassing all neutral letters and, most notably, vowels.

Letter-shape differences are particularly significant for an incomplete lexical representation. For instance, external features provide greater salience for lexical access, while internal features are not sufficiently relevant (Beech & Mayall, 2005). Based on this observation, the *external features hypothesis* claims the superiority of external over internal features (Figure 2). As the internal layer appears to have redundant information for human cognition, the "incompleteness" or gap of information should allegedly be located right there. Moreover, these observations support a consonantal superiority in information storage (New et al., 2008; Soares et al., 2014),

indicating that vowels are the most affected letters in memory. It is important to clarify that the external features hypothesis should not be confused with the *word envelope hypothesis*, as the latter focuses on the outline shape of the word (cf., Beech & Mayall, 2005).

Figure 2

Vertical Axis (from top-to-bottom)



Note. The added numbers represent the presence or absence of information gaps. The lower the number, the more salient the feature is for the perceptual system. Conversely, higher numbers indicate a more significant lack of information, represented by darker shades of gray. There is no comparative evidence regarding the relative superiority between the superior and inferior layers, so they hold the same value.

The superior storage of external features is further supported by studies that note an advantageous effect of consonantal for lexical access and a disadvantage when relying on vowels. For example, in a study of French-speaking adults, low-frequency words with ascending/descending letters elicited faster responses than those without them (Lété & Pynte, 2003). Also, consonants contain sufficient information to guide word recognition, even when vowels are omitted (e.g., Mehler et al., 2006; Perea et al., 2009, for eye-tracking evidence). Similarly, Duñabeitia and Carreiras (2011) observed that when a word is primed³ in a lexical decision task (e.g., *CASINO*), primes composed solely of consonants (e.g., *csn*) are more effective than those comprised

³ Priming refers to the activation of an idea or behavior triggered by an associated element that serves as a cue (Bargh, 2006; Bermeitinger, 2016). In this context, primes are the stimuli that trigger the activation of a target.

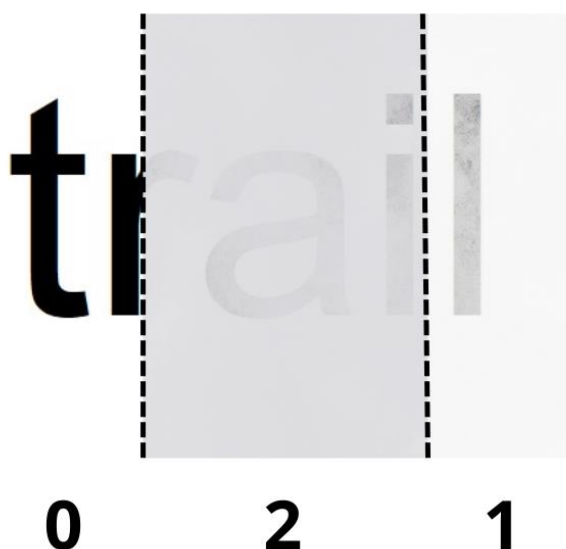
only by its vowels (e.g., *aio*).

It is also interesting to note that there are differences in processing some subtleties over others. For instance, Paap et al. (1984) found that the modification between *than* and *tban* went undetected, while *than* and *tdan* were successfully distinguished. This nitpicking indicates that not only does the visual perception system fail to notice internal features (between ⟨b⟩ and ⟨h⟩), but it can also be precise enough to appreciate the difference between ⟨b⟩ and ⟨d⟩. In line with this neutral letter indiscernibility, Martin (2017) showed a higher misspelling rate for vowels than for consonants at higher proficiency levels.

Regarding the *horizontal axis*, some studies argue for a superiority of the beginning letters and, to a lesser extent, final letters in word recognition (e.g., Davis, 1999; Fischer-Baum et al., 2010). In contrast, letters in the middle of the word are as "unsalient" as the internal features of the vertical axis. An interesting example is the *typoglycemia phenomenon*, where transposed nonwords can still be read by misperceiving them as the original word. In a brief report by Rayner et al. (2006) titled "*Raeding Wrods with Jubmled Letters: There Is a Cost*," the authors noted that reading speeds were faster when transposed letters were in the middle of the word. Conversely, they found that the slowest reading speed occurred when the transpositions were located at the beginning of the word, seconded by the ending (Figure 3). This observation suggests that transpositions that modify the middle of the word are perceived as more similar to the original word (e.g., *reading* & *raeding*), in contrast to those occurring at their edges (e.g., *reading* & *erading*). In line with this evidence, Fischer-Baum et al. (2010) proposed the "both-ends" letter-position coding, stating that letter encoding goes from left-to-right and, at the same time, from right-to-left. All this emphasizes the apparent lack of consideration for the middle portions in the word recognition process while focusing on the selective privilege of the initial letters, followed by the final ones (see Rayner & Pollatsek, 1989). By the same token, David (1999) argued that spelling errors more frequently affect the letters in the middle of the word, while those at the beginning are less affected, alluding to a *bow-shape serial position function* between the start and end of the word. In his words, "*the accuracy rate is greatest for letters in the first position, and then decreases from left to right for internal letters, but increases for the final letter*" (David, 1999, p. 95).

Figure 3

Horizontal Axis (from left-to-right)



Note. The hierarchical salience between all positions is represented through different numbers and shades of gray.

The importance of the beginning letter is evident in the *Tip-of-the-Tongue phenomenon*, i.e., the feeling of knowing a word that cannot be recalled at the moment (Schwartz & Metcalfe, 2011). When recalling a lexeme, there is often a selective retrieval of its first letter/sound of the missing word (Koriat & Lieblich, 1974). Moreover, among Indo-European languages, a transposition at the initial position seems more disruptive than one occurring in the middle of the word (see White et al., 2008). This selective nitpick implies a great focus and positional relevance of the beginning letters, which may alter how new lexemes are learned. For example, when attempting to learn nonwords with "illegal" doublets at the beginning (e.g., *rrug*), their recalling end up with transposition errors towards more "legal" positions (e.g., *rugg*) (Wright & Ehri, 2007; Pacton et al., 2013). In other words, linguistic processing tends to adhere more strictly to the "legality" of initial positions than medial or final ones.

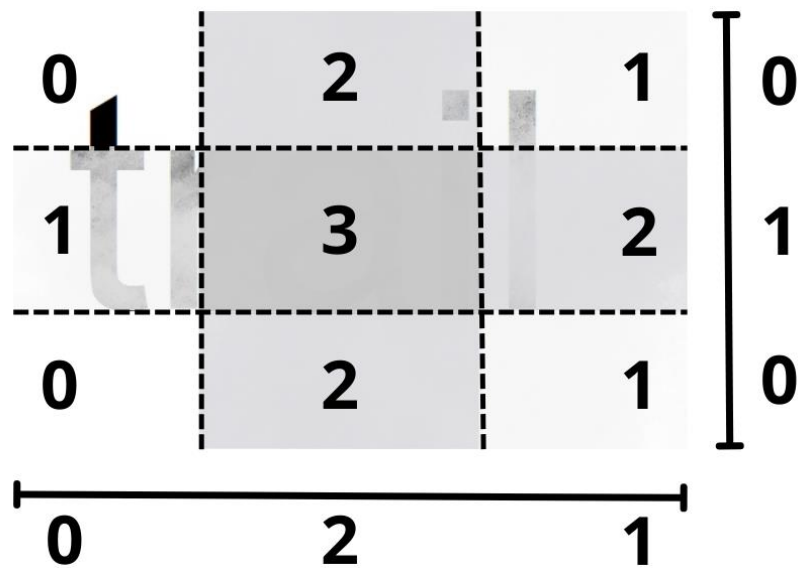
Finally, a study by Vergara-Martínez et al. (2013) —which did not focus on letter position but instead on the modification type— showed that transposition pairs (e.g., *mother/mohter*) seem more similar than substitution ones (e.g., *mother/mopher*). While both pairs involved changes in the middle of the word, the pair with less similarity (i.e., substitution pair) altered an ascending consonant for a descending one, which is perceptually more noticeable. Thus, interpreting these findings can be enriched if the

analysis considers letter-shape and letter-position.

Based on the above, when combining the vertical and horizontal axes, the nature of the salient traits becomes more apparent (Figure 4). The priority of the word perception system lies within the entire periphery of the word, with a straightforward inclination towards the left over the right. By contrast, the word's frequently forgotten and least privileged portion is located in the inner-middle portion. Such a description coincides with the description of the incomplete lexical representation's gap of information. The selectivity of some salient features over others argues for a peripheral bias in the perception process and suggests a bias in the learning process of new lexical representations. Conversely, the gap can be complementarily "filled" by using the motor-kinesthetic aspect of handwriting.

Figure 4

Alleged Nature of the Incomplete Orthographic Representation



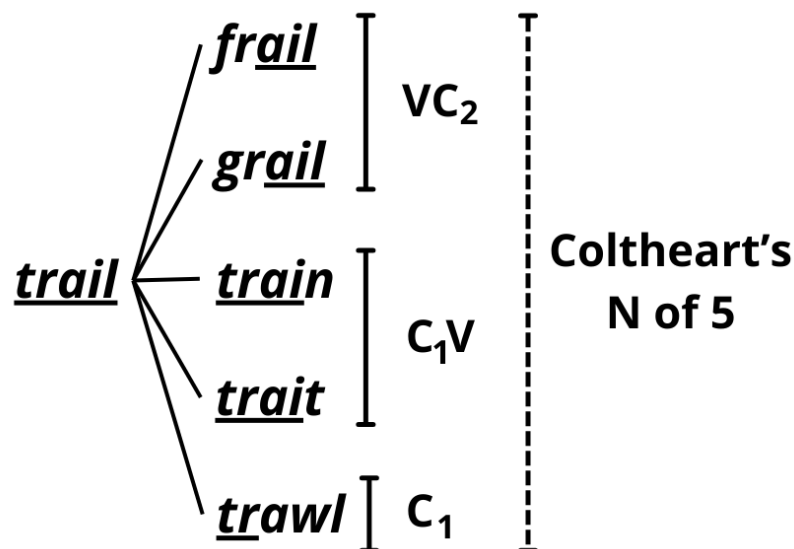
After examining the processing of individual lexical representations, it is essential to understand how similar representations relate to one another in the lexicon.

1.2. Lexical Neighborhood

The primary operationalization of "word similarity" in psycholinguistics is the *lexical neighbor*, i.e., a word "similar" to another given target. By extension, a *similarity neighborhood* comprises the whole number of neighbors of that particular word (Hammond, 2004). Coltheart et al. (1977) proposed the *neighborhood size* or *density* metric —nowadays referred to as *Coltheart's N* (Figure 5)— as the number of neighbors a given word can create after substituting a single letter or phoneme. This metric was previously suggested by Havens and Foote (1963), who referred to neighbors as the *word's competitors*. Since neighbors can only be obtained through substitutions, all words that arise from these alterations necessarily share the same length and letter/phoneme positions as the original word. This limitation was coined as the *slot-based position coding*.

Figure 5

Substitution-only Restriction (Coltheart's N)



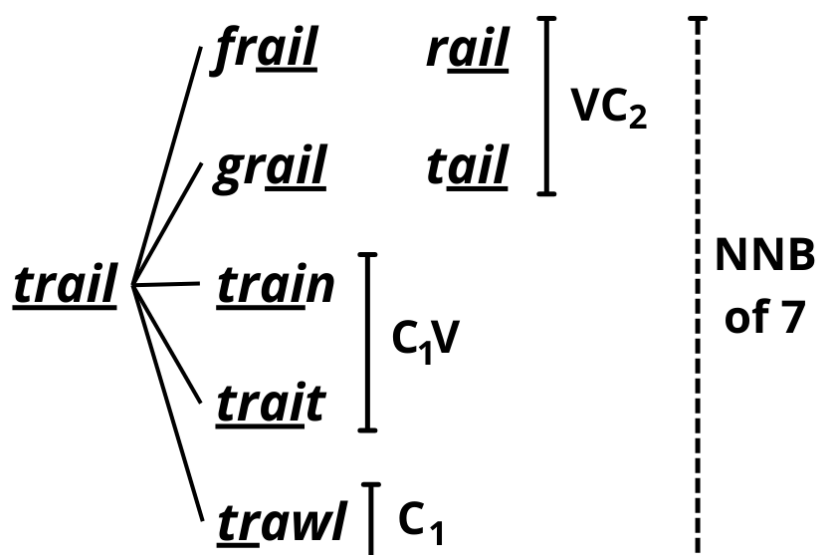
Note. The words considered here as neighbors and those of the following figures were taken from the *Hoosier Mental Lexicon* (Nusbaum et al., 1984). In this image, the features shared between the neighbors and their original target word are referred to by the CVC (Consonant-Vowel-Consonant) abbreviated nomenclature (Treiman et al., 1995).

Following the first boom of the lexical neighborhood concept, doubts arose regarding whether substitution was the only modification possible for creating a

neighborhood. This discussion led to the inclusion of additions and omissions as legitimate modifications, challenging the restrictions of size and position inherent in the slot-based position coding. The inclusion of additions and omissions allowed neighbors to have different lengths (e.g., *art* & *cart*). To accurately determine neighborhood size, authors such as Luce (1986) and Hammond (2004) employed the metric of *single, minimum, or standard edit distance metric*. This procedure involves making substitutions, additions, or omissions to the target's letters or phonemes in a left-to-right manner, thereby obtaining the total number of possible neighbors without skipping any. Bailey and Hahn (2001) referred to this neighborhood size subtype as the *number of such neighbors (NNB)* (Figure 6).

Figure 6

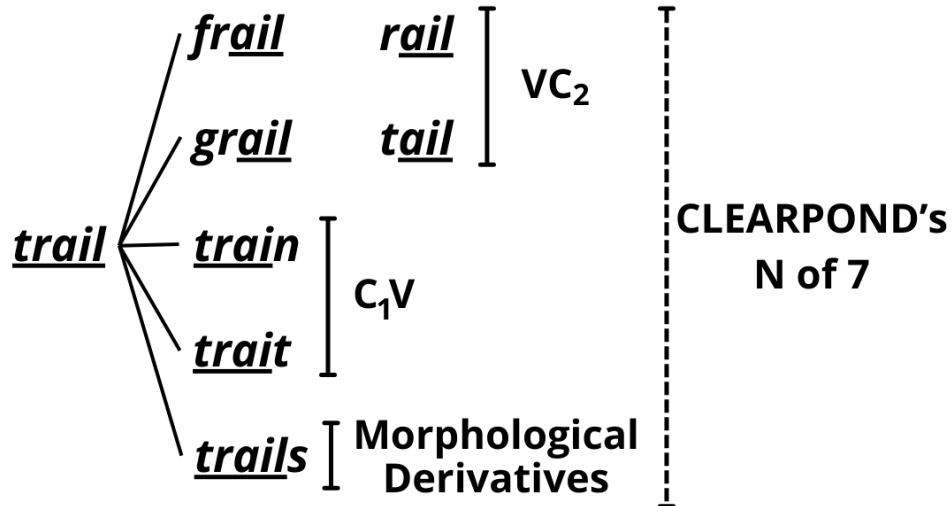
Substitution & Addition/Omission Restriction (NNB)



Years later, Marian et al. (2012) introduced the *Cross-Linguistic Easy Access Resource for Phonological and Orthographical Neighborhood Densities (CLEARPOND)*. This is a free-access database designed for the calculation of neighborhood size based on the inclusion of additions and omissions, rather than only allowing substitutions (<https://clearpond.northwestern.edu/>). However, unlike NNB, it also includes morphological derivatives of the target as plausible neighbors (e.g., *mean* & *means*), a subtype referred to as CLEARPOND's N (Figure 7).

Figure 7

Substitution, Addition/Omission & Morphological Derivatives Restriction
(CLEARPOND's N)



Note. Unlike the other neighborhood size subtypes, CLEARPOND's database does not consider *trawl* a neighbor, possibly because this word is not present in it.

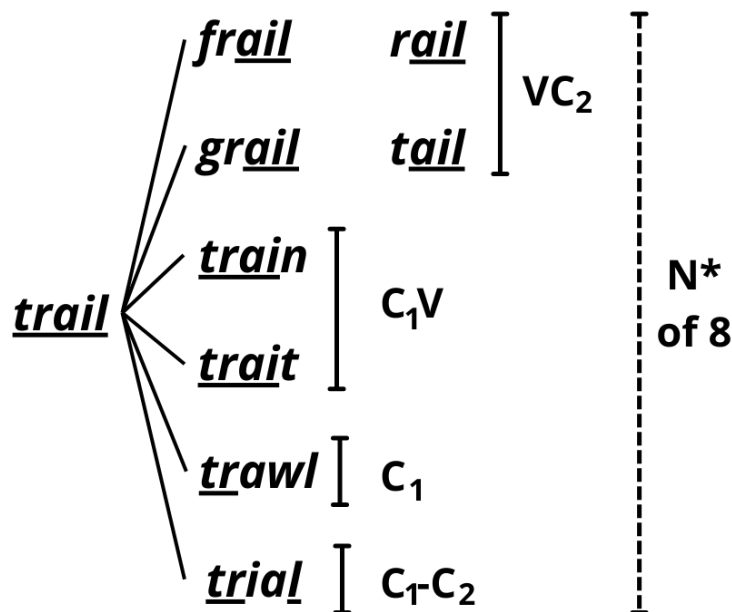
To the author's knowledge, there has hitherto been no overt discussion over including morphological derivatives as legitimate neighbors, nor are there other authors that have counted them as such for the neighborhood size metric. The inclusion of morphological derivatives is redundant in terms of lexical neighbors. The relationship between a word and its morphological derivatives is based on "sameness" rather than "similarity," as one is already fully encompassed within the other, both lexically and semantically. In any case, a thorough discussion of the inclusion of morphological derivatives is made in Chapter 3.

Although the concept of the lexical neighborhood first emerged in 1977 (Coltheart et al., 1977), Bruner and O'Down (1958) had already noted that neighbors arising from transpositions were subjectively perceived as very similar (e.g., *aviation* vs. *avitaion*). These transposition neighbors (e.g., *mother* & *mohter*) even surpassed substitution neighbors (e.g., *mother* & *mopher*) in terms of similarity (e.g., Perea & Fraga, 2006; Gomez et al., 2008; Vergara-Martínez et al., 2013; Perea et al., 2023). This higher similarity was maintained even when the transposition occurred between two non-contiguous letters—for example, *caniso* and *casino* were considered more similar than *caviro* and *casino* (Perea & Lupker, 2004). Andrews (1996) coined the transposition-

related similarity as *specific similarity*, whereas the one originating from substitutions, additions, and omissions was named *general similarity*. Based on the Generalized Neighborhood Model of Bailey and Hahn (2001), the proposed *subset lattice distance metric* included all four modifications as legitimate. Moreover, Davis (2005) designated the neighborhood size subtype that allowed transpositions as N^* (Figure 8). With the inclusion of transpositions as plausible modifications, Coltheart's positional restriction was already discredited.

Figure 8

Substitution, Addition/Omission & Transposition Restriction (N^)*



Authors like Yarkoni et al. (2008) retorted that the one-modification-only restriction for creating neighbors was unacceptable. They objected that it binary and categorically eliminates all other types of words that are also similar to the target. Since most English words have a length of four or five letters, longer words typically exhibit a low number of neighbors —if any at all. This fact is especially problematic for languages with a longer average length (e.g., Perea & Carreiras, 1998). To complicate matters, some argued that the neighborhood effect was not limited by the number of modifications between two words (e.g., Wurm & Samuel, 1997). This idea would lead Bailey and Hahn to assert that "*all relevant items are neighbors to some degree*" (2001, p. 572). Thus, in contrast to the categorical definition of neighborhood, a more gradient and continuous scale of similarity was proposed (Nosofsky, 1986), a discussion that

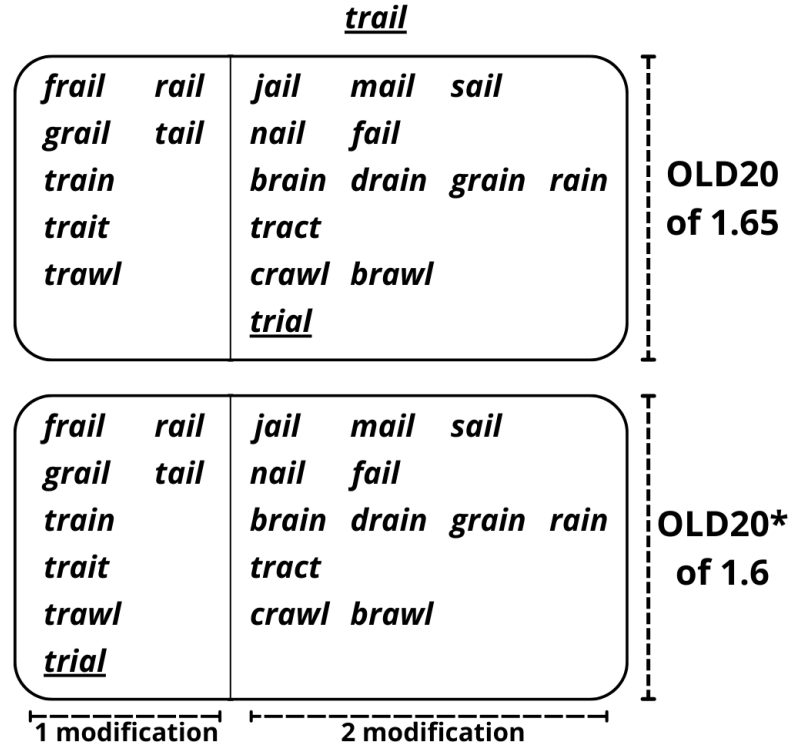
undermined the paradigms of what was considered "similar" at the time.

Nevertheless, a methodological problem hindered the breakthrough. How can similarity be measured if all words can be neighbors to each other? The categorical restriction of the original neighborhood definition provides clear boundaries for what is similar and not, facilitating the metric's calculation. In contrast, the absence of limitations in the gradient paradigm makes it impossible to represent similarity across such a large sample of words numerically. To address this difficulty, Yarkoni et al. (2008) created the *Orthographic Levenshtein Distance 20 (OLD20)*, while Suárez et al. (2011) developed a similar metric for the phonological domain (*POLD20*). Both alternatives were based on the *Levenshtein Distance (LD)*, defined as the number of additions, omissions, and substitutions required to transform one word into another target (Levenshtein, 1965). For instance, to change the word *smile* to *similes*, the addition of ⟨i⟩ and ⟨s⟩ is required, resulting in a distance of 2. OLD20 and POLD20 are defined as the average LD needed to obtain the 20 nearest orthographic or phonological neighbors of a given word, respectively. The lower the OLD20 number, the easier it is to find orthographically similar words to the target word, indicating a higher overall similarity within the language as a whole. This measure became the new standard by successfully combining the gradient neighbor paradigm with a plausible operationalization. Its use allows for the measurement of similarity in longer words, as it considers the existence of neighbors with a distance greater than one.

However, the new measures presented problems. Perea stated that the OLD20 measure "*weights single substitutions more heavily than letter transpositions*" (2015, p. 79). This assertion stems from the fact that the OLD20 calculation treats each transposition as two separate substitutions, which goes against the previously mentioned superiority of transposition over substitution concerning similarity. For example, consider the substitution neighbors *trail* and *train* —where ⟨n⟩ replaces ⟨l⟩—, compared to the transposition neighbors *trail* and *trial* —where ⟨i⟩ swaps with ⟨a⟩. The former would have an LD of 1, while the latter would have an LD of 2 —due to the substitution of ⟨a⟩ for ⟨i⟩ and the substitution of ⟨i⟩ for ⟨a⟩. Figure 9 illustrates how the OLD20 calculation is affected when transpositions are present or not as a single modification. Moreover, it compares a version of OLD20 that does, referred to as *OLD20**.

Figure 9

Substitution & Addition/Omission Restriction (OLD20) vs Substitution, Addition/Omission & Transposition Restriction (OLD20)*



Despite the evidence surrounding categorical versus gradient neighborhood definitions, there is still no consensus on which should be considered the core paradigm. Some authors still avoid the gradient neighborhood view and exclude transpositions as genuine modifications (e.g., Han et al., 2016; Jones et al., 2021).

Anecdotally, Chan and Vitevitch created the *clustering coefficient*, an indicator based on graph theory that represents the proportion of a target's neighbors that are also neighbors with each other (2009; Goldstein & Vitevitch, 2014). This metric emphasizes the idea that not all neighbors within a particular similarity neighborhood are necessarily similar to one another. Alternatively, following the gradient definition, not all neighbors within the same neighborhood exhibit the same degree of similarity between each other.

In this way, lexical similarity has changed its theoretical restrictions over the decades. The divergence of what constitutes or does not constitute a lexical neighbor continues to this day, as authors often take the "general" similarity idea for granted without delving into its complexities. This is why researchers tend to build their work on

older and less complicated definitions (e.g., Ansteeg et al., 2022, as an example of a recent study).

The fact that transpositions create more similar neighbors than substitutions presents a problematic observation for some (e.g., Vergara-Martínez et al., 2013). The controversy lies in the inclusion of transpositions as legitimate single modifications, as it rules out all the other measures that do not consider them (Coltheart's N, NNB, CLEARPOND's N, OLD20) while reinforcing those that do (N* and OLD20*). Furthermore, in line with the lexical gap of information, Perea (1998) found that substitution neighbors are subjectively perceived as more similar when modifications are made in the middle of a word (e.g., *lint* & *list*) rather than at the beginning or end (e.g., *lint* vs *line*). This observation implies that the impact of a modification not only depends on the type of modification (substitution vs transposition) but also on the position where such modification occurs (e.g., beginning versus middle letters). Thus, the body of research on lexical neighborhoods is compatible with that of incomplete lexical representation.

Interestingly, the idea that preceded lexical neighbors —the "word's competitors" (Havens & Foote, 1963)— did consider the position of the substituted letter. They observed that modifying letters at the beginning or ending entailed less similar pairs than when changes occurred in the middle of the word. To date, the newer and more complex metric of orthographic similarity —the OLD20 (Yarkoni et al., 2008)— does not include transpositions as a single type of modification (cf., Perea, 2015). It also does not discriminate between the spatial aspects of the word, as it fails to recognize the greater salience of peripheral features and the higher incidence of issues with inner-middle ones. Therefore, Havens and Foote are the pioneers of the lexical neighborhood concept and the first to propose that similarity depends on positional biases. Instead of asking "which" and "how many" modifications should be allowed, they focused on "where." With this new perspective, two neighbors are perceived as more similar as long as the range of modifications that differentiate them occurs within the "gap of information" of the word. "Vowel per vowel" modifications are less likely to be perceived, as these features fall within the gap of information of the incomplete lexical representation. Conversely, changes in ascending or descending consonants are more easily distinguishable, as they contain highly salient features for the visual perception of words.

In conclusion, the nature of lexical representations and what is a lexical neighbor is now defined. What remains to be addressed is how these neighbors interact with

each other during activation, a question that requires the introduction of lexical frequency.

1.3. *Confusability Approach of Similarity*

The concept of similarity is ill-defined in psycholinguistics (Luce & Pisoni, 1998; Bybee, 2010), an issue that will be addressed in detail in Chapter 3. Nevertheless, one of the most frequently mentioned proposals regarding lexical similarity is its association with confusability, i.e., the degree to which one stimulus can be confused with another (Luce, 1986). Under this perspective, the *confusability approach of similarity* (*C-similarity*) posits that when the similarity between two different stimuli is high, the likelihood of confusing one for the other is higher.

However, how can this perceptual confusion be explained? Considering sections 1.1 and 1.2, the answer to this question must consider the incomplete nature of lexical representations. This consideration becomes particularly significant when observing the consequences of such confusability: errors in comprehension (i.e., misperceptions) or production (i.e., misspellings and mispronunciations).

More importantly, why do visual perception systems tend to generate selective uncertainty? What cognitive benefit comes from having an incomplete and error-prone lexical system? A plausible answer to this problem could relate to an evolutionary cognitive advantage in processing changes and nuances. If human cognition had a complete lexical (or semantic) representation, it would lead to categorical perception akin to a computer's artificial vision: a binary identification of "yes" (100%) and "no" (0%). With such rigidity in the system, in order to recognize different letter-shapes as the same, the lexicon would require prior storage of many representations of the same idea. In this way, the identification process could weigh all possible forms a word could take, whether due to different handwriting styles or font sizes (i.e., shape invariance). The issue with this approach is that such extensive storage would consume an enormous amount of cognitive resources, and the brain always seeks a more economical solution, given its limited memory capacity (Leibold & Kempter, 2006). Another alternative would be to suggest that cognition has many representations at the letter level, allowing for invariant recognition based on the analysis of isolated letters (see Grainger & Dufau, 2012).

Such challenges are not an issue for machine learning. A computer can effortlessly store thousands of resources simply through the exposure to a huge number of examples with various features, which are then integrated to the data storage to process future stimuli. If the dataset of high-quality examples is sufficiently diverse —collectively known as *Big Data*—, machine learning shifts from a simplistic binary perception —100% "yes" and 0% "no" that leads to categorical errors— to an

algorithm capable of discerning nuances that are impossible for human perception to distinguish (see Asogwa et al., 2019 for a review of machine learning techniques). In other words, the programmed perceptual adaptability of computers comes from the sheer quantity of datasets.

Conversely, a more frugal network system —such as human cognition— cannot store such a vast amount of information, much less with the same resolution. Therefore, the incomplete nature of lexical representations may account for overcoming this limitation (see Fig. 4). These representations provide sufficiently relevant information to recognize an object (e.g., a word). Nevertheless, such information is not entirely detailed, allowing different forms of the same stimulus to be misperceived as identical, i.e., a perception unaffected by its shape (cf., Perea & Rosa, 2002). This justification also explains the human tendency to misinterpret one object as if it were another, i.e., misperceptions caused by confusability. Thus, lowering the perceptual resolution compromises effectiveness in exchange for the ability to process more with less. In other words, rather than infinitely increasing the amount of data (i.e., quantity), as occurs with the machine learning of computers (Fadaee et al., 2017), the human memory system adjusts the quality of its representations (Baroni et al., 2007), which compromises its discriminative power.

There are cases where the misperception of a stimulus can lead to better text comprehension. An example is *LEET words*, defined as lexemes that have replaced some of their letters with digits that share similar perceptual features (e.g., *R3BOUND* instead of *rebound*, *484NDON* instead of *abandon*). Studies have shown that using LEET words as primes for their real counterparts enhances recognition (Perea et al., 2008; Kinoshita et al., 2014). The same can be said for the misperception of a nonword with transposed letters as if it were its real neighbor (see Chambers, 1979; O'Connor & Forster, 1981).

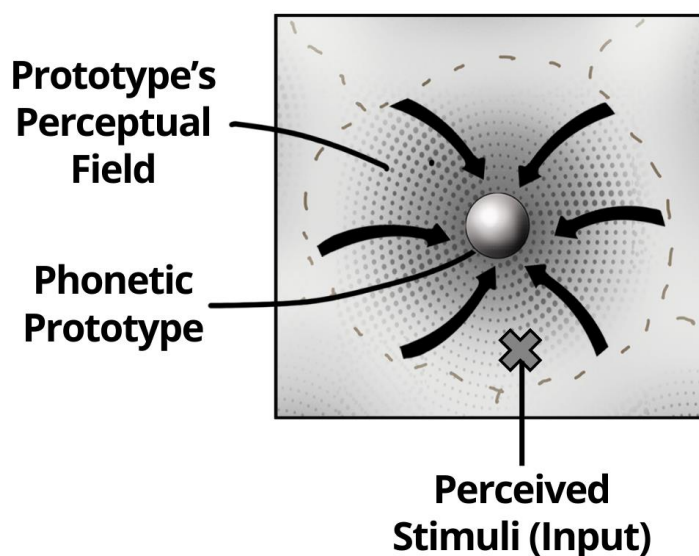
According to the lexical gap of information (see Fig. 4), initial consonants are the most important features when processing lexical similarity (Sendlmeier, 1987; Bailey & Hahn, 2001), indicating a prioritization of the first position. Due to this perceptual importance, misrecognition due to confusability is more significant when these elements remain unchanged. An example of this is the *typoglycemia phenomenon*, which elicits a correct and rapid activation and recognition of a target word as long as the differential transposition of the pseudoword occurs at its middle (Rayner et al., 2006).

The distorted perception of language is reminiscent of what Kuhl (1994)

described as the *perceptual magnetic effect*. She focused on the phonological domain to explain the decline in phonetic discrimination during the first year of life (e.g., Polka & Werker, 1994). To this end, Kuhl proposed the *prototype* concept, a phonetic example that is representative of an individual's linguistic environment. After repeated encounters with the stimulus, the child acquires a new phonetic prototype. This prototype distorts the original perceptual space to form a unique and specific perceptual field. Thanks to this distortion, the prototype becomes a lens that biases the perception of future stimuli. According to Kuhl, these perceptual fields cause different interpretations of the same input between people of different mother tongues. When a heard stimulus falls within the prototype's perceptual field, the individual will perceive the sound as if it were the prototypical phoneme of their respective linguistic environment (Figure 10). The closer the perceptual distance between the stimulus and the stored prototype, the greater the likelihood of distorting the input and interpreting it as the prototype (Iverson & Kuhl, 1994). The perceptual magnetic effect claim attempts to answer why children can discriminate between a wide range of sounds at birth, as their phonetic prototypes have not yet fully formed, resulting in minimal perceptual distortion. At the same time, it also provisionally explains why they slowly lose such an ability, as the formation of new prototypes creates distortions of perceptual space.

Figure 10

Perceptual Magnetic Effect



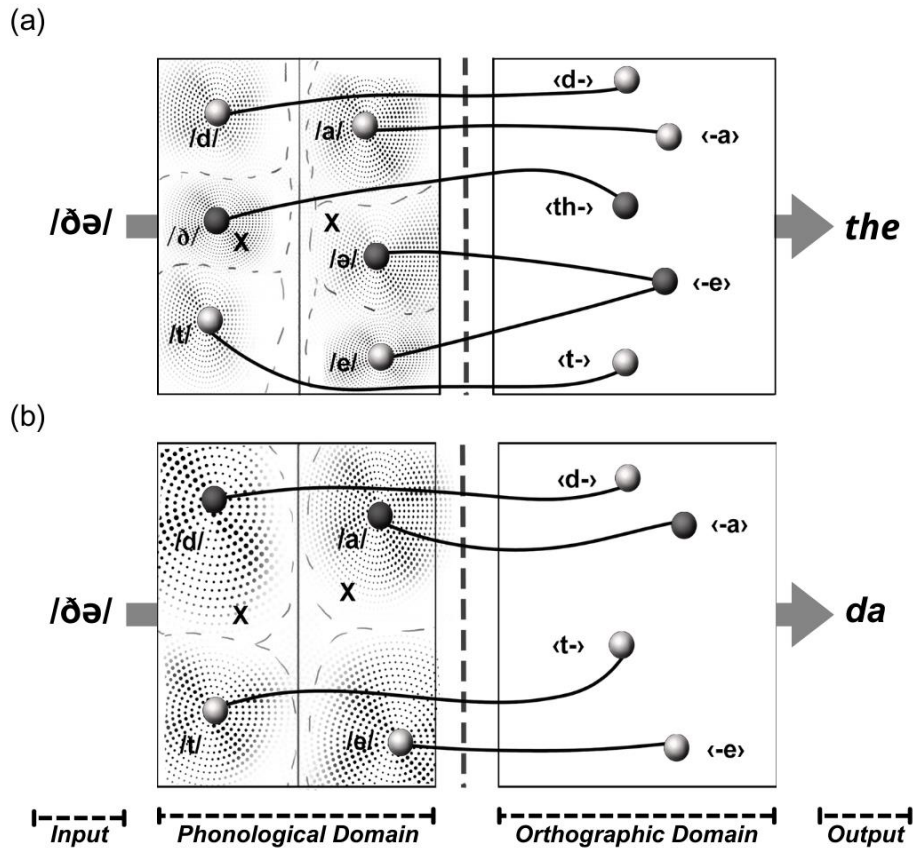
Note. The image shows that the prototype (gray sphere) distorts the perceptual area to create its perceptual field (delimited by the dotted line).

The X represents the perceived stimulus (input), misperceived as the prototype due to the perceptual magnetic effect (black arrows). The figure attempts to account for the effect by comparing it to the gravitational pull.

To better understand this, suppose a thought experiment compares an English speaker with a Spanish speaker when they hear the English word /ðə/ (the). Since the phonemes /ð/ and /ə/ are common in English, the English speaker's cognition will have these phonemic prototypes already developed. Thus, the English speaker will identify the phonemes /ð/ and /ə/, distinguishing them from their neighboring phonemes /d/ and /t/, and /e/ and /a/, respectively. By contrast, /ð/ and /ə/ are foreign phonemes for Spanish speakers, meaning there is no specific phonetic prototype for these sounds in their cognition. Due to this absence, the perceived stimulus will inevitably fall into the perceptual field of the closest prototype developed in Spanish. In this example, the phoneme /ð/ falls within the perceptual field of /d/, while the phoneme /ə/ falls within the perceptual field of /a/, leading the Spanish speaker to misperceive the original stimulus as the syllable /da/, or <da> if written (Figure 11).

Figure 11

Distortion due to the Vernacular Phonetic Prototypes



Note. From the left, the phonological input falls into the perceptual field of a prototype (gray spheres), which causes the activation of the specific phonemic prototype (black spheres) and its orthographic counterpart on the right. This results in an orthographic output that reflects what has been interpreted by each listener. In (a) it is shown the process for English speakers while in (b) it is illustrated for Spanish speakers.

Kuhl's proposal aims to explain the misinterpretation caused by the similarity between two phonemes. Phonemic similarity relates to what can be coined as a "sublexical similarity," a topic this work will not develop as it falls outside its scope. Nevertheless, her model allows for understanding the perceptual distances between prototypes within the same linguistic domain. Thus, its basis should also be applicable at the lexical level. Following this logic of the previous sections, if two neighbors diverge due to a modification that predominantly affects the lexical gap of information (e.g., *trail* & *trill*), they are more likely to be confused because of their higher similarity,

resulting in a smaller perceptual distance. On the same token, if a pseudoword is misperceived as if it were a real lexical prototype (Figure 12), this is known as a *lexicalization error* (An et al., 2023). Conversely, suppose the neighbors differ due to a modification in a more peripheral location (e.g., *trail* & *train*). In that case, the similarity will be lower, and the perceptual distance will be greater, making it easier to distinguish them as independent units. Therefore, the size of the perceptual distance is equivalent to the degree of similarity between the stimuli (Goldstone & Son, 2005).

Figure 12

Lexicalization Error

N

ot

e.

T

h

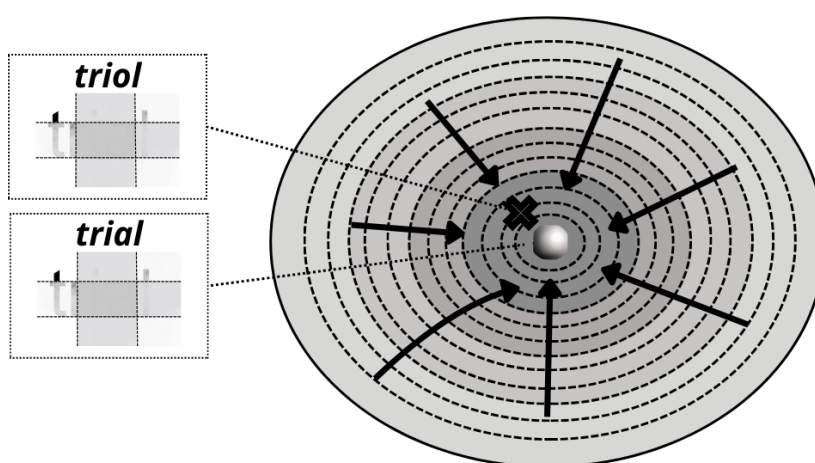
e

s

c

h

e



matic illustrates how the pseudoword *triol*, depicted as an X, is misinterpreted as the word *trial* due to the high similarity between their incomplete lexical representations. This misperception occurs because the only difference between the two stimuli lies in the inner-middle portion, the most overlooked word segment during fast lexical processing.

After establishing the parallel between C-similarity and the perceptual magnet effect, the question is: How can lexical representation impose its perceptual magnetic field over its neighbors'? As Iverson and Kuhl (1994) stated, a prototype is "representative of the linguistic environment." Such representativeness is only granted after it has been notably frequent in a specific spatiotemporal communication area. This aspect can be linked to the *lexical token frequency (L-TF)*, the number of occurrences of a given word within a particular context of communication. This metric divides the neighborhood into *frequently-used* neighbors —those that exert a more

significant influence over the entire neighborhood— and *rarely-used* ones— those inhibited by the former (Treiman et al., 1995; Berg, 2014).

In this sense, the higher the L-TF, the greater the likelihood it will develop a lexical prototype and increase the size of its perceptual field, which will later clash with that of its neighbors. The word with the highest L-TF in the neighborhood creates interference over its lower L-TF neighbors, an inhibitory effect formalized as the *principle of lexical inhibition* (Grainger & Jacobs, 1996). Such inhibitory impact predisposes the occurrence of more lexicalization errors —since pseudowords have an L-TF of zero— and interferes with neighbors of lower L-TF, slowing their activation.

Substantial evidence supports the idea that the L-TF influences the appearance of lexicalization errors. For example, there is a higher rate of lexicalization errors for transposed-letter pseudowords, particularly when these come from high L-TF words compared to those derived from low L-TF words (see Andrews, 1996; O'Connor & Forster, 1981; Perea et al., 2005). Additionally, in electrophysiological studies, the N400⁴ component is evoked not only during the recognition of high L-TF words but also when processing their neighboring transposed-letter pseudowords (Vergara-Martínez et al., 2013).

Due to lexicalization errors, transposed neighbors (e.g., *caniso*) can serve as primes to facilitate the processing of their real counterparts (e.g., *casino*) (Perea & Lupker, 2004). Sometimes, the magnitude of perceptual distortion is such that even a pseudoword with glaring spelling errors (e.g., *apgricfot*) can activate the desired lexical representation (e.g., *apricot*) (van Assche & Grainger, 2006). Therefore, the visual word recognition system appears to ignore and overlook these intrusive elements. Both the lexical gap of information and the "lexical" perceptual magnetic effect can explain this. The former provides the adaptability to many shapes while the latter allows for the activation of similar, albeit more typical, lexical representations.

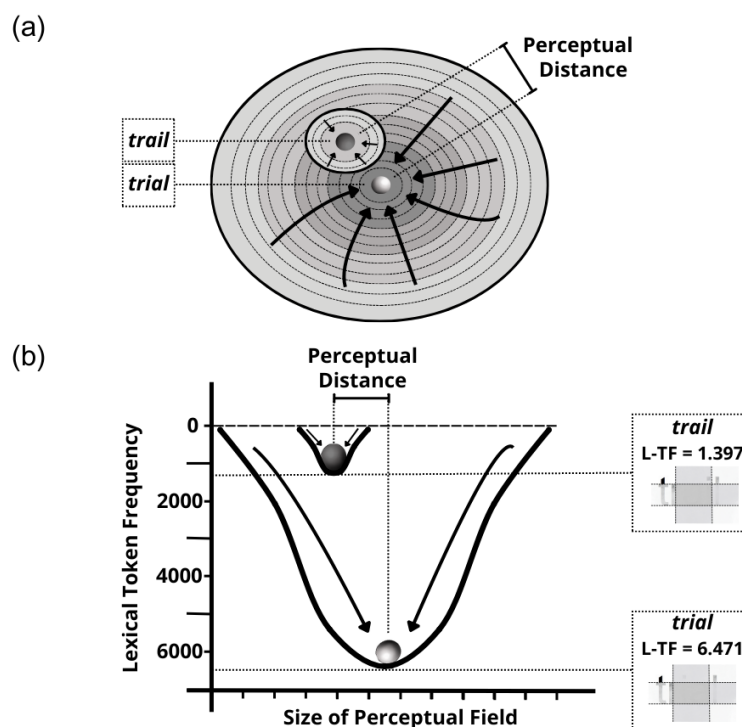
It should be noted that there are a significant number of studies and models supporting the inhibitory effect exerted by high L-TF neighbors (e.g., McClelland & Rumelhart, 1981; Luce, 1986; Grainger et al., 1989; Luce et al., 1990; Luce & Pisoni, 1998; Perea & Rosa, 2002; see Perea, 2015, for a review), dating back to the first study on lexical neighborhoods (e.g., Coltheart et al., 1977).

4 The N400 component is a wave with a negative deflection at a latency of 400 ms, characteristic of evoked potential studies. It reflects lexical-semantic processing in visual word recognition studies, as it appears when the stimulus potentially carries meaning.

This inhibitory impact can be illustrated through a hypothetical example. Suppose an individual must accurately identify either *trail* or *trial* using fast guessing, without any prior context for priming. In that case, both words are likely to compete against each other due to their high C-similarity (or small perceptual distance). In this case, the neighbor with the higher L-TF will experience fewer difficulties in quick identification (Figure 13). In comparison, the word with the lower L-TF will have its activation interfered with by its neighbor of higher L-TF, which could hinder the speed of its recognition or even lead to the incorrect identification of the neighbor with the higher L-TF.

Figure 13

Lexical Token Frequency and Perceptual Field



Note. (a) shows a schematic view of the perceptual distance between the lexical prototypes/representations *trial* and *trail*. The size of their perceptual fields is proportional to their respective L-TF, whose values were obtained from the frequency database of the *British National Corpus (BNC)* (Keuleers et al., 2012). The black arrows indicate the direction of the perceptual magnetic effect of each lexical prototype. (b) shows the same concepts from a "more sagittal" perspective, illustrating the overlap

between the perceptual fields of the prototypes and how they distort the perceptual space. The x-axis represents the size of the perceptual distance and perceptual field for each prototype, while the y-axis graphs the L-TF of each word and its impact on the distortion of the perceptual mapping. As an analogy, L-TF is herein equivalent to the intensity of the "gravitational pull."

The inhibitory influence of high L-TF neighbors will occur with all words within the similarity neighborhood. However, the impact varies depending on the perceptual distance between the neighbors. That is to say, the greater the C-similarity between stimuli —meaning less modification at the periphery of the word— the higher the likelihood of inhibition. For example, *trail* and *trial* will have greater C-similarity than *trail* and *train*, since the latter pair modifies the final letter —a position that is more salient than that of the middle (see Fig. 4). Additionally, when replacing an ascending consonant with a neutral one (e.g., replacing <-l> with <-n>), the superior layer is modified, making the change more salient.

Concerning the "filling" of the information gap, Beech and Mayall (2005) conducted an experiment in which the superior and inferior features of ascending and descending letters were used as masked primes. The authors noted that participants illusorily perceived the primes as complete words, as if they had "filled the empty space" of the visual stimulus. This phenomenon has been described as the *Gestalt principle of closure*, which asserts that gaps between stimuli are perceptually "filled" through the extrapolation of simple features (e.g., Wertheimer, 1950b; Donnelly et al., 1991). This effect did not occur when internal features were shown, i.e., those within the hypothetical gap of information. Unfortunately, Beech and Mayall (2005) regarded the illusion as anecdotal, which explains why they did not elaborate on the phenomenon in detail. Considering the aforementioned, it is reasonable to presume that participants "filled the gap" with the word of the highest L-TF that best matched those external features. Therefore, this effect may stem from a projection of misrecognition, a "filling" or "closure" that illusorily distorts the most vulnerable elements of the word, namely those confined within the lexical gap of information.

In conclusion, this section has explained the inhibitory effects among lexical neighbors. Nevertheless, there is evidence of other effects on lexical processing, such as facilitatory influences and those arising from the grapho-phonological consistency of words and sequences. To explain both topics, a formal introduction to the sublexical domain is necessary, along with a detailed understanding of the language acquisition

process and the underlying memory systems.

CHAPTER 2

Appearance of the Sublexicon

2.1. *Memory Systems and the Dual-Route Model*

Children are gifted with the ability to rapidly acquire complex sequences of processes at a high level of proficiency, as can be seen when they learn to play a musical instrument. The relationship between age and language acquisition is a common topic of psycholinguistics, especially the hypothetical *critical period*, which accounts for how infants are more capable of effortlessly developing linguistic skills (Hartshorne et al., 2018). Once this threshold is crossed, this ability begins to sharply decline during late adolescence (Hartshorne & Germine, 2015), a physiological decline that is noticeable in feral children. These infants grow up in contexts of language deprivation, and once they reach a certain age, they lose the opportunity to passively acquire the language (Vyshedskiy et al., 2017). This limitation shows that children and adults assimilate language differently, and such a difference cannot be explained solely by how language is learned. To understand this, it is necessary to know how the different memory types function according to age and what linguistic domain is first developed as a consequence.

Regarding long-term memory systems, Endel Tulving (1983) originally proposed that they could be divided into two subtypes: a) autobiographical events and b) a network of isolated meanings. The first, called the *episodic memory*, is described as a system that encompasses all information related to recent or past experiences. These autobiographical "episodes" can be imagined as a sort of mental movie, with events organized narratively. In other words, they represent a collection of interconnected meanings that are temporally and spatially contextualized (Cohn-Sheehy et al., 2022). Such "movies" are internally created from all kinds of personal experiences (Tang et al., 2016), which do not necessarily need to feature the person as the protagonist of the narrative (e.g., watching a movie or listening to a story).

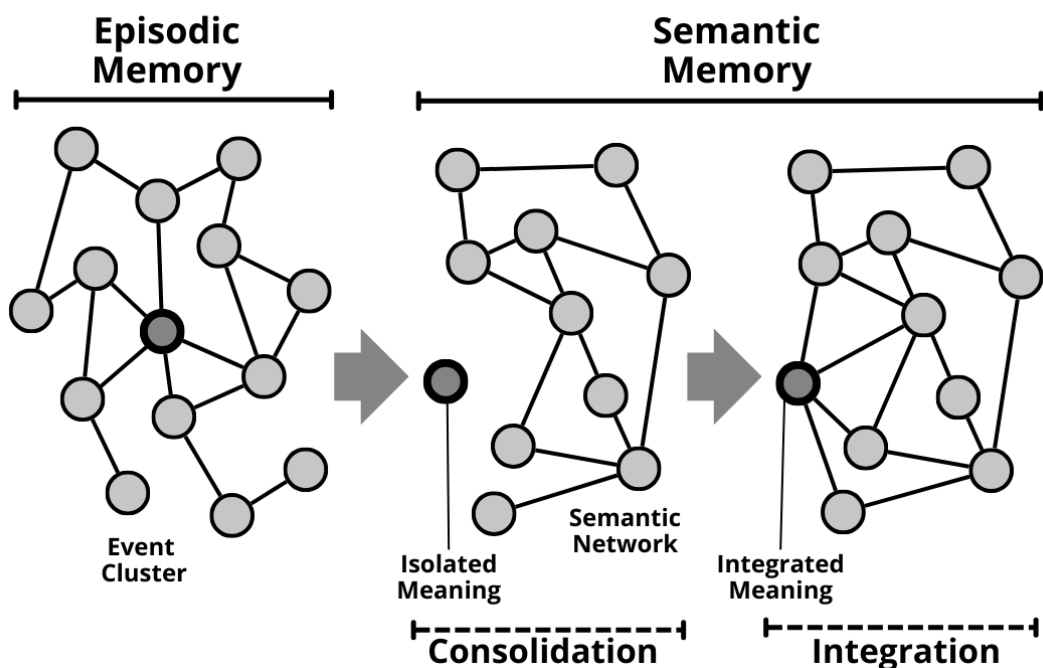
This episodic memory deteriorates over time due to the forgetting mechanism (Yalnizyan-Carson & Richards, 2022), a lifespan that can be extended through repetitive reconsolidation processes that cement them within the memory storage. Nevertheless, these reconsolidatory interventions not only reinforce and update memories, but they can also alter the original memories (Agren, 2014). Due to the

sensibility to these modifications, the reconsolidation process can create gaps of false memories, which, when pathologized, turn into a symptom called *confabulation* (see Schneider, 2007).

In order to avoid the alteration of these unstable memories, the system shifts the encoding of each element of the episodic "movie" into a more stable format (Figure 14). First, the whole episodic cluster shatters into its multiple isolated units —those included within the original event— to encode them afterward into a new network of isolated meanings, i.e., a *consolidation process*. Second, interconnections are created between the new unit and the previously stored network of semantic associations, i.e., the *integration process* (Takashima, 2017).

Figure 14

Episodic-to-Semantic Memory System

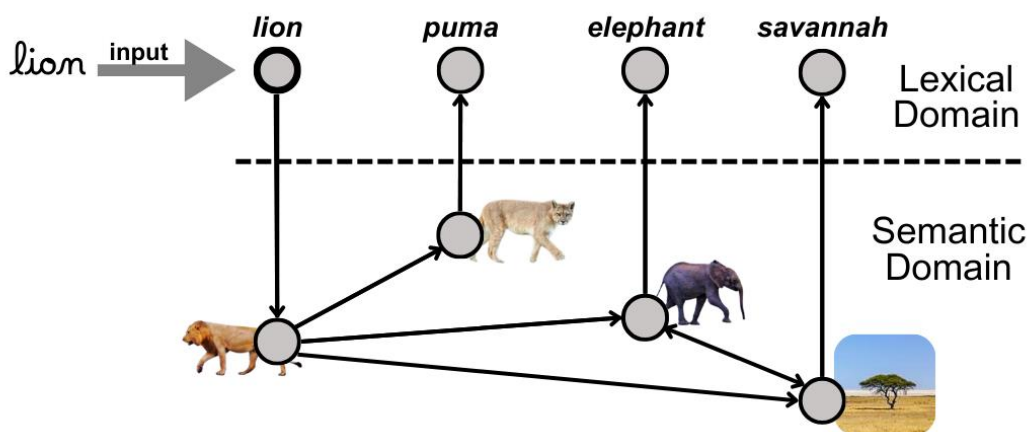


Unlike its episodic counterpart, the *semantic memory* is more resistant to deterioration, functioning as a safeguard for all perceived meanings. The retrieval of semantic memories is a daily practice, as it is the system we use the most when recognizing meanings in our environment. In contrast, recalling an episodic memory requires more cues to trigger it. Retrieving information from semantic memory has a passive component, as meanings are interconnected in the same network, ending up

with the activation of their adjacent representations through a cascaded effect (e.g., Altvater- Mackensen & Mani, 2013), even when those meanings are irrelevant. Thus, reading the word *lion* not only activates its phonological form but also elicits the recall of other felines (e.g., *cougar*, *tiger*, *serval*), related mammals (e.g., *elephant*, *giraffe*), or its habitat (i.e., *savannah*). Furthermore, it is possible to activate episodic memories associated with the activated representation (e.g., remembering watching Disney's *The Lion King* movie). Thanks to this cascading activation, it can be explained how one word can effectively prime and facilitate the processing of semantically-related words (Figure 15).

Figure 15

Cascaded Activation of the Lexical and Semantic Domain



Tulving (1983) said that the network of semantic representations can be imagined as a *thesaurus*. Like the lexicon, it stores isolated facts and meanings that shape the individual's world knowledge. The thesaurus contains semantic representations whose primordial experience is often forgotten, as they do not have the space-time label that characterizes episodic memory. Because of this, someone may know what a *lion* is, but not necessarily remember when they saw the animal or its name for the first time.

The associations between semantic representations are created and shaped from the episodic memories containing the interconnection cues. For example, an individual living in the Southern Cone of Latin America will likely associate the meaning of *lion* with that of *puma*. By contrast, if that individual lives in Southern Africa, the word *lion* is more likely to be linked to the *serval*. In other words, personal experiences in one's environment condition how the network of meanings is assembled.

The influence between the systems is bidirectional, as the semantic network can also influence the episodic one. The reconsolidations that the episodic memory needs to prevent forgetting rely on the information stored in semantic memory. Therefore, the lost gaps of episodic memory can be "filled" with what is coherent for the semantic memory (Hyman & Loftus, 1998). This assertion explains why reconstruction is not always an exact reproduction of the original autobiographical episode (Lee et al., 2020).

The bidirectional influence between both types of memory has led to the claim that they constitute a *complementary learning system* (McClelland et al., 1995; Arani et al., 2021). Since the information stored in the episodic and semantic memories can be conveyed through language, they are often grouped under the term *declarative* or *explicit memory*.

Moreover, the mental lexicon can be accounted for as a special semantic subgroup within the thesaurus. In line with this statement, the lexicon is also a storehouse of interconnected representations that serve as the individual's lexical knowledge. Like other semantic representations, a word is known without necessarily remembering what specific experience made its learning possible.

Whenever a word is read or heard, the word recognition process activates the representations from the lexical neighborhood. Then, the most relevant lexical representation activates its associated meanings in a cascaded manner—from the lexicon to the thesaurus. This sequence may or may not lead to the pre-activation of other lexical representations with similar meanings (see Fig. 15)—from the thesaurus back to the lexicon (Altvater- Mackensen & Mani, 2013). The meaning of the chosen lexical representation is held within the semantic working memory, where it assembles the context needed to filter future lexical or semantic representations (Yeari, 2017). This context-related filter constitutes what is referred to as *top-down processing* (Angosto et al., 2013). Not only that, the semantic working memory also functions as the reconstructive center for new episodic memories. Using semantic representations as cognitive material, reading a book can become a cognitive experience and later be internalized as a genuine episodic cluster (Martin, 2021). In line with the expression "*a reader lives a thousand lives before he dies*" (Martin, 2013, p. 495), it should come as no surprise that reading a book enhances the performance of episodic memory (Stine-Morrow et al., 2022).

The other side of the coin is *procedural memory*, which stores sequences or algorithms of actions. Most of these memories are acquired unintentionally and

unconsciously (Nagao & Kitazawa, 2008), including all automatic processes that cannot be explained but shown, such as motor (e.g., playing the piano, writing) or cognitive processes (e.g., reading). Due to their ineffable nature, they are often called *implicit or non-declarative memories*.

On the topic of language processing, Michael Ullman (2001; Ullman & Lovelett, 2018) proposed the *Declarative/Procedural Model*, which divides the linguistic domain of control between the two memory systems. From this perspective, declarative memory stores lexical and semantic information, while procedural memory controls unconscious processes such as grammar and sequence recognition (see Ullman et al., 2020). These memory systems compete with each other to monopolize the process of general acquisition, a proposal known as the *see-saw effect* (Ullman, 2004; Boucher & Anns, 2018). Since the declarative memory is still underdeveloped in the early years of life—which explains infantile amnesia in early childhood (Alberini & Travaglia, 2017)—, procedural memory operates at full capacity. This monopoly of the procedural memory allows for the quick acquisition of implicit processes, such as the passive internalization of grammatical rules and the distinction of relevant sublexical patterns. However, as soon as a child enters puberty, the increase in estrogen and testosterone triggers a shift from an overreliance on procedural memory to a greater reliance on declarative memory. The age at which this shift occurs coincides with the end of the aforementioned critical period for language learning. From this point onward, the fast and unconscious learning of the procedural memory begins to decline. Such a shift forces adults to acquire new linguistic information more explicitly, so it may later automate into the procedural memory after numerous repetitions. Thus, while acquiring new lexical and semantic information becomes more manageable for adults than children, the automation of grammatical and sublexical knowledge is much more difficult.

Up until now, it has been claimed that the declarative memory primarily controls the lexical and semantic domains. Likewise, the sublexical and grammatical domains (morphology and syntax) are mainly stored—or guided—by the procedural memory. If the grammatical domain is ruled out—as it is beyond the scope of this thesis—the lexical and sublexical domains converge into what is known as the *dual-route model*, also called the *dual-stage model* (e.g., Coltheart, 2006; Diependaele et al., 2010; Hautala et al., 2022). This model is a foundational proposal in psycholinguistics, as it accounts for word identification and speech/spelling production. The routes or stages are two parallel and independent pathways designed to accommodate different but

complementary inputs. Briefly said, the lexical route allows for the rapid identification of already-known words. In contrast, the sublexical route offers the cognitive means to handle unknown words and pseudowords effectively —albeit slowly.

On the one hand, the *lexical route* makes use of the mental lexicon to quickly search for a representation that matches the input, so that its meaning can be attained in the semantic domain (see Fig. 15). In line with the idea of an incomplete lexical representation (see Section 1.1), it is essential to clarify that this process is carried out in an undetailed manner, given that the word is identified through a *holistic processing* that focuses on the stimulus as a whole (Allen et al., 2009; Hirshorn et al., 2020). The lack of details in this route could be behind word misperceptions, a claim that fully aligns with the evidence and theoretical proposals presented in Chapter 1. These lexicalization errors occur in the orthographic domain (e.g., typoglycemia phenomenon) (Rayner et al., 2006; Sakaguchi et al., 2016) and the phonological domain. As an example of the latter, the *Ganong effect* alludes to misperceiving an ambiguous sound, as if it were the phoneme needed to complete a real word rather than a pseudoword (Gianakas & Winn, 2016). The corrective lexicalization error requires a fast identification process that struggles to distinguish the specific details. Because of that, the rate of errors decreases when sufficient processing time is provided, as it allows the more specific route to accurately distinguish what the fast and holistic guess failed to do (Bakker, 2006).

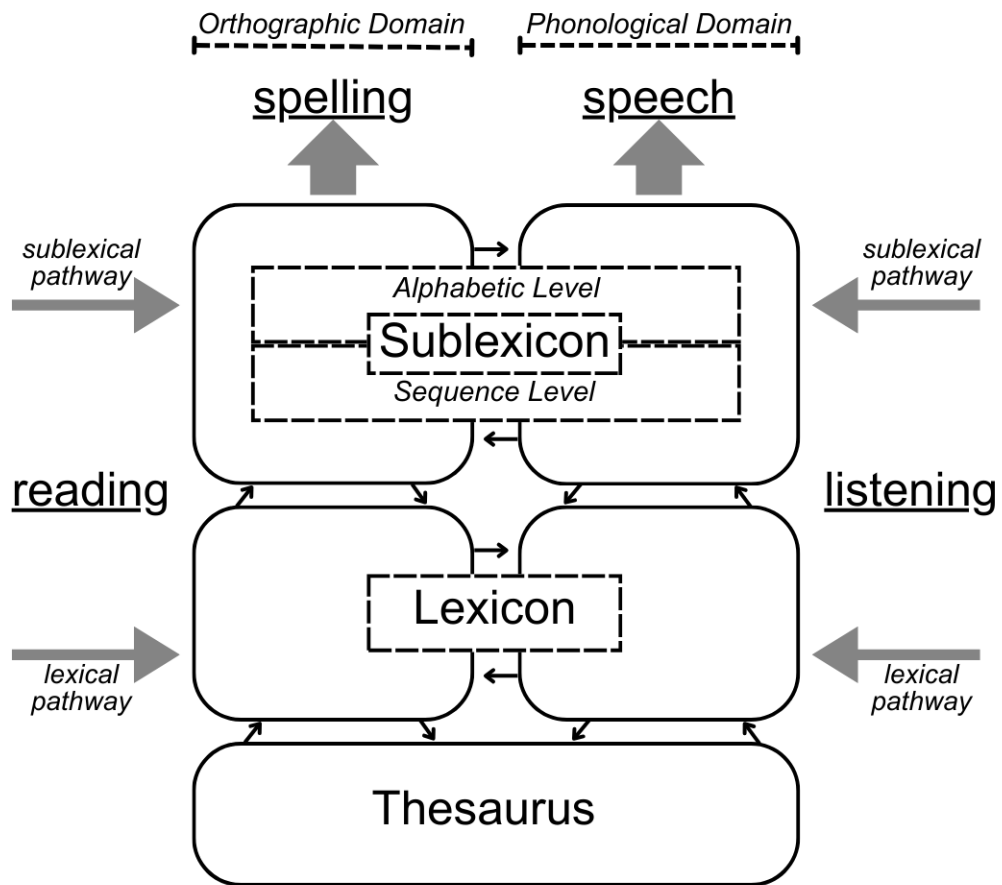
On the other hand, the *sublexical route* is commonly described as a letter-to-phoneme route. This route uses a slower but more *analytical processing*, providing a more detailed activation of sublexical representations (Allen et al., 2009). This dictionary of sublexemes —henceforth the *sublexicon*— contains smaller fragments of whole words and their corresponding grapho-phonological mappings. In contrast to morphological representations (e.g., word roots, prefixes, or suffixes), sublexical units (sublexemes) do not have a specific semantic or grammatical function (Pacton et al., 2005).

That said, sublexical units provide information about the relative frequency of segments within a word. This knowledge allows for the distinction of which sequences are more frequent at the word's beginning (e.g., <sph->) or at its ending (e.g., <-ck>). Through an extrapolation process of all perceived stimuli, a learner creates sublexical units that facilitate the processing of new lexemes and future pseudowords. Only after internalizing such patterns are newborns capable of distinguishing individual words through the speech segmentation process (Altwater-Mackensen & Mani, 2013).

The identity of the stored sublexical sequence is a controversial topic, as there are two major perspectives regarding the sequence's length. Some view the sublexical domain as an *alphabetic level* containing isolated letters and phonemes (e.g., Perry et al., 2010). Others support the idea of a more complex *sequence level* (e.g., Seidenberg & McClelland, 1989), whose length could either be based on its functionality (e.g., syllabic or subsyllabic) or on a fixed number of letters (e.g., bigram, trigram). In line with the latter proposal, letters are better recognized within words and regularized sequences than when isolated (Orbán et al., 2008). However, an integrative statement is the *integration of multiple pattern theory* (Treiman & Kessler, 2014). This framework suggests that experienced spellers store sublexical information simultaneously as individual units and context-sensitive patterns (e.g., phonotactic and orthotactic patterns). Considering both perspectives, it seems reasonable to assume that both the alphabetic and sequence levels coexist within the sublexicon. According to this logic, if the linguistic input is very different from the average form of the native language, it will tend to be processed more slowly, letter by letter, starting from the alphabetic level. By contrast, if the input resembles many other words in the native lexicon, it is more likely to be quickly recognized through longer and more specific patterns from the sequence level (Figure 16).

Figure 16

Dual-Route Model



From another perspective, not all sublexemes in a given language are necessarily stored as independent units, since some are not considered "significant enough" to warrant an individual storage. For instance, the English consonant sequence <dj-> —as in the foreign word *djinn*— is less likely to be internalized as a whole sublexeme compared to the consonant sequence <kn->. The <dj-> sequence does not appear in other English words, nor is it found in frequently-used words, whereas the opposite is true for <kn->. For a Spanish equivalent, the long syllable <hioi-> —as in the word *hioides* (hyoid)— also appears in only a tiny number of low L-TF words. In both cases, the sequences <dj-> and <hioi-> are not sufficiently relevant to allocate specific cognitive resources for their creation as individual sublexemes.

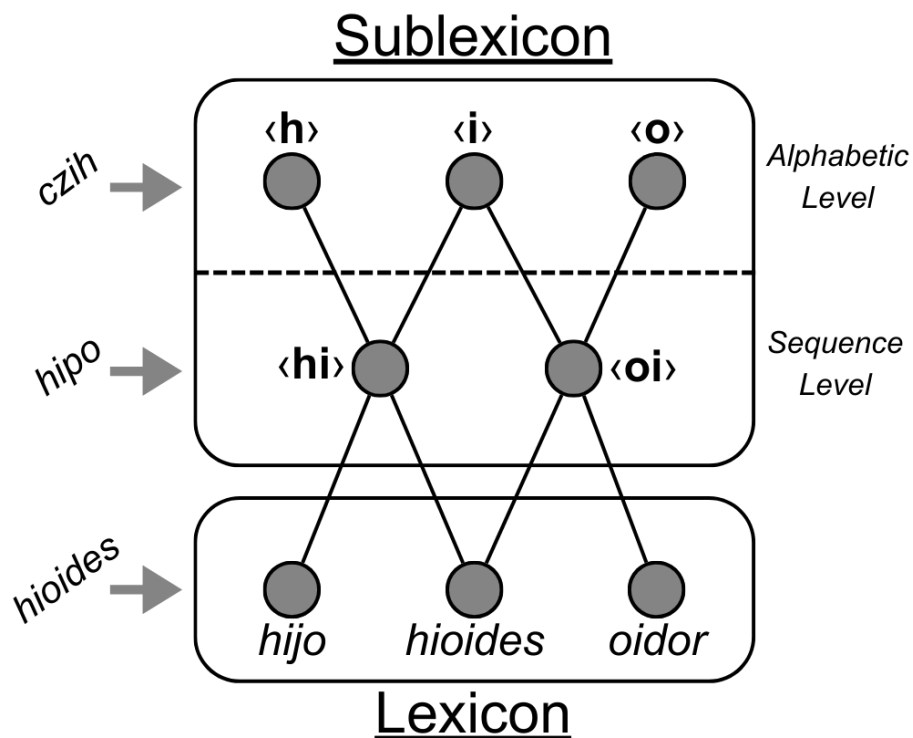
Based on the above, we propose the *Sublexical Irrelevancy Hypothesis*, claiming that the sublexicon does not create an independent sequence representation to account for only one or a pair of uncommon words. It is more frugal to allocate the

cognitive resources to more meaningful sublexical representations, such as the <d> and <j> in the alphabetic level for English speakers, or the digraphs <hi-> and <oi-> in the sequence level for Spanish speakers (Figure 17).

Even better, rather than focusing on the sublexical identification of weirdly constructed words (e.g., *djinn* or *hioides*), it is more pragmatic to internalize and process them through the lexical route. After all, although they have a low L-TF, they exhibit a low C-similarity, which leads to a low likelihood of confusion with other words, ergo, a close to non-existent inhibitory effect. The only sequences to extrapolate into the sublexicon are those with a higher probability of appearing in daily life, i.e., those with a higher sublexical typicality.

Figure 17

Extrapolated Sublexical Representations from the Lexicon



Note. This illustration is a schematic example of the cognition of a Spanish speaker who receives different orthographic inputs (left side). The lexical domain already stores the word *hioides*, which is why it is processed through the lexical route. Nevertheless, the word *hipo* is unknown to our hypothetical Spanish speaker, so the sequence level is required, as it contains the necessary sublexical units for the analysis. Finally, the

pseudoword *czih* also goes through the sublexical route. However, it will use the alphabetic level since it is composed of a sequence entirely foreign to the Spanish language.

Regarding the criticisms of the holistic vs. analytic dual-route model, the most significant is the shape invariance problem (e.g., Perea & Rosa, 2002). The lexical route provides a fast alternative for recognizing known words, a pathway primarily employed for words with moderate-to-high L-TF (see Allen et al., 2009). By contrast, to avoid the inhibitory pressure from high L-TF neighbors, words with low L-TF tend to be analyzed through the slower sublexical route. Since the holistic processing of the lexical route depends on the salience of external features (see Fig. 4), it can be deduced that the words requiring this route are more affected by changes in format (e.g., size alternation, full or mixed use of uppercase letters). However, as Perea and Rosa (2002) pointed out, this is untrue. The high L-TF words that employ the lexical route are unaffected by the format, whereas the low L-TF words are.

Nevertheless, the authors never considered the incomplete nature of the lexical representation, which allows for the recognition of the many shapes that a given stimulus can take (see Section 1.3). The adaptability of the lexical domain provides a solution to the shape invariance problem. Conversely, the sublexical representations' more detailed and fixed nature clashes with the diversity of formats, making it more challenging to recognize rare stimuli, such as low L-TF words or pseudowords (cf., Kinoshita, 1987).

The high processing speed of high L-TF words stems not only from the speed of the lexical route. Repetitive interactions with the same stimulus imply a greater likelihood of encountering different formats of the same word. Such repetition enables the lexical representation of a high L-TF word to improve its adaptability (Wong et al., 2011) and, therefore, its processing speed. In this way, the format will affect the processing of words that depend on the more analytic sublexical route (i.e., low L-TF) but will have less of an impact on words that rely on the holistic lexical route (i.e., high L-TF).

The lexical versus sublexical dichotomy manifests not only within the psycholinguistic framework but also in neurobiology. Another parallel network also aligns with the characteristics of the dual-route model's pathways, a system referred to as the *multistream model* (Allen et al., 2009).

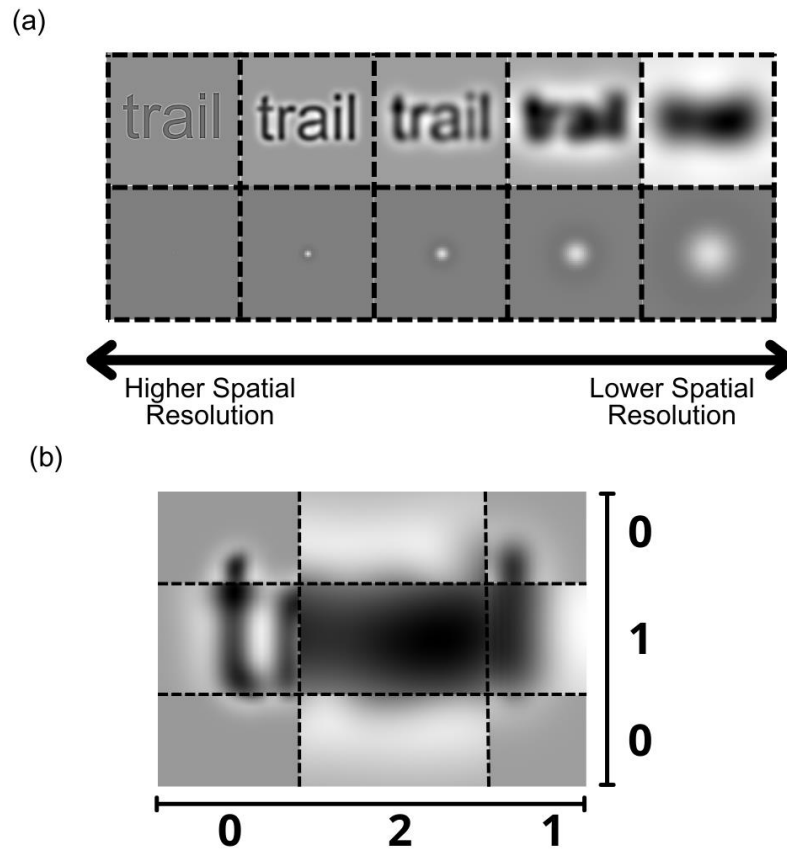
The first stream, known as the *magnocellular pathway*, has been associated with

the perception of contrast differences, low spatial frequencies, high temporal frequencies, and motion (Livingstone & Hubel, 1987). In other words, it rapidly processes the stimulus, albeit with a more ephemeral and less detailed spatial resolution. Some authors have even determined that this pathway projects to the orbitofrontal cortex, enabling the generation of "initial guesses" (Kveraga et al., 2007). As described in Section 1.3, the "fast guessing" is the foundation for lexical misperceptions due to high C-similarity (e.g., Coltheart et al., 2001). When there is more time to process the lexical stimulus, the accuracy of word identification improves because it can effectively use the sublexical route, complementing and enhancing the resolution of the lexical processing. Furthermore, the excentricity of the word —the quality of being an exception word, which could either include rare sequences (e.g., *djinn*) or inconsistent grapho-phonological correspondences (e.g., the ⟨-ough⟩ sequence)— has been associated with increased activation of the magnocellular pathway (Ciavarelli et al., 2021). This finding is consistent with the claim that the lexical route handles the exceptions, while the sublexical route provides the rules (Kang et al., 2009; Peterson et al., 2013). The second stream, the *parvocellular pathway*, serves as the equivalent of the sublexical route (Figure 18), as it processes the stimulus with low temporal resolution (i.e., slowly) but with high spatial resolution (Livingstone & Hubel, 1987; Masri et al., 2020).

Figure 18

Alleged Magnocellular versus Parvocellular Spatial Resolution

Not
e.
On
the
one
han
d,
(a)
sho
ws
the
wor
d
trail
filte
red
into
its



low and high spatial frequency forms, illustrating how the magnocellular and parvocellular pathways might process the same stimulus. We have used the convolution process from the study by Edward, Goodhew, and Badcock (2021), who also employed this technique to clarify the resolution differences between the two pathways. On the other hand, (b) combines each of the different spatial resolutions of *trail* with the presumed nature of the incomplete lexical representation (see Fig. 4).

Although the parallelism with the magnocellular and parvocellular pathways may seem convincing, these pathways remain debatable today (cf., Orban et al., 2004; Skottun, 2014, 2015). Therefore, it is relevant to present an alternative neurological model: the *dual-stream neurobiological model* (see Kearns et al., 2019; Turker & Hartwigsen, 2021).

First, the *ventral visual stream* enables object recognition, bestowing the intrinsic

physical properties of the object to identify it (Almeida et al., 2010; Cloutman, 2013). However, these unrefined object representations require the other stream to stay updated (van Polanen & Davare, 2015). In other words, it allows for the recognition of the identity of an object, but it carries less detailed information than its complementary stream. Interestingly, the information transmitted through the ventral visual stream comes from the magnocellular and parvocellular pathways (Tootell & Nasr, 2017). Among the perceptual objects it processes are written words, fulfilling the same task as the lexical route (Kearns et al., 2019; Turker & Hartwigsen, 2021).

Second, the *dorsal visual stream* provides action-oriented information, focusing on recognizing manipulable objects in a visuomotor sense, i.e., it transforms visual cues into relevant motor knowledge. However, this information does not enable object identification, as it requires the "*integration of stored knowledge about object function, identity, and motor programs associated with it*" (Almeida et al., 2010, p. 6). In other words, it relies on the ventral visual stream to identify the object. Similar to the sublexical route, the dorsal visual stream does not capture the semantic identity of the object, so it needs to activate the specific lexical representation associated with it.

Since this thesis focuses primarily on word recognition rather than word production —i.e., the motor automatizations involved in writing and speech— it is more challenging to establish a direct parallel between the dorsal stream, the sublexical route, and motor programs. As a brief reference, the sublexeme —usually depicted as the isolated letter or phoneme unit (i.e., the alphabetic level)— is often considered by other psycholinguistic models as the final step before producing a word, whether orally or in writing (e.g., Perry et al., 2010; Grainger & Ziegler, 2011; Hendrix et al., 2019). In other words, there is a tacit consensus that the sublexical domain functions as an interface that simplifies the association between lexical information and the motor articuleme⁵ that controls the mouth and hand movements. According to Longcamp et al. (2003), the representation of letters does not seem to be strictly visual, as it also

⁵ An *articuleme* is defined by Michon et al. (2019) as the motor sequence of the articulatory movements required to produce a specific phoneme. We have expanded this definition to also include the articulatory engrams required for handwriting. This articuleme is controlled by the motor procedural memory, specifically located in the surroundings of Exner's area for hand movements (Roux et al., 2009) and Broca's area for mouth movements (Basilakos et al., 2018). In particular, Exner's area is also associated with the processing of grapho-phonological conversions (Keller & Meister, 2014).

depends on networks constructed through writing, i.e., motor aspects. This notion supports the association of sublexical representations with embodied memory (Mangen & Balsvik, 2016). Thus, the sublexical cues of the visual lexeme serve as a bridge between the word's visual representation and the motor networks needed for its articulation, directed to pronouncing or spelling it.

Furthermore, sublexical regularities are stored in the fronto-temporal areas (Regev et al., 2024), which are the exact locations that govern production, syntax, and lexical functions. Thus, the sublexical processes are related to the procedural and the semantic memories. If we add that the cortical areas associated with the dorsal visual stream are strongly connected to the processing of unknown words and pseudowords (Kearns et al., 2019; Turker & Hartwigsen, 2021), the association with the sublexical route becomes even more apparent. However, the dorsal visual stream is composed of only magnocellular and not parvocellular information (Tootell & Nasr, 2017), which explains the earlier controversy related to the multistream model (Allen et al., 2009).

Before concluding, it is essential to explain the attentional requirement of the sublexical route (e.g., Reynolds & Besner, 2006). Since sublexical processing is analytical and detailed, it demands more cognitive resources to process it with a high spatial resolution. This higher demand helps to explain why the processing of smaller sublexemes—such as those of the alphabetic level—is not resource-efficient, as it needs a higher focusing power to distinguish the individual elements from the whole word.

A sequence level provides an alternative "subroute" to reduce the consumption of cognitive resources, especially if inconsistencies between the orthographic and phonological domains are considered (e.g., Seidenberg & McClelland, 1989; Perry et al., 2002). The importance of attentional processes for the sublexical route is evident whenever there is a deficit in brain areas related to these functions. For example, there is evidence that when the inferior parietal lobe and its supramarginal gyrus are damaged, it causes a selective sublexical deficit. At the same time, the lexical route appears to be unscathed (e.g., Facoetti et al., 2006, 2008).

In sum, the lexical route involves a fast recognition process of known words, allowing for semantic activation. The imprecise nature of this route aligns with the already described incomplete lexical representation. By contrast, the sublexical route is a slow, analytical, and high-consumption process that analyzes new and unfamiliar stimuli. Additionally, it serves as a linguistic interface for the motor articulation of the hand and mouth. In line with Ullman's model, the declarative memory governs the

lexical-semantic aspects, while the procedural memory provides the foundations for processing sequences specific to the sublexical aspect. That being said, how is the acquisition of the first phonological information explained? And more importantly, what happens to this primordial phonological information when the orthographic knowledge starts appearing?

2.2. Sublexical Development and Consistency Degree

From the intrauterine period, babies perceive the mother's voice and prosodic patterns, a stimulus that biases the child's linguistic preference. Therefore, when born, they tend to respond better to words that contain typical sublexical sequences than those that do not (Jusczyk et al., 1993). These subsegmental patterns are called *phonotactic regularities*, and together, they form what is known as *phonotactic knowledge* (Frisch et al., 2000). This knowledge can be thought of as the phonological domain of the sequence level of the sublexicon (see Fig. 16).

Before being able to segment the speech stream into words, which are later connected with meanings, the child needs to distinguish when such words begin and end. To identify and store the phonotactic cues of the language without the prerequisite of meaning, the infant's brain must function as a 'pattern recognition machine' (cf., Cairns et al., 1997). The procedural memory passively achieves this task, a system at its peak during the first years of life. Procedural acquisition of phonotactic cues coincides with evidence that links the speech segmentation process to the anterior caudate and putamen of the basal ganglia (Karuza et al., 2013; see Ullman et al., 2020). The child's parents tend to facilitate this acquisition through *motherese*, a type of speech that provides the child with simple, isolated, and repetitive input (Aslin, 1993). The child is influenced by this emotionally-directed speech, even during the intrauterine period (Parlato-Oliveira et al., 2021). Only after developing the ability to discriminate the phonotactic patterns of the L1 can the child distinguish isolated words and, as a result, manage to associate them with specific meanings linguistically.

That said, how are these complex sublexemes in English? What is their nature, their length, their boundaries? To answer these questions, it is necessary to know that the English syllable is subdivided into multiple subsyllabic structures, all according to a tripartite or bipartite perspective (Gross et al., 2000). On the one hand, the *tripartite perspective* divides the syllable into an initial set of consonants —the *onset* or C_1 —, the vowels — *peak*, *nucleus*, or V —, and the final consonants —*coda*, *offset*, or C_2 . On the other hand, the *bipartite perspective* combines the peak with one of the two consonants. Thus, the combination of the onset and the peak forms the C_1V unit (*body*), while the fusion of the peak and the coda creates the VC_2 unit (*rhyme*). Based on this, the monosyllabic word /treɪl/, whose orthographic representation is *trail*, can be subdivided into /tr/ (onset), /eɪ/ (peak), /l/ (coda), /treɪ/ (body), and /eɪl/ (rhyme). It is important to note that, despite this terminology, some authors have chosen alternative names for the C_1V and VC_2 units (Table 1).

Table 1

Alternative bipartite subsyllabic nomenclature

OTHER C ₁ V NAMES	OTHER VC ₂ NAMES	AUTHORS
Vowel + Beginning Initial Consonant- Vowel Segment	Vowel + Ending -----	<i>Taraban & McClelland (1987)</i> <i>Stanback (1992)</i>
Beginning Analogies Initial Consonant plus Vowel Unit -----	Ending Analogies Vowel plus Final Consonant Unit Word Body	<i>Goswami & Mead (1992)</i> <i>Treiman et al. (1995)</i> <i>Patterson & Coltheart (1987)</i>
Antibody -----	Body Spelled Body	<i>Forster & Taft (1994)</i> <i>Ziegler et al. (1997)</i>

Note. The names in bold warp the previously defined nomenclature, making them a cause of confusion.

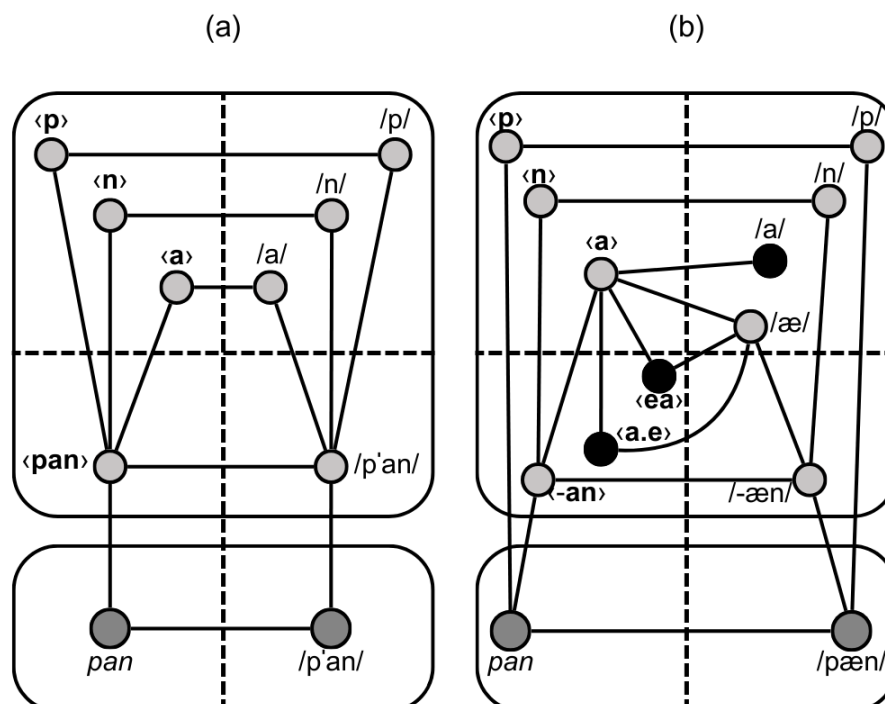
The relevance of subsyllabic structures lies in the fact that English speakers have a greater preference and sensitivity to rhyme units (Treiman & Chafetz, 1987; Treiman et al., 1995), a bias that is even exhibited among pre-literate children (Treiman, 1992). Also, the rhyme unit facilitates reading acquisition (Treiman et al., 1995) and spelling accuracy (Nation, 1997). This observation has led some authors to assert that English requires an onset-rhyme processing (Perry et al., 2002), a claim that provides an answer to the nature of the complex sublexemes of the sublexicon.

The onset and rhyme units become relevant when considering the topic of grapho-phonological consistency. According to the *Orthographic Depth Hypothesis* (Frost, 1994), a *shallow* or *transparent orthography* is a writing system in which phonemes and graphemes have a one-to-one correspondence. Thus, a sound tends to be represented by a single letter, while a letter represents a single sound. In contrast, a *deep* or *opaque orthography* involves a more complex grapho-phonological mapping. This complexity opens the door to bidirectional inconsistencies, as there may be multiple pronunciations for the same orthographic unit (heterophony) or multiple spellings for the same sound (heterography) (Ziegler & Goswami, 2005). When this dichotomy classifies real languages, most do not fall in at the extremes but somewhere in between. Some languages have a deeper orthography, while others exhibit a higher consistency between their grapho-phonological correspondences (Figure 19). There is a tendency to create inconsistencies over time due to the natural evolution of

languages. This increase in complexity happens because speech is constantly changing, while the spelling is more conservative, a difference that causes the pronunciation to diverge from its original orthography (Treiman & Kessler, 2006).

Figure 19

Transparent versus Opaque Networks



Note. Both images illustrate the lexical and sublexical representations required to process a lexeme. (a) shows the Spanish-speaking network for the word *pan* (*bread*) in Spanish, while (b) displays the English-speaking network of the English word *pan*, a false cognate. The high consistency of Spanish means that grapho-phonological correspondences are mostly one-to-one. These consistent networks contrast with English, where vowels develop a more complex network of convergent correspondences (black nodes), which could lead to plausible misspellings and mispronunciations.

Given that the English language has a deep orthography, many of its sublexical units exhibit grapho-phonological divergence. However, not all sublexemes are equally inconsistent. For example, the most inconsistent sublexeme within the English syllable is the peak unit, that is, its vowel (Treiman et al., 1995). In contrast, the onset's consonants remain the most consistent subsyllabic units, leaving the consonants of the

coda as an intermediate point in terms of consistency. However, some authors have noted that when the more inconsistent subsyllabic units (i.e., peak and coda) are interpreted as a single cluster (i.e., rhyme unit), the uncertainty of their processing is reduced (Treiman et al., 1995; Grainger & Ziegler, 2008 for a review). The combined consistency of the rhyme aligns with both Treiman (1992), who proposed that specific phoneme sequences tend to be processed as single units, and with Apel, who refers to English orthography as a "*highly regular, pattern-based system*" (2011, p. 593). It would be logical to assume that the rhyme unit may be the simplest way to disambiguate the correspondences of the peak and coda.

Given that the onset and rhyme units have the best consistency compared to other subsyllabic divisions, cognition appears to process English patterns in an onset-rhyme manner (Perry et al., 2002). These sublexemes better address inconsistent correspondences, which explains why the models that deal with these topics include them (e.g., Seidenberg & McClelland, 1989). This onset-rhyme bias also provides the rationale for why the two sublexemes facilitate reading acquisition and spelling accuracy (e.g., Treiman et al., 1995; Nation, 1997), both in pre-literate and literate children (e.g., Treiman, 1992; Ziegler & Perry, 1998).

Regarding the consequences of the inconsistencies, *regularization errors* are plausible mispronunciations of words with exceptional correspondences (Treiman et al., 1995). For example, the rhyme unit <-ave> has two phonological correspondences: the exceptional /-æv/ (e.g., *have*) and the regular /-eiv/ (e.g., *gave*, *slave*, *pave*). If the pronunciation of *have* is /heiv/ instead of /hæv/, it would be considered a regularization error. Contrary to lexicalization errors, regularization errors occur due to the prioritization of the sublexical route over the lexical route (Diependaele et al., 2010). In other words, the sublexeme and the more frequent correspondence can prevail, regardless of whether the exception word has a high L-TF. While regularization errors can occur under physiological conditions, a damaged lexical route produces more, as in the cases of surface dyslexia and lexical dysgraphia (Ellis, 1982; Binder et al., 2016).

It is interesting to note that the positional order of grapho-phonological consistency aligns with the previously mentioned incomplete lexical representation (see Fig. 4). The onset monopolizes the most salient portion of the horizontal axis (e.g., Fischer-Baum et al., 2010) and it has the highest consistency. In contrast, the peak is located in the middle portion of the word, an area that aligns with the location of the gap of information. Therefore, not only are vowels the most inconsistent sublexemes in the word, but they are also the most affected by lexical incompleteness. Finally, the

coda serves as the midpoint between the onset and the peak (Mooshammer et al., 2019), both in terms of consistency and salience in the lexical representation.

The complex network of sublexical correspondences results from multiple cognitive instances of interaction with the orthographic stimuli. The first interactions with the written world trigger its emergence (cf., Pollo et al., 2008). Given that inconsistencies exist in a language, the orthographic and phonological domains represent two sources of information that conflict with each other, slowing down the process of literacy acquisition (Ziegler & Goswami, 2005; Treiman & Kessler, 2014). As a result, this process requires children to understand two ideas tacitly. First, the written symbols are equivalent to phonological elements. Second, the written and phonological domains do not necessarily have a logical association, at least not without a philological-genetic explanation.

When the original phonotactic knowledge was acquired, the child never needed to understand it consciously. As with all procedural memories, children acquire and use this information without the need to explicitly know its rules, much like how they use syntax. However, with the appearance of alphabetic orthographic processing, a more refined fragmentation of sublexical representations is required to understand isolated and clustered letters. Thus, the literacy process forces the individual to understand original phonological knowledge more explicitly. This metacognitive development leads to what is known as *phonological awareness*, defined as the ability to break down spoken language into smaller segments in order to manipulate them (Adams, 1994; Snow et al., 1998).

Furthermore, due to written stimuli, sublexical knowledge of the most common orthographic patterns begins to take shape, i.e., the *orthotactic regularities*. The combination of these regularities —stored at the sequence level of the sublexicon (see Fig. 16)— forms the *orthotactic knowledge*. Finally, one reaches an *orthographic awareness* only after acquiring a more conscious understanding of this knowledge (Cassar & Treiman, 1997).

Depending on the language, the written input may vary to a greater or lesser extent regarding its original phonological information. This informational divergence triggers a series of differential cognitive changes, allowing for the development of linguistic awareness and the facilitation of language processing. We posit the *sublexical reconfiguration process* as the cognitive changes that occur during the literacy process. This process involves the reformulation of original phonological sublexemes —those that were only functional in the absence of an orthographic

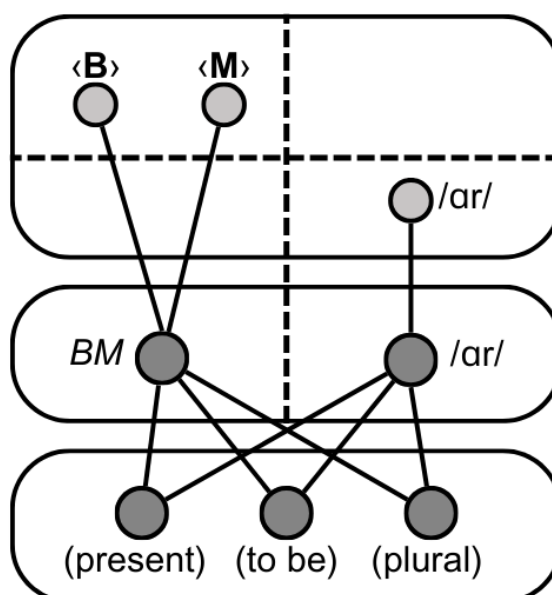
domain— into restructured units better suited to the existence of both sensory modalities. Therefore, this includes the creation of fragmented phonemes —through *phonemic segmentation*— and helps to overcome the typical grapho-phonological inconsistencies of the language at the sequence level, whether with the syllabic sublexeme or the subsyllabic one (Perea & Carreiras, 1998; Ziegler & Perry, 1998). Thanks to this process, the sublexical route gains the versatility to handle all kinds of future stimuli of the language with minimal cognitive resources.

The literacy process allows for the observation of the impact of orthographic exposure. In particular, misspellings and mispronunciations provide a lens through which changes in sublexical representations can be observed. Concerning English, Richard Gentry (1982) proposed the *phonological perspective* to understand children's literacy development phases. His phase model is based on the case of an English-speaking child who learned to read and write without any formal instruction (Bissex, 1980), using various strategies to represent the sounds of words. It also attempts to understand how cognition handles the inconsistencies of English orthography by analyzing misspellings. The aforementioned phases are as follows: a) pre-alphabetic phase; b) partial alphabetic phase; c) full alphabetic phase; d) consolidated/automatic alphabetic phase; and e) conventional phase (see Gentry & Ouellette, 2019 for a review). As a clarification, the *non-alphabetic phase* will be omitted in this description, as it refers to the period that precedes literacy acquisition.

First, the *pre-alphabetic phase* —formerly known as the *pre-communicative phase*— shows the initial use of actual letters in spelling. However, the letters are random, as the child has not yet formed an association between orthography and phonology (e.g., *quick* is spelled as *HS*). These writings are tantamount to an "orthographic babbling." There is still no understanding that orthography is a visual representation of sound, and therefore, grapho-phonological correspondences have not yet emerged (Figure 20).

Figure 20

The Symbolic Pre-Alphabetic Phase



Note. The semantic, lexical, and sublexical domains are illustrated from bottom to top. The sublexical level is subdivided into the alphabetic and sequence levels. The word that will exemplify Gentry's phases is the conjugated verb *are* (the present tense of *to be* in the second person singular and all plurals). This word is selected because its pronunciation converges with that of the letter <R>, which will be helpful later on. Additionally, due to its monosyllabic nature, it simplifies the different representations of each phase. Disyllabic lexemes with highly inconsistent correspondences (e.g., *eagle*) have been avoided, as they exceed the complexity addressed in this work.

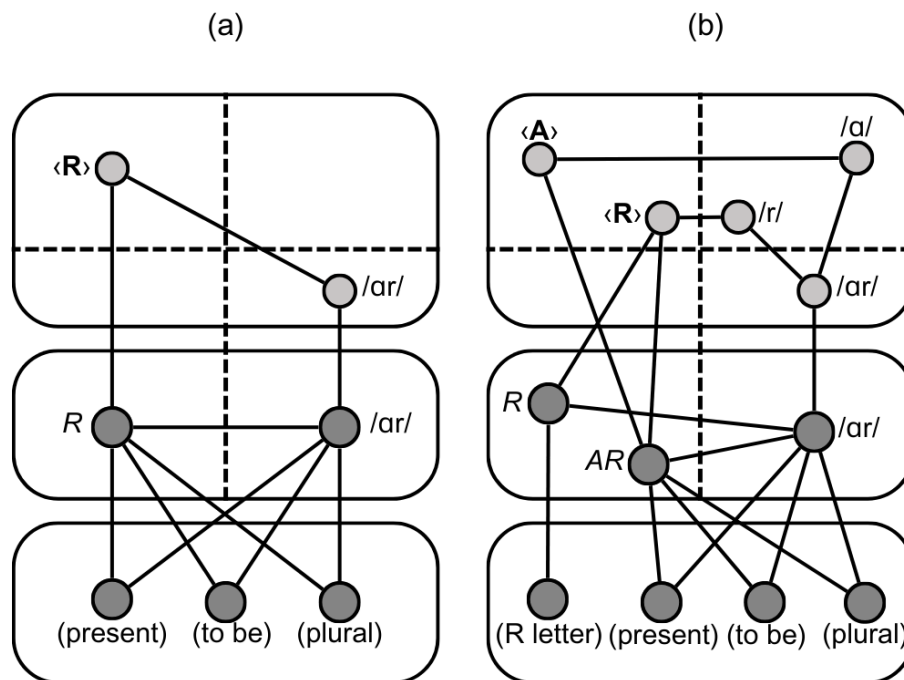
Second, the *partial alphabetic phase*—previously known as the *semiphonetic phase*—shows a greater awareness of the relationships between sounds and letters. However, this association is not based on the relationship between individual letters (or groups of letters) and phonemes. As a result, children tend to use a *letter-name strategy* (Treiman, 1994), which equates a single letter with more than one phoneme or even with a complete lexical segment (e.g., *RUDF* instead of *are you deaf?*). In other words, phonemic segmentation is not yet complete. Interestingly, the letters most frequently used in this phase are the consonants from the onset and coda (e.g., *TL* instead of *tell*), the two most consistent subsyllabic units. Because of this observation,

this phase is also called the *consonantal stage* (Kamii et al., 1990). Conversely, vowels are only written when their name matches the word's pronunciation (e.g., *U* instead of *you*).

Third, the *full alphabetic phase* —previously known as the *phonetic phase*—demonstrates a fully achieved phonemic segmentation process for both consonants and vowels. The child's spelling represents the entirety of the word's phonological form, showing a fully developed phonological awareness and mastery. Nearly all phonemes have a one-to-one correspondence between letters and sounds (Figure 21). However, due to the orthographic inconsistency of English, most of the child's spelling errors can be understood as regularization errors (e.g., *KAR* instead of *car*). As a result, some orthotactic patterns are still not accounted for, as seen with apparently silent vowels (e.g., *tip* instead of *type*), vowel digraphs (e.g., *kep* instead of *keep*), pre-consonantal nasals (e.g., *stap* instead of *stamp*), among others (cf., Gentry & Ouellette, 2019). During this process, the child is still aware of phonemic similarities that adults no longer detect (e.g., /tr/ versus /tʃ/). Because of this, they create tentative spellings that adults do not perceive as phonetically plausible (e.g., *CHRAC* instead of *truck*), referred to as *invented spellings* (Read, 1986). Longer words are still challenging, as the child keeps processing them sound-by-sound or letter-by-letter.

Figure 21

The Phonemic Segmentation Phases

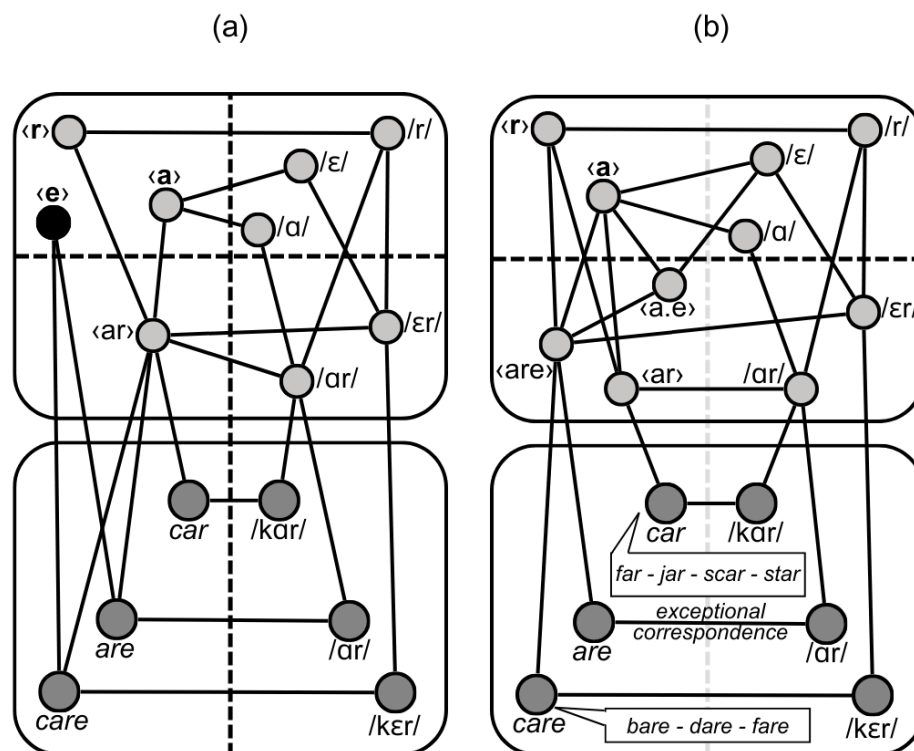


Note. The word *are* is written as *R* in (a) the partial alphabetic phase and as *AR* in (b) the full alphabetic phase. After the phonemic segmentation, the individual can distinguish between the name of the letter (lexical) and the specific sound of the letter (sublexical).

Fourth, the *consolidated/automatic alphabetic phase*—previously known as the *transitional phase*—begins the development of the sublexical and morphological fragments required to facilitate the processing of long words. This new improvement significantly reduces the memory load during reading (decoding) and writing (encoding), as it surpasses the rudimentary sound-by-sound and letter-by-letter processing (Gentry & Ouellette, 2019). On the one hand, onsets, rhymes, and even syllables—especially words derived from Latin and Greek (Moats, 2006)—are gradually stored to regularize inconsistencies. On the other hand, the use of prefixes, roots, and suffixes (e.g., -ing & -ed) becomes more precise. From this point on, individuals can expand their orthographic knowledge self-taught. Finally, reading and writing expand the lexical knowledge to attain the *conventional phase*—formerly called the *correct phase*. This final phase attains full orthographic awareness, and the literacy acquisition process concludes (Figure 22).

Figure 22

Overcoming the Inconsistency Problem



Note. (a) shows the consolidated/automatic alphabetic phase. Although some inconsistencies are resolved, there are still other correspondences that remain unaddressed. For example, if the individual wishes to spell the word *care* from the phonological input, they still require the lexical route. The lack of a significant correspondence with the silent <-e> (black node) complicates the spelling of the rhyme unit <-are>. In contrast, (b) illustrates the conventional phase, which addresses the <-are> issue by prioritizing the whole rhyme unit. Thanks to the increased vocabulary (in boxes), the sublexicon can distinguish the most resource-efficient orthotactic patterns. Likewise, exception words and their irregular correspondences are finally recognized as such (e.g., *are*).

The choice of rhyme units as the final sublexemes in Fig. 22 is not a casual decision but rather a consequence of their reliability in representing the inconsistent grapho-phonological correspondences. However, the onset and rhyme preference is not always the case in English. There is evidence that some English inconsistencies

are better understood when the body (C₁V) is focused rather than the rhyme (Venezky, 1970); the same holds for other languages. Yoon et al. (2002) observed that, unlike English, Korean speakers showed greater sensitivity to the body and the coda (see Li et al., 2022 for a more recent study). This inclination coincides with Dutch (Geudens & Sandra, 2003), Hebrew (Share & Blum, 2005), and Chinese speakers (Wang & Cheng, 2008). Based on this observation, Yoon et al. (2002) argued that the sublexical unit of choice depends on the particular traits of the language, a proposal known as the *language-specific sublexical dependence hypothesis*.

Moreover, as already mentioned, new stimuli tend to be processed at the longest and most complex sublexical level over a letter-by-letter level. While the sequence level is thrifty, the alphabetical level requires more cognitive and attentional resources. The length of the most complex sublexical unit also varies depending on the levels of consistency within the language. The *psycholinguistic grain size theory* posits that a higher grapho-phonological consistency entails higher reliability on the longer sublexemes (Ziegler & Goswami, 2005; Gottardo et al., 2016). For example, while subsyllabic units better address English inconsistency, more consistent languages like Spanish rely more on syllabic units (Carreiras et al., 1993; Perea & Carreiras, 1998; Ziegler & Perry, 1998; Jiménez et al., 2007).

The selection of one fragment or another occurs during the literacy process through the sublexical reconfiguration process. This procedurally-related choice aims to create a sublexicon capable of overcoming the highest number of inconsistencies in the language in the most resource-efficient way. Thus, we propose three main factors that guide the sublexical reconfiguration process: a) availability, b) synthesis, and c) accuracy.

First, the term "grain size" from the theory of Ziegler and Goswami (2005) is tantamount to the sublexical length. The authors postulated that when the size is larger, the child can grasp it more efficiently, meaning it is "more available" to cognition. In contrast, smaller grain sizes —such as the individual phoneme— are more challenging to process and distinguish from the whole stimulus, meaning they are "less available." In other words, distinguishing 'something' from a 'whole' requires a more significant amount of attentional resources and memory load, which makes processing more challenging. Therefore, the psycholinguistic grain size theory proposes *availability* as the sensitivity to a given grain size, i.e., the degree of conscious accessibility to different sublexical lengths. The greater the sensitivity, the fewer the extra resources required to process it, and vice versa. Because the sublexical reconfiguration process

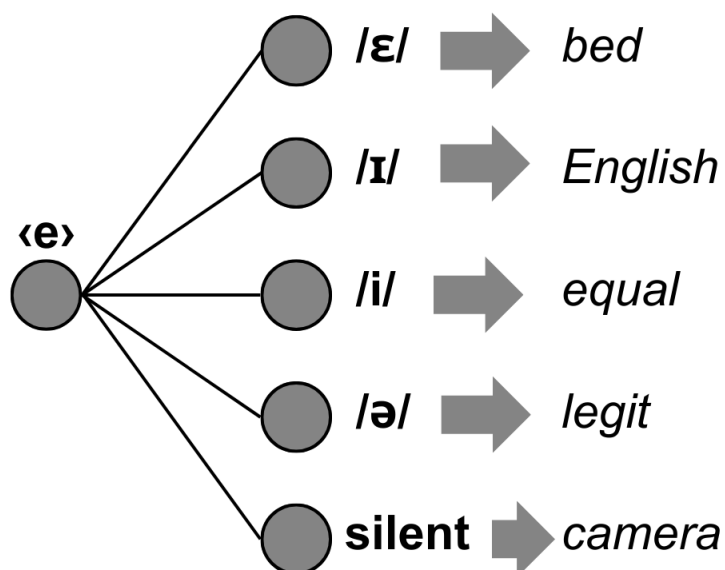
chooses the most resource-efficient sublexemes to process new words, the alphabetic level is avoided whenever possible. Conversely, the sequence level is usually the golden choice.

Second, the term *synthesis* describes the number of grapho-phonological correspondences needed to overcome a specific number of inconsistencies. When there are fewer correspondences, the number of sublexical representations required is smaller, allowing for processing more stimuli with fewer memory elements. In contrast, when the number of sublexical correspondences needed to deal with a given inconsistency increases, the number of sublexical representations also increases. A higher number of sublexemes implies an increase in memory load, making the sublexicon less synthetic and, therefore, less resource-efficient. In this way, synthesis prioritizes smaller grain sizes, being maximal at the alphabetic level. Conversely, availability prefers larger grain sizes, being maximal at the lexical level. The tension between these two factors forces the sublexical reconfiguration process to select intermediate-sized patterns. That is, neither as long as whole words nor as short as individual letters and phonemes.

Third, *precision* addresses the number of grapho-phonological correspondences that diverge from —or converge into— a single sublexical unit. The degree of inconsistency in a language directly influences this factor. A lower precision, i.e., more divergences or convergences, results in more heterophonies, heterographies, and regularization errors. The grain size is directly proportional to the number of correspondences (Treiman et al., 1995). Therefore, precision is maximal at the lexical level and minimal at the alphabetic level. The higher complexity of the alphabetic level explains why letter-by-letter processing is slow and challenging, especially considering the high heterophony of English vowels (Figure 23). If precision is to be improved, a pattern longer than the minimum must be selected, a choice that inevitably compromises the synthesis factor.

Figure 23

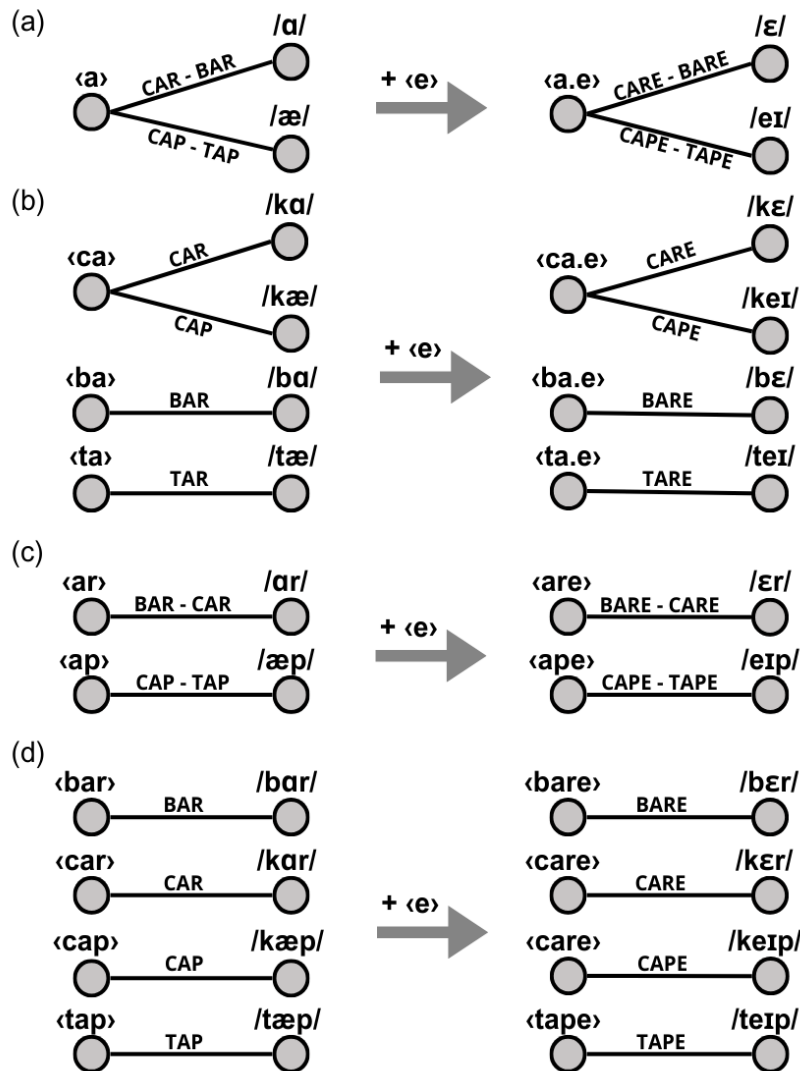
Lack of Precision in the English Alphabetic Level



Note. The corresponding phonemes were taken from American English.

Thus, the sublexical reconfiguration process selects the ideal length for each sublexical unit based on its availability, synthesis, and precision, all to create a complex mapping that overcomes inconsistencies. As an example, Figure 24 illustrates how words with the same vowel achieve their correct phonological correspondences at different sublexical units. Four words share the vowel $\langle a \rangle$ (*bar, car, cap, tap*), while the other four have added the letter $\langle e \rangle$ to create additional complexities (*bare, care, cape, tape*). On the one hand, the smallest grain size (i.e., alphabetical level) has the highest number of convergences —minimal availability and precision— and the fewest correspondences and representations —maximum synthesis. On the other hand, the largest grain size (i.e., syllabic level) shows no convergence —maximum availability and precision. However, it has the highest number of correspondences and representations —minimal synthesis. Since all these words are monosyllabic, the number of correspondences at the syllabic level is equivalent to the lexical level, making it redundant for the synthetic goal of the sublexicon.

Figure 24



The Sublexical Advantages of the English Rhyme Unit

Note. (a) shows the alphabetical and peak level (vowel set), (b) the body level, (c) the rhyme level, and (d) the syllabic level. The spheres represent the sublexical units of the orthographic and phonological domains, while the lines connecting them are their grapho-phonological correspondences.

In this way, the subsyllabic unit is a midpoint between both extremes, as it has the best availability after the lexical level. Figure 24 shows that the rhyme unit has the fewest correspondences and does not experience any convergence —maximum synthesis and precision—. In contrast, the body unit combines the worst aspects of the alphabetical and syllabic levels —minimum synthesis and precision. In light of this

specific example, when attempting to regularize vowel correspondences, the rhyme is the sublexical unit that saves the most cognitive resources. Despite having moderate availability —compared to the syllabic level— it has the highest precision and synthesis among all the options.

The sublexical bias toward the subsyllabic unit —or sometimes to the complete syllable— not only influences how daily stimuli are processed but can also alter the time required to achieve full literacy. As Ziegler and Goswami stated, "*children generally mastered (...) syllable-level skills before onset-rime skills*" (2005, p. 4), a syllabic bias that can be traced back to Gentry's full alphabetic phase (see Fig. 21). This may help explain the differences in time required per language to achieve a successful sublexical reconfiguration. For instance, Spanish —a consistent, syllabic-oriented language— allows for faster literacy acquisition than English (Goswami et al., 1998).

Beyond the greater transparency of Spanish, a detailed approach cannot label it as a purely consistent language. In contrast to English, most of its inconsistencies are found in consonants, while vowels are the most consistent letters in the alphabet (de Manrique & Signorini, 1998). Since Spanish is present across a vast portion of the Americas, many of the inconsistent correspondences of consonants are phenomena dependent on the regional varieties of the language. For example, in most varieties of Latin American Spanish, the letters <b-> and <v-> are pronounced as the phoneme /b/, making them homophones. The letters <s-> and <c-> can be pronounced as /s-/ (*seseo*) or /θ/ (*ceceo*), the latter being the original phoneme of the letter <z-> (Santana Marrero, 2016). Another example is the phonological convergence of the letter <y-> and the digram <ll->, which can correspond to /j/ (*yeísmo*), /ʎ/ (*lleísmo*), /z/ (*zheísmo*), or /ʃ/ (*sheísmo*) (Quijada van der Berghe et al., 2023).

In addition to regionalisms, some heterographies and heterophonies are not region-locked but are instead a standard across the Spanish language. For example, the phoneme /k/ can be written as <c->, <k->, or <qu->; the phoneme /x/ can be written as either <j-> (e.g., *exijo*) or <-g-> (e.g., *elige*); while the letter <g> can be either /g/ (e.g., *goma*) or /x/ (e.g., *exige*). Similarly, the phoneme /r/ is written as <r-> at the beginning of words or as <-rr-> in the medial position, while the letter <r-> is pronounced as /r/ at the beginning or as /r/ in the middle.

As a result of the phonological variety of Spanish, most regularization errors of exception words or borrowed words from other languages are interpreted as varietal differences rather than mispronunciations. This interpretation of phonological

differences may explain why the diagnosis of surface dyslexia is not conclusive in Spanish when examined through the reading of exception words (e.g., Difalcis et al., 2021), in contrast to English-speaking dyslexics. This pathology impairs the lexical route and increases reliance on the sublexical route, leading to pathological regularization errors. The caveat is that pathological regularization errors in the speech of Spanish speakers can go unnoticed, as they can merely be physiological regionalisms. Conversely, since the spelling rules of Spanish are more universally accepted and followed, the identification of misspellings is simplified, making the diagnosis of lexical dysgraphia more straightforward.

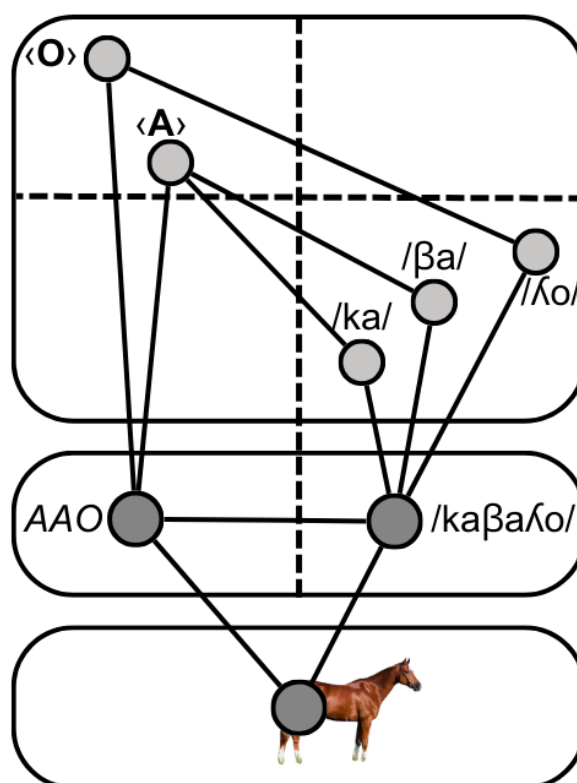
As a result of all these differences with English, Gentry's literacy development phases are incompatible with the Spanish language. Thus, other scholars addressed the literacy acquisition phases in more consistent languages, such as the Argentinian Emilia Ferreiro (Ferreiro & Teberosky, 1979; 1982; Vernon & Ferreiro, 1999; Ferreiro, 2006). Her *constructivist perspective* is not limited to Spanish as it is also the predominant paradigm for all consistent Romance languages (e.g., Pontecorvo & Zucchermaglio, 1988; Silva & Alves-Martins, 2002). Based on Piagetian constructivism, Ferreiro proposes that children already know about writing before beginning any formal literacy instruction. Contrary to Gentry's proposal, the constructivist perspective divides the literacy process into four stages rather than five: 1) the pre-syllabic stage, 2) the syllabic stage, 3) the syllabic-alphabetic stage, and 4) the alphabetic stage. The *pre-communicative stage* is omitted because it refers to the period prior to literacy acquisition, tantamount to Gentry's non-alphabetic phase.

First, the *pre-syllabic stage* is similar to Gentry's pre-alphabetic phase, as children have not yet associated orthography with phonology (see Fig. 20). However, Ferreiro diverges from Gentry by proposing that the use of letters do follow a specific logic. When a sequence of letters does not contain three or more letters (e.g., *BC*), the child will not consider it a word (i.e., *minimum quantity hypothesis*). Similarly, if a sequence of letters is a mere repetition of the same letter (e.g., *BBB*), it will also not be considered a word (i.e., *variation hypothesis*). Finally, despite the lack of grapho-phonological correspondences, a strong association is formed between the orthographic and semantic domains. Therefore, children tend to write the names of small animals like *mariposa* (*butterfly*) with fewer letters than the names of larger animals like *vaca* (*cow*) (i.e., *name hypothesis*) (Stella & Biancardi, 1990).

Second, the *syllabic stage* evidences the first rudimentary link between the orthographic and phonological domains. In line with Gentry's partial alphabetic phase,

children use a single letter to represent a set of phonemes, indicating an incomplete phonemic segmentation. However, rather than using consonants, Spanish-speaking children use vowels, making it a sort of *vocalic stage* (Ferreiro & Teberosky, 1982; Kamii et al., 1990). For example, the words *palo* (*stick*) and *sapo* (*toad*) are both spelled as AO (Figure 25). This observation supports that the first sublexemes learned in the literacy process are the most consistent.

Figure 25



Vocalic Syllabic Stage

Third, the *syllabic-alphabetic stage* addresses the final requirements before achieving a complete phonemic segmentation by overcoming the problem of inconsistency. This step is equivalent to a combination of Gentry's total alphabetic and consolidated/automatic phases (see Figures 21 and 22). The child reaches the alphabetic stage only after achieving phonological and orthographic awareness. This final step is equivalent to Gentry's conventional phase (see Figure 22).

The comparison between the literacy phases in English and Spanish allows for

understanding the differential sublexical changes triggered by written information. The letter-name strategies used by both English and Spanish speakers provide a comprehensive view of the early recognition of grapho-phonological consistency in children. In this way, a "consonantal stage" emerges in English and a "vocalic stage" in Spanish. Similarly, misspellings tend to appear in the segments with the highest inconsistency, such as vowels in English (Kamii et al., 1990; Varnhagen et al., 1999) and consonants in Spanish (e.g., Ferreiro & Teberosky, 1982; Zhang et al., 2021).

Interestingly, based on the aforementioned examples, English-speaking children tend to spell monosyllabic words with two letters (e.g., *tell* as *TL*). In contrast, Spanish-speaking children use one letter per syllable (e.g., *palo* as *AO*) (Pollo et al., 2005). This discrepancy supports the existence of divergent sublexical reconfiguration processes, which in turn emphasizes the importance of the language-specific sublexical dependency hypothesis (Yoon et al., 2002) and the psycholinguistic grain size theory (Ziegler & Goswami, 2005). Gentry's higher number of literacy phases aligns with the longer time required to complete the process (Goswami et al., 1998). The phonological awareness reached in Gentry's full alphabetic phase can be interpreted as the penultimate phase in consistent languages, corresponding to Ferreiro's syllabic-alphabetic stage. However, due to the high number of inconsistencies in English, Gentry adds the consolidated/automatic alphabetic phase to resolve the higher grapho-phonological inconsistency of the language.

Unsurprisingly, the two most well-known pedagogical methods for early literacy in both languages align with the outcome of sublexical reconfiguration. The *Word Family Method* involves presenting the English-speaking child with a series of words that share the same pattern, which coincidentally corresponds to the rhyme unit (e.g., *cat*, *mat*, *sat*). This method aims to facilitate the identification of regularities between letters and phonemes (Johnston, 1999). Conversely, Gedike and Heinicke's *Syllabic Method* involves teaching all the vowels to Spanish-speaking children (<a>, <e>, <i>, <o>, <u>), and then syllabically presenting them with various consonant combinations (e.g., <ca>, <ce>, <ci>, <co>, <cu>). This technique allows the child to explicitly discriminate letters, forcing an early phonetic segmentation process while also simplifying inconsistencies (Cano et al., 2017). While the Word Family Method encourages the speaker to understand language in a subsyllabic way, Gedike and Heinicke's method promotes a syllabic perspective. Both are strategies that help the implicit memory to navigate through the entire sublexical reconfiguration process.

After addressing the origin and development of the English sublexical

representations, one might ask: How do they influence the linguistic processing of similar or familiar stimuli?

2.3. *Gestaltian Approach of Similarity*

In Section 1.3, C-similarity explained the inhibitory effects caused by confusing two similar stimuli (Luce, 1986). The absence of a detailed lexical representation and a holistic and fast lexical route leads to misperception. Such errors tend to activate high L-TF words to the detriment of their low L-TF neighbors (see Fig. 13). This is a quick guessing process that only reduces its error rate if the target is allowed to be processed for a longer time (Coltheart et al., 2001).

Under this logic, the confusability approach does not account for any true "facilitation" that the psycholinguistic research has shown (cf., Strauß et al., 2022). The C-similarity differences in processing speed can be explained as a dichotomy of unaffected stimuli —high L-TF neighbors with fast processing— versus inhibited stimuli —low L-TF neighbors with slower processing (e.g., Williams et al., 2006). Furthermore, since the lexical route is already the fastest, the words that benefit the most from lexical processing —those with the highest L-TF— will always exhibit the highest possible speed. Therefore, the only words that could benefit from a facilitatory speeding-up effect would be words with the lowest L-TF.

A new perspective on similarity is needed to account for the facilitation between neighbors, given that C-similarity cannot do it. The alternative is herein named the *Gestaltian approach of similarity (G-similarity)*. The Gestalt principle of similarity proposes that stimuli are perceived as similar when a group of visual features overlap between items (Wertheimer, 1950a). The presence of repetitive elements among the stimuli allows for their grouping, which economizes the cognitive resources needed for their processing in the visual working memory. This results in facilitation —rather than inhibition— of the word recognition process (e.g., Peterson & Berryhill, 2013). Some authors have argued that these specific features related to facilitation are the specific sublexical segments shared between lexemes (e.g., Vitevitch & Luce, 1998; Ziegler & Perry, 1998; Hendrix, 2016). Stella (2018) further added that a higher-order structural relationship likely causes facilitation, an idea that aligns with the sublexical domain. This emphasis on patterns implies more analytical processing than the one used in C-similarity, leading to an association between the lexical and sublexical domains.

Also, the linguistic pattern used for facilitation is not random. For example, the existence of a shared onset unit within a group of words, collectively referred to as an *onset* or *cohort neighborhood*, has been associated with inhibitory effects and longer latencies (see Magnuson et al., 2007; Ulicheva & Coltheart, 2015; Donnelly & Kidd, 2021). As explained earlier, the holistic lexical processing relies on salient patterns that

are primarily highlighted at the beginning of the word, especially those ascending and descending consonants (see Fig. 4). Since the onset is particularly prone to being used as the salient feature in holistic processing, it should come as no surprise that this sublexical unit associates with C-similarity inhibition.

By contrast, the vowel (or peak) encompasses most of what is within the lexical void of information. Due to the vowel irrelevancy for the lexical route and its enhanced sublexical consistency when combined with the coda, the rhyme unit is the most valuable and specific pattern for the sublexical route in English.

Cohen-Goldberg (2012) noted that onsets are activated and selected before activating rhyme units, implying that the two sublexical units never compete. This latency difference between the onset and the rhyme is also equivalent to that between the lexical and sublexical routes. The lexical versus sublexical inclination of the onset and the rhyme can be explainable in diagnostic terms.

On the one hand, the onset offers a "sensitivity screening," a lexical criterion that quickly filters the set of words to eliminate those less likely to be relevant. When this process occurs in isolation, it can lead to false positives—that is, lexicalization errors due to quick guesses. On the other hand, the rhyme unit may be better suited for a "specificity screening," a slower sublexical criterion that identifies the stimulus that best matches the target from the remaining stimuli. This way, resources are allocated to detect the stimulus's vernacular constraints accurately.

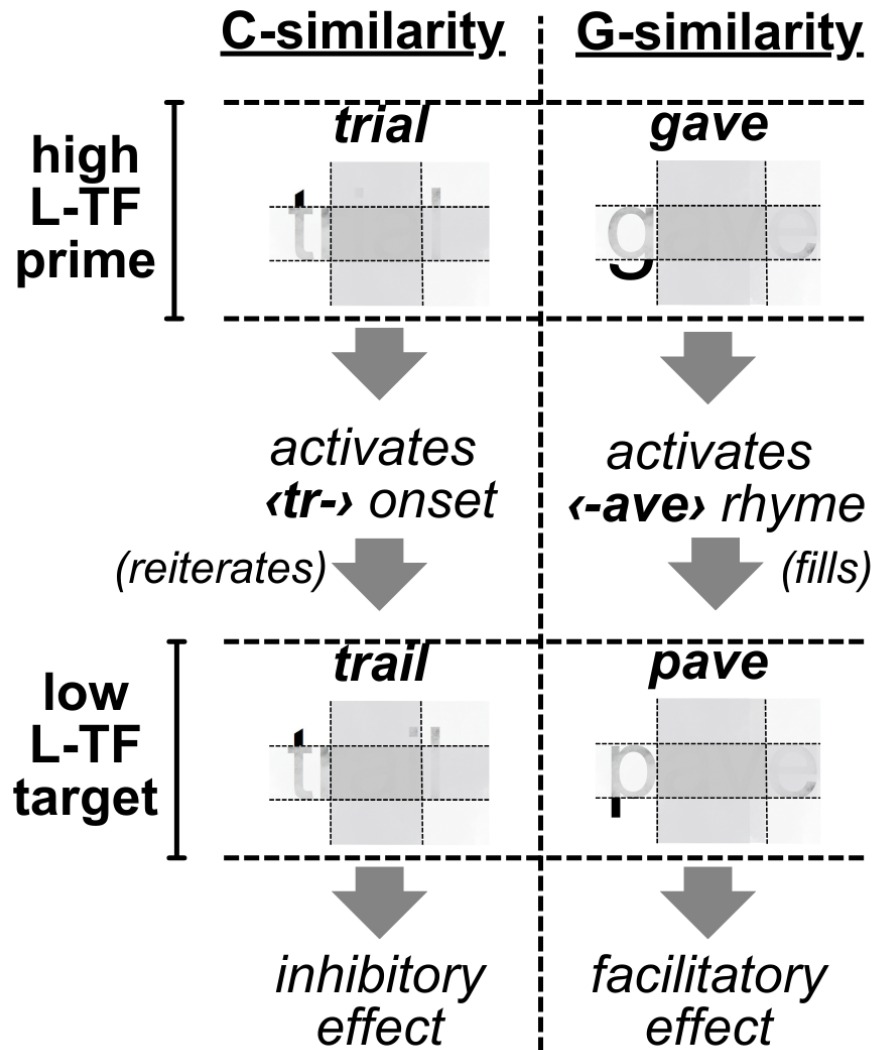
When priming one word with another, the final effect depends on the nature of the shared feature. For instance, if the two words share a rhyme unit, the first word (prime) pre-activates the sublexeme of the rhyme, which "fills" the lexical void of information when interacting with the second word (target). Since they do not share the same onset, the "lexical screening" will not consider them as C-similar, meaning the inhibitory effect will be minimal. The facilitation effect, born from the aforementioned sublexical pre-activation, becomes more significant if the target has a low L-TF (e.g., Mander et al., 2020). Facilitation is even better when the rhyme unit has a high frequency of occurrence in the language.

Conversely, suppose the prime and the target share the onset. In that case, the "lexical screening" does not quickly distinguish between the two stimuli. Therefore, rather than facilitate, the prime with the higher L-TF inhibits the target with the lower L-TF (e.g., Mander et al., 2020; Donnelly & Kidd, 2021). The activation of the onset triggered by the presentation of the prime serves as a redundant identification for the lexical route (Figure 26), which leads to a greater allocation of attentional resources to

the onset, to the detriment of the rhyme unit. In other words, the inhibitory effect of the onset is a consequence of C-similarity exacerbated by G-similarity (cf., Seitz, 2023).

Figure 26

C-Similarity Inhibition versus G-Similarity Facilitation in Priming



In line with G-similarity, the *rhyme neighborhood* —or *body neighborhood* (see Table 1)— is the number of words that share the same VC₂. This metric is relevant because it is associated with facilitative effects (Ziegler & Perry, 1998). When priming a target (e.g., *tint*) with a rhyme neighbor (e.g., *mint*), the performance improves and the error rate decreases (Seidenberg & McClelland, 1989). Under this new neighborhood type, words with more than one modification difference within the onset remain neighbors (e.g., *strength* & *length*), which differentiates it from the neighborhoods

associated with C-similarity (e.g., Coltheart's N, NNB, CLEARPOND N, and N*).

The *English Crowdsourcing Project* —a dataset of word recognition times for 62,000 English lexemes (Mandera et al., 2020) offered further evidence for the facilitatory association of rhyme neighborhoods. The authors observed that facilitative neighborhood effects are more prevalent in words with many rhyme neighbors. In contrast, inhibitory effects tend to be more apparent in lexemes with lots of high L-TF neighbors.

Based on Andrews's (1997) review, Ziegler and Perry (1998) noted that most facilitative effects are associated with English (e.g., Andrews, 1989, 1992), while languages such as Spanish and French generate inhibitory effects (e.g., Perea & Carreiras, 1998). From this, they pointed out that the rhyme unit is the most important differential factor for facilitative effects, a sublexical pattern that is not as significant in Romance languages (Ziegler & Perry, 1998). This finding supports the language-specific sublexical dependency hypothesis (Yoon et al., 2002) and suggests that not all neighbors influence the linguistic space similarly (Peereman & Content, 1997). Furthermore, it challenges the original definitions of neighborhood, as they do not distinguish the type of pattern that neighbors share with the target. For example, if we consider the words in Figure 8, the word *trail* is inhibited by its neighbors *train*, *trawl*, and *trial*, whereas the neighbors *frail* and *grail* exert a facilitative influence on *trail*.

After noting the specific impact of the sublexical domain, an attempt appeared to measure the specific relevance of each sequence under the umbrella of the phonotactic and orthotactic. Based on the Greek word *τακτικός* (i.e., fit for ordering or arranging), the term *tactic* seeks to account for all combinatory possibilities of a given language. As said before, phonotactic and orthotactic knowledge are the sublexical information associated with patterns within the mental sublexicon —i.e., phonotactic and orthotactic regularities. Through the sublexical reconfiguration process, orthotactic knowledge is not a mere reflection of phonotactic knowledge but rather a completely different, albeit complementary, sublexical domain.

A more abstract way of characterizing *tactic constraints* is to consider them a ruleset that correctly assembles morphemes and words (Frisch et al., 2000). In other words, they distinguish between the "legal" and "illegal" patterns in a given language. Thus, they form a grammar of probabilistic sublexemes located at the sequence level of the sublexicon, enabling a tacit identification of the language (Malmkær, 1991; Casaponsa et al., 2020). The "legality" of a language often operationalizes as a probability, given that rarer patterns tend to be "illegal" sequences (Pierrehumbert,

1994). The *phonotactic* and *orthotactic probabilities* are the most commonly used metrics to measure this, albeit some authors prefer to refer to them as *sublexical probability* (e.g., Luce & Large, 2001), *subcomponent frequency* (Nuerk et al., 2000) or simply *frequency* (Hammond, 2004). While L-TF represents the frequency of occurrence of a lexeme, tactic probability metrics indicate the frequency of occurrence of the extrapolated sublexical constituent (e.g., Apel et al., 2013; Bartolotti & Marian, 2014), that is, the likelihood that a sequence will appear in everyday language.

That said, the method for calculating tactic probability presents significant divergence among its proponents. This issue arises from differences in the equations and databases used, the number of phonemes/graphemes considered for each pattern, and whether empty spaces at the boundaries of words are independent characters or not as, among other things.

The most popular operationalization is the bigram/biphone probability metric (e.g., Vitevitch & Luce, 1998; Bailey & Hahn, 2001; Crossfield & Damian, 2021), defined as the simple co-occurrence of two graphemes or phonemes within a given corpus, which later averaged or transformed into a logarithmic scale. For instance, the word *back* has the digraphs <ba>, <ac>, and <ck>. Each bigram probability is calculated by tallying the number of real words that share these sequences from a lexical corpus. The geometric average of all the sequences is obtained to calculate an overall bigram/biphone probability for the entire word (Bailey & Hahn, 2001). By contrast, Apel et al. (2006) divide the tallied numbers by the total number of bigram sequences, that is, by 3 in the *back* example, to later express the stimulus's tactic probability as the logarithm of the product of all bigram probabilities. Some authors argue that word counting should not only consider the entries that share the sequence (e.g., <ac> in *back*). It should also match the position of the bigram/biphone in regards to the target (e.g., *pack*, *lack*, *lace*, *act*) (Davis, 2005; Davis & Perea, 2005).

Akin to the slot-based position coding (see Section 1.2), the positional constraint rules out many words containing the sequence but in a different locus (e.g., *acetal*). It is important to note that this approach does not consider sublexical boundaries. For example, when counting the bigram <ng> in *strength* and *avenger*, it does not distinguish that these sequences belong to different syllables.

As an alternative way to calculate this metric, Ring et al. (2023b) transformed the bigrams/biphones of sequences into logarithmized probabilities before summing them. Then, they normalized the sum of the logarithmized probabilities by dividing it by the total number of letters in the word rather than the number of bigrams/biphones. They

even counted the boundary gaps at the word's edges as individual characters, which increased the three original bigrams of *back* (<ba>, <ac>, and <ck>) to five due to the inclusion of the bigrams <_b> and <k_>.

The biphone/bigram probability is popular among psycholinguists but is not the only constraint employed to measure tactic probabilities. Some authors prefer the use of other sizes and limits, such as the triphone/trigram, the individual phoneme/grapheme (e.g., Davis & Perea, 2005; Clopper et al., 2018), or the onset-rhyme probability (e.g., Coleman & Pierrehumbert, 1997; Bailey & Hahn, 2001). This arbitrary divergence creates a significant discrepancy in the operationalization between the studies that aim to represent sublexical influence.

Nevertheless, the results associated with the tactic probabilities have served as a tool to study facilitatory effects in various language-related processes. For instance, in a study by Vitevitch and Luce (1998), the authors created pseudowords with small neighborhood sizes and, thus, a low number of lexical competitors. Their results showed facilitatory effects in word recognition when there were high phonotactic probabilities, suggesting that this reflected sublexical activation. In the same vein, 9-month-old infants are more sensitive to pseudowords with higher phonotactic "legality" for their L1 (Jusczyk et al., 1994). A greater knowledge of phonotactic probabilities is associated with larger vocabularies in 19-month-old children (Graf Estes et al., 2016; see Sundara et al., 2022 for a meta-analysis on the influence of phonotactic regularities on children). Likewise, preschoolers learn words more quickly when their sequences have a higher phonotactic probability (Storkel, 2001).

Concerning printed stimuli, sensitivity to orthotactic regularities emerges between the ages of 5 and 7 (Krasa & Bell, 2021) due to the beginning of the literacy acquisition process. When dealing with high orthotactic probability words, there is greater spelling accuracy (Apel et al., 2013) and longer fixation times among 6- and 7-year-olds (Fisher, 2019). Both phonology and orthotactic probability appear to influence the use of single or double consonants (Treiman & Wolter, 2018). The facilitatory power of orthotactic probability appears to be independent of the language, as evidenced by Finnish and German speakers, who also show a preference for pseudowords with higher orthotactic probability (Lehtonen & Bryant, 2005; Ise et al., 2014). Even in Spanish, children between 7 and 12 years old prefer pseudowords with frequent bigrams (e.g., <bu>) rather than infrequent ones (e.g., <vu>) when spelling (Carrillo & Alegría, 2014).

The facilitatory bias toward words with higher tactic probability is especially

significant in the early stages of lexical acquisition (Wright & Ehri, 2007). Not many words with "illegal" sequences are stored in memory during infancy. This lack of "rarely-assembled" vocabulary prevents the extrapolation of sublexical representations with lower tactic probabilities (i.e., the Sublexical Irrelevancy Hypothesis), meaning there is no facilitation or affinity for words with such sequences at these ages.

After developing considerable sublexical knowledge, cognition facilitates learning vocabulary with unusual patterns rather than those that conform to sublexical rules (Pacton et al., 2013). This tendency to facilitate the acquisition of words with low tactic probability sequences is not evident in earlier stages (Wright & Ehri, 2007). The system prioritizes acquiring lower tactic probability words only after attaining a fundamental level of sublexical knowledge. In the same line, "*whereas low orthographic probability contributed positively to incidental acquisition of novel word meanings in first language (L1), high orthographic probability affected positively the second language (L2) intentional learning*" (Bordag et al., 2017, p. 1). In summary, low tactic probability is relevant in incidental and passive lexical learning, while high tactic probability is relevant in intentional and active lexical learning.

As an anecdote, when recalling a pseudoword with low tactic probability sequences, its lexical representation seems to be "sublexically filled" with the most plausible sublexemes. This surmise accounts for why the most "illegal" pseudowords are more prone to regularization errors (e.g., Sobaco et al., 2015). This predisposition to regularize generates misrecalling of the original rhyme unit and can even distort the onset to make it more "legal" (e.g., *nnumar* is misrecalled as *nummar*). In cases like these, the most plausible sublexical representations warp the memory of a lexical representation that has been weakly internalized.

Beyond the facilitatory influence on recognition and retrieval, sublexical patterns also contribute to labeling a sequence as "legal" within a given language (Dankovicova et al., 1998). This "legality" is linked to *orthographic distinctiveness* (Zechmeister, 1972) and *language tagging*, defined as the process of recognizing the language to which a word belongs (Casaponsa et al., 2019). Recognizing a pseudoword as part of a specific language is often expressed as a gradient judgement of "*acceptability*." This acceptability perception—coined as *wordlikeness*, *wellformedness*, or *goodness*—has been proposed as the relative similarity of pseudowords to real words (Coleman & Pierrehumbert, 1997; Hammond, 2004; Frisch et al., 2000; Bartolotti & Marian, 2014; Brea-Spahn et al., 2020). Since this "similarity" arises from the shared sublexemes rather than holistic confusability, it frames itself into G-similarity. Therefore, the

wordlikeness rating is associated with the tactic probability of each sublexeme that makes up the pseudoword (Treiman et al., 1996; Vitevitch et al., 1997; Coleman & Pierrehumbert, 1997; Frisch et al., 2000). If this claim is accepted, all C-similarity effects supposedly caused by "high wordlikeness" —for example, the claim that *trial* and *trail* are similar due to high wordlikeness (cf., Vergara-Martínez et al., 2013; Perea, 2015)— must be considered as terminological inaccuracies. Wordlikeness is the concept that underlies the speakers' conscious and accurate intuition regarding what can or cannot be considered a typical sublexical sequence for a given language (Bailey & Hahn, 2001).

According to the results of tactic probabilities, higher wordlikeness judgements positively correlate with increases in speed and accuracy in both recognition/production tasks and learning in general (Frisch et al., 2000; Bartolotti & Marian, 2014; Do & Lai, 2020 for a Cantonese study). When learning an L2, it is easier to learn words with higher wordlikeness in a given L1 (Ring et al., 2023a), as happens with lexical cognates (Tribushinina et al., 2023).

As the dual-route model claims, the sublexical route processes new words and pseudowords. However, when a new stimulus's sequence is not sufficiently distinct to avoid misperceptions (i.e., high C-similarity), it can still occur a misuse of the lexical route (e.g., typoglycemia, LEET words) (Rayner et al., 2006; Kinoshita et al., 2014), namely, a lexicalization error. Likewise, when the pseudoword is sufficiently distinctive from other lexemes, its processing depends almost exclusively on the sublexical route.

Thus, combining pseudowords and wordlikeness is a valuable tool for studying the sublexical domain. Nevertheless, there are two questions worth answering. First, how is the pseudoword's wordlikeness measured? Second, how are these non-existent stimuli assembled? The answers to these questions are the wordlikeness judgement task of pseudowords and the stochastic grammar, respectively.

The wordlikeness judgement task involves the subjective rating of pseudowords on a scale from 1 to 7 according to their plausibility of belonging to a given language (Coleman & Pierrehumbert, 1997; Frisch et al., 2000; Bailey & Hahn, 2001; Kahng & Durvasula, 2023). A score of 1 is equivalent to an impossible word in English (e.g., *źdźbło*), while a score of 7 indicates a high likelihood of being a real word from the language (e.g., *cand*). A score of 4 is considered the midpoint, with 2-3 indicating "unlikely" and 5-6 representing "likely." Since these results are language-dependent, a pseudoword may be perceived as unwordlike in one language but wordlike in another. Wordlikeness is a gradient that ranges from "ill-formed" to "well-formed." Even when a

binary categorical response system is implemented (i.e., "yes" or "no"), the results cannot be explained without a gradient perspective (Coleman & Pierrehumbert, 1997). The existence of such a range explains why the primary method for its measurement is a 1-to-7 scale.

Traditional phonological models (e.g., generative phonological model, optimality theory) did not conceive the idea of a pseudoword that could be at a midpoint in terms of wordlikeness. For these authors, a pseudoword was automatically 'ill-formed' when there was a sequence with low tactic probability (see Hammond, 2004 for discussion). However, wordlikeness judgements are proportionally affected by the tactic probability of sublexemes. A wordlike sublexeme can compensate for the unwordlikeness of another (Coleman & Pierrehumbert, 1997).

If all sublexemes can independently modify or compensate for wordlikeness, then it becomes possible to control the process of pseudoword assembly using sublexical metrics. This view gave rise to what is known as *stochastic grammar*, which enables pseudoword assembly based on its sequence probabilities (Coleman & Pierrehumbert, 1997; Kumagai & Kawahara, 2018 for its adaptation to Japanese). Under the stochastic grammar paradigm, sublexical constituents follow the onset-rhyme structure, whose probability depends on the prosodic position within the syllable/word. The primary metric for this grammar is the *cumulative phonotactic probability*, calculated as the product of the probabilities of each sublexeme and its logarithm. Coleman and Pierrehumbert (1997) noted that this metric is one of the best predictors of wordlikeness judgements, as it outperforms the logarithm of neighborhood size and the lowest tactic probability of all sublexemes in a word.

In summary, G-similarity emphasizes the role of the sublexical analytical route rather than the holistic lexical route. This similarity type highlights the relevance of the sublexical aspect in facilitating lexical processing. This segment has emphasized pseudowords' role in the study of the sublexical domain. Also, it has provided the framework needed to assemble them in a controlled manner through the onset-rhyme structure.

However, the notions of similarity and typicality seem interconnected whenever these topics arise, often leading to an interchangeable use. For example, are the facilitatory effects of lexical neighborhood based on G-similarity or sublexical typicality? Do tactic probabilities accurately operationalize sublexical typicality? If this is the case, does the wordlikeness of pseudowords depend on their relative similarity to real words, or can it also be explained by sublexical typicality? Finally, is there any relevance to

this conceptual nitpicking? The answer to this last question will be the focus of the next chapter.

CHAPTER 3

Resolution of the Conceptual Conundrum

3.1. *Similarity, Typicality, and Familiarity*

Similarity is a key concept in psycholinguistics, explicitly or implicitly referenced in its models and metrics, up to the point that judging when it does not play a role is challenging. As shown in the previous chapters, some metrics are evident in their focus on lexical similarity, such as similarity neighborhood (Hammond, 2004), OLD20 (Yarkoni et al., 2008), and the clustering coefficient (Chan & Vitevitch, 2009). However, other concepts have also been associated with similarity, albeit they do not specify any nuance or distinction as the ones mentioned above. For instance, wordlikeness is the relative similarity of pseudowords to real words (e.g., Frisch et al., 2000). This definition leads to its direct association with the similarity neighborhoods and the tactic probabilities (Bartolotti & Marian, 2014), making it imprecise.

Luce and Pisoni stated that similarity is "*an ill-defined concept in research*" (1998; p. 1), a point later expanded by Bybee (2010), who noted that it is empirically problematic to identify the specific features that generate such similarity. For this reason, it is essential to control for several variables when addressing the topic, ensuring that the information is free from bias. For example, within this thesis, semantic and morphological similarities had to be controlled so that the lexical and sublexical influences could be isolated (cf. Gil Vallejo et al., 2018 for a review on verbal similarity).

Moreover, the psycholinguistic discussion of inhibitory versus facilitatory effects is not yet finished, complicating the emergence of a universally accepted definition of similarity and its effects. This topic aggravates because similarity is often equated with or subordinated to typicality (e.g., Bailey & Hahn, 2001) or familiarity (e.g., Perea & Rosa, 2002), further confusing the interpretation of findings. As long as the concepts of similarity and typicality are not adequately defined and delimited, there is a risk of bias in any work that uses them.

Given this issue, a study by Goldstone and Son (2005) distinguishes four types of similarity: a) geometric approach, b) feature set-based approach, c) structure mapping-based approach, and d) transformational approach. This section will only discuss the geometric approach and the feature set-based approach.

First, the *geometric approach* posits that linguistic representations are arranged in a representational space and claims that similarity depends on the perceptual

distance. In other words, the "closer" two lexical representations are to each other, the greater their perceived similarity (Shepard, 1962). This idea of similarity, thought of in terms of "perceptual proximity," aligns with the LD metric, which equates "perceptual distance" with the "number of modifications required to transform one neighbor into another." According to the logic of the OLD20 metric, the target word with the best similarity will be the one whose 20 closest neighbors differ from the target by only a single modification. The higher the number of modifications required to obtain the neighbors, the lower the target's overall similarity to the language's lexicon. Moreover, some lexical neighborhood models, such as the Generalized Context Model (Nosofsky, 1986) and its descendant, the Generalized Neighborhood Model (Bailey & Hahn, 2001), assert that neighbors vary along a continuous similarity scale. For both models, the degree of similarity depends on the perceived psychological distances of the stimulus, operationalized as the *D term*. Since the perceptual distance is the underlying idea behind similarity neighborhood and OLD20, this approach extends to C-similarity.

The concept of perceptual magnetic effect (Kuhl, 1994), which uses a spatial explanation to account for similarity (see Fig. 10), is also aligned with this notion of representational space. All prototypes curve the perceptual mapping to facilitate their activation, creating a sort of "gravitational attraction," whose impact depends on the L-TF of the prototype (see Fig. 12). If the differences between the two stimuli remain within the gap of information, the perceptual distance is shorter. In contrast, the distance is longer when the modification modifies the periphery. If the distance is small enough, the perceptual area of one lexical representation will overlap with that of the other, forcing both to compete for activation (see Fig. 13). This effect, mentioned in Section 1.3 as an explanation for C-similarity, aligns with the geometric approach of Goldstone and Son (2005). This way, the inhibitory effect alters a less relevant stimulus (i.e., low L-TF), causing it to be misperceived as its most frequent neighbor (i.e., high L-TF).

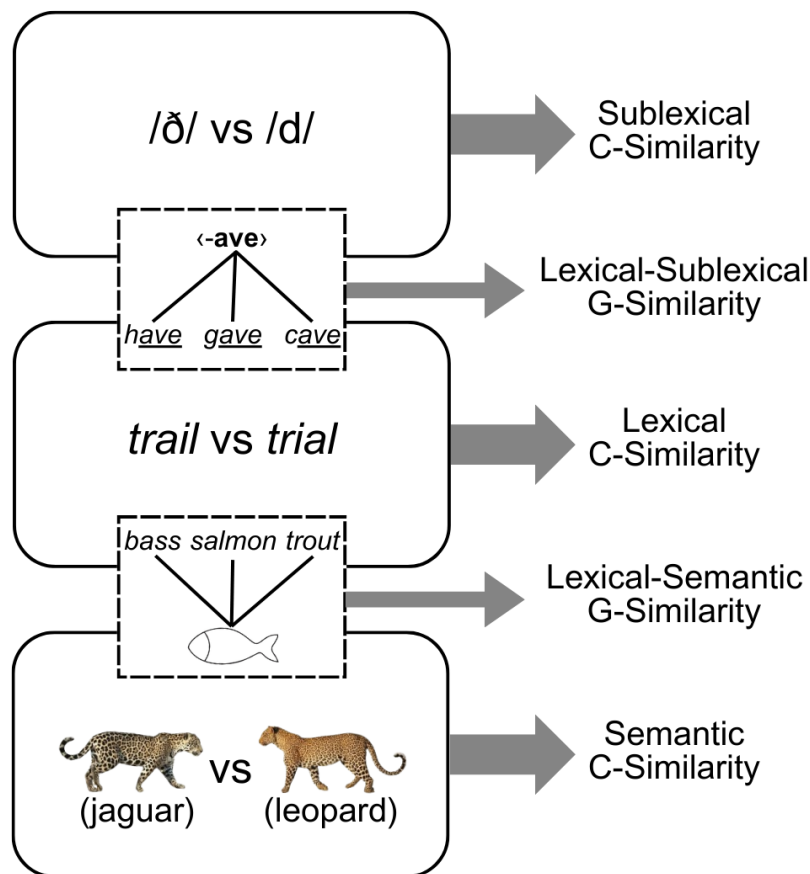
It is important to note that the original perceptual magnetic effect (Kuhl, 1994) dealt with phonemic similarity rather than lexical similarity. However, the lexical application of Kuhl's model becomes relevant when considering that this effect only addresses similarities within the same linguistic domain. In other words, Kuhl attempts to explain phonological misperceptions within the sublexical domain, i.e., it focuses on sublexical similarity (see Fig. 11). In contrast, this work has adapted the model to understand the misperceptions that occur within the lexical domain.

Second, the *feature set-based approach* asserts that similarity depends on

shared and distinctive features between stimuli (Tversky, 1977; Falkowski et al., 2021). Rather than advocating for spatial mappings and distances, this approach focuses on the sublexical feature shared by a group of lexical items, which strongly connects it with G-similarity (Figure 27).

Figure 27

The Domains of C-Similarity and G-Similarity



In line with Figure 26, Seitz (2023) observed that geometric similarity increases with feature set-based similarity. This observation is consistent with the inhibitory effect observed between onset neighbors, given that their use as a prime only serves to pre-activate the shared representation and allocate cognitive resources to lexically redundant information.

After identifying the parallelism between the two similarity approaches, it is necessary to define similarity as a general concept. Thus, both types of similarity imply an association between different mental objects due to the coincident presence of

shared elements (e.g., Gil Vallejo et al., 2018). In other words, there are two underlying ideas behind similarity: a *comparison* process and an *overlapping* between stimuli.

On the one hand, "something" is similar to what? The perception of similarity suggests a cognitive judgement when processing the properties of two or more perceived objects. Therefore, it is impossible to perceive similarity if only a single target is processed in isolation. This fact is why the word "similar" is paired with the preposition "to," a structure that presupposes the existence of an "additional other." This way, similarity always requires at least two elements to compare to.

On the other hand, what does it mean to be "similar" and not the "same"? If the quality of being "similar" had to be numerically modulated, an infinite increase in similarity would eventually result in the transformation of two "different but similar" objects into the "same" object. Therefore, the overlapping concept is paramount to comprehend similarity (e.g., Iniesta et al., 2021). To reach a "similar" judgement, the overlap must necessarily be less than 100% since a complete overlapping is tantamount to saying "identical," that is, an "absolute match."

Our preliminary definition of *similarity* is *the subjective perception of partial overlapping between two or more compared stimuli*. This new conceptualization includes the two similarity types described so far. Thus, depending on the amount of modification that occurs outside the lexical gap of information, C-similarity will have more or less overlapping. The greater the overlap, the smaller the perceptual distance between the representations. The lexical overlap leads to inhibitory interactions between the two stimuli and induces the misperception of the less frequent word as if it were the more frequent one, that is, as a "same." Likewise, G-similarity requires a comparison between lexical representations. However, the overlapping element takes the form of the sublexical unit, which pre-activates the sublexical representation and allows for the facilitation of a neighbor that also contains it.

This new definition of similarity makes it possible to address various conceptual confusions. One example is the neighbor generation task (Luce & Large, 2001), which involves presenting a target pseudoword orally and later asking the participants to name all the words with a similar structure (e.g., /fin/ → *feet, phone, bin*). As a caveat, this work does not delve into lexical neighbors based on semantic similarities (e.g., *nurse* → *doctor, medicine*). The authors coined the number of recalled words by this task as the *entropy degree*, understood by them as an equivalent to the neighborhood size metric, i.e., a measure of C-similarity. They illustrated one of their results in a table, where the recalled words and their percentage of occurrence are shown (Table

2).

Table 2

Luce & Luce (2001) Neighbor Generation Task Example⁶

Target Pseudoword Example: /fin/						
Example of Word Response	<i>Feet</i>	<i>Mean</i>	<i>Phone</i>	<i>Foam</i>	<i>Meat</i>	<i>Bin</i>
Percentage of Word Responses	34%	27%	10%	8%	7%	3%
* <i>BNC frequency</i>	14100	44098	8126	687	3600	1008
* <i>CELEX word-form frequency</i>	4104	7547	1414	122	1280	91

Note. The rows with asterisks are not part of the original table and show the L-TF of each recalled word, taken from the BNC and the *Center for Lexical Information* (CELEX) (Bober et al., 2021; van der Wouden, 1990). Both L-TFs were obtained from the British Lexicon Project (BLP) database (Keuleers et al., 2012).

The Pearson correlation coefficient was obtained based on the percentage of recallings per word provided in Table 2 and the L-TFs per word. The results showed a correlation of 0.6963 ($p = 0.124$) with the *BNC frequency* and 0.8332 ($p = 0.039$) with the *CELEX word-form frequency*, a very high and significant positive correlation with the latter. The CELEX lexical corpus is an American database, so it is reasonable that the percentages of recallings per word—from the American study of Luce and Large (2001)—show a higher correlation with the CELEX L-TF. Responses below 3% were omitted in the original table, so the calculation lacks a complete picture, and consequently, this analysis is incomplete. Nevertheless, the underlying question raised by these results is the following: What is the basis for the results of the neighbor generation task? Is the entropy degree equivalent to the neighborhood size and, therefore, to C-similarity? Or are there other causal variables to consider?

Luce and Large (2001) argued that when a target has a higher number of neighbors—a larger neighborhood size—the neighbor generation task will necessarily retrieve a higher number of words—a higher entropy degree. However, as our correlation indicates, the neighbors recalled in the results could also reflect their L-TF

⁶ When requesting the datasheet from the original study, the authors mentioned that they no longer had it in their possession. Therefore, Table 2 only addresses the examples from Luce and Large's article.

rather than their C-similarity. In other words, the participants might generate the most frequent words of the neighborhood rather than a reflection of the actual neighborhood size itself.

To highlight the conflict, consider a thought experiment. If a large neighborhood were filled with low L-TF words (e.g., technical jargon, archaisms), logically, the task should generate fewer recalled words (Table 3). This result would imply that a high C-similarity could be associated with low entropy, breaking the parallelism. Conversely, a small neighborhood containing only high L-TF words should more easily recall its neighbors, entailing a higher entropy degree.

Table 3

Lexical Typicality and the Neighbor Generation Task

Target Word	Lexical Neighbors	BNC frequency	CELEX word-form frequency
leave	<i>lease</i>	2212	74
	<i>heave</i>	217	75
	<i>weave</i>	330	73
	<i>cleave</i>	42	28
eyne	<i>eye</i>	9575	2284
	<i>dyne</i>	0	1
	<i>hyne</i>	not found	not found
	<i>syne</i>	not found	not found

Note. The table shows two targets for the same task, the word *leave* and the archaism *eyne* (*eye*). Both targets have the same NNB (C-similarity), although their neighbors have different L-TF (lexical typicality). Since the neighborhood of *eyne* contains technical jargon (e.g., *dyne*) and archaisms (e.g., *hyne*, *syne*), lexical recall (entropy degree) should not be the same. All L-TF values were obtained from the BLP corpus (Keuleers et al., 2012).

This thought experiment shows how the results of specific tasks are mistakenly associated with C-similarity. In the case of the neighbor generation task, the degree of entropy cannot be explained solely by neighborhood size but rather by lexical typicality (i.e., L-TF). This task alone cannot reveal the actual neighborhood size, as it could easily lead to underestimations —if there are many low L-TF neighbors— or

overestimations —if there are many high L-TF neighbors. Although the study by Luce and Large (2001) is over 20 years old, the use of the neighbor generation task and its interpretation as equivalent to C-similarity persists without discussion to this day (e.g., Castro & Vitevitch, 2023).

The conclusions of Luce and Large (2001) highlight the need to exclude the influence of typicality, especially when analyzing similarity. These influences can be confounded, and the issue becomes even more complicated when high degrees of similarity positively correlate with high typicality (e.g., Frauenfelder & Schreuder, 1992; Bailey & Hahn, 2001). Based on such observations, it could mistakenly establish a non-rigorous causal association between the two concepts. If this were the case, could it be suggested that having a higher number of similar words is tantamount to saying they are "more typical"? Before answering this in the next section, it is pertinent to clarify the term "*typicality*."

A typical object is "something" likely to occur within a specific time and place, a claim that associates typicality with the repetition of the same object or feature. This statement presents a significant difference from similarity, as the latter requires comparing two or more overlapping objects that "are not the same." In contrast, typicality does not require any comparison mechanism, as it focuses on how often the same idea repeats, which may or may not be within different objects. Furthermore, something can be typical in a particular space-time context, but that does not necessarily make it known to all individuals who interact in that environment; instead, it makes it more likely to be encountered by them. In mathematics, typicality equates to *frequency* —when the calculation is retrospective— or *probability* —when the calculation is prospective.

Typicality has been referred to several times throughout this thesis as word frequency or tactic probabilities. The former is the number of times a word repeats in a controlled environment. The latter —also alluded to as a frequency (Hammond, 2004)— refers to the number of times a specific sequence or set of sequences has appeared in a particular database. Along the same lines, Treiman et al. (1995; Berg, 2014) distinguish two typicality types: token and type frequency. *Token frequency* is the number of times the same stimulus occurs within a controlled environment. Based on this description, we have used the term lexical token frequency (L-TF) to refer to the occurrence of the same word. By contrast, *type frequency* counts the number of different words that share the same sequence in a given database. This definition closely aligns with tactic probabilities, as they also focus on repeating the same

sublexeme across different words in a language (e.g., Apel et al., 2006; Apel, 2011; Bartolotti & Marian, 2014). However, as will be developed in the following section, even though type frequency and tactic probability conceptually allude to typicality, they do not represent it in practice.

Finally, a third term is also associated with similarity and typicality. Depending on the author, the vaguely defined term "*familiarity*" has been used interchangeably as equivalent to a measure of similarity (e.g., Perea & Rosa, 2002) or as one of typicality (e.g., Boden & Giaschi, 2009; Lien et al., 2019). However, what distinguishes the perception of familiarity from the other two concepts is its "non-retrieved" nature. Familiarity is "felt" when a cognitive representation is inaccessible, while typicality and similarity require a more conscious recognition of what is being repeated or compared. As Besner and Johnson stated, the mechanism of familiarity consists of the ability to "*recognize a pattern (...) without identifying it, that is, without retrieving what the pattern is,*" which is equivalent to "*the sense one gets of having seen a person before, but not knowing who the person is*" (1989, p. 303).

In a study by Brown et al. (2013), the authors analyzed how individuals behave when words are unsuccessfully retrieved by measuring the confidence felt during their recovery. Higher confidence levels during retrieval were associated with the tip-of-the-tongue phenomenon, while lower levels generated a sense of "familiarity" or "lack of familiarity" regarding the unrecalled stimulus. In both the tip-of-the-tongue phenomenon and the familiarity cases, the individual metacognitively "feels" that there is "something" internalized in their memory storage, something of sufficient relevance to resonate with the target (Schwartz & Metcalfe, 2011). However, while the tip-of-the-tongue phenomenon carries a high certainty that words are retrievable, the sense of familiarity implies a lower certainty of achieving it. Therefore, the target that causes the feeling of familiarity could indicate that the stimulus is already stored—a tendency towards typicality—or that there is something resembling it—a tendency towards similarity.

Thus, the term "familiarity" can be caused by typicality or similarity, which is why it is often associated with both concepts. More concisely, familiarity can be defined as *the subjective feeling of total or partial overlapping between a perceptible stimulus and a known representation, albeit a "consciously inaccessible" one.*

Without theoretical boundaries between similarity and typicality, this section has presented a proposal to overcome the terminological conflict. However, confusing nuances still undermine the operationalization of sublexical metrics. The following section will tackle theoretical divergences, contradictory results, and operationalization

problems related to similarity and typicality. This framework will be needed to distinguish a sublexical typicality metric for controlled pseudoword assembly. Finally, regarding the neighbor generation task, this thesis proposes to reiterate it by comparing the influence of lexical typicality with C-similarity.

3.2. *Lexical-Sublexical Metrics*

So far, the field of psycholinguistics has motivated the use of metrics that allegedly cover all conceptual gaps. However, the idea that there are two similarity types (i.e., C-similarity and G-similarity) introduces hermeneutic-theoretical difficulties in interpreting the previous psycholinguistic instruments. Since the other metrics were created with an unspecified idea of similarity, attempting to explain them through the new theoretical paradigm is complicated.

The L-TF metric is undoubtedly a metric of lexical typicality, an axiom not questioned within this work. The classical neighborhood size and OLD20 are lexical similarity metrics, aligning with the underlying logic of C-similarity (i.e., geometric approach). However, a problem arises when dealing with tactic probabilities, which have been used hitherto as the mainstream metrics of sublexical typicality (e.g., Vitevitch et al., 1997; Frisch et al., 2000; Bartolotti & Marian, 2014; Bordag et al., 2017). Could these tactic metrics be more aligned with G-similarity rather than sublexical typicality? If this were the case, what would happen with the rhyme neighborhood? Does it represent a differential theoretical core compared to the one addressing tactic probabilities?

These questions open the door to another problem: There is a dichotomy between lexical metrics (e.g., L-TF, neighborhood size, and OLD/POLD20) versus sublexical metrics (e.g., tactic probabilities). To this day, there is still debate regarding the association of the former with inhibitory effects and the latter with facilitators in language processing and learning (e.g., Vitevitch & Luce, 2016; Ring et al., 2023b). This pairing supports the equivalence of tactic metrics as a measure of sublexical typicality, an interpretation that conflicts with that of G-similarity.

As mentioned in Section 3.1, typicality metrics were divided into the occurrence of a word —the token frequency— and the occurrence of a shared element across the entire lexicon —the type frequency (Treiman et al., 1995; Berg, 2014). It is true that these two frequencies positively correlate under certain circumstances, such as when a large number of words in the lexicon adhere to the same average length (e.g., Berg et al., 2022). However, researchers should not undervalue that they can differ markedly (Berg, 2014). While L-TF can be interpreted as a "pure" lexical typicality metric, type frequency is more difficult to associate with a single linguistic domain. In fact, type frequency shares the same operationalization as tactic probabilities, as both count the number of words that share a specific sequence (e.g., Bailey & Hahn, 2001; Davis, 2005; Davis & Perea, 2005; Dal Ben et al., 2021). Some authors even combine both

concepts, such as Davis (2005), who proposed the *Average Bigram/Trigram Type Frequency (BF_TP)*. To calculate BF_TP, the type frequency of each bigram/trigram in the stimulus is summed and then divided by the total number of bigrams/trigrams in the entire word. In other words, it is an alternative version of the classic tactic probability.

Section 3.1 postulated that typicality counts the occurrence of the same target within a specific space-time. In contrast, similarity counts the number of different targets that share the same object or pattern. Under this assumption, neither tactic probability nor type frequency are interpretable as sequence/sublexical typicality metrics. Neither represents the daily occurrence of a sequence but rather the number of times a given pattern appears within a list of non-repeating words (i.e., a database). These measures represent the number of different lexemes that share the same sublexeme, therefore, they are G-similarity metrics (see Fig. 27). To better specify, we propose that tactic probability and type frequency are both *lexical-sublexical metrics*.

The new nomenclature considers all measures of G-similarity that encompass both the lexical and sublexical domains. Instead of stating that sublexemes are more "typical" or "atypical" within a given database, it should be reformulated as there being "larger" or "smaller" neighborhoods for a given sublexical pattern. Thus, the counting of words with the rhyme unit <-ave> is not only equivalent to the type frequency <-ave> and the isolated trigram probability <-ave>, but also to the rhyme neighborhood <-ave> (see Section 1.3). The fundamental difference between these three metrics —and others not mentioned, such as the *subcomponent frequency* (Nuerk et al., 2000)— lies in two aspects. First, they differ in terms of the length constraints of the sequence in the neighborhood (Table 4): arbitrary and fixed (e.g., bigram or trigram) or functional (e.g., syllabic or subsyllabic). Second, they diverge in the number of distinct sequence neighborhoods considered in the same metric: one neighborhood per measure (e.g., type frequency, rhyme neighborhood) or the average of all the neighborhoods of the sequences of a target word (e.g., tactic probability).

Table 4

Arbitrary and Functional Constraints in Lexical-Sublexical Metrics

Arbitrary Constraint		Functional Constraint	
<i>Ortho/Phonotactic Probability</i>	Apel et al. (2013)	<i>Onset (C₁) Neighborhood</i>	Donnelly & Kidd (2021)
<i>Type Frequency</i>	Treiman et al. (1995)	<i>Vowel (V) Neighborhood</i>	Peereman & Content (1997)
<i>Unigram/Uniphone Probability</i>	Gao et al. (2023)	<i>Coda (C₂) Neighborhood</i>	Treiman et al. (1995)
<i>Bigram/Biphone Probability</i>	Crossfield & Damian (2021)	<i>Body (C₁V) Neighborhood</i>	Kim (2007)
<i>Trigram/Triphone Probability</i>	Gao et al. (2023)	<i>Rhyme (VC₂) Neighborhood</i>	Ziegler & Perry (1998)
<i>Subcomponent Frequency</i>	Nuerk et al. (2000)	<i>Syllabic Neighborhood</i>	Chetail & Mathey (2011)
<i>Average Bigram/Trigram Type Frequency (BF_TP)</i>	Davis (2005)		

We propose the term *functional type frequency* to unify all the measures with a functional constraint under a single nomenclature. In this way, when this metric focuses on rhyme, it becomes the rhyme neighborhood, while when it focuses on the onset, it is equivalent to the onset neighborhood.

As mentioned in Section 2.3, the rhyme unit is primarily relevant in the sublexical domain, as its information hides within the lexical gap of information. Therefore, the presentation of a rhyme neighbor pre-activates the rhyme unit and, consequently, facilitates the lexical processing of the target through G-similarity (see Fig. 26). In contrast, the onset unit is not only present at the sublexical level, but it is also the most salient segment of the incomplete lexical representation (see Fig. 4). Since the shared features monopolize the scarce attentional resources (Seitz, 2023), the pre-activation of the shared onset unit hinders the recognition of the rhyme unit. Hence, the shared onset deepens the competitive interaction between the neighbors due to C-similarity (see Fig. 26). Considering both observations, the functional type frequency encompasses both the onset and rhyme type frequencies, even though these exert divergent effects on word processing.

Introducing the new nomenclature for the G-similarity metric seeks to overcome

the criticisms that target the classic lexical-sublexical metrics, especially tactic probabilities. For example, some authors consider tactic probabilities nonspecific when attempting to predict the outcome of processes dependent on the sublexical domain (e.g., Bailey & Hahn, 2001; Ring, Leoné & Dijkstra, 2023b). Others claim that tactic probabilities are nothing more than more complex C-similarity metrics (e.g., Frauenfelder et al., 1993; Hammond, 2004). These statements conflict with the assumption that the tactic probability metrics represent sublexical typicality. At the same time, it blurs the boundaries between G-similarity and C-similarity.

Three topics are discussed to understand the theoretical and operational problems of classical lexical-sublexical metrics: 1) the lack of boundaries between G-similarity and C-similarity metrics, 2) the use of arbitrary constraints when delimiting sublexical lengths and positions, and 3) the bias created by the databases.

First, G-similarity metrics show a high correlation with C-similarity metrics, so words and pseudowords with a high tactic probability tend to have a large neighborhood size, and vice versa (e.g., Landauer & Streeter, 1973; Frauenfelder & Schreuder, 1992; Charles-Luce & Luce, 1995; Gathercole & Martin, 1996; Vitevitch & Luce, 1998, 1999; Frisch et al., 2000; Apel et al., 2013). This correlation makes it difficult to disentangle the C-similarity/inhibition versus G-similarity/facilitation dichotomy in a controlled manner and, in turn, distorts the interpretation of results. For example, if this proportionality were absolute, a high tactic probability word would be slowly processed, as it would also have a sizable lexical neighborhood. Conversely, a low tactic probability word would be quickly processed, as it would also have a small lexical neighborhood.

Also, Luce and Large (2001) noticed that known words are more facilitated than pseudowords when sharing the same tactic probability. This observation contradicts the claim that the sublexical aspect facilitates the processing of stimuli with lower L-TF. The same authors noted that pseudowords with a small lexical neighborhood and high tactic probability increased the inhibitory effect, reversing the dichotomy between C-similarity and G-similarity.

Furthermore, Forsten and Shen (1996) stated that the neighborhood size metric has a facilitatory impact on words and simultaneously exerts an inhibitory effect on pseudowords in visual word recognition. Others argue that the C-similarity metric is equal to (e.g., Frisch et al., 2000; Hammond, 2004) or even superior to the tactic probabilities in predicting outcomes related to pseudowords (e.g., Bailey & Hahn, 2001; Janse & Newman, 2013; Apel et al., 2013). Finally, Bartolotti and Marian (2014)

claimed that both C-similarity and G-similarity measures reflect the "degree of closeness" between sublexical patterns, a description that is more aligned with geometric similarity than with gestalt similarity.

Due to these numerous contradicting examples, more radical authors have questioned the existence of the sublexical domain in its entirety (e.g., Bailey & Hahn, 2001). They argue that all the facilitating effects associated with tactic probability — and, by extension, all G-similarity metrics— solely emerge from the lexical domain. A more moderate perspective does not dismiss the existence of sublexical processes but attributes sublexical importance to C-similarity measures, such as the Coltheart N (e.g., Strauß et al., 2022). In this regard, Hammond suggested that "*neighborhood density would seem to be simply an extreme version of phonotactic probability*" (2004; p. 6).

The blurry boundaries between the tactic probability metric and the C-similarity question the former's operationalization, its usefulness as a representative reflection of the sublexical domain, or even to doubt the existence of G-similarity as such. Using tasks whose results depend on sublexical processing may be helpful to shed some light on these issues. This way, using wordlikeness judgement tasks for pseudowords has mainly been associated with sequence (sublexical) typicality and the sublexical route (Bailey & Hahn, 2001). This association makes it the most essential task for analyzing sublexical processing and the legitimacy of its metrics. Since this task requires a slow and detailed analysis of the pseudoword features, the fast and holistic processing of the lexical route does not interfere. Consequently, the effects of C-similarity are minimal. As for the results of this task, the authors noted that neighborhood size had greater predictive power than tactic probabilities (Bailey & Hahn, 2001). They also clarified that both metrics could only predict 38% of the results. These conclusions allow us to conclude two things. On the one hand, the tactic probability metric is an unreliable reflection of sublexical typicality. On the other hand, it is a very unspecific lexical-sublexical metric, as even the sublexical-unrelated C-similarity metrics better predicted sublexically-related results.

Second, both type frequency and tactic probability lack theoretical support regarding their positional and fixed length constraints necessary for counting. For example, what is the rationale for choosing a bigram constraint when calculating the tactic probability of *back*? Why not use a trigram constraint? A three-letter constraint would at least respect the length of the rhyme unit <-ack>. Are subsyllabic and syllabic functional constraints insignificant for sublexical calculation? Or better yet, is it legitimate to consider sequences like <_b> and <k_> from *back* as plausible bigrams?

With such arbitrariness in the rules, the operationalization method of lexical-sublexical metrics will create discrepancies between authors (cf., Davis, 2005; Apel et al., 2006; Ring et al., 2023b), if not divergences in the results.

The lack of sublexical functionality in the constraints opens the door to many conclusions that would be unthinkable, even for laypeople. Ring et al. (2023b) exemplify that the Polish word *dobrze* —translated as *well* or *good*— is orthotactically "legal" in English if calculated with a bigram model. Thus, this foreign word has a higher-than-low tactic probability in English despite its low L-TF and implausible consonantal sequence (<brz>). If such a result were legal, the word *dobrze* would exhibit a high G-similarity and zero C-similarity —a small neighborhood size and an OLD20 higher than one— which, by definition, leads to a facilitatory effect devoid of inhibitory influences. Due to this implausibility, the authors argued that the rarity or typicality of the sequence —i.e., its sublexical typicality— cannot depend on tactic probability (Ring et al., 2023b). This assertion is now understandable, as the tactic probabilities measure G-similarity, not sublexical typicality. The case of the word *dobrze* clearly illustrates the problem of arbitrary and fixed lengths, even when calculating G-similarity. Moreover, when using prefixed bigrams or trigrams, it becomes impossible to integrate these results with those addressing grapho-phonological inconsistencies, as they omit the existence of the onset and rhyme.

Calculating bigram/trigram probabilities becomes even more problematic when considering the sequence's position. Given that the same sequence may appear in a different position within the word, positional exclusion criteria hamper the count of some entries that contain it. The issue is that, in the absence of restrictions based on functional aspects (e.g., syllabic, subsyllabic), excluding some sequences over others becomes highly arbitrary. For example, when calculating the bigram probability of <ge> from the word *edge*, the word *George* is counted only once, despite appearing twice in the whole stimulus. If subsyllabic restrictions are irrelevant, why should the <ge> at the end be excluded compared to the one at the beginning? Is the sublexical information of the bigram <ge> stored twice in memory to account for its occurrence in both positions? The proposal of multiple storages for the same bigram/trigram depending on its position is neither synthetic nor available —as it reduces the grain size (see Section 2.2)—, leading to an increase in memory load (see Fig. 24). Positional restriction unnecessarily introduces an arbitrary rule to compensate for the lack of natural boundaries such as the syllable and subsyllabic structures. Moreover, contrary to functional constraints, it does not explain sublexical inconsistency.

The most problematic case of a positional restriction is observable when morphological derivatives are included in the calculation of tactic probability as if they were different entries, a situation that often occurs. Continuing with the example of the bigram probability ⟨ge⟩ from *edge*, can words like *edges*, *edged*, *edgewise*, or compound words like *edgeway* be considered legitimate entries? As their bigrams ⟨ge⟩ are not in the last two letters, a strictly applied positional restriction cannot count these words as legitimate. Nevertheless, the bigram ⟨an⟩ in *man* does appear in *human*, *jazzman*, and *postman*. How can the inclusion of some morphological derivatives and the exclusion of others be justified? Is it logical to treat them as individual entries? This problem only worsens with the third topic.

Third, since the lexical-sublexical metrics derive from lexical entries, any biases inherent in the database's origin can affect the final value. Given the difficulty in collecting all lexical entries used in a language, it is natural for the databases' authors to prioritize including the most important entries (see Keuleers & Marelli, 2020 for a review on lexical resources). However, all databases exhibit questionable omissions and irrelevant additions upon thorough analysis. For example, the *Hoosier Mental Lexicon* (Nusbaum et al., 1984) —an online version of Webster's Pocket Dictionary with 20,000 words— includes irrelevant entries like *rauwolfia*. Also, it omits the word *although* while it does contain its informal counterpart *altho*. This apparent incoherence raises the question: Is it acceptable to omit formal written language entries and simultaneously include informal spoken language entries when calculating the lexical-sublexical metric? The sensory modality of the original text —or its country of origin— notably affects the words that appear or do not appear. Therefore, all studies should adequately justify the choice of any database.

Due to the significant divergence in linguistic usage, morphological derivatives only increase the problem. High L-TF verbs entail the inclusion of all their derivatives in the database (e.g., *edge*, *edges*, *edged*), while those with lower L-TF have none (e.g., *interlay*, *interlays*, *interlaid*). Sometimes, the morphological derivative is more important than the base word. For instance, the BLP database —which includes 28,780 monosyllabic and disyllabic English words— contains the entry *clueless* but not its root word *clue* (Keuleers et al., 2012). As Coleman and Pierrehumbert (1997) pointed out, including multiple entries for the same root word significantly affects the tactic probability value when not adequately filtered. This problem distorts the values of G-similarity metrics and those of C-similarity. An example of this is the aforementioned CLEARPOND's N (Marian et al., 2012), which includes all morphological derivatives of

the target (e.g., *edge*) as legitimate neighbors (e.g., *edges*, *edged*). Therefore, this discussion concerns all similarity measures.

Whenever a lexical database is selected, most authors accept it as it is, a decision that risks including and counting some morphological derivatives over others. This approach inevitably leads to many parasitic variables that can collectively modify the values of the lexical-sublexical metrics and their predictive power. A real solution to this problem would be to create an ideal lexical database for each study, a challenge that is impractical for most researchers. Alternatively, it is more plausible to introduce exhaustive exclusion criteria that standardize the conditions of the lexical entries when calculating the sequence's type frequency.

That said, the proposal of the functional type frequency seeks to overcome the operational issues of traditional lexical-sublexical metrics. This resolution is achieved by dividing stimuli based on functional patterns (i.e., onset and rhyme), avoiding arbitrary lengths or positional restrictions. Suppose a database contains many forms of the same root word that could bias the calculation. Exclusion criteria can quickly rule out these morphological derivatives (e.g., no plurals or compound words, only infinitive verbs). Onset and rhyme units are the best sublexical patterns to address and explain inconsistencies in English, making them essential for combining G-similarity and the notion of grapho-phonological correspondences. Moreover, this section has already developed the pairing of onset/inhibition and rhyme/facilitation. Thus, G-similarity has been theoretically distinguished from C-similarity while acknowledging and understanding their specific points of convergence (e.g., onset unit).

Nevertheless, some lingering questions remain regarding the relationship between G-similarity and C-similarity. Beyond the metrics' imprecision, how can the high correlation between G-similarity and C-similarity be explained? After all, modifying one while keeping the other constant is no simple task. How can they be experimentally decoupled? Most importantly, why are C-similarity metrics sometimes associated with facilitatory effects rather than inhibitory ones?

Understanding the mechanisms behind word formation can account for the high correlation between G-similarity and C-similarity. Aside from neologisms assembled or borrowed from other languages (e.g., *vodka*), most new words are composed of typical and "legal" sublexemes. For example, neologisms can arise from morphological modifications that do not transgress any native sublexical rules: affixation (e.g., *babble* → *cyberbabble*), backformation (e.g., *cathartic* → *cathart*), clipping (e.g., *prescription* → *script*), or blending (e.g., *light* + *radar* → *lidar*) (see Fang, 2021 for a review on

neologism formation). Thanks to these practices, the assembly of new words is naturally guided by the sublexical constraints contained within the morphemes. As a result, the neologism increases the general similarity of the neologism to the rest of the language.

A thought experiment can help in understanding this claim. To better represent the process, the pseudoword *trood* will be formed by blending *troll* —a supernatural being from Scandinavian folklore— and *hood* —a neighborhood. Thus, *trood* means "a neighborhood of trolls." Because this new monosyllabic pseudoword emerges from the onset ⟨tr⟩ and the rhyme ⟨ood⟩, it automatically belongs to both the onset neighborhood ⟨tr⟩ and the rhyme neighborhood ⟨ood⟩. When calculating the bigram/trigram probability of ⟨oo⟩, ⟨od⟩, or ⟨ood⟩ after creating this new neologism, an increase of one would be evident for all of these sequences, resulting in a higher value of G-similarity metrics. Furthermore, since the overall length of *trood* aligns with that of other monosyllabic English words, it will likely increase the size of an already established neighborhood (e.g., *brood*, *troop*, *trod*), thus increasing C-similarity. Unless the creation of the neologism follows a mechanism that does not guarantee adherence to sublexical constraints (e.g., foreign borrowings, acronyms), the new entry should simultaneously modify both similarity types.

If the natural creation of neologisms tends to increase both similarity metrics, it becomes a problem to distinguish the influence of each on a given result. So, how can the values of G-similarity and C-similarity be decoupled? The answer is simple: by controlling word length. Most English words have a length of four to five letters, with an average of 5.1 (Bochkarev et al., 2015), implying that the language has abundant monosyllabic words. This fact means that classic C-similarity metrics (e.g., neighborhood size) can only be high when the word length is close to the average. However, when this number is exceeded, the number of neighbors of the word decreases, and it may even cause the appearance of a second or third syllable. Up until this thesis, there has been no discussion regarding the legitimacy of neighbors with different syllable counts, excluding the CLEARPOND's N metric (e.g., *cleave* & *cleaver*, *place* & *palace*) (Marian et al., 2012). Without careful reflection, word syllabification could introduce another bias in calculating both similarity types.

In line with the gradient metric of OLD20, as the number of modifications that distinguish two words increases, the perceived similarity decreases through the holistic processing of the lexical route. This issue aggravates when considering the one-modification-only restriction of the classic C-similarity metrics, as most long words

inevitably have Coltheart's N, NNB, and N* values of zero. As Perea and Carreiras (1998) observed, classic neighborhood size metrics only apply when the average word length is close to the English average, a prerequisite that hinders their use in other languages. By contrast, since G-similarity focuses on the shared feature rather than the whole word, length differences between neighbors are not a problem (e.g., *goat* vs *throat*). Thus, C-similarity may decrease when working with longer experimental stimuli, but G-similarity remains constant.

Regarding inhibitory and facilitatory effects, the apparent lack of correspondence between G-similarity and facilitation could be due to the nonspecific nature of its most conventional metric, the tactic probability. However, if C-similarity is associated with confusability and lexical competence, how can it occasionally be related to facilitatory effects? One answer to this question is the lack of consistency in the use of the terminology, as some authors generically refer to the classic lexical neighborhood but end up using the rhyme neighborhood metric (e.g., Cortese et al., 2004; Justi & Jaeger, 2017).

Nevertheless, a more thorough observation indicates that the classic lexical neighborhood metrics —i.e., one-modification-only C-similarity metrics— also suffer nonspecificity. These metrics do not distinguish which sublexemes are shared between the target and its neighbors. Due to this nuance, C-similarity metrics overlook that a neighborhood can contain more onset or rhyme neighbors. For example, of the 4 Coltheart's neighbors of the word *brood* (*blood*, *broad*, *broom*, and *brook*), 3 are ⟨br⟩ onset neighbors. In contrast, of the 3 Coltheart's neighbors of the word *brawl* (*crawl*, *bawl*, and *brawn*), 2 are ⟨awl⟩ rhyme neighbors. By taking Coltheart's N values of 4 and 3, respectively, all this information is invisible, so it omits the association of the onset neighborhood with inhibitory effects and the rhyme neighborhood with facilitatory ones. In other words, the classic C-similarity metrics can function as equivalents of either onset or rhyme neighborhoods, that is, as metrics of G-similarity.

Because of all of the above, the conventional metrics of G-similarity and C-similarity are highly nonspecific with respect to their original purposes, making them unreliable measures for predicting outcomes. Their lack of theoretical consistency has led some authors to argue that the tactic probability metric is purely lexical (Bailey & Hahn, 2001) or purely sublexical (e.g., Vitevitch et al., 1999; Strauß et al., 2022), while neighborhood size behaves as a lexical-sublexical metric (e.g., Luce & Large, 2000; Strauß et al., 2022). In this regard, a worthwhile alternative to these inaccuracies is using OLD20/POLD20 for C-similarity and the functional type frequency for G-similarity.

This recommendation offers more substantial theoretical support than the classic pair.

In summary, this section has examined in detail the topic of the lexical-sublexical metrics of G-similarity. These have been defined and addressed with an emphasis on specific operationalization problems and their close association with wordlikeness judgements and C-similarity. Due to the highly nonspecific nature of the tactic probability metric, a new alternative has been proposed: the functional type frequency. However, sublexical processes can no longer be explained by G-similarity and C-similarity since both focus on some aspect of the lexical domain. So, what measure can accurately represent sublexical typicality?

3.3. Sublexical Token Frequency

Chapter 2 proposed the sublexicon as the domain whose representations are obtained through extrapolating repetitive patterns from the lexicon. In other words, sublexical representations and their respective sublexical typicality derive from analyzing variables related to the lexical domain. In line with this, wordlikeness judgement—a task presumably associated with sublexical typicality—positively correlates with the individual's vocabulary size (Bartolotti & Marian, 2014). Due to the close relationship between metrics in the lexical domain and pure sublexical processes, several authors have defined "wordlikeness" as a "relative similarity of pseudowords to real words" (Coleman & Pierrehumbert, 1997; Frisch et al., 2000; Hammond, 2004), rather than as the "subjective perception of compounded sublexical typicality." However, as mentioned in the previous section, neither G-similarity nor C-similarity can fully predict wordlikeness judgement results (Bailey & Hahn, 2001), so a satisfactory answer has yet to be found. Contrary to the lexical typicality metric (i.e., L-TF), which only requires tallying a word's occurrences in speech or written form, sublexical typicality demands a more refined calculation.

According to the first sublexical development processing of the infant (Cairns et al., 1997), *statistical learning* is defined as the unconscious process of acquiring patterns without the mediation of explicit instruction (Aslin, 2017). This idea is curiously akin to the passive internalization of phonotactic patterns through procedural memory (Ullman et al., 2020). From this emerged the *statistical-learning perspective*, which emphasizes that the choice of patterns to be learned is not random, as it does reflect features of the perceived input (see Pollo et al., 2008 for a review; Spencer et al., 2015). This new perspective conflicts with Gentry and Ferreiro's, arguing that the sublexical reconfiguration process occurs simultaneously and at different speeds for each pattern, not through a sequence of phases or stages. The term that unifies this new perspective is the *statistic*, which refers to the frequency of a given event or object that cannot be explained by chance. Thus, as the number of times a statistical pattern appears increases, it is more likely to be extrapolated as an independently stored representation.

The frequency of the statistic could refer to the number of words that share the same sequence, i.e., the lexical-sublexical metric (or G-similarity). If G-similarity were the only factor influencing sublexical acquisition, no word with low G-similarity should influence sublexical processes. However, there is evidence that a small number of repeatedly used words can significantly impact the occurrence of sublexically motivated

errors. For example, the sequences within a child's name tend to cause intrusive regularization errors when other words are read or spelled (Treiman & Broderick, 1998; Treiman et al., 2001). Instead of focusing solely on the number of different words with the same sequence (lexical-sublexical metrics), this observation emphasizes the importance of the repetition of each of these words (their respective L-TF). This perspective supports those who argue that similarity metrics better reflect wordlikeness judgements and real speech statistics if L-TF is weighted into the equation (cf., Bailey & Hahn, 2001; Fuchs, 2018).

The isolated number of entries with a shared pattern from a database (i.e., G-similarity) is not the same as the daily occurrence of a sequence (i.e., sublexical typicality). To correctly count the precise occurrence of a pattern in a given time space, it is necessary to include the number of times each entry was individually repeated (i.e., its summed lexical typicality). Therefore, a "pure" sublexical typicality metric is possible when two metrics related to the lexical domain combine, creating a *sublexical token frequency* (SL-TF).

$$\text{Sublexical Typicality} = \Sigma \text{Lexical Typicality} \times \text{G-Similarity}$$

While G-similarity provides the number of different lexical entries that share the same sequence, the lexical typicality of each word contributes to the overall importance of the pattern. Thus, creating a new sublexical representation is based on the sublexical typicality of the sequence, obtained by calculating the lexical relevance of all the words that share the pattern. The lack of consideration of lexical typicality in similarity metrics helps explain why they cannot fully predict sublexical processes. Nevertheless, G-similarity and the summed lexical typicality do not influence SL-TF similarly.

On the one hand, if a sublexical segment contains many archaic words that no one uses—a high G-similarity and very low summed lexical typicality—its sublexical typicality will be zero or close to it. In other words, this pattern might be present in many database entries, but it will not be typical in a real linguistic environment. There are not enough repetitions of each entry to consolidate the sequence as an independent sublexical representation. Since most lexical databases have already discarded most archaic entries, a situation of high G-similarity and low lexical typicality is rare.

On the other hand, if a pattern only appears in one frequently used word—a low

G-similarity but high summed lexical typicality— the final sublexical typicality should also be zero. Considering the Sublexical Irrelevancy Hypothesis (see Chapter 2), it is inefficient for the sublexicon to allocate resources into independent sublexical representations that only address one or a few words. In these cases, it is more resource-efficient—in terms of cognitive resources—to store the whole word in the lexicon and process it through the lexical route rather than from the sublexicon. Therefore, it is convenient to identify the threshold from which the G-similarity is sufficiently high to be relevant, justifying the creation of a sublexical representation. This "relevancy threshold" will be addressed in Section 4.3.2 of Chapter 4.

Also, as mentioned in Section 3.2, lexical databases do not always include typical words for a given environment. This omission can result in overestimating the number of words individuals know, reflected as a higher G-similarity (i.e., functional type frequency), which could anomalously increase the SL-TF of a rare sublexical segment. Thus, smaller G-similarity values are reduced through the natural logarithm (\ln) to diminish the impact of the overestimation. The higher the number of words with the same sequence (i.e., type frequency), the more significant the exponential increase of SL-TF.

$$SL-TF = \sum L-TF \times \ln (\text{Functional Type Frequency})$$

The calculation of SL-TF might seem straightforward if it were not for the multiple biases that lexical databases carry in their assembly. These biases lead to the emergence of other parasitic influences after the corpus selection process. For decision-making, it is helpful to consider two main qualities: the original nature of the information and the presence of morphological entries.

First, the lexical database can come from different sources, so its entries and values can diverge significantly from the linguistic context of a new study. Regarding L-TF, there are several databases: the *Kučera and Francis frequency corpus* (1967; Brysbaert & New, 2009); the *CELEX word-form frequency* (Baayen et al., 1993; Baayen et al., 1995); the *Subtitle (SUBTLEX) word-form frequency* (Brysbaert et al., 2012; van Heuven et al., 2014); and the *BNC frequency* (Bober et al., 2021).

Section 3.2 highlighted that the orthographic or phonological origin of the information biases the type of lexical entries included in the database. Nevertheless, the original sensory nature of the information influences L-TF even more significantly. For instance, there are cases where formal words—primarily associated with the

orthographic domain— are used orally, allowing their automatic addition as a lexical entry in the database. However, the entry's infrequent nature for this sensory modality will only be evident in the lexical typicality metric. If the sensory modality of the chosen database is not identical to that of the object of study, the values of L-TFs will be biased. In the same vein, the original time-space context from which the information was obtained (e.g., country, year of collection) can drastically modify lexical usage. Due to this issue, the selected lexical databases require, as much as possible, columns of L-TF that are more up-to-date. Another provisional solution is selecting databases that include L-TFs from many sources and countries to compensate for the lack of specificity. This alternative is a mainstay in some databases, such as the English Lexicon Project (Balota et al., 2007) or the BLP corpora (Keuleers et al., 2012).

Second, the existence of morphological derivatives as separate entries is an even bigger problem for L-TF than for G-similarity metrics, as it fragments the same root word into multiple entries, dividing its true lexical typicality. A criterion of exclusion for all morphological derivatives alleviates the issues of G-similarity metrics. However, it undermines the specific calculation of the sublexical typicality of the shared pattern. For example, if the SL-TF of the ⟨-air⟩ rhyme were to be calculated, the word *stair* should be considered, whose L-TF is 369/50 (BNC/CELEX). The problem lies in the fact that both its plural form *stairs* and its compound form *downstairs* have higher L-TF values than its singular form, with 805/3647 for the former and 442/1919 for the latter. By filtering them out of the database, the L-TF of *stair* cannot represent the true lexical typicality of the sequence ⟨-air⟩, as it only reflects the least typical form of the word. Also, what if the database lacks the standard form of the word? The BLP corpus does not contain the lexical entry *clue*, but it has its derivative *clueless*, which could be discarded due to the exclusion criterion (Keuleers et al., 2012). Since L-TF is highly influential in the calculation of SL-TF, any modification in its quantification will affect the accuracy of the sublexical metric. Currently, there is no simple solution to this problem. If the corpus is left intact, the presence of morphological derivatives alters the G-similarity metric. By contrast, the criterion that excludes all morphological derivatives biases the lexical typicality metric. This bias would diminish if someone undertakes the tedious task of merging all the L-TF values of the same word family into a single, representative entry. This procedure will not be carried out in this thesis, as it requires a deeper exploration of the morphological domain. Until a perfected database appears, the decision to include or exclude morphological derivatives will continue to alter, respectively, the G-similarity or lexical typicality values.

As a final remark, there have been previous attempts to combine a similarity metric with L-TF into a single weighted metric. However, they suffer from the same operationalization problems described in Section 3.2. For instance, Luce (1986) combined the L-TF of the neighborhood size metric (i.e., C-similarity) into the *frequency-weighted neighborhood probability rule*. Similarly, Bailey and Hahn (2001) also combined L-TF with C-similarity in the form of the *frequency-weighted neighborhood similarity* metric. These authors also tested the inclusion of L-TF in G-similarity using a *frequency-weighted bigram phonotactic probability*. The issue with the latter lies in its use of arbitrary constraints of tactic probabilities (e.g., bigram, trigram), meaning the weighted L-TF is taken from words that do not necessarily share the functional sublexical unit with the target. To illustrate, when calculating the bigram ⟨la⟩ for the word *place*, the L-TF of *claim* and *flair* are considered, even though these words do not share the onset or rhyme with *place*. Or take the words *avenger* and *strength*, whose shared bigram ⟨ng⟩ does not even belong to the same syllable. How can the L-TF of these words be justified in the calculation? Is there any sublexical foundation that legitimizes the arbitrariness of the bigram? As argued in Section 3.2, this lack of specificity only introduces biases into the metric, diminishing its predictive power. Therefore, it should come as no surprise that the inclusion of weighted frequencies was utterly irrelevant in predicting wordlikeness judgements in the study of Bailey and Hahn (2001).

An interesting take on the topic is the *Average Bigram/Trigram Token Frequency (BF_TK)* of Davis (2005; see Davis & Perea, 2005). The L-TFs of all words containing the same bigram/trigram are summed and later divided by the number of bigrams/trigrams in the target. For example, the word *edge* has three bigrams (⟨ed⟩, ⟨dg⟩, and ⟨ge⟩), while the L-TF of each of these bigrams is 81, 79, and 343, respectively. By summing all these L-TFs and dividing them by 3 (i.e., the number of bigrams), a BF_TK of 167 is obtained. Due to its operationalization, BF_TK becomes a peculiar metric of lexical typicality, as it represents the average of all L-TFs obtained from a sublexical constraint in the form of bigrams or trigrams. Interestingly, the same author also addresses type frequency (i.e., G-similarity) with his other metric, the aforementioned BF_TP metric.

Beyond these attempts to combine lexical typicality and sublexical constraints, no other metric has combined L-TF and a functional G-similarity metric. In the present day, SL-TF is the only proposed sublexical typicality metric that has been subjected to exhaustive theoretical and conceptual rigor. It might become a new tool for controlling

and predicting sublexical processes while also holding significance for pseudoword assembly based on stochastic grammar (Coleman & Pierrehumbert, 1997). Moreover, thanks to the functional type frequency, the SL-TF metric is integrable into the usually-overlooked study of grapho-phonological inconsistencies.

Finally, thanks to the introduction and support of the new SL-TF and functional type frequency metrics, understanding processes related to pseudowords, such as the subjective perception of wordlikeness, becomes more feasible. There are several issues regarding lexical databases, which may undermine the calculation of an accurate SL-TF. Nevertheless, through the use of pseudowords assembled in a controlled manner, the new sublexical typicality metric is put to the test as a predictor of wordlikeness judgements.

SECOND PART:
EXPERIMENTAL RESEARCH

CHAPTER 4

Methodological Design

4.1. *Overall Design and Sampling Strategy*

This work is framed within a quantitative experimental design, as the examined variables are metrics numerically collected (Hernández Sampieri & Mendoza, 2018). It also has a correlational scope, comparing the differential relationships between these variables. In addition to controlling the level of vocabulary knowledge and previous bilingual experiences in the pre-experimental phase, the study is conducted cross-sectionally, as only a single instance obtains all the information.

Regarding the sampling process, all participants need to have a level of English proficiency close to native speakers. Therefore, this study only considers participants with a high performance level in the language, as sublexical orthographic knowledge develops with a bigger vocabulary size and through repeated use and interaction with the language (see Chapter 2) (Bartolotti & Marian, 2014; Bordag et al., 2017). Additionally, the processing of opaque languages (i.e., English) at the native level tends to rely more on orthographic information than phonological information (Holm & Dodd, 1996; Nassaji & Geva, 1999). This fact is relevant, as this thesis focuses on orthographic processing, which requires controlling and reducing phonological influence in all tasks.

Due to the scarcity of native English-speaking communities in Argentina, this study's participants must have an English level similar to native proficiency. To reach this population sample, we work with students from the following careers at the *Universidad Nacional de Cuyo* (UNCuyo): *Profesorado Universitario en Lengua Inglesa* (University Teaching Degree in English), *Licenciatura en Filología Inglesa* (Bachelor's Degree in English Philology), and *Traductorado Público en Inglés* (Public Translator Degree in English). In order to register the effects of different vocabulary sizes in this community, the sampling has students from the first and final years of their careers. This selection is not random, i.e., it is non-probabilistic (Hernández et al., 2004). As an exclusion criterion, no native English speakers, individuals with cognitive or linguistic disorders, or those with knowledge of other Germanic languages as L2 are included.

4.2. *Pre-experimental Tests*

Before starting any experimental test, all participants must go through a pre-

experimental phase that consists of three segments: an informed consent that must be accepted, a test to verify vocabulary size (Nation & Beglar, 2007), and a questionnaire to document experiences and individual traits relevant to bilingualism (Li et al., 2019).

This phase will last approximately 50 minutes; thus, it accounts for 65% of the time required for data collection.

4.2.1. *Vocabulary Size Test (VST)*

Nation and Beglar (2007) created a revised version of the *Vocabulary Size Test (VST)*, a free online test (my.vocabularysize.com/session/evstxx20) that contains 100 multiple-choice items. The VST measures the English receptive written vocabulary size for L1 and L2 learners. To do so, it measures written and decontextualized knowledge of words, which helps predict the current lexical storage required for processes such as reading. However, the VST does not provide any information on their daily correct usage in speaking and writing. Test takers must choose a sentence's best definition or translation by selecting one of four options. All these exercises represent various levels of lexical frequency based on the *BNC* (Bober et al., 2021). This database suits our study, as most of the English learning books used at UNCuyo come from the United Kingdom. The test takes approximately 20 minutes to complete.

Although vocabulary size is not a very important variable in this study, it may significantly impact the results of the tasks. The size of the lexicon directly influences the size of sublexical knowledge (Bartolotti & Marian, 2014) and, consequently, could affect the results dependent on sublexical processes. Similarly, vocabulary size can influence all processes that depend on lexical typicality. In other words, with a larger language knowledge, low L-TF words could be subjectively perceived and remembered as more typical. Finally, vocabulary size helps ensure that the division between groups with different knowledge levels serves its intended purpose (i.e., first-year versus last-year students).

4.2.2. *Language History Questionnaire 3 (LHQ3)*

This questionnaire is an instrument designed to measure an individual's linguistic history and performance, a necessary tool for research related to L2 (Li et al., 2019). It has been translated into several languages, including Spanish, and takes approximately 20-30 minutes to complete, making it the longest test in this study. Some of the variables recorded or calculated here include the following: personal characteristics (age, gender, education level), native language proficiency (L1

proficiency), and knowledge of foreign languages (L2 proficiency, multilingual language diversity score, knowledge of consistent, inconsistent, and Germanic L2s).

4.3. *Creation of the Experimental Tests*

During the remaining 10 to 15 minutes, participants must complete the three tasks focused on in this study. The first is a modified version of the neighbor generation task by Luce and Large (2001), while the other two address the sublexical process of wordlikeness perception of pseudowords. It is important to note that all the stimuli used for these tasks require careful selection based on their respective SL-TF, obtained after filtering the database.

4.3.1. *Step 1: Filtering of the BLP Database*

When working with lexical databases, numerous issues exist regarding their use in calculating SL-TF (see Chapter 3). First, G-similarity values (e.g., type frequency and tactic probability) can be influenced by the omission of important words (e.g., *clue*, *although*) and the inclusion of foreign or irrelevant entries (e.g., *rauwolfia*, *suave*). Second, the L-TF of a stimulus may not be present or may be too outdated to be considered current and representative of the present linguistic environment. Third, the counting of G-similarity and L-TF is affected by morphological derivatives of the same word as independent entries (e.g., *edge*, *edges*, *edged*, *edgeway*). This fact forces the researcher to take two measures: 1) select the most representative lexical entry, with the drawback that it loses significant L-TF values (e.g., plurals that are more frequent than their singular forms) —control of G-similarity and bias in lexical typicality—; 2) include all lexical entries so that it comprises all L-TF values in the calculation, with the drawback that the same word family might be counted multiple times —control of lexical typicality and bias of G-similarity. Lastly, the impact that polysyllabic processing may have, compared to monosyllabic stimuli processing, is unknown, so its inclusion should be adequately justified.

The inability to fully control the biases mentioned forces a decision when using the database to obtain the SL-TF. In order to achieve standardization, a filtering process was carried out on a database. This work used the *BLP* corpus (Keuleers et al., 2012), which provides valuable columns: Coltheart N, OLD20, number of syllables, number of letters, and L-TF metrics. This database provides 14,365 monosyllabic and disyllabic entries obtained from British participants, which aligns with UNCuyo's

preference for British material. From the BLP, its lexical entries are filtered based on six factors: 1) syllable, 2) word length, 3) L-TF, 4) onset, 5) morphological derivatives, and 6) foreign words from Spanish.

First, all lexical entries with unidentified syllables—which encompasses all pseudowords in the BLP—and those with more than one syllable are eliminated. Thus, only monosyllabic entries remain to control the influence of phonological length, accent, and other prosodic cues.

Second, based on the modified methodological simplification of the work by Jacobs et al. (1998), only monosyllabic words with a length of 3 to 6 letters will be used. This decision is based on two arguments. 1) Emphasis is placed on words with the three subsyllabic structures (i.e., onset, peak, and coda), which is only possible with words of more than 2 letters. It reduces the impact that the absence of any of these structures could have on the results. 2) The average length of words in the English language is 5.1 letters (Bochkarev et al., 2015), so 4-to-6-letter words are included to incorporate as many words as possible. However, given that there is a large number of 3-letter words with high L-TF, this length has also been considered. By selecting the range of 3-6 letters, all sequences with relevant wordlikeness are included to assemble monosyllabic pseudowords.

Third, all lexical entries with an uncategorized or zero L-TF—in terms of BNC frequency (BNC L-TF) and word-form frequency in CELEX (CELEX L-TF)—are eliminated. This decision reduces the number of archaic and foreign words while also discarding all words that do not influence the calculation of SL-TF.

Fourth, all monosyllabic entries without an onset are ruled out, preventing any effects caused by their absence.

Fifth, all plural forms (e.g., <-s>, <-es>) and all verbs or adjectives formed by regular verb suffixes (e.g., <-s>, <-es>, <-ed>, <-ied>) are removed, which inevitably creates a bias in the calculation of L-TF. Nevertheless, this filter reduces the morphological influence on calculating G-similarity and SL-TF.

Finally, all foreign entries with an evident Spanish origin (e.g., *suave*) are discarded, given that individuals in the population sample are Spanish speakers.

4.3.2. Step 2: SL-TF Calculation and Sublexical Irrelevancy Hypothesis

The BLP corpus offers three different types of metrics for L-TF, of which only the BNC and CELEX are considered (van der Wouden, 1990; Keuleers et al., 2012). The first was obtained from a British population, while the second was from an American

community. An averaged L-TF metric (BNC-CELEX L-TF) accounts for the influence of a British-American L-TF. As for a *frequency per-million* measure, the BLP only includes that of the BNC (Keuleers et al., 2012) but not of CELEX. The frequency per-million metric is not considered, as it lacks an American equivalent and because its calculation includes all lexical entries filtered for this study (e.g., disyllabic words, morphological derivatives).

Once the BLP is filtered, word lists organize them based on their onsets or rhymes, allowing for the calculation of the functional type frequency and total L-TF (BNC Σ L-TF, CELEX Σ L-TF, BNC-CELEX Σ L-TF) for each subsyllabic unit.

$$\text{Sublexical Typicality} = \Sigma \text{Lexical Typicality} \times \ln(\text{G-Similarity})$$

In Chapter 2, we proposed the *Sublexical Irrelevancy Hypothesis*, which states that the sublexical domain does not extrapolate a new representation in the sublexicon for a sequence that is shared in a few words, regardless of whether these words are used frequently or not. Only when a certain number of different words sharing the same pattern (i.e., G-similarity) is stored does the system consider the sequence sufficiently relevant to justify as an independent sublexical unit. Otherwise, words or new words whose sublexical patterns do not surpass this "*relevancy threshold*" must necessarily rely on more rudimentary comparison processes at the lexical level. Since this is the first time the sublexical irrelevancy and relevancy threshold are mentioned, it is unclear from which G-similarity value (i.e., functional type frequency) a sequence becomes "relevant."

In order to identify the relevancy threshold, four different methods of calculating SL-TF are tested, each with a different G-similarity restriction value. The first has no restriction (*SL-TF (no restriction)*), the second has a restriction of 2 (*SL-TF (>2)*), the third has a restriction of 4 (*SL-TF (>4)*), and the last one has a restriction of 6 (*SL-TF (>6)*). This restriction corresponds to the G-similarity relevancy threshold, below which the SL-TF is always equal to 0. For example, if a rhyme has a type frequency of 3, it will be relevant in SL-TF (>2) but irrelevant in SL-TF (>4).

$$\text{SL-TF} = \Sigma \text{L-TF} \times \ln(\text{Functional Type Frequency})$$

The different forms of SL-TF reflect distinct relevancy thresholds, with values that will be tested as predictors of the results of the sublexically-related tasks. Under this

criterion, the thresholds most in line with reality should show better predictive power regarding the results of these tasks.

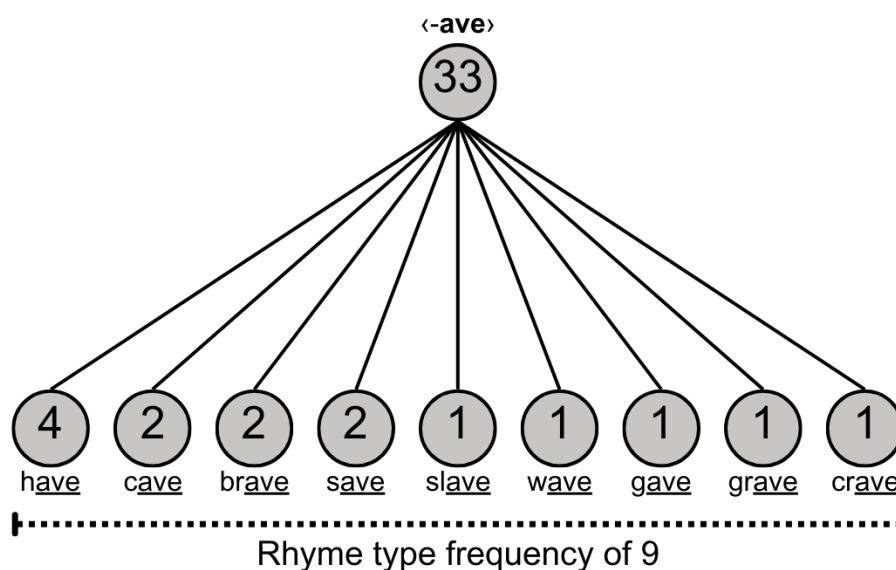
In summary, the steps required to obtain the SL-TF in the English language are as follows. First, a subsyllabic pattern (i.e., onset or rhyme) is selected as a target sequence. Second, the selected functional type frequency is calculated from the chosen database, and its final value is later transformed into logarithmic form. Third, the L-TF of all the entries considered in the previous step must be summed (Σ L-TF). Finally, the products of the second and third steps are multiplied to obtain the sequence's SL-TF. During this final calculation, the four relevancy thresholds are applied for each of the three L-TF databases (BNC, CELEX, BNC-CELEX), resulting in twelve different types of SL-TF. Figure 28 illustrates the SL-TF (>2) for the rhyme unit <-ave>, its rhyme type frequency, and the L-TF of each entry, obtained from a text, not a specific database.

Figure 28

Sublexical Token Frequency

In front of the **cave** one of the boys reflected: "I should **have** known that I was not as **brave** as I thought. If I continue on this path, it will be my **grave**." To **crave** for recognition was something that the boys rarely felt, but, as the girl he liked asked for his help, one of them was now a **slave** to his own pride. In spite of the **waves** dangerously upswinging into the entrance, they still had to **save** the cat. "W... we **have** to do it." said one. If they **gave** up and returned empty-handed from the **cave**, they would become the mockery of the town. "We **have** to finish the job." said another. "Let's muster our courage and **brave** against the odds!" finally said the scared child. "We **have** to **save** it!"

<-ave>									
Lexical Entries	have	cave	brave	save	slave	wave	gave	grave	crave
L-TF	4	2	2	2	1	1	1	1	1
Rhyme Type Frequency	9								
SL-TF	32.9584								



Note. (a) is a fabricated text with multiple repetitions of words with the rhyme <-ave> (underlined and in bold). Based on the text, (b) shows the calculated L-TF values, the rhyme type frequency, and the SL-TF (>2), while (c) represents them as lexical and sublexical nodes. As can be seen, the word wave is included in its singular form despite appearing in its plural form (waves) within the text.

Once the SL-TF of the subsyllabic sequences is calculated, they are organized in descending order to distinguish the most typical sublexemes from the rarest ones. The BNC-CELEX SL-TF (>2) metric will be the organizing keyword.

4.3.3. Step 3: *Alphabetic and Numerical Sublexical Hierarchy*

Since this work focuses on the sublexical aspect of language, the emphasis is placed on the rhyme, as the onset is equally relevant to the lexical domain (see Fig. 26).

All onsets have a medium SL-TF, i.e., between the 35th and 65th percentiles of all onsets greater than 0 in the BNC-CELEX SL-TF (>2) column. Controlling the SL-TF values of the onset ensures that the task results are predominantly influenced by the rhyme and, thus, by sublexical processing. The reason for selecting such a wide range for the onset is due to its limited number in the complete list. Expanding the range prevents the same onsets from being used multiple times in the pseudoword assembly process. Also, it broadens the range of possibilities when combined.

Regarding the rhymes, five levels (A, B, C, D, and E) are created based on both the BNC-CELEX SL-TF (>2) and the BNC-CELEX Σ L-TF. Levels A, B, and C are the percentiles of entries with a BNC-CELEX SL-TF (>2) greater than 0. *Level A* corresponds to a high SL-TF rhyme —between the 85th and 100th percentiles—, *level B* corresponds to a medium SL-TF rhyme —between the 55th and 70th percentiles—, and *level C* corresponds to a low SL-TF rhyme —between the 15th and 30th percentiles. On the other hand, levels D and E have a BNC-CELEX SL-TF (>2) of 0 and are only distinguished by their BNC-CELEX Σ L-TF. This choice aligns with the statement that when the sequence of a stimulus does not reach sublexical relevance, it relies on lexical measures and comparisons to process it (see Section 2.1). Thus, *level D* corresponds to high Σ L-TF rhymes —between the 80th and 90th percentiles—, and *level E* corresponds to low Σ L-TF rhymes —between the 10th and 20th percentiles.

This *alphabetical hierarchy* aligns with the Sublexical Irrelevancy Hypothesis (Section 2.1), as it prioritizes the importance of rhymes presumably already stored in the sublexical domain (i.e., levels A, B, and C). At the same time, it organizes the sublexically-irrelevant rhymes based on the Σ L-TF of the lexical domain (i.e., levels D and E). The alphabetical hierarchy categorically divides rhymes into categorical groups, allowing for the selection of rhymes for this study. However, it is necessary to conceive a hierarchy that nuances the value of rhymes in a non-categorical manner so that it

may offer a way to analyze the multiple SL-TFs better.

Thus, the *numerical hierarchy* consists of the enumeration or ranking of rhymes based on their SL-TF and, in cases where the SL-TF is equal to 0, on their Σ L-TF. Since there are 12 types of rhyme SL-TF —due to the three databases of Σ L-TF (i.e., BNC, CELEX, BNC-CELEX) and four relevancy thresholds (i.e., no restriction, >2, >4, >6)—, there are 12 variations of the numerical hierarchy that differentially rank the sequences. This hierarchy becomes critical in Section 5.2, where its use in pseudoword wordlikeness tasks is specified.

4.3.4. Step 4: Stochastic Assembly of Pseudowords

As mentioned in Section 2.3, the original claim of Coleman and Pierrehumbert's stochastic grammar (1997) states that a compound sublexical frequency (i.e., cumulative phonotactic probability) can predict the perception of wordlikeness. This compounded metric is obtained by multiplying the sublexical (tactic) probabilities of the functional sublexemes (e.g., onset, rhyme, syllable) that make up the analyzed pseudowords of the task.

This work adopts the methodology proposed by stochastic grammar for pseudoword assembly, albeit it changes the metric used by the authors for the calculation. Instead of tactic probabilities, the SL-TF values are used to identify the compound sublexical frequency of the pseudowords.

However, two factors in this study could affect the cumulative SL-TF's predictive power of wordlikeness results. First, levels D and E have an SL-TF of 0; therefore, the compound sublexical frequency will also be 0. This fact makes it impossible to predict wordlikeness at these levels using this metric alone. Second, since the selected onsets all have a medium SL-TF (35th and 65th percentiles), the differences in results should respond more significantly to the rhyme's SL-TF than to the onset's. Thus, at least under the restrictions of this study, the cumulative SL-TF is equivalent to the SL-TF of the rhyme.

When selecting sublexical units for the pseudoword assembly, the onset must have a length of 2 letters, while the rhymes should be 3 or 4 letters long. The coda must be present in every selected rhyme. The assembled pseudowords exhibit a length of 5 to 6 letters, which allows for the disruption of the proportionality between C-similarity and G-similarity (see Section 3.2). Most English words have a length of 4 to 5 letters; therefore, only words within this range exhibit the largest neighborhoods in the language (e.g., Perea & Carreiras, 1998). By contrast, longer words exhibit fewer

neighbors, a property used in this study to homogenize the impact of C-similarity metrics (i.e., Coltheart's N, NNB, and OLD20). Bailey and Hahn (2001) stated that C-similarity metrics have a predictive power similar to sublexical typicality metrics regarding pseudoword wordlikeness. For this to be true, C-similarity metrics must show a higher correlation with the results, regardless of the disadvantage created by word length.

After assembling the pseudoword, a search on the Google engine verifies if the stimulus exists. It will not be included if it appears as a real and valid word. However, if the search concludes that the pseudoword is a proper noun of an unknown person, a business name, or an archaism from Middle or Old English, it will be accepted nonetheless. It is unlikely that the Mendocinean community has interacted with such specific information. In any case, after the experimental phase, participants will be asked on the final sheet whether they have previously interacted with any of the pseudowords.

After assembling 20 pseudowords, 4 for each level (A, B, C, D, and E), the values of classic C-similarity (i.e., Coltheart's N, NNB, and OLD20) and G-similarity metrics (e.g., orthotactic probability, BF_TP) are calculated. These metrics serve as a comparative framework against the values of numerical hierarchy.

On the one hand, Coltheart's N (Coltheart et al., 1977) and NNB (Bailey & Hahn, 2001) are obtained by the simple edit distance metric (Hammond, 2004). Both C-similarity measures represent neighborhood size, with the distinction that the former only considers substitution neighbors, while the latter also includes neighbors obtained by addition and deletion. Each pseudoword's neighbor tallied is recorded in a LibreOffice 24.2 Calc sheet so that it may provide a visual record of the calculation process for each of these metrics. This method also allows for the methodical calculation of OLD20 (Yarkoni et al., 2008).

On the other hand, given the lack of consensus regarding the operationalization of standard G-similarity metrics, two different calculation methods are used: the *Orthotactic Probability* by Ring et al. (2023b) and the *Average Bigram Type Frequency (BF_TP)* by Davis (2005). The former is characterized by including empty characters in the bigrams at the word boundaries—the word *back* has five bigrams: ⟨_b⟩, ⟨ba⟩, ⟨ac⟩, ⟨ck⟩, and ⟨k_⟩— and by dividing the sum of the bigram probabilities by the original number of letters in the word—the summed probability of the bigrams in *back* is divided by 4. By contrast, the BF_TP omits the semi-empty bigrams of the first method—*back* has three bigrams: ⟨ba⟩, ⟨ac⟩, and ⟨ck⟩— and divides the summed bigram

probability by the number of bigrams —the summed bigram probability of *back* is divided by 3. While the metric of Ring et al. (2023b) better reflects the original definition of the tactic probability, BF_TP emerges from Davis's (2005) recognition that there are two types of frequency: token frequency and type frequency. However, in both cases, the bigram restrictions are fixed and arbitrary, as they do not consider the functional salience of the onset or rhyme units. The only change we made to the original calculation of these metrics is that their values are not transformed into logarithms, which should not affect their final correlation with the results. This decision is taken to homogenize both measures of G-similarity, as the transformation into logarithm occurs at different steps in their respective calculations.

CHAPTER 5

Analysis and Results

5.1. *Modified Orthographic Neighbor Generation Task*

5.1.1. *Description and Hypothesis*

As stated in Section 3.1, the original neighbor generation task (Luce & Large, 2001) is a behavior-based test. It involves presenting a phonological target (e.g., /fin/) and then asking participants to recall and write down all the words that show structural similarity with the target (e.g., *feet, mean, phone, foam, meat, bin*). The authors referred to the total number of neighbors recalled in this task as *entropy*, a value equated with neighborhood size, i.e., C-similarity. According to this logic, the larger the neighborhood size of the stimulus, the higher its entropy.

However, upon further analysis of some of the results provided by the authors (see Table 2), the neighbors with a higher occurrence of being recalled were those with higher L-TF within the entire neighborhood. This favoritism towards high L-TF could imply that the recall is not based on lexical similarity but on lexical typicality. A hypothetical case can exemplify this point. If there were a neighborhood with 17 lexical neighbors but filled with low L-TF words except for one high L-TF word (i.e., a large but infrequent neighborhood), it is likely that the individual would not recall all the words (i.e., low relative entropy). By contrast, if a neighborhood of 6 neighbors only had high L-TF words (i.e., a small but frequent neighborhood), it is more likely that all the words would be recalled (i.e., high relative entropy). The example questions the equivalence between entropy and neighborhood size (i.e., C-similarity).

This observation justifies the repetition of the task, albeit modified, to simplify some of the complexities of the first study. Instead of orally presenting a pseudoword that triggers the recall of neighbors, a written rhyme is given to trigger the recall of the rhyme neighbors. On the one hand, this makes it easier to delimit the words to be recalled and simplifies the subjective appreciation of "structural similarity" between the whole pseudoword and other real words. On the other hand, by working with the orthographic domain instead of the phonological one, the stimulus is temporally fixed, allowing for more time to analyze and recall similar stimuli.

To reduce the influence of G-similarity in the results —and focus solely on C-similarity versus L-TF— their values are controlled and homogenized. Thus, the

selected rhymes as targets must have a rhyme type frequency ranging from 11 to 17 (percentile 85-95). The use of rhymes that could be considered Anglo-Saxon names (e.g., *ike*) or words (e.g., *and*, *ace*, *imp*, *old*, *oak*) in themselves is not allowed.

If the entropy measure depended on similarity to the target, it should show a higher correlation with the values of C-similarity or even with those of G-similarity. Conversely, to verify the hypothesis, the results should be better explained by the L-TF values of each neighbor in the rhyme neighborhood.

With that said, the objective of the present experiment is the following:

- To compare the differential impact between lexical typicality (i.e., L-TF) and lexical similarity (i.e., C-similarity) in the orthographic neighbor generation task results in a population of non-native Argentinian English speakers.

Based on this, the hypothesis is the following:

- The number of words recalled in the orthographic neighbor generation task shows a stronger correlation with lexical typicality than lexical similarity in a population of non-native Argentinian English speakers.

5.1.2. *Participants*

54 native Spanish-speaking students from UNCuyo participated in the study (38 women, 15 men, and 1 non-binary) in April 2024. The age range extended from 17 to 29 years (mean = 20.8 ± 0.4). Of these, 49 identified as right-handed, 4 as left-handed, and 1 as ambidextrous. None of the participants reported any neurological issues related to language. Among the participants, 5 enrolled in the Bachelor's Degree in English Philology, 42 in the Public Translator Degree in English, and 11 in the University Teaching Degree in English. Two participants pursued two degrees simultaneously, while one enrolled in all three. Regarding bilingual knowledge, 16 participants know a consistent L2 (e.g., Italian, Korean, Portuguese), while 12 know an inconsistent L2 (e.g., French, Japanese). Of these participants, 27 are first-year students—plus one in the second year—and the remaining 26 are in their third and fourth years, defining the groups with lower and higher proficiency in English, respectively. All participants provided written consent and were informed about the procedure before participating in the study.

5.1.3. Materials and Methods

The experimental items in the neighbor generation task consist of four rhyme units, two with a high SL-TF (A level) and two with the lowest SL-TF values above zero (C level). The rhyme type frequency is kept between 11 and 17 for all four stimuli, so the differences in SL-TF only differ in Σ L-TF.

Due to the difficulty in finding two rhymes with both these restrictions at level C, a rhyme between levels B and C was selected (i.e., B-C). Additionally, it was impossible to find rhymes with these characteristics that maintained the same letter length. Therefore, to keep this variable under control, the selection considers one group with 2 letters and another with 3 letters. The level and letter length of the rhyme were subjected to an orthogonal analysis with respect to the results (Table 5).

Table 5

Selected Rhymes

Alphabetic Hierarchy	Stimulus	Number of Letters	Rhyme Type Frequency	BNC Σ L-TF	CELEX Σ L-TF	BNC-CELEX Σ L-TF
A	UT	2	15	541626	114304	327965
A	OOK	3	12	130300	26727	78513,5
B-C	AB	2	17	5718	1187	3452,5
C	OOP	3	12	2660	498	1579

After selecting the rhymes, they were randomly organized in the questionnaire in an uppercase format to avoid the influence of ascending and descending consonants on the results. This decision aims to make sublexical processing more difficult and force dependence on lexical processing (see Section 2.1 on shape invariance). An example was placed at the top of the page (e.g., <-ave> → *have, gave*), along with a note prohibiting the inclusion of words that functionally modify the rhyme (e.g., *heave, paved*).

The task implies observing the rhyme and writing down all the words that contain it within their structure. Any recalled word with more than one syllable was omitted.

Data from eight participants were excluded during this task: four for knowing other Germanic languages, one for providing an incorrect vocabulary size, and three for not completing the neighbor generation task. Not completing the task could be considered a conscious and deliberate response by the participants. However, in the end, these eight cases were nevertheless omitted. A total of 46 participants were

included in this task.

5.1.4. Analysis

For the data analysis, Pearson's correlation coefficient was calculated using the *PSPP* software (GNU Project, 2015), a free alternative to *IBM SPSS Statistics*. This correlation between the independent variables and the dependent variables was obtained. The analysis is divided into three segments: per participant, per recalled word, and per rhyme.

The *analysis per participant* focuses on detecting the influence of the participants' characteristics that could affect the results of the analysis per recalled word. On the one hand, this includes the following independent variables obtained from the pre-experimental data collection: academic year, age, variables from the LHQ3 (L1/L2 proficiency, multilingual language diversity score, knowledge of other consistent/inconsistent L2s), and vocabulary size. On the other hand, the dependent variables include the following: the *percentage of recalled words per rhyme*, their *average percentage per rhyme level*, and *per rhyme's number of letters*.

The *analysis per recalled word* focused on the differential impact of the lexical typicality and C-similarity metrics of the rhyme neighbors concerning their recall in the neighbor generation task. In this case, the independent variables were those derived from the rhyme neighbors: the number of letters of these words, C-similarity (Coltheart's N, OLD20), and all L-TF types (BNC, CELEX, BNC-CELEX). The analysis of these variables was subject to groupings by rhyme, level, and number of letters of the rhymes. The dependent variable was the *percentage of recall per word*.

The *analysis per rhyme* focused on identifying the existence of sublexical influences on the results of the analysis per recalled word. The independent variables were as follows: number of letters in the rhyme, measures of G-similarity with arbitrary constraints (orthotactic probability and BF_TP) (Ring et al., 2023b; Davis, 2005) and functional measures (rhyme type frequency), Σ L-TF (BNC, CELEX, BNC-CELEX), and all varieties of SL-TF. The dependent variable was the *percentage of recalling per rhyme*.

5.1.5. Results

In the analysis per participant, vocabulary size was the only significant variable across all results. It showed a moderate positive correlation with the percentage of recalling for neighbors with A level rhymes —UT (Pearson 0.462; $p = 0.001$), OOK

(Pearson 0.419; $p = 0.004$)— and with OOP from level C (Pearson 0.495; $p < 0.001$). However, the strongest correlation was observed with the recalling of neighbors with the AB rhyme (Pearson 0.636; $p < 0.001$), which is the B-C level rhyme. This high correlation explains the higher correlations with the average recalling of neighbors with B-C/C level (Pearson 0.608; $p < 0.001$) and 2-letter rhymes (Pearson 0.604; $p < 0.001$), while the correlations were lower for level A (Pearson 0.511; $p < 0.001$) and 3-letter rhymes (Pearson 0.516; $p < 0.001$).

Regarding the academic year and age, the correlations were mostly non-significant. In the few instances where significance was attained, the correlations were very weak and did not indicate any particular pattern. On the one hand, the academic year showed a correlation of 0.298 ($p = 0.044$) in recalling words with OOK (A level). On the other hand, age showed a correlation of 0.303 ($p = 0.041$) in the recalling of words with AB (level B-C), 0.324 ($p = 0.028$) for the average recalling of level A, and 0.314 ($p = 0.034$) and 0.293 ($p = 0.048$) for the recalling of words with 2- and 3-letter rhymes, respectively.

Several participants did not follow the instructions correctly when completing the LHQ3, which compromised the validity of the data on L1 proficiency, L2 proficiency, and the multilingual language diversity score. Knowledge of other L2s —whether consistent or inconsistent— did not correlate significantly with the results.

In the analysis per recalled word, lexical typicality was the only variable that showed significance in retrieving all four words. Of the three L-TF databases, the one with the best values was BNC-CELEX —AB from the B-C level (Pearson 0.728; $p = 0.001$) and BNC for UT (Pearson 0.536; $p = 0.040$), OOK (Pearson 0.811; $p = 0.001$), and OOP (Pearson 0.834; $p = 0.001$). Regarding the C-similarity metrics, significant and moderate/strong correlations were found for UT —the only instance where Coltheart's N exceeded L-TF as the most important correlation (Pearson 0.620; $p = 0.014$)— and OOK, i.e., the two A-level rhymes. Conversely, results for the AB rhyme showed a significant correlation only with OLD20 (Pearson -0.502; $p = 0.040$) —a negative value, given that similarity is higher when OLD20 is lower— but no significance for Coltheart's N. The results for the OOP rhyme showed a lack of significance for both. Anecdotally, the number of letters in words was only significant in word recalling with OOK at level A (Pearson -0.700; $p = 0.011$). The details of the L-TF and C-similarity values are presented in Table 6.

Table 6*Correlations of Words per Rhyme with L-TF and C-Similarity*

Rhyme of Generated Neighbors	Coltheart's N	OLD20	BNC L-TF	CELEX L-TF	BNC-CELEX L-TF	
UT	0.620	-0.569	0.536	0.531	0.535	Pearson
	0.014	0.027	0.040	0.042	0.040	p
OOK	0.765	-0.711	0.811	0.808	0.811	Pearson
	0.004	0.010	0.001	0.001	0.001	p
AB	0.460	-0.502	0.705	0.676	0.728	Pearson
	0.063	0.040	0.002	0.003	0.001	p
OOP	0.192	-0.270	0.834	0.690	0.831	Pearson
	0.550	0.395	0.001	0.013	0.001	p

All correlations were significant when the analysis was made based on the number of rhyme letters. The data indicate that the neighbor generation with 2-letter rhymes correlates better with C-similarity than with L-TF, particularly OLD20 (Pearson -0.543; $p = 0.001$). Conversely, results of 3-letter rhymes show a stronger correlation with L-TF, especially BNC (Pearson 0.684; $p < 0.001$). When the analysis delved into rhyme levels, level A showed a significant correlation with both C-similarity and L-TF, Coltheart's N being the highest (Pearson 0.595; $p = 0.001$). By contrast, the only significant correlation at the B-C/C level was with L-TF, especially with BNC-CELEX (Pearson 0.685; $p < 0.001$). The details of these values are presented in Table 7.

Table 7*Correlations of Words per Level/N° of Letters with L-TF and C-Similarity*

Analysis Criteria	Coltheart's N	OLD20	BNC L-TF	CELEX L-TF	BNC-CELEX L-TF	
2-Letter Rhymes	0.540	-0.543	0.467	0.463	0.466	Pearson
	0.001	0.001	0.007	0.008	0.007	p
3-Letter Rhymes	0.559	-0.542	0.684	0.678	0.683	Pearson
	0.005	0.006	0.000	0.000	0.000	p
A Level Rhymes	0.595	-0.541	0.433	0.426	0.432	Pearson
	0.001	0.004	0.024	0.027	0.024	p
B-C/C Level Rhymes	0.214	-0.287	0.681	0.560	0.685	Pearson
	0.264	0.131	0.000	0.002	0.000	p

Regarding the analysis per rhyme, none of the correlations were significant, suggesting that generating neighbors based on rhymes did not have sublexical influences.

5.1.6. Preliminary Discussion

Among the results of the analysis per participant and per rhyme, the only variable that became important in interpreting the results was vocabulary size. It correlated significantly and strongly with word recalling of the AB rhyme (Pearson 0.636; $p < 0.001$), a short and infrequent rhyme. It is important to mention that the AB rhyme had the highest rhyme type frequency of the four used (i.e., 17). The second most significant rhyme was OOP (Pearson 0.495; $p < 0.001$), a long and infrequent rhyme. Thus, vocabulary size —equivalent to lexical knowledge— showed moderate/strong correlations with word recalling of infrequent rhymes (i.e., low SL-TF due to low L-TF). This finding is coherent, as the words with high Σ L-TF —those with the A level rhymes— are better known and recalled by all types of individuals, even novices. In contrast, words with low L-TF —those with the B-C or C level rhymes— are more likely to be recalled if the participant has had enough interactions with the language. Therefore, a greater knowledge of the language should entail a higher influence on its recalling.

The rest of the variables showed a lack of significance, and in the correlations that did attain significance, there was a lack of well-defined patterns. This observation

could indicate that the 46 participant entries and the 4 rhymes did not provide enough information to draw unambiguous conclusions about the influence of the individual and the sublexical factor on the neighbor generation task. The analysis per participant was particularly affected hereby, as some of its variables got correlations that were possibly products of chance. The incorrect data collection from the LHQ3 did not alter this fact, as the lack of significance also affected the correctly obtained variables (e.g., knowledge of consistent and inconsistent L2). It did not help that the few significant correlations were all weak or very weak.

Furthermore, the years of study did not show a significant correlation with vocabulary size (Pearson 0.217; $p = 0.139$). This finding implies that the participants' entry to their degree programs pre-requires a high level of language proficiency. Thus, the distinction between the first and final years did not reveal significant differences in the results. Due to these findings, the remaining analyses in this thesis will exclude the academic year as an independent variable. For future studies, samples with more unequal language knowledge should be considered.

Regarding the analysis per recalled word, the L-TF was the only measure that maintained a significant correlation across all four rhymes, with moderate values for UT (Pearson 0.536; $p = 0.040$), strong values for AB (Pearson 0.728; $p = 0.001$), and very strong values for OOK (Pearson 0.811; $p = 0.001$) and OOP (Pearson 0.834; $p = 0.001$). In contrast, the C-similarity measures (Coltheart's N and OLD20) showed non-significant correlations for rhymes with the lowest SL-TF (AB and OOP), except for OLD20, which demonstrated a significant and moderate correlation with the results of AB (Pearson -0.502; $p = 0.040$).

From these results, three conclusions can be drawn. First, the C-similarity metrics—i.e., measures related to the lexical neighborhood—cannot predict all the outcomes of the neighbor generation task when a sublexical stimulus is used. Second, the entropy measure of Luce and Large (2001) showed a more consistent correlation with OLD20 than with classic C-similarity measures such as Coltheart's N, a fact not considered by the original authors. Third, the lexical typicality of each generated neighbor rather than by C-similarity influences the entropy measure, confirming our hypothesis.

When analyzing the data based on the number of letters of the rhymes, the C-similarity of the retrieved neighbors showed a stronger correlation with recalling words of 2-letter rhymes. In contrast, L-TF more strongly correlated with neighbors of 3-letter rhymes. Although recalling neighbors with AB did not significantly correlate with

Coltheart's *N*, this indicates that shorter rhymes —ergo, shorter words— are more influenced by C-similarity than longer ones. This conclusion aligns with the assertion that classic C-similarity measures are helpful as long as the length of the stimuli is not too long (Perea & Carreiras, 1998).

Nevertheless, it is logical to assume that lexical similarity (C-similarity) should influence the impact of a lexical stimulus on other similar words within the lexicon. Conversely, C-similarity should not influence as much when a sublexical pattern (rather than a word) triggers the retrieval of words that contain it. Considering that the task is a "sublexical to lexical" process, the correlation with G-similarity should be higher than that with C-similarity. Despite controlling the rhyme type frequencies by only using those obtained in the 85-95 percentile, the 2-letter rhymes also showed the highest G-similarity values —15 for UT and 17 for AB. Conversely, both 3-letter rhymes —OOK and OOP— had a rhyme type frequency of 12. Since C-similarity and G-similarity are directly proportional to each other for shorter lengths (see Section 3.2), the higher correlation with C-similarity may indicate that there is also a correlation with G-similarity. This observation was not adequately addressed in this study, as the aim was to control for G-similarity to prevent it from influencing the process.

It is also interesting that recalling 2-letter rhymes shows a higher correlation with C-similarity, while vocabulary size correlates with the AB rhyme in the analysis per participant (Pearson 0.636; $p < 0.001$). As language knowledge increases, so does the number of neighboring words to a given target. Thus, the influence of the vocabulary size on the effects caused by lexical and lexical-sublexical similarity should not be underestimated. Classical C-similarity metrics have been characterized as a strong predictor of vocabulary size at the early ages (e.g., Stokes, 2010). Similarly, vocabulary size modulates and is associated with G-similarity in sublexically-mediated processes (e.g., Munson et al., 2005; Graf Estes et al., 2016). Beyond these clarifications, there is little to conclude on this topic since the influence of vocabulary size on C-similarity or G-similarity values was not the focus of this study. Future research should address the exact nature of the relationship between rhyme length, G-similarity, C-similarity, and vocabulary size.

Three conclusions can be drawn based on the results per number of letters in the rhymes. First, although L-TF is the measure that predicted the results in all cases, its importance is more significant when the rhymes/words are longer. Second, similarity metrics appear to predict better the results related to short rhymes/words, albeit it is impossible to say whether this is due to lexical similarity (i.e., C-similarity) or lexical-

sublexical similarity (i.e., G-similarity). This finding is consistent, given that a smaller word size reduces the perceptual distance between the neighbors and increases the proportionality between both similarities (see Section 3.2). Third, the methodology used to control the G-similarity values of the rhymes was insufficient. More rhymes should be included in a future study to fully clarify their influence in the neighbor generation task. This way, different G-similarity values would be considered, improving the significance of the analysis per rhyme.

When analyzing the data based on rhyme levels, L-TF was once again the only metric that showed significance in the A and B-C/C levels. In contrast, C-similarity metrics only showed significance for the A level. What stands out is that the correlation between the A level results and the C-similarity was higher than with the L-TF. Comparing the highest values for each A level rhyme, we see that OOK showed a Pearson of 0.811 ($p = 0.001$) with BNC L-TF versus 0.765 ($p = 0.004$) for Coltheart's N, while UT showed a Pearson of 0.536 ($p = 0.040$) with BNC L-TF versus 0.620 ($p = 0.014$) for Coltheart's N. That is, OOK was more correlated with L-TF —L-TF with longer rhymes/words— while UT was more correlated with Coltheart's N —C-similarity with shorter rhymes/words. However, it should not be ruled out the possibility that responses triggered by high SL-TF rhymes may rely more on C-similarity, in contrast to rhymes with low SL-TF.

Three conclusions stem from the results associated with the rhyme level. First, L-TF is very important for retrieving neighbors with low SL-TF rhymes. Thus, only the most frequent words in these rhyme neighborhoods are recalled, while the less frequent ones are not. Second, L-TF is less relevant for retrieving neighbors with high SL-TF rhymes. These rhyme neighborhoods contain a higher percentage of words with high L-TF, leading to a more homogeneous sample of words, reducing the L-TF's statistical power. Third, C-similarity appears only relevant at higher SL-TF levels, with a greater preference for shorter rhymes/words. However, it is essential to note that the correlations between the A level responses with C-similarity and with L-TF are moderate, with C-similarity being slightly stronger by about 0.1.

Regarding the correlations with the L-TF, a higher correlation appeared for the BNC L-TF values, followed by BNC-CELEX. This finding suggests that the study's participants prefer British and mixed material, which aligns with the characteristics of the chosen population.

Finally, it is important to remember that morphological derivatives were removed from the list due to the database filtering, meaning that the L-TF values do not account

for their usage in these cases (see Section 3.3). Therefore, unifying the L-TF values of the word family into a single lexical entry would be beneficial to improve further the predictive power of lexical typicality.

5.2. *Wordlikeness Tasks for Pseudowords*

5.2.1. *Description and Hypothesis*

Chapter 1 presented the lexical domain with its measures of L-TF and lexical neighborhood, while Chapter 2 focused on the sublexical domain and its measures of tactic probability. However, Chapter 3 discussed how frequently the boundaries between these theoretical and operational divisions become blurred, causing problems and controversies in interpreting the results. In this regard, the sublexical aspect is the most affected, as it is the least understood and most underexplored area of linguistic processing. Tactic probability —the most commonly used sublexical metric by psycholinguists to date— has shown great proportionality and equivalence with C-similarity values, leading some to argue that it is simply a more elaborate version of the lexical neighborhood (see Sections 3.1 and 3.2). Without a doubt, the lexical aspect remains one of the most studied topics in psycholinguistics, alongside the semantic and morphological domains. Conversely, the sublexical domain remains shrouded in an assumed simplicity, relegated to mere processes of "grapheme-to-phoneme" or "phoneme-to-grapheme" conversions, without recognizing a greater complexity beyond the individual letter or phoneme (e.g., Macchi et al., 2019; Kaestner, 2022).

Under the framework of the dual-route model (see Section 2.1), the sublexical route processes new words and pseudowords by using previously stored patterns and units in the sublexicon. Nevertheless, it is essential to recognize that the lexical route can influence the processing of pseudowords. For example, when a pseudoword is sufficiently similar to a known word, it can generate lexicalization errors that depend on the lexical route (see Section 2.3). Furthermore, when a pseudoword exhibits a highly atypical pattern for the language —meaning it has no sublexical relevancy (see the Sublexical Irrelevancy Hypothesis)— the processing of the stimulus shifts into a smaller level than that of sequences (i.e., the alphabetic level). However, if the individual knows other words that share the atypical pattern, the lexical information (e.g., L-TF) may be used as a reference frame (see Section 4.3.3).

By controlling the lexical influences in sublexical processing, pseudowords become a powerful tool for studying the sublexical route. This revisit of the stochastic grammar highlights the importance of a controlled and meticulous assembly of

pseudowords (Coleman & Pierrehumbert, 1997). The selected stimuli cannot be too similar to an existing word but must not be so different as if it were a foreign one. To achieve this balance, the pseudowords used in future studies of sublexical processing must have a controlled wordlikeness. Thus, sublexical metrics capable of accurately predicting this aspect are required.

Wordlikeness tasks for pseudowords are an excellent tool for observing the effects of sublexical frequency (Bailey & Hahn, 2001). The subjective perception of wordlikeness involves a non-categorical weighting of plausibility or "legality" of sequences. Such weighting depends on the typicality of the sublexical representations that make up the stimulus (e.g., onset, rhyme, syllable). Additionally, since the wordlikeness rating process is modality-specific—meaning that orthotactic regularities are different and independent from phonotactic regularities—it is possible to study the orthographic sublexicon without considering its conversion to phonology.

So far, the tactic probability and the lexical neighborhood metrics have been used as predictors of wordlikeness, but their results have been unsatisfactory, as they only explain 38% of the outcomes (Bailey & Hahn, 2001). The theoretical reasons why these variables do not predict the outcomes have already been detailed in Section 3.2, with the most critical factor being that neither truly reflects the sublexical typicality of pseudowords. However, after designing a new measure of sublexical typicality—the sublexical token frequency (SL-TF)—it is necessary to test its effectiveness as a predictive tool for the results obtained in the wordlikeness task for pseudowords.

That said, the goal of the following experiments is:

- To examine the predictive power of the new sublexical typicality metric, compared to conventional lexical and sublexical metrics, in visual wordlikeness tasks for pseudowords in a population of non-native Argentinian English speakers.

With this goal in mind, the numerical hierarchy has been proposed (see Section 4.3.3), an ordering of pseudowords based on their SL-TF—for sublexically relevant sequences (i.e., levels A, B, and C)—and their Σ L-TF—for sublexically irrelevant sequences (i.e., levels D and E). With that said, the hypothesis is as follows:

- The numerical hierarchy—organized based on sublexical typicality (for relevant sequences) and lexical typicality (for irrelevant sequences)—is the best

predictor of the results of the visual wordlikeness task for pseudowords in a population of non-native Argentinian English speakers.

The method for measuring the pseudowords' wordlikeness will be divided into two task types: wordlikeness judgement and wordlikeness arrangement.

5.2.2. Participants

The participants for this task were the same as those for the orthographic neighbor generation task (see Section 5.1.2).

5.2.3. Wordlikeness Judgement

5.2.3.1. Materials and Methods

Based on the 20 pseudowords stochastically assembled in Section 4.3.4, ten stimuli are selected, two per level (A, B, C, D, and E). Table 8 shows the ten pseudowords with their respective BNC-CELEX SL-TF (>2) and BNC-CELEX Σ L-TF.

Table 8

Pseudowords for Wordlikeness Judgement

Stimulus	Onset BNC-CELEX SL-TF (>2)	Rhyme BNC-CELEX SL-TF (>2)	BNC-CELEX Σ L-TF	Alphabetic Hierarchy	Numerical Hierarchy (BNC-CELEX >2)
KNADE	319636.05	177955.35	71614.5	A	9
DRITH	185425.21	431633.23	392889.5	A	10
TWEAL	297141	32605.38	13597.5	B	8
PRATCH	181965.03	38265.51	16618.5	B	7
GRELVE	506953.97	4401.04	3476	C	6
CLIRTH	300877.81	4818.76	4006	C	5
PLESK	300314.97	0	2994	D	4
TRULF	263812	0	1803.5	D	3
FLIEL	146615.38	0	25	E	2
BLYRRH	185527.06	0	21	E	1

Once the pseudowords are selected, they are randomly listed in the questionnaire in uppercase format. While homogenizing the format may make sublexical processing more difficult, avoiding any bias related to ascending and descending letters is preferable.

The participant has to observe the pseudoword and rate it on a scale from 1 to 7, based on its "English-like" nature (see Section 2.3). According to Coleman and

Pierrehumbert's (1997) methodology, 1 corresponds to impossible pseudowords in English; 2 and 3 are for unlikely ones; 4 represents an intermediate wordlikeness; 5 and 6 are for very likely ones; and 7 is for the most "English-like" pseudowords.

Data from nine participants were excluded from this task: four for knowing other Germanic languages, one for providing an incorrect vocabulary size, one for omitting all the wordlikeness tasks, and three for specifically omitting wordlikeness judgement tasks. A total of 45 participants were included in the final analysis for this task.

5.2.3.2. Analysis

The Pearson correlation coefficient was calculated using the PSPP software, with the independent variables correlated with the dependent variables. The data analysis is divided into two parts: per participant and per judged pseudoword.

The *analysis per participant* focused on detecting individual influences that could affect the analysis results per judged pseudoword. On the one hand, this analysis considers the following independent variables obtained from the pre-experimental data collection: age, variables from the LHQ3 (L1/L2 proficiency, multilingual language diversity score, knowledge of other consistent/inconsistent L2s), and vocabulary size. On the other hand, it includes the following dependent variables: the *average wordlikeness by level* (A, B, C, D, E), the *weighted average based on the alphabetical hierarchy* —A(5), B(4), C(3), D(2), and E(1)— and the *weighted average based on the numerical hierarchy* —10, 9, 8, 7, 6, 5, 4, 3, 2, and 1. The latter has 12 variants due to combining the three L-TF databases —BNC, CELEX, and BNC-CELEX— and the four types of sublexical relevancy thresholds— no restriction, >2, >4, and >6.

The *analysis per judged pseudoword* focuses on identifying the variable that best predicts the wordlikeness judgement results. The independent variables from each pseudoword are the following: number of letters, C-similarity (Coltheart's N, NNB, OLD20), G-similarity with arbitrary constraint (orthotactic probability and BF_TP) (Ring et al., 2023b; Davis, 2005) and functional constraint (onset and rhyme type frequency), Σ L-TF —BNC, CELEX, BNC-CELEX—, the 36 variants of SL-TF —onset, rhyme, and cumulative; BNC, CELEX, and BNC-CELEX; no restriction, >2, >4, and >6—, the alphabetical hierarchy —A(5), B(4), C(3), D(2), E(1)—, and the 12 variants of the numerical hierarchy —BNC, CELEX, and BNC-CELEX; no restriction, >2, >4, and >6. The dependent variable is the *average wordlikeness of the pseudoword*.

5.2.3.3. Results

In the analysis per participant, the two dependent variables that showed more significant correlations were the weighted averages based on the alphabetical and numerical hierarchies. However, all these significant values correspond to correlations with the years of study, a metric now irrelevant to this study (see Section 5.1.6).

As with the orthographic neighbor generation task, the values obtained in the LHQ3 (L1 proficiency, L2 proficiency, and multilingual language diversity score) are compromised, so they are not detailed.

Regarding the wordlikeness averages per level, level A significantly correlated with vocabulary size (Pearson 0.298; $p = 0.046$) and the knowledge of other inconsistent L2s (Pearson 0.302; $p = 0.044$). Level D showed a significant (although weak) correlation with age (Pearson 0.312; $p = 0.037$).

Regarding the analysis per judged pseudoword, only three types of variables showed a correlation with the results. First, the rhyme type frequency had a Pearson correlation coefficient of 0.781 ($p = 0.008$). Second, the alphabetical hierarchy —A(5), B(4), C(3), D(2), and E(1)— showed a correlation of 0.812 ($p = 0.004$). Finally, the strongest correlations were obtained with the variants of the numerical hierarchy, being particularly high for BNC-CELEX at the relevancy thresholds of >4 and >6 (Pearson 0.905; $p < 0.001$). The details of the correlations between the results and the variants of the numerical hierarchy, alongside those of the C-similarity and classic G-similarity metrics, are presented in Table 9.

Table 9

Correlation between the Average Wordlikeness Judgement and the Predictor Variables

C-Similarity	Coltheart's N	NNB	OLD20	Pearson p
	0.463	0.574	-0.609	
	0.178	0.083	0.062	
G-Similarity	Orthotactic Probability	BF_TP	Rhyme Type Frequency	Pearson p
	0.461	0.606	0.77	
	0.18	0.063	0.009	
Lexical Typicality (Σ L-TF)	BNC	CELEX	BNC-CELEX	Pearson p
	0.129	0.128	0.129	
	0.722	0.725	0.723	
Numerical Hierarchy (SL-TF & Σ L-TF)	BNC + no restriction	CELEX + no restriction	BNC-CELEX + no restriction	Pearson p
	0.804	0.804	0.804	
	0.005	0.005	0.005	
Numerical Hierarchy (SL-TF & Σ L-TF)	BNC + >2	CELEX + >2	BNC-CELEX + >2	Pearson p
	0.804	0.804	0.804	
	0.005	0.005	0.005	
Numerical Hierarchy (SL-TF & Σ L-TF)	BNC + >4	CELEX + >4	BNC-CELEX + >4	Pearson p
	0.9	0.893	0.905	
	0	0	0	
Numerical Hierarchy (SL-TF & Σ L-TF)	BNC + >6	CELEX + >6	BNC-CELEX + >6	Pearson p
	0.9	0.893	0.905	
	0	0	0	

At the end of the task, all participants indicated that they had never interacted with the pseudowords used in the study.

5.2.3.4. Preliminary Discussion

In the analysis per participant, most of the variables lack significance, and there is no clear pattern in the values that do reach significance. This situation is similar to that of the orthographic neighbor generation task, where the analysis per participant also did not provide clear data.

Two conclusions can be drawn from these results. First, it reiterates that the

decision to divide participants by academic year was ineffective, as it did not reveal any statistically significant differences in the knowledge of English. We recommend that future studies consider more heterogeneous samples in terms of language knowledge to report significant correlations regarding the influence of vocabulary size in wordlikeness tasks. Second, it is also likely that the lack of significance is a consequence of the small number of students (45), a sample size that is too small for this type of analysis. Given this, it is not possible to draw definitive conclusions about the individual influence on the results.

In comparison, the analysis per judged pseudoword provides a more insightful result. First of all, the paradigmatic metrics of both C-similarity (i.e., Coltheart's N, NNB, and OLD20) and G-similarity with arbitrary constraint (i.e., orthotactic probability and BF_TP) showed moderate correlation values that did not reach significance. These results confirm that these measures are too nonspecific concerning wordlikeness judgements for pseudowords, as they do not reflect the sublexical (or lexical) typicality needed to ponder the sequences. However, there was a correlation with the rhyme type frequency. This fact may indicate that even though this metric does not directly reflect sublexical typicality, the lexical-sublexical similarity used to calculate SL-TF retains some predictive power. When comparing the G-similarity metrics with arbitrary constraints (i.e., orthotactic probability and BF_TP) against those with functional ones (i.e., rhyme type frequency), the insignificance of the former allows us to infer that they do not reliably reflect the sublexical processing, while the latter does.

The best correlation with the results —and, therefore, the best predictor of wordlikeness judgements—is provided by the BNC-CELEX numerical hierarchy with relevancy thresholds of >4 and >6 (Pearson 0.905; $p < 0.001$). The most representative consequence can be seen with the pseudoword *DRITH*, initially ranked as A level and numerically as 10, i.e., this task's most "English-like" pseudoword. Despite its high SL-TF, its rhyme type frequency is 3. Because of this, when calculating the hierarchies with the >4 and >6 relevancy thresholds, the pseudoword drops to level D and to the 7th position, a change reflected in the results. Across all the other L-TF databases, the >4 and >6 relevancy thresholds consistently showed the strongest correlations, both in BNC (Pearson 0.9; $p < 0.001$) and in CELEX (Pearson 0.893; $p < 0.001$).

From these results, six conclusions can be drawn. First, the classic measures of C-similarity and G-similarity poorly predict the results of wordlikeness judgements for pseudowords. Second, the functional constraint (e.g., syllabic, sub-syllabic) is more closely related to sublexical processing than the arbitrary constraint (e.g., bigrams and

trigrams). Third, individuals show a preference for mixed material and, secondarily, for British material. This inclination is consistent with the characteristics of the study's population. Fourth, as posited in Section 2.1, the Sublexical Irrelevancy Hypothesis is confirmed, given that a relevancy threshold higher than 0 is required to obtain the strongest correlations. Thus, the relevancy threshold labeled as no restriction—which represents a null restriction that does not allow for the existence of irrelevant sequences—showed one of the weakest correlations in the numerical hierarchy. Fifth, the relevancy threshold seems to lie close to the values of 4 and 6. Since this study only used even-numbered thresholds—and only up to a threshold of 6—the precise restriction might be 3 or higher. Sixth, our hypothesis is confirmed that the numerical hierarchy—organized on both the new SL-TF for relevant sequences and the Σ L-TF for irrelevant sequences—is the best predictor of wordlikeness judgement for pseudowords. This evidence theoretically and experimentally supports the new sublexical typicality metric. Moreover, it contradicts Bailey and Hahn's (2001) claim that lexical typicality is irrelevant to the perception of wordlikeness.

5.2.4. Wordlikeness Arrangement

5.2.4.1. Materials and Methods

From the 20 pseudowords stochastically assembled in Section 4.3.4, the remaining ten stimuli—not used in the wordlikeness judgement task—were selected, two per level (A, B, C, D, and E). Table 10 presents the ten pseudowords alongside their respective BNC-CELEX SL-TF (>2) and BNC-CELEX Σ L-TF values.

Table 10

Pseudowords for Wordlikeness Arrangement

Stimulus	Onset BNC-CELEX SL-TF (>2)	Rhyme BNC- CELEX SL-TF (>2)	BNC-CELEX Σ L-TF	Alphabetic Hierarchy	Numerical Hierarchy (BNC-CELEX >2)
PRELL	181965.03	320978.44	115768.5	A	10
TROOD	263812	200094.77	83446	A	9
KNORD	319636.05	44142.68	24636.5	B	8
BRELT	271004.51	47051.02	20434	B	7
CHAITH	465785.34	3449.64	3140	C	6
CLUDE	300877.81	2893.77	1798	C	5
FLUGE	146615.38	0	4917.5	D	4
QUILST	140543.82	0	3095	D	3
BLAUB	185527.06	0	19	E	1
DRULLE	185425.21	0	27.5	E	2

After determining the ten stimuli, two groups of five pseudowords each were created, one for each level (A, B, C, D, and E). These pseudowords were then randomly listed in the questionnaire in uppercase format.

The participant's task was to observe the five pseudowords and rank them hierarchically —from most wordlike (very English-like) to least wordlike (impossible as an English word). Unlike the classic wordlikeness judgement task, the wordlikeness arrangement task requires the participant to evaluate each stimulus while simultaneously rating the other pseudowords, meaning there is a mandatory comparison between the stimuli. The results for each pseudoword were subsequently quantified on a scale of 1 to 5 based on its position in the task. This task is of original design and has not been used in previous studies.

The data from six participants were excluded from this task: four due to knowledge of other Germanic languages, one for providing an incorrect vocabulary size, and one for omitting all the wordlikeness tasks. A total of 48 participants were included in this task.

5.2.4.2. *Analysis*

The analysis followed the guidelines described in Section 5.2.3.2 for the wordlikeness judgement tasks, so it will not be reiterated. The distinction between an *analysis per participant* and one *per arranged pseudoword* is maintained.

5.2.4.3. *Results*

In the analysis per participant, the only independent variables that showed any significant correlation were the academic year —which will not be detailed here (see Section 5.1.6)— and the vocabulary size. In both cases, the dependent variable they correlated with was the average wordlikeness of level C, showing a negative correlation of -0.323 for vocabulary size ($p = 0.025$). Again, since the values obtained in LHQ3 are compromised, they are not addressed in this section.

In the analysis per arranged pseudoword, only three variables correlated with the results. First, the lowest significant correlation was found for the rhyme type frequency (Pearson 0.71; $p = 0.021$). Second, the alphabetical hierarchy showed a correlation of 0.857 ($p = 0.002$). Finally, the highest correlations were obtained from the variants of the numerical hierarchy, with the most important being CELEX with the >4 relevancy threshold (Pearson 0.889; $p = 0.001$). Table 11 presents the details of the correlations

between the task results regarding the variants of the numerical hierarchy, along with the classic C-similarity and G-similarity metrics.

Table 11

Correlation between the Average Wordlikeness Arrangement and the Predictor Variables

	Coltheart's N	NNB	OLD20	
C-Similarity	0.322	0.224	-0.439	Pearson
	0.364	0.533	0.205	p
	Orthotactic Probability	BF_TP	Rhyme Type Frequency	
G-Similarity	0.324	0.595	0.71	Pearson
	0.36	0.069	0.021	p
	BNC	CELEX	BNC-CELEX	
Lexical Typicality (Σ L-TF)	0.487	0.501	0.49	Pearson
	0.153	0.14	0.151	p
	BNC + no restriction	CELEX + no restriction	BNC-CELEX + no restriction	
	0.795	0.868	0.822	Pearson
	0.006	0.001	0.003	p
	BNC + >2	CELEX + >2	BNC-CELEX + >2	
Numerical Hierarchy (SL-TF & Σ L-TF)	0.795	0.868	0.822	Pearson
	0.006	0.001	0.003	p
	BNC + >4	CELEX + >4	BNC-CELEX + >4	
	0.794	0.889	0.844	Pearson
	0.006	0.001	0.002	p
	BNC + >6	CELEX + >6	BNC-CELEX + >6	
	0.687	0.808	0.737	Pearson
	0.028	0.005	0.015	p

As with the wordlikeness judgement task, all participants indicated that they had never interacted with the pseudowords used in this task.

5.2.4.4. Preliminary Discussion

As with all analyses per participant of this study, there is a lack of significance in most of the correlations. However, this time, the insignificance is even greater than that

seen with the wordlikeness judgements. The only difference with the previous task is that the academic year and vocabulary size showed a shared significance pattern in the average wordlikeness value for level C. Such a result may suggest that at lower levels of knowledge —whether in terms of the academic year or vocabulary size— the pseudowords in level C were rated as "more English-like." However, this coincidence is interpreted as noise rather than a defined pattern due to the lack of correlation between both variables. Additionally, it is essential to note that the number of participants must increase to improve the significance of the results, as the 48 participants included here do not appear to have been sufficient.

However, what is curious is that the results from the wordlikeness arrangement task showed fewer significant correlations than the wordlikeness judgement task. The erratic significance of some correlations in the wordlikeness judgements —academic year, age, vocabulary size, and knowledge of another inconsistent L2— may have occurred due to the lack of comparison in each individual rating. This fact could have led to more fluctuating ratings than if they were restricted by comparison with other pseudowords (comparative weighting). In contrast, the wordlikeness arrangement task requires a comparative analysis between other pseudowords, which limits the subjectivity with which the stimulus can be rated. A future study could explore in more detail which of these two wordlikeness tasks generates results with less bias and more alignment with actual patterns.

Regarding the analysis per arranged pseudoword, the results were equivalent to those observed in the analysis per judged pseudoword. Once again, the classic measures of C-similarity and G-similarity with arbitrary constraints did not show significance in the results. In addition to the significant correlation of rhyme type frequency (Pearson 0.71; $p = 0.021$), these results confirm the lack of specificity of C-similarity and G-similarity with arbitrary constraints as predictors of sublexical processes. Moreover, this reinforces the assertion that the sublexical domain is guided and stored based on functional constraints (e.g., subsyllabic and syllabic units) rather than arbitrary bigrams or trigrams.

In the context of the wordlikeness arrangement task for pseudoword, the numerical hierarchy showed (once again) the best correlation with the results, becoming their best predictor. Here, the most critical L-TF database was CELEX — followed by BNC-CELEX in second place— suggesting that the participants' knowledge is more American than British. The American inclination contradicts both the expectations and results observed in the other tasks unless one accepts the possibility

that the participants' profile is somewhat mixed. This conclusion aligns with the significant importance of BNC-CELEX throughout all the tasks.

The most notable finding that distinguished the results of the wordlikeness arrangement task was that the relevancy threshold with the best results was only >4 , whereas >6 showed much lower values. This finding is highly relevant, as it not only reinforces the validity of the conclusions obtained in the wordlikeness judgement task but also further narrows the potential relevancy thresholds to around 4 ± 1 .

Based on these results, all the conclusions mentioned in the previous discussion are corroborated. First, the classic measures of C-similarity and G-similarity are too unspecific to predict sublexical processes. Second, functional constraints (e.g., syllabic, subsyllabic) better reflect sublexical processing than arbitrary constraints (e.g., bigrams and trigrams). Third, participants from UNCuyo showed a preference for American materials and, secondarily, for materials of mixed origin. Fourth, the Sublexical Relevancy Hypothesis seems essential in predicting the results of sublexical processes. Fifth, the relevancy threshold —the limit that G-similarity must surpass to store a sublexeme in the sublexicon— must lie above 2 and below 6, i.e., 4 ± 1 . Sixth, the hypothesis that the numerical hierarchy —organized based on the new SL-TF for relevant sequences and Σ LTF for irrelevant sequences— is the best predictor of the results of the wordlikeness task for pseudowords is confirmed.

CHAPTER 6

General Discussion

The division between the lexical and sublexical domains is a topic that has not been extensively addressed in previous literature. It is common to read texts discussing the lexical, semantic, and morphological domains. However, the sublexical domain has historically been relegated to a secondary role, a sort of "wild card" used to simplify the explanation of the processing of unknown stimuli.

Thanks to dual-route models, the sublexicon can be placed on a pedestal due to its adaptability to new words and pseudowords. Nevertheless, these models have failed to delve into their underlying complexities, reducing them to a mere "grapheme-phoneme" or "phoneme-grapheme" conversion, depending on the modality in which the input is received and the output is produced. To this day, it is still the standard to refer to the sublexical route as a "phonological route." Such a label does not reflect reality and it even suggests that the lexical route —also known as the "superficial", "surface" or "orthographic route"— is incapable of reaching its phonological counterpart without passing through the semantic domain.

Thanks to the study of wordlikeness of pseudowords, the concept of sublexical frequency gained importance, shaping the development of orthotactic probability metrics. However, the use of arbitrary constraints instead of functional ones only resulted in a lack of significance, sparking countless debates about their validity. Additionally, their count of lexical entries has been interpreted as a mere complicated replica of the lexical neighborhood size metric. As a result, similarity is tackled without distinguishing between the different similarity types. At the same time, similarity metrics are sometimes used to represent typicality.

Thanks to the challenges posed by inconsistent languages, we can recognize that not all sequences are the same: some facilitate grapho-phonological processing, while others slow it down. This proposal, while reasonable, lacks recognition in works addressing sublexical processing. As a result, grapho-phonological inconsistency is usually tackled in detail but in isolation, without a psycholinguistic model that integrates all its mechanisms and explores its implications to their full extent.

The knowledge that underlies the ideas of similarity, typicality, and the sublexical domain has been proposed numerous times, sometimes even in pathological contexts. However, they have never been combined with the level of detail presented in this

work, nor has the validity of the paradigmatic measures used to represent such concepts been questioned.

The following sections present a synthesis of the observations and global findings so readers can understand their relevance for future work.

6.1. *Analysis per-Participant*

Although the analysis per participant was not the focus of this work, but rather the analysis per word or pseudoword, it remains relevant as it provides information about external influences. However, the conclusions drawn from this analysis were significantly compromised due to three issues.

First, the data collection with the LHQ3 was carried out with a multiplicity of errors made by the participants, which undermines the reliability of the variables obtained from it. The questionnaire was physically delivered, meaning that the accuracy of the students' answers directly depended on how carefully they read each instruction. As a result, reading comprehension errors were the leading cause of bias. For example, some questions required a numerical answer based on the daily interaction time with the language, but many participants answered with an "X," as if it were a checklist. Because of this, the only relevant LHQ3 variables worthy of consideration were those indicating some knowledge of other L2s (consistent or inconsistent). Still, these metrics did not delve into each participant's proficiency level per language. A virtual questionnaire would have solved this problem, as its interface explicitly pre-defines the type of response the participant should give.

Second, the data would have been even more enriching regarding lexical and sublexical processing if individuals had had more divergent levels of language proficiency. Lower proficiency in the language might have shown a lower rate of words recalled in the orthographic neighbor generation task. This hypothetical scenario could have shown a stronger correlation between results and their L-TF. Or perhaps as a less typical wordlikeness rating of pseudowords, less in line with the numerical hierarchy. Unfortunately, the division between high or low language proficiency was operationalized based on the students' academic years at UNCuyo. This criterion did not produce the significantly unequal differences that were sought, evidenced by the lack of a significant correlation between the academic year and participants' vocabulary size. Given the high entrance requirements for English-related degrees at UNCuyo, it is preferable to interpret these results as if all participants formed a homogeneous group with near-native language knowledge. Future studies should consider more diverse

population samples in terms of language proficiency —whether native vs. non-native or university students vs. schoolchildren. By polarizing the proficiency levels, the division based on language proficiency will show greater significance concerning participants' vocabulary size.

Third, and probably the most critical issue of the three, is this study's small number of participants. This fact impacted the analyses of participant characteristics across the three tasks, resulting in mostly non-significant correlations, and for those that were significant, they seemed random. The low sample size also affected the analysis per rhyme in the orthographic neighbor generation task. The four rhymes were insufficient to generate significance and identify whether there were lexical-sublexical influences in retrieving lexemes from a sublexical cue. The sample issue can be addressed by increasing its size —both the number of participants and the number of rhymes—, a solution that could not be implemented due to time and funding constraints in this study.

Interestingly, UNCuyo has been characterized by a strong inclination toward using British-related material in English teaching. However, when comparing the results from the three lexical typicality databases, BNC-CELEX provided the most consistent correlations with the results. This finding suggests that the students have a more mixed body of knowledge than purely British. Additionally, despite Mendoza being surrounded by Spanish-speaking regions, it is not easy to find university students who only possess bilingual Spanish-English knowledge. Based on the participants of the study, most have had multiple interactions with a third or even a fourth foreign language. As human relationships become increasingly globalized through the internet, finding "pure" monolinguals or bilinguals who meet all the exclusion criteria for future studies is becoming more difficult.

Finally, vocabulary size was the only variable obtained in the pre-experimental phase that showed clear patterns of significance in the orthographic neighbor generation task. Unlike LHQ3, this task was performed online, so the data collection did not show any omitted questions or reading comprehension errors. The participants' vocabulary size exhibited a moderate-to-strong correlation with the recalling of words with infrequent rhymes. It is important to remember that all four rhymes selected had a high rhyme type frequency, so the greater or lesser frequency of SL-TF depends entirely on their Σ L-TF. With this in mind, it is reasonable to think that as rhymes become more infrequent —i.e., the lexical typicality of the neighbors containing them decreases— a greater language knowledge will be required to recall them —i.e., larger

vocabulary size.

6.2. *Analysis per-Word*

The conclusions of the orthographic neighbor generation task can be orderly addressed by delving into its most important ideas: 1) C-similarity versus G-similarity and 2) L-TF.

First, the original authors of the task placed significant emphasis on the idea of entropy—the total number of neighbors recalled in this task (Luce & Large, 2001)—as being equivalent to the lexical neighborhood of the stimulus, i.e., its C-similarity. Based on the results provided in this work, it becomes evident that it is not consistently possible to explain entropy in terms of C-similarity. Among the C-similarity metrics, only OLD20 proved significant in most cases—a measure not considered by the original authors—while Coltheart's N showed inadmissible significance in half of the cases. In the instances where it did predict the results, C-similarity mainly correlated with the generation of neighbors with shorter rhymes. This finding is consistent with the literature, as shorter word sizes yield better C-similarity values than longer words.

However, as mentioned in Section 3.2, as word length decreases, C-similarity becomes proportional to G-similarity, making it difficult to determine which similarity type is truly behind the influence over the results. This question becomes especially problematic, as the two chosen shortest rhymes coincidentally have the highest rhyme type frequency among the four alternatives. Given that the analysis per rhyme encompassed only 4 rhymes (a very low N), the results did not show any significant findings, making it impossible to definitively rule out whether the G-similarity with functional constraint influences the results or not. Therefore, this study should be replicated with more rhymes to resolve the uncertainty. Additionally, rhymes with varying G-similarity values should be included rather than homogenizing the rhyme type frequencies.

When proposing this study, we expected that L-TF would be the only significant metric regarding the neighbor generation results. Therefore, the fact that both similarity types also correlated significantly with the task's results is an unexpected finding. Since the stimulus is a rhyme (a sublexical element) and the response consists of words with the same rhyme (a lexical element), it can be hypothesized that G-similarity measures (i.e., lexical-sublexical) are what underlie this influence. Conversely, if this situation were to occur using the original methodology of Luce and Large (2001)—a whole pseudoword as the stimulus and "structurally similar" words as responses—C-similarity measures would be more likely to be dominant.

Second, this study contradicted the original authors and proposed that entropy is more influenced by lexical typicality rather than C-similarity. Based on the results, the hypothesis presented in this work was confirmed, as L-TF was the only metric that showed a significant correlation —ranging from moderate to very strong— across all results. Unlike C-similarity, L-TF correlates better with the results of the neighbor generation task that involves longer rhymes, as if it were compensating for the lack of influence from C-similarity. Furthermore, L-TF increased its predictive power when the rhymes had a lower SL-TF, meaning when Σ L-TF was lower (as the rhyme type frequency was homogenized). This finding makes sense because a lower Σ L-TF implies a higher number of infrequent words, resulting in an L-TF that is more impactful when recalling the words, compared to a situation where all neighborhood words are frequent.

It is important to remember that the calculation of L-TF and Σ L-TF was not fully precise. Since morphological derivatives from the BLP were excluded from the list to rule their influence out of the results, the L-TF values per word did not consider their usage in all their forms (see Section 3.3). If the database were to be modified to unify the L-TF of all morphological derivatives into a single lexical entry, the obtained L-TF would better reflect reality and further increase the predictive power of the variable in the analysis per word.

Finally, few studies address lexical activation from sublexical stimuli, so this methodology could become a valuable tool for future research.

6.3. *Analysis per-Pseudoword*

The conclusions provided by the wordlikeness tasks for pseudowords are the most enriching of this thesis and can be divided into the following topics: 1) similarity metrics; 2) criteria of sublexical constraint; 3) SL-TF, Σ L-TF, and the Sublexical Irrelevancy Hypothesis; and 4) subjectivity in pseudoword weighting.

First, all wordlikeness research on pseudowords had been limited until this study, given that the neighborhood size and the tactic probability were the only metrics employed to predict it. The former is referred to as the quintessential lexical similarity metric, while the latter has been the paradigmatic measure to represent sublexical typicality. This study analyzed the theoretical foundations of similarity and typicality in the psycholinguistic context, redefining the boundaries of these concepts. As a result, two similarity types were distinguished: one dependent on perceptual distance within the same domain (C-similarity) and another dependent on the shared presence of a

factor originating from a different domain (G-similarity).

C-similarity refers to a resemblance between words (lexical similarity), meanings (semantic similarity), or sublexemes (sublexical similarity). Meanwhile, G-similarity refers to the presence of the same sublexeme across different words (lexical-sublexical similarity) or the same meaning across multiple words (lexical-semantic similarity). Based on this, it was determined that the neighborhood size is tantamount to C-similarity, while the tactic probabilities are measures of G-similarity. This conceptual association explains why they do not fully reflect sublexical typicality and cannot consistently predict wordlikeness-related outcomes.

The results of this study confirmed the validity of the new terminological division, as the classic measures of C-similarity and G-similarity were unable to predict any of the results in both wordlikeness tasks. This assertion indicates that these metrics do not reliably represent the sublexical aspect, an inaccuracy that prevented them from showing any correlations with a task requiring sublexical weighting.

Second, the tactic probability metric was not only insignificant regarding the wordlikeness of pseudowords, but it is also not a good measure of lexical-sublexical similarity (G-similarity). Conversely, the alternative measure of G-similarity proposed in Section 3.2, the functional type frequency, did show notable significance and predictive power regarding the results. The key difference between the two metrics lies in the criterion used to select the sequences in the database. The classic tactic probability tallies based on bigrams and trigrams positionally-restricted within the word. This metric has an arbitrary criterion that does not consider the functional division of words, leading to situations where even subtle syllabic nuances between words are overlooked (e.g., the bigram <ng> in *aven-ger* and *strength*). By contrast, the functional type frequency respects and complies with the functional division that facilitates English processing, the onset-rhyme structure. This trait allows for its implementation in studies that address the grapho-phonological inconsistency of the language and offers a new interpretation for studies related to rhyme neighborhoods. Moreover, it aligns with the conclusions of the integration of multiple pattern theory of Treiman and Kessler (2014) while reinforcing the approach taken in this study. In other words, it seems that the sublexicon indeed possesses a sequence level that stores patterns that best fit the characteristics of the language.

Third, since the tactic probability was no longer considered a measure of sublexical typicality, it was necessary to create a new metric to replace it. Thus, the new alternative for sublexical typicality is SL-TF, calculated from the values of G-

similarity —the number of words that share the same sublexeme— and Σ L-TF —the sum of the lexical typicality of all the words considered by G-similarity. Not considering the Σ L-TF entails that only "sublexical typicality within a database" is calculated rather than the occurrence of the pattern in a real context. This metric is more akin to what tactic probability represents and, therefore, to G-similarity.

However, one question remained: Are all rhymes of the language stored within the sublexicon, or only some of them? Based on this, we proposed the Sublexical Irrelevancy Hypothesis, stating that not all sequences deserve to be stored, and this "relevance" depends on the number of words that contain that pattern. In this way, if a sequence appears in only a few words, it will not be stored sublexically, and consequently, its processing will be delegated to the lexical domain. This proposal implied two things: 1) the processing of "relevant" sequences depends on sublexical factors, while the processing of "irrelevant" sequences depends on lexical factors; 2) there exists a G-similarity value beyond which a sequence is either deemed worthy or unworthy of sublexical storage, i.e., a "relevancy threshold."

The results were decisive and aligned with the hypothesis of this work. The variable that best predicted the pseudoword wordlikeness results in both tasks was the numerical hierarchy with a relevancy threshold of 4 ± 1 , which showed very strong and highly significant correlations. Since the classical metrics exhibited weak-to-moderate correlations that did not reach significance, this validates SL-TF as the best measure of sublexical typicality to date. Also, it provides experimental support for the Sublexical Irrelevancy Hypothesis. For this reason, it also contradicts the claims of Bailey and Hahn (2001) that lexical typicality is irrelevant to the perception of wordlikeness since it is indeed relevant —though only when the sequence has an SL-TF of zero. Therefore, unless one is solely working with sublexically-relevant or irrelevant sequences, it is impossible to fully predict wordlikeness judgements with a single metric of typicality.

Initially, we thought that the relevancy threshold was 1 or 2, justifying how the pseudowords were assembled for this study. Therefore, the fact that it is higher than those values is another unexpected finding. Based on the SL-TF/ Σ L-TF measures with a relevancy threshold of 4 ± 1 , a better future design of pseudowords with controlled wordlikeness can be made for psycholinguistic, computational linguistic, or neurolinguistic studies.

Finally, it is worth reiterating that although the wordlikeness judgement and arrangement tasks manipulate the same object of study, the second task might be more reliable than the first. The lack of a comparison framework in the wordlikeness

judgements can lead to a much more impulsive and fluctuating assessment. Conversely, the comparative weighting of the wordlikeness arrangement task forces the participant to analyze each stimulus based on the other pseudowords, which may better restrict the values that cause noise in the statistics.

CONCLUSIONS

This thesis's contribution is important for interpreting future studies regarding similarity, typicality, and the sublexical domain.

Multiple novel findings have emerged from this work. The incomplete nature of lexical representation was explored for the first time, enabling the creation and understanding of the C-similarity concept. The emergence of the sublexical process was examined, along with the changes that occur during literacy, the inner workings of the sublexicon, and its corresponding sublexical representations. The terminological conundrum regarding similarity, typicality, and familiarity was clarified. Such a topic has been overlooked by various researchers, in spite of being factors that explain numerous inconsistencies in the field. For the first time, lexical similarity was distinguished from lexical-sublexical similarity, allowing for the framing and understanding of measures that previously aimed to reflect sublexical typicality. It was demonstrated for the first time that the results of the neighbor generation task are better predicted by lexical typicality than by neighborhood size, contrary to the claims of the original authors. The Sublexical Irrelevancy Hypothesis was proposed for the first time, and a new metric for sublexical typicality was designed: the sublexical token frequency. Building on this last point, a method for assembling pseudowords with controlled wordlikeness was created, a milestone previously attempted with limited success.

The theoretical reconceptualizations of the terms similarity and typicality are transferable to the study of the semantic domain, as their new definitions are justified by studies in visual perception, making them coherent even outside the linguistic realm. With the acceptance of the sublexical representation as a legitimate entity, alongside the ability to calculate its sublexical frequency, the simplified "grapheme-phoneme conversion" view is surpassed. This new perspective allows for a more complex sublexicon model, envisioned as a network. When parameters such as availability, synthesis, and precision are added, it becomes possible to include all this in other fields such as computational linguistics, graph theory, and artificial intelligence. Due to the deepening of the requirements that lead to the extrapolation of a sublexical representation, these explanations also account for the emergence of prototypes that exert the perceptual magnetic effect in Patricia Kuhl's original work. As a caveat, we only delved into the relevancy threshold required to store at the sequence level. It is

still unknown whether the threshold for internalizing sublexical representations at the alphabetic level is the same or higher. In any case, these conclusions align with the statistical-learning perspective (see Section 3.3).

So far, the sublexical domain has often been confused with the morphological domain, even though only the latter is associated with meaning. However, by creating a measure capable of designing pseudowords with controlled wordlikeness, it becomes possible to study the pathways and cortices related to the processing of sequences devoid of semantic activation. For instance, future studies focusing on the sublexical route or procedural memory can use pseudowords made solely of relevant sublexemes. Conversely, if pseudowords are assembled with real but irrelevant sublexemes, the influence of the lexical route and the entire declarative memory cannot be ruled out. Wordlikeness tasks hold diagnostic potential for language-related pathologies, making its application a fertile ground for future studies.

Despite the above, it is important to mention several limitations regarding this study. First, this work focuses on the processing of monosyllables due to simplicity, so it remains unclear how these findings will hold when multiple syllables are introduced. Second, it focused on the visual-orthographic aspect of psycho/neurolinguistic processing, and while there are similarities to the phonological aspect, they are necessarily different. Third, it is yet to be seen how these results vary across different individual profiles. The study should be replicated in native contexts, bilingual contexts with varying proficiency levels, or even within pathological contexts. Fourth, the study was restricted to English, a highly inconsistent language. According to the language-specific sublexical dependence hypothesis (Yoon et al., 2002), the solutions found for the opacity of English spelling are not necessarily transferable to the study of other languages. Fifth, future studies should include more participants to elucidate individual-dependent influences accurately. Finally, this thesis used a suboptimal database to obtain lexical typicality for the word family. Therefore, to improve the process of pseudoword assembly and better predict their wordlikeness, it is essential to have a database that unifies the L-TFs of all morphological derivatives into a single lexical entry.

LIST OF REFERENCES

- Adams, M.J. (1994). *Beginning to read: Thinking and learning about print*. MIT Press.
- Agren T. (2014). Human reconsolidation: a reactivation and update. *Brain Research Bulletin*, 105, 70–82. <https://doi.org/10.1016/j.brainresbull.2013.12.010>
- Alberini, C. M., & Travaglia, A. (2017). Infantile Amnesia: A Critical Period of Learning to Learn and Remember. *The Journal of Neuroscience: the official journal of the Society for Neuroscience*, 37(24), 5783–5795. <https://doi.org/10.1523/JNEUROSCI.0324-17.2017>
- Allen, P. A., Smith, A. F., Lien, M. C., Kaut, K. P., & Canfield, A. (2009). A multistream model of visual word recognition. *Attention, Perception & Psychophysics*, 71(2), 281–296. <https://doi.org/10.3758/APP.71.2.281>
- Almeida, J., Mahon, B. Z., & Caramazza, A. (2010). The role of the dorsal visual processing stream in tool identification. *Psychological Science*, 21(6), 772–778. <https://doi.org/10.1177/0956797610371343>
- Altwater-Mackensen, N., & Mani, N. (2013). Word-form familiarity bootstraps infant speech segmentation. *Developmental Science*, 16(6), 980–990. <https://doi.org/10.1111/desc.12071>
- Altwater- Mackensen, N., & Mani, N. (2013). The impact of mispronunciations on toddler word recognition: Evidence for cascaded activation of semantically related words from mispronunciations of familiar words. *Infancy*, 18(6), 1030–1052. <https://doi.org/10.1111/infa.12022>
- An, S. K., Jang, H., Kim, H. J., Na, D. L., & Yoon, J. H. (2023). Linguistic, visuospatial, and kinematic writing characteristics in cognitively impaired patients with beta-amyloid deposition. *Frontiers in Aging Neuroscience*, 15, 1-13. <https://doi.org/10.3389/fnagi.2023.1217746>

- Andrews, S. (1989). Frequency and neighborhood effects on lexical access: Activation or search? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 15(5), 802–814. <https://doi.org/10.1037/0278-7393.15.5.802>
- Andrews, S. (1992). Frequency and neighborhood effects on lexical access: Lexical similarity or orthographic redundancy? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 18(2), 234–254. <https://doi.org/10.1037/0278-7393.18.2.234>
- Andrews, S. (1996). Lexical retrieval and selection processes: Effects of transposed-letter confusability. *Journal of Memory and Language*, 35(6), 775–800. <https://doi.org/10.1006/jmla.1996.0040>
- Andrews, S. (1997). The effect of orthographic similarity on lexical retrieval: Resolving neighborhood conflicts. *Psychonomic Bulletin & Review*, 4(4), 439–461. <https://doi.org/10.3758/BF03214334>
- Angosto, A., Sánchez, P., Álvarez García, M., Cuevas, I., & León, J. A. (2013). Evidence for Top-Down Processing in Reading Comprehension of Children. *Psicología Educativa*, 19(2), 83-88. [https://doi.org/10.1016/S1135-755X\(13\)70014-9](https://doi.org/10.1016/S1135-755X(13)70014-9)
- Ansteeg, L., Leoné, F., & Dijkstra, T. (2022). Characterizing the semantic and form-based similarity spaces of the mental lexicon by means of the multi-arrangement method. *Frontiers in Psychology*, 13, 1-15. <https://doi.org/10.3389/fpsyg.2022.945094>
- Apel, K. (2011). What is orthographic knowledge?. *Language, Speech, and Hearing Services in Schools*, 42(4), 592–603. [https://doi.org/10.1044/0161-1461\(2011/10-0085\)](https://doi.org/10.1044/0161-1461(2011/10-0085))
- Apel, K., Wolter, J. A., & Masterson, J. J. (2006). Effects of phonotactic and orthotactic probabilities during fast mapping on 5-year-olds' learning to spell. *Developmental Neuropsychology*, 29(1), 21–42. https://doi.org/10.1207/s15326942dn2901_3

- Apel, K., Brimo, D., Wilson-Fowler, E. B., Vorstius, C., & Radach, R. (2013). Children develop initial orthographic knowledge during storybook reading. *Scientific Studies of Reading, 17*(4), 286–302. <https://doi.org/10.1080/10888438.2012.692742>
- Arani, E., Sarfraz, F., & Zonooz, B. (2021). Learning fast, learning slow: A general continual learning method based on complementary learning system. *International Conference on Learning Representations*. <https://doi.org/10.48550/arXiv.2201.12604>
- Aslin, R. N. (1993). Segmentation of fluent speech into words: Learning models and the role of maternal input. In B. de Boysson-Bardies, S. de Schonen, P. W. Jusczyk, P. McNeilage, & J. Morton (Eds.), *Developmental neurocognition: Speech and face processing in the first year of life* (pp. 305–315). Kluwer Academic/Plenum Publishers. https://doi.org/10.1007/978-94-015-8234-6_25
- Aslin, R. N. (2017). Statistical learning: a powerful mechanism that operates by mere exposure. *Wiley interdisciplinary reviews. Cognitive Science, 8*(1-2), 1-12. <https://doi.org/10.1002/wcs.1373>
- Asogwa, D. C., Anigbogu, S. O., Onyesolu, M. O., & Chukwuneke, C. I. (2019). Review of Some Machine Learning Techniques for Big Data Their Uses Strengths and Weaknesses. *International Journal of Trend in Research and Development, 6*(5).
- Baayen, R. H., Piepenbrock, R., van Rijn, H. (1993). The CELEX lexical database [CD-ROM]. *Linguistic Data Consortium*, University of Pennsylvania.
- Baayen, R. H., Piepenbrock, R., & van Rijn, H. (1995). The CELEX lexical database (Release 2) [CD-ROM]. *Linguistic Data Consortium*, University of Pennsylvania
- Bailey, T. M., & Hahn, U. (2001). Determinants of wordlikeness: Phonotactics or lexical neighborhoods? *Journal of Memory and Language, 44*(4), 568–591. <https://doi.org/10.1006/jmla.2000.2756>

- Bakker D. J. (2006). Treatment of developmental dyslexia: a review. *Pediatric Rehabilitation, 9*(1), 3–13. <https://doi.org/10.1080/13638490500065392>
- Balota, D. A., Yap, M. J., Cortese, M. J., Hutchison, K. A., Kessler, B., Loftis, B., Neely, J. H., Nelson, D. L., Simpson, G. B., & Treiman, R. (2007). The English Lexicon project. *Behavior Research Methods, 39*(3), 445–459. <https://doi.org/10.3758/BF03193014>
- Bargh J. A. (2006). What have we been priming all these years? On the development, mechanisms, and ecology of nonconscious social behavior. *European Journal of Social Psychology, 36*(2), 147–168. <https://doi.org/10.1002/ejsp.336>
- Baroni. M., Lenci, A., & Onnis, O. (2007). ISA meets Lara: An incremental word space model for cognitively plausible simulations of semantic learning. In *Proceedings of the Workshop on Cognitive Aspects of Computational Language Acquisition* (pp. 49–56). Association for Computational Linguistics. <https://doi.org/10.3115/1629795.1629802>
- Bartolotti, J., & Marian, V. (2014). Wordlikeness and Novel Word Learning. In *Proceedings of the 36th Annual Meeting of the Cognitive Science Society, CogSci 2014* (pp. 146-151). The Cognitive Science Society.
- Basilakos, A., Smith, K. G., Fillmore, P., Fridriksson, J., & Fedorenko, E. (2018). Functional Characterization of the Human Speech Articulation Network. *Cerebral Cortex, 28*(5), 1816–1830. <https://doi.org/10.1093/cercor/bhx100>
- Beech, J. R., & Mayall, K. A. (2005). The word shape hypothesis re-examined: Evidence for an external feature advantage in visual word recognition. *Journal of Research in Reading, 28*(3), 302–319. <https://doi.org/10.1111/j.1467-9817.2005.00271.x>

- Berg, T. (2014). On the Relationship between Type and Token Frequency. *Journal of Quantitative Linguistics*, 21(3), 199–222. <https://doi.org/10.1080/09296174.2014.911505>
- Berg, T., Zörnig, P., & Lehr, C. (2022). The effects of type and token frequency on word length: A cross-linguistic study. *Glottology. International Journal of Theoretical Linguistics*, 13(2), 173-209. <https://doi.org/10.1515/glott-2022-2007>
- Bermeitinger, C. (2016). Priming. In *Psychology and mental health: Concepts, methodologies, tools, and applications* (pp. 42–88). Information Science Reference/IGI Global. <https://doi.org/10.4018/978-1-5225-0159-6.ch003>
- Besner, D., & Johnston, J. C. (1989). Reading and the mental lexicon: On the uptake of visual information. In W. D. Marslen-Wilson (Ed.), *Lexical representation and process* (pp. 291-316). MIT Press.
- Binder, J. R., Pillay, S. B., Humphries, C. J., Gross, W. L., Graves, W. W., & Book, D. S. (2016). Surface errors without semantic impairment in acquired dyslexia: a voxel-based lesion-symptom mapping study. *Brain: a journal of neurology*, 139(5), 1517–1526. <https://doi.org/10.1093/brain/aww029>
- Bissex, G. I. (1980). *Guys at wrk*. Harvard University Press.
- Bober, N., Kapranov, Y., Kukarina, A., Tron, T., & Nasalevych, T. (2021). British National Corpus in English language teaching of university students. *International Journal of Learning, Teaching and Educational Research*, 20(6), 174–193. <https://doi.org/10.26803/ijlter.20.6.9>
- Bochkarev, V. V., Shevlyakova, A. V., & Solovyev, V. D. (2015). The Average Word Length Dynamics as an Indicator of Cultural Changes in Society. *Social Evolution & History*, 14(2), 153–175. <https://doi.org/10.48550/arXiv.1208.6109>

- Boden, C., & Giaschi, D. (2009). The role of low-spatial frequencies in lexical decision and masked priming. *Brain and Cognition*, 69(3), 580–591. <https://doi.org/10.1016/j.bandc.2008.12.002>
- Bordag, D., Kirschenbaum, A., Rogahn, M., & Tschirner, E. (2017). The role of orthotactic probability in incidental and intentional vocabulary acquisition L1 and L2. *Second Language Research*, 33(2), 147-178. <https://doi.org/10.1177/0267658316665879>
- Borghesani, V., Hinkley, L. B. N., Ranasinghe, K. G., Thompson, M. M. C., Shwe, W., Mizuiri, D., Lauricella, M., Europa, E., Honma, S., Miller, Z., Miller, B., Vossel, K., Henry, M. M. L., Houde, J. F., Gomo-Tempini, M. L., & Nagarajan, S. S. (2020). Taking the sublexical route: brain dynamics of reading in the semantic variant of primary progressive aphasia. *Brain: a Journal of Neurology*, 143(8), 2545–2560. <https://doi.org/10.1093/brain/awaa212>
- Boucher, J., & Anns, S. (2018). Memory, learning and language in autism spectrum disorder. *Autism & Developmental Language Impairments*, 3(2). <https://doi.org/10.1177/2396941517742078>
- Brea-Spahn, M. R., Frisch, S., & Bryant, J. (2020). Wordlikeness and nonword repetition in Spanish-speaking bilingual children. In L. Fangfang, K. Pollock, & R. Gibb (Eds.), *Research in child second language acquisition: Towards an integrated understanding* (pp. 275-302). John Benjamin Publishing. <https://doi.org/10.1075/bpa.10.13bre>
- Brown, A. S., Burrows, C. N., & Croft Caderao, K. (2013). Partial word knowledge in the absence of recall. *Memory & cognition*, 41(7), 967–977. <https://doi.org/10.3758/s13421-013-0321-x>
- Bruner, J. S., & O'Dowd, D. (1958). A note on the informativeness of parts of words. *Language and Speech*, 1(2), 98–101. <https://doi.org/10.1177/002383095800100203>

- Brysbaert, M., & New, B. (2009). Moving beyond Kucera and Francis: A Critical Evaluation of Current Word Frequency Norms and the Introduction of a New and Improved Word Frequency Measure for American English. *Behavior Research Methods*, 41(4), 977–90. <https://doi.org/10.3758/BRM.41.4.977>
- Brysbaert, M., New, B., & Keuleers, E. (2012). Adding part-of-speech information to the SUBTLEX-US word frequencies. *Behavior research methods*, 44(4), 991–997. <https://doi.org/10.3758/s13428-012-0190-4>
- Burt, J. S., & Tate, H. (2002). Does a reading lexicon provide orthographic representations for spelling? *Journal of Memory and Language*, 46(3), 518–543. <https://doi.org/10.1006/jmla.2001.2818>
- Bybee, J. (2010). *Language, usage and cognition*. Cambridge University Press.
- Cairns, P., Shillcock, R., Chater, N., & Levy, J. (1997). Bootstrapping word boundaries: A bottom-up corpus-based approach to speech segmentation. *Cognitive Psychology*, 33(2), 111–153. <https://doi.org/10.1006/cogp.1997.0649>
- Cano, S., Palta, A., Posso, F., & Peñeñory, V. M. (2017). Towards designing a serious game for literacy in children with moderate cognitive disability. *Proceedings of the XVIII International Conference on Human Computer Interaction*, (19), 1-5. <https://doi.org/10.1145/3123818.3123835>
- Carreiras, M., Alvarez, C. J., & de Vega, M. (1993). Syllable frequency and visual word recognition in Spanish. *Journal of Memory and Language*, 32(6), 766–780. <https://doi.org/10.1006/jmla.1993.1038>
- Carrillo, M. S., & Alegría, J. (2014). The development of children’s sensitivity to bigram frequencies when spelling in Spanish, a transparent writing system. *Reading and Writing: An Interdisciplinary Journal*, 27(3), 571–590. <https://doi.org/10.1007/s11145-013-9459-y>

- Casaponsa, A., Thierry, G., & Duñabeitia, J. A. (2019). The Role of Orthotactics in Language Switching: An ERP Investigation Using Masked Language Priming. *Brain Sciences*, *10*(1), 1-21. <https://doi.org/10.3390/brainsci10010022>
- Cassar, M., & Treiman, R. (1997). The beginnings of orthographic knowledge: Children's knowledge of double letters in words. *Journal of Educational Psychology*, *89*(4), 631–644. <https://doi.org/10.1037/0022-0663.89.4.631>
- Castro, N., & Vitevitch, M. S. (2023). Using Network Science and Psycholinguistic Megastudies to Examine the Dimensions of Phonological Similarity. *Language and speech*, *66*(1), 143–174. <https://doi.org/10.1177/00238309221095455>
- Chambers, S. M. (1979). Letter and order information in lexical access. *Journal of Verbal Learning & Verbal Behavior*, *18*(2), 225–241. [https://doi.org/10.1016/S0022-5371\(79\)90136-1](https://doi.org/10.1016/S0022-5371(79)90136-1)
- Chan, K. Y., & Vitevitch, M. S. (2009). The influence of the phonological neighborhood clustering coefficient on spoken word recognition. *Journal of Experimental Psychology: Human Perception and Performance*, *35*(6), 1934–1949. <https://doi.org/10.1037/a0016902>
- Charles-Luce, J., & Luce, P. A. (1995). An examination of similarity neighbourhoods in young children's receptive vocabularies. *Journal of Child Language*, *22*(3), 727–735. <https://doi.org/10.1017/s0305000900010023>
- Chetail, F., & Mathey, S. (2011). Effect of syllabic neighbourhood as a function of syllabic neighbour length. *Psychonomic Bulletin & Review*, *18*(5), 951-957. <https://doi.org/10.3758/s13423-011-0138-1>
- Choi, S. J., & Sim, H. S. (2019). Grapheme-phoneme conversion in Korean word recognition: evidence from event-related potentials. *Communication Science & Disorders*, *24*(3), 647–655. <https://doi.org/10.12963/csd.19644>

- Ciavarelli, A., Contemori, G., Battaglini, L., Barollo, M., & Casco, C. (2021). Dyslexia and the magnocellular-parvocellular coactivation hypothesis. *Vision Research*, 179, 64–74. <https://doi.org/10.1016/j.visres.2020.10.008>
- Clopper, C. G., Turnbull, R., & Steindel Burdin, R. (2018). Assessing predictability effects in connected read speech. *Linguistic Vanguard*, 4(2), 1-13. <https://doi.org/10.1515/lingvan-2017-0044>
- Cloutman L. L. (2013). Interaction between dorsal and ventral processing streams: where, when and how?. *Brain and Language*, 127(2), 251–263. <https://doi.org/10.1016/j.bandl.2012.08.003>
- Cohen-Goldberg, A. M. (2012). Phonological competition within the word: Evidence from the phoneme similarity effect in spoken production. *Journal of Memory and Language*, 67(1), 184-198. <https://doi.org/10.1016/j.jml.2012.03.007>
- Cohn-Sheehy, B. I., Delarazan, A. I., Crivelli-Decker, J. E., Reagh, Z. M., Mundada, N. S., Yonelinas, A. P., Zacks, J. M., & Ranganath, C. (2022). Narratives bridge the divide between distant events in episodic memory. *Memory & cognition*, 50(3), 478–494. <https://doi.org/10.3758/s13421-021-01178-x>
- Coleman, J., & Pierrehumbert, J. (1997). Stochastic phonological grammars and acceptability. In J. S. Coleman (Eds.). *Computational Phonology. Third Meeting of the ACL Special Interest Group in Computational Phonology. Proceedings of a workshop sponsored by the Association for Computational Linguistics* (pp. 49-56). Association for Computational Linguistics.
- Coltheart, M. (2006). Dual route and connectionist models of reading: An overview. *London Review of Education*, 4(1), 5-17. <https://doi.org/10.1080/13603110600574322h>
- Coltheart, M., Davelaar, E., Jonasson, J. T., & Besner, D. (1977). Access to the internal lexicon. *Attention and Performance VI* (1st ed.). Academic Press.

- 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>
- Cortese, M. J., Watson, J. M., Wang, J., & Fugett, A. (2004). Relating distinctive orthographic and phonological processes to episodic memory performance. *Memory & Cognition*, *32*(4), 632–639. <https://doi.org/10.3758/bf03195854>
- Crossfield, E., & Damian, M. F. (2021). The role of valence in word processing: Evidence from lexical decision and emotional Stroop tasks. *Acta Psychologica*, *218*, 1-15. <https://doi.org/10.1016/j.actpsy.2021.103359>
- Dal Ben, R., Souza, D. H., & Hay, J. F. (2021). When statistics collide: The use of transitional and phonotactic probability cues to word boundaries. *Memory & Cognition*, *49*(7), 1300–1310. <https://doi.org/10.3758/s13421-021-01163-4>
- Dankovicova, J., West, P., Coleman, J., & Slater, A. (1998). Phonotactic grammaticality is gradient. *Poster presented at the 6th International Conference on Laboratory Phonology*, University of York.
- Davis, C. J. (1999). *The self-organising lexical acquisition and recognition (SOLAR) model of visual word recognition* (Doctoral dissertation). University of New South Wales.
- Davis C. J. (2005). N-watch: a program for deriving neighborhood size and other psycholinguistic statistics. *Behavior Research Methods*, *37*(1), 65–70. <https://doi.org/10.3758/bf03206399>
- Davis, C. J., & Perea, M. (2005). BuscaPalabras: a program for deriving orthographic and phonological neighborhood statistics and other psycholinguistic indices in Spanish. *Behavior Research Methods*, *37*(4), 665–671. <https://doi.org/10.3758/bf03192738>

- De Manrique, A. M. B., & Signorini, A. (1998). Emergent writing forms in Spanish. *Reading and Writing: An Interdisciplinary Journal*, 10(6), 499–517. <https://doi.org/10.1023/A:1008019206946>
- Diependaele, K., Ziegler, J. C., & Grainger, J. (2010). Fast phonology and the Bimodal Interactive Activation Model. *European Journal of Cognitive Psychology*, 22(5), 764–778. <https://doi.org/10.1080/09541440902834782>
- Difalcis, M., Morello García, F., Abusamra, V., & Ferreres, A. (2021). Types of acquired dyslexia in Spanish-speaking patients with aphasia. *Cognitive Neuropsychology*, 38(4), 283-301. <https://doi.org/10.1080/02687038.2020.1734530>
- Do, Y., & Lai, R. K. Y. (2020). Incorporating tone in the modelling of wordlikeness judgements. *Phonology*, 37(4), 577–615. <https://doi.org/10.1017/S0952675720000287>
- Donnelly, S., & Kidd, E. (2021). Onset Neighborhood Density Slows Lexical Access in High Vocabulary 30-Month Olds. *Cognitive science*, 45(9), 1-19. <https://doi.org/10.1111/cogs.13022>.
- Donnelly, N., Humphreys, G. W., & Riddoch, M. J. (1991). Parallel computation of primitive shape descriptions. *Journal of Experimental Psychology: Human Perception and Performance*, 17(2), 561–570. <https://doi.org/10.1037/0096-1523.17.2.561>
- Duñabeitia, J. A., & Carreiras, M. (2011). The relative position priming effect depends on whether letters are vowels or consonants. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 37(5), 1143–1163. <https://doi.org/10.1037/a0023577>
- Edwards, M., Goodhew, S. C., & Badcock, D. R. (2021). Using perceptual tasks to selectively measure magnocellular and parvocellular performance: Rationale and a

- user's guide. *Psychonomic Bulletin & Review*, 28(4), 1029–1050.
<https://doi.org/10.3758/s13423-020-01874-w>
- Ellis, A. W. (1982). Spelling and writing (and reading and speaking). In A. W. Ellis (Ed.), *Normality and pathology in cognitive functions* (pp. 113-146). Academic Press.
- Facoetti, A., Ruffino, M., Peru, A., Paganoni, P., & Chelazzi, L. (2008). Sluggish engagement and disengagement of non-spatial attention in dyslexic children. *Cortex*, 44(9), 1221–1233. <https://doi.org/10.1016/j.cortex.2007.10.007>
- Facoetti, A., Zorzi, M., Cestnick, L., Lorusso, M. L., Molteni, M., Paganoni, P., Umiltà, C., & Mascetti, G. G. (2006). The relationship between visuo-spatial attention and nonword reading in developmental dyslexia. *Cognitive Neuropsychology*, 23(6), 841–855. <https://doi.org/10.1080/02643290500483090>
- Fadaee, M., Bisazza, A., & Monz, C. (2017). Data Augmentation for Low-Resource Neural Machine Translation. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)* (pp. 567–573). Association for Computational Linguistics. <https://doi.org/10.18653/V1/P17-2090>
- Falkowski, A., Sidoruk, M., Olszewska, J., & Jabłonska, M. (2021). Positive—negative asymmetry in evaluation of natural stimuli: Empirical study in the contrast model of similarity extended to open sets. *The American Journal of Psychology*, 134(1), 1–11. <https://doi.org/10.5406/amerjpsyc.134.1.0001>
- Fang, Y. Q. (2021). A Glimpse at the Study of English Neologism. *Business Prospects*, 2(2), 78-84. <https://doi.org/10.52288/bp.27089851.2021.12.13>
- Ferreiro, E. (2006). La escritura antes de la letra. *CPU-E, Revista de Investigación Educativa*, 3, 1-52. <https://doi.org/10.25009/cpue.v0i3.136>
- Ferreiro, E. & Teberosky, A. (1979). *Los sistemas de escritura en el desarrollo del niño*. Siglo XXI Editores.
- Ferreiro, E., & Teberosky, A. (1982). *Literacy before schooling*. Heinemann.

- Fisher, E. (2019). Orthographic and Phonological Processing in Beginning Readers. *Senior Theses*, 1-29.
- Fischer-Baum, S., McCloskey, M., & Rapp, B. (2010). Representation of letter position in spelling: evidence from acquired dysgraphia. *Cognition*, *115*(3), 466–490. <https://doi.org/10.1016/j.cognition.2010.03.013>
- Forster, K. I., & Shen, D. (1996). No enemies in the neighborhood: absence of inhibitory neighborhood effects in lexical decision and semantic categorization. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*(3), 696–713. <https://doi.org/10.1037//0278-7393.22.3.696>
- Forster, K. I., & Taft, M. (1994). Bodies, antibodies, and neighborhood-density effects in masked form priming. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*(4), 844–863. <https://doi.org/10.1037/0278-7393.20.4.844>
- Frauenfelder, U. H., & Schreuder, R. (1992). Constraining psycholinguistic models of morphological processing and representation: The role of productivity. In G. Booji & J. van Marle (Eds.), *Yearbook of Morphology 1991*. Kluwer.
- Frauenfelder, J. H., Baayen, R. H., & Hellwig, F. M. (1993). Neighborhood Density and Frequency Across Languages and Modalities. *Journal of Memory and Language*, *32*(6), 781-804. <https://doi.org/10.1006/jmla.1993.1039>.
- Frisch, S. A., Large, N. R., & Pisoni, D. B. (2000). Perception of Wordlikeness: Effects of Segment Probability and Length on the Processing of Nonwords. *Journal of Memory and Language*, *42*(4), 481–496. <https://doi.org/10.1006/jmla.1999.2692>
- Frost, R. (1994). Prelexical and postlexical strategies in reading: Evidence from a deep and a shallow orthography. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*(1), 116–129. <https://doi.org/10.1037/0278-7393.20.1.116>

- Fuchs, M. (2018). Antepenultimate stress in Spanish: In defense of syllable weight and grammatically-informed analogy. *Glossa: A Journal of General Linguistics*, 3(80), 1-34. <https://doi.org/10.5334/gjgl.433>
- Gao, C., Shinkareva, S. V., & Desai, R. H. (2023). SCOPE: The South Carolina psycholinguistic metabase. *Behavior Research Methods*, 55(6), 2853–2884. <https://doi.org/10.3758/s13428-022-01934-0>
- Gathercole, S. E., & Martin, A. J. (1996). Interactive processes in phonological memory. In M. A. Conway (Ed.), *Cognitive models of memory*. Psychology Press/MIT Press.
- Gentry, J. R. (1982). An analysis of developmental spelling in GNYS AT WRK. *Reading Teacher*, 36(2), 192–200.
- Gentry, J. R., & Ouellette, G. P. (2019). *Brain Words: How the Science of Reading Informs Teaching* (1st ed.). Routledge. <https://doi.org/10.4324/9781032680743>
- Geudens, A., & Sandra, D. (2003). Beyond implicit phonological knowledge: No support for an onset-rime structure in children's explicit phonological awareness. *Journal of Memory and Language*, 49(2), 157–182. [https://doi.org/10.1016/S0749-596X\(03\)00036-6](https://doi.org/10.1016/S0749-596X(03)00036-6)
- Gianakas, S. & Winn, M. (2016). Exploiting the Ganong effect to probe for phonetic uncertainty resulting from hearing loss. *Journal of the Acoustical Society of America*. 140(4), 3440-3441. <https://doi.org/10.1121/1.4971092>.
- Gil Vallejo, L., Coll-Florit, M., Castellón Masalles, I., & Turmo, J. (2018). Verb similarity: Comparing corpus and psycholinguistic data. *Corpus Linguistics and Linguistic Theory*, 14(2), 275-307. <https://doi.org/10.1515/cllt-2016-0045>
- GNU Project (2015). *GNU PSPP for GNU/Linux* (Version 0.8.5) [Computer Software]. Free Software Foundation.

- Goldstein, R., & Vitevitch, M. S. (2014). The influence of clustering coefficient on word-learning: how groups of similar sounding words facilitate acquisition. *Frontiers in Psychology, 5*, 1-6. <https://doi.org/10.3389/fpsyg.2014.01307>
- Goldstone, R. L. & J. Y. Son. (2005). Similarity. In Holyoak, K. J., & Morrison, R. G., *The Cambridge handbook of thinking and reasoning* (pp. 13–36). Cambridge University Press.
- Gómez, P., Ratcliff, R., & Perea, M. (2008). The overlap model: A model of letter position coding. *Psychological Review, 115*(3), 577–600. <https://doi.org/10.1037/a0012667>
- Goswami, U., & Mead, F. (1992). Onset and rime awareness and analogies in reading. *Reading Research Quarterly, 27*(2), 152–162. <https://doi.org/10.2307/747684>
- Goswami, U., Gombert, J. E., & de Barrera, L. F. (1998). Children's orthographic representations and linguistic transparency: Nonsense word reading in English, French, and Spanish. *Applied Psycholinguistics, 19*(1), 19–52. <https://doi.org/10.1017/S0142716400010560>
- Gottardo, A., Pasquarella, A., Chen, X., & Ramirez, G. (2016). The impact of language on the relationships between phonological awareness and word reading in different orthographies: A test of the psycholinguistic grain size theory in bilinguals. *Applied Psycholinguistics, 37*(5), 1083–1115. <https://doi.org/10.1017/S0142716415000508>
- Graf Estes, K., Gluck, S. C., & Grimm, K. J. (2016). Finding patterns and learning words: Infant phonotactic knowledge is associated with vocabulary size. *Journal of Experimental Child Psychology, 146*, 34–49. <https://doi.org/10.1016/j.jecp.2016.01.012>
- Grainger, J., & Dufau, S. (2012). The front end of visual word recognition. In J. S. Adelman (Eds.), *Visual word recognition: Models and methods, orthography and phonology* (pp. 159–184). Psychology Press.

- Grainger, J., & Jacobs, A. M. (1996). Orthographic processing in visual word recognition: A multiple read-out model. *Psychological Review*, 103(3), 518–565. <https://doi.org/10.1037/0033-295X.103.3.518>
- Grainger, J., & Ziegler, J. C. (2008). Cross-code consistency in a functional architecture for word recognition. In E. L. Grigorenko & A. J. Naples (Eds.), *Single-word reading: Behavioral and biological perspectives* (pp. 129–157). Lawrence Erlbaum Associates Publishers.
- Grainger, J., & Ziegler, J. C. (2011). A dual-route approach to orthographic processing. *Frontiers in Psychology*, 2, 54. <https://doi.org/10.3389/fpsyg.2011.00054>
- Grainger, J., O'Regan, J. K., Jacobs, A. M., & Segui, J. (1989). On the role of competing word units in visual word recognition: the neighborhood frequency effect. *Perception & psychophysics*, 45(3), 189–195. <https://doi.org/10.3758/bf03210696>
- Gross, J., Treiman, R., & Inman, J. (2000). The role of phonology in a letter detection task. *Memory & Cognition*, 28(3), 349–357. <https://doi.org/10.3758/bf03198550>
- Hammond, M. (2004). Gradience, Phonotactics and the Lexicon in English Phonology. *International Journal of English Studies*, 4(2), 1-24. <https://doi.org/10.6018/ijes.4.2.47961>
- Han, M. K., Storkel, H. L., Lee, J., & Cox, C. (2016). The Effects of Phonotactic Probability and Neighborhood Density on Adults' Word Learning in Noisy Conditions. *American Journal of Speech-language Pathology*, 25(4), 547–560. https://doi.org/10.1044/2016_AJSLP-14-0165
- Hartshorne, J. K., & Germine, L. T. (2015). When does cognitive functioning peak? The asynchronous rise and fall of different cognitive abilities across the life span. *Psychological Science*, 26(4), 433–443. <https://doi.org/10.1177/0956797614567339>

- Hartshorne, J. K., Tenenbaum, J. B., & Pinker, S. (2018). A critical period for second language acquisition: Evidence from 2/3 million English speakers. *Cognition*, *177*, 263–277. <https://doi.org/10.1016/j.cognition.2018.04.007>
- Hautala, J., Hawelka, S., & Aro, M. (2021). Dual-stage and dual-deficit? Word recognition processes during text reading across the reading fluency continuum. *Reading and Writing: An Interdisciplinary Journal*, *35*(3), 663–686. <https://doi.org/10.1007/s11145-021-10201-1>
- Havens, L. L., & Foote, W. E. (1963). The effect of competition on visual duration threshold and its independence of stimulus frequency. *Journal of Experimental Psychology*, *65*(1), 6–11. <https://doi.org/10.1037/h0048690>
- Hendrix, P. (2016). Experimental explorations of a discrimination learning approach to language processing (Doctoral dissertation). Eberhard Karls Universität Tübingen. <https://doi.org/10.15496/publikation-9333>
- Hendrix, P., Ramscar, M., & Baayen, H. (2019). NDRA: A single route model of response times in the reading aloud task based on discriminative learning. *PloS One*, *14*(7), 1-63. <https://doi.org/10.1371/journal.pone.0218802>
- Hernández Sampieri, R. & Mendoza, C (2018). *Metodología de la investigación. Las rutas cuantitativa, cualitativa y mixta*. Editorial Mc Graw Hill Education.
- Hernández Sampieri, R., Fernández Collado, C., & Baptista Lucio, P. (2002). *Metodología de la Investigación* (4th ed.). McGraw Hill.
- Hirshorn, E. A., Simcox, T., Durisko, C., Perfetti, C. A., & Fiez, J. A. (2020). Unmasking individual differences in adult reading procedures by disrupting holistic orthographic perception. *PloS One*, *15*(5), 1-17. <https://doi.org/10.1371/journal.pone.0233041>
- Holm, A., and Dodd, B. (1996). The effect of first written language on the acquisition of English literacy. *Cognition*, *59*(2), 119–147. [https://doi.org/10.1016/0010-0277\(95\)00691-5ca](https://doi.org/10.1016/0010-0277(95)00691-5ca)

- Hyman, I. E., Jr, & Loftus, E. F. (1998). Errors in autobiographical memory. *Clinical Psychology Review, 18*(8), 933–947. [https://doi.org/10.1016/s0272-7358\(98\)00041-5](https://doi.org/10.1016/s0272-7358(98)00041-5)
- Iniesta, A., Rossi, E., Bajo, M. T., & Paolieri, D. (2021). The Influence of Cross-Linguistic Similarity and Language Background on Writing to Dictation. *Frontiers in Psychology, 12*, 1-19. <https://doi.org/10.3389/fpsyg.2021.679956>
- Ise, E., Arnoldi, C. J., & Schulte-Körne, G. (2014). Development of orthographic knowledge in German-speaking children: A 2-year longitudinal study. *Journal of Research in Reading, 37*(3), 233–249. <https://doi.org/10.1111/j.1467-9817.2012.01535.x>
- Iverson, P., & Kuhl, P. K. (1994). Tests of the perceptual magnet effect for American English /r/ and /l/. *The Journal of the Acoustical Society of America, 95*(5). <https://doi.org/10.1121/1.408983>
- Jabeen, R., & Shahzad, D. K. (2023). Organization of Words in the Mental Lexicon: A Psycholinguistic Study. *Journal of English Language, Literature and Education, 5*(4), 24-48. <https://doi.org/10.54692/jelle.2023.0504200>
- Jacobs, A. M., Rey, A., Ziegler, J. C., & Grainger, J. (1998). MROM-p: An interactive activation, multiple read-out model of orthographic and phonological processes in visual word recognition. In J. Grainger & A. M. Jacobs (Eds.), *Localist connectionist approaches to human cognition* (pp. 147–188). Erlbaum.
- Janse, E., & Newman, R. S. (2013). Identifying nonwords: effects of lexical neighborhoods, phonotactic probability, and listener characteristics. *Language and Speech, 56*(4), 421–441. <https://doi.org/10.1177/0023830912447914>
- Jiménez, J. E., Hernández-Valle, I., García, E., Ramírez, G., Ortiz, M. R., Rodrigo, M., et al. (2007). Computer speech-based remediation for reading disabilities: The size

- of spelling-to-sound unit in a transparent orthography. *The Spanish Journal of Psychology*, 10(1), 52-67. <https://doi.org/10.1017/S1138741600006314>
- Johnston, F. R. (1999). The Timing and Teaching of Word Families. *The Reading Teacher*, 53(1), 64-75.
- Jones, G., Cabiddu, F., Andrews, M., & Rowland, C. F. (2021). Chunks of phonological knowledge play a significant role in children's word learning and explain effects of neighborhood size, phonotactic probability, word frequency and word length. *Journal of Memory and Language*, 119, 1-13. <https://doi.org/10.1016/j.jml.2021.104232>
- Jusczyk, P. W., Luce, P. A., & Charles-Luce, J. (1994). Infants sensitivity of phonotactic patterns in the native language. *Journal of Memory and Language*, 33(5), 630–645. <https://doi.org/10.1006/jmla.1994.1030>
- Jusczyk, P. W., Friederici, A. D., Wessels, J. M., Svenkerud, V. Y., & Jusczyk, A. M. (1993). Infants' sensitivity to the sound patterns of native language words. *Journal of Memory and Language*, 32(3), 402–420. <https://doi.org/10.1006/jmla.1993.1022>
- Justi, F. R., & Jaeger, A. (2017). Orthographic neighborhood effects in recognition and recall tasks in a transparent orthography. *Journal of experimental psychology. Learning, Memory, and Cognition*, 43(4), 565–578. <https://doi.org/10.1037/xlm0000327>
- Kaestner, E., Wu, X., Friedman, D., Dugan, P., Devinsky, O., Carlson, C., Doyle, W., Thesen, T., & Halgren, E. (2022). The precentral gyrus contributions to the early time-course of grapheme-to-phoneme conversion. *Neurobiology of Language*, 3(1), 18–45. https://doi.org/10.1162/nol_a_00047
- Kahng, J., & Durvasula, K. (2023). Can you judge what you don't hear? Perception as a source of gradient wordlikeness judgements. *Glossa: a Journal of General Linguistics*, 8(1), 1-29. <https://doi.org/10.16995/glossa.9333>

- Kamii, C., Long, R., Manning, M., & Manning, G. (1990). Spelling in kindergarten: A constructivist analysis comparing Spanish-speaking and English-speaking children. *Journal of Research in Childhood Education*, 4(2), 91–97. <https://doi.org/10.1080/02568549009594790>
- Kang, S. H., Balota, D. A., & Yap, M. J. (2009). Pathway control in visual word processing: converging evidence from recognition memory. *Psychonomic Bulletin & Review*, 16(4), 692–698. <https://doi.org/10.3758/PBR.16.4.692>
- Karuz, E. A., Newport, E. L., Aslin, R. N., Starling, S. J., Tivarus, M. E., Bavelier, D. (2013). The neural correlates of statistical learning in a word segmentation task: an fMRI study. *Brain and Language*, 127(1), 46–54. <https://doi.org/10.1016/j.bandl.2012.11.007>
- Kearns, D. M., Hancock, R., Hoefft, F., Pugh, K. R., & Frost, S. J. (2019). The neurobiology of dyslexia. *Teaching Exceptional Children*, 51(3), 175–188. <https://doi.org/10.1177/0040059918820051>
- Keller, C., & Meister, I. G. (2014). Agraphia caused by an infarction in Exner's area. *Journal of Clinical Neuroscience: Official Journal of the Neurosurgical Society of Australasia*, 21(1), 172–173. <https://doi.org/10.1016/j.jocn.2013.01.014>
- Keuleers, E., and Marelli, M. (2020). Resources for Mental Lexicon Research: A Delicate Ecosystem. In V. Pirrelli, I. Plag, & W. U. Dressler (Eds.), *Word Knowledge and Word Usage, A Cross-Disciplinary Guide to the Mental Lexicon* (pp. 167-188), De Gruyter. <https://doi.org/10.1515/9783110440577-005>
- Keuleers, E., Lacey, P., Rastle, K., & Brysbaert, M. (2012). The British Lexicon Project: lexical decision data for 28,730 monosyllabic and disyllabic English words. *Behavior Research Methods*, 44(1), 287–304. <https://doi.org/10.3758/s13428-011-0118-4>

- Kim, Y. S. (2007). Phonological awareness and literacy skills in Korean: An examination of the unique role of body-coda units. *Applied Psycholinguistics*, 28(1), 69–94. <https://doi.org/10.1017/S014271640707004X>
- Kinoshita, S. (1987). Case alternation effect: Two types of word recognition?. *The Quarterly Journal of Experimental Psychology: A Human Experimental Psychology*, 39(4), 701–720. <https://doi.org/10.1080/14640748708401810>
- Kinoshita, S., Robidopux, S., Mills, L., & Norris, D. (2014). Visual similarity effects on masked priming. *Memory & Cognition*, 42(5), 821–833. <https://doi.org/10.3758/s13421-013-0388-4>
- Koriat, A., & Lieblich, I. (1974). What does a person in a “TOT” state know that a person in a “don’t know” state doesn’t know. *Memory & Cognition*, 2(4), 647–655. <https://doi.org/10.3758/BF03198134>
- Krasa, N., & Bell, Z. (2021). Silent word-reading fluency is strongly associated with orthotactic sensitivity among elementary school children. *Journal of Experimental Child Psychology*, 205, 1-23. <https://doi.org/10.1016/j.jecp.2020.105061>
- Kučera, H., & Francis, W. (1967). *Computational analysis of present day American English*. Brown University Press
- Kuhl, P. K. (1994). Learning and representation in speech and language. *Current Opinion in Neurobiology*, 4(6), 812–822. [https://doi.org/10.1016/0959-4388\(94\)90128-7](https://doi.org/10.1016/0959-4388(94)90128-7)
- Kumagai, G., & Kawahara, S. (2018). Stochastic Phonological Knowledge and Word Formation in Japanese. *Gengo Kenkyu: Journal of the Linguistic Society of Japan*, 153, 57-83. https://doi.org/10.11435/GENGO.153.0_57
- Kveraga, K., Boshyan, J., & Bar, M. (2007). Magnocellular projections as the trigger of top-down facilitation in recognition. *The Journal of Neuroscience: the Official Journal*

of the Society for Neuroscience, 27(48), 13232–13240.
<https://doi.org/10.1523/JNEUROSCI.3481-07.2007>

Landauer, T. K., & Streeter, L. A. (1973). Structural differences between common and rare words: Failure of equivalence assumptions for theories of word recognition. *Journal of Verbal Learning & Verbal Behavior*, 12(2), 119–131.
[https://doi.org/10.1016/S0022-5371\(73\)80001-5](https://doi.org/10.1016/S0022-5371(73)80001-5)

Large, N., & Luce, P. (2000). Separable effects of neighborhood density and phonotactic probability on word recognition in speech. *Journal of the Acoustical Society of America*, 107(5), 2857–2858. <https://doi.org/10.1121/1.429254>.

Lee, H., Bellana, B., & Chen, J. (2020). What can narratives tell us about the neural bases of human memory?. *Current Opinion in Behavioral Sciences*, 32, 111–119.
<https://doi.org/10.1016/j.cobeha.2020.02.007>

Lee, B., Martinez, P. M., Midgley, K. J., Holcomb, P. J., & Emmorey, K. (2022). Sensitivity to orthographic vs. phonological constraints on word recognition: An ERP study with deaf and hearing readers. *Neuropsychologia*, 177, 1–21.
<https://doi.org/10.1016/j.neuropsychologia.2022.108420>

Lehtonen, A., & Bryant, P. (2005). Doublet challenge: form comes before function in children's understanding of their orthography. *Developmental Science*, 8(3), 211–217. <https://doi.org/10.1111/j.1467-7687.2005.00409.x>

Leibold, C., & Kempter, R. (2006). Memory capacity for sequences in a recurrent network with biological constraints. *Neural Computation*, 18(4), 904–941.
<https://doi.org/10.1162/089976606775774714>

Lété, B. & Pynte, J. (2003). Word-shape and word-lexical-frequency effects in lexical-decision and naming tasks. *Visual Cognition*, 10(8), 913–948.
<https://doi.org/10.1080/13506280344000112>

- Levenshtein, V.I. (1965). Binary codes capable of correcting deletions, insertions, and reversals. *Soviet physics. Doklady*, 10(8), 707-710.
- Li, C., Wang, M., Kim, S. Y., Bolger, D. J., & Wright, K. (2022). Phonological Preparation in Korean: Phoneme, or Syllable or Another Unit?. *Language and Speech*, 65(2), 337-353. <https://doi.org/10.1177/00238309211020039>
- Lien, M. C., Allen, P. A., & Ruthruff, E. (2019). Multiple routes to word recognition: evidence from event-related potentials. *Psychological Research*, 85(1), 151–180. <https://doi.org/10.1007/s00426-019-01256-5>
- Li, P., Zhang, F., Yu, A., & Zhao, X. (2020). Language History Questionnaire (LHQ3): An enhanced tool for assessing multilingual experience. *Bilingualism: Language and Cognition*, 23(5), 938-944. <https://doi.org/10.1017/S1366728918001153>
- Livingstone, M. S., & Hubel, D. H. (1987). Psychophysical evidence for separate channels for the perception of form, color, movement, and depth. *The Journal of Neuroscience : the official journal of the Society for Neuroscience*, 7(11), 3416–3468. <https://doi.org/10.1523/JNEUROSCI.07-11-03416.1987>
- Longcamp, M., Anton, J. L., Roth, M., & Velay, J. L. (2003). Visual presentation of single letters activates a premotor area involved in writing. *Neuroimage*, 19(4), 1492–1500. [https://doi.org/10.1016/s1053-8119\(03\)00088-0](https://doi.org/10.1016/s1053-8119(03)00088-0)
- Luce, P. A. (1986). *Neighborhoods of words in the mental lexicon* (Doctoral dissertation), Indian University.
- Luce, P. A., & Large, N. R. (2001). Phonotactics, density, and entropy in spoken word recognition. *Language and Cognitive Processes*, 16(5-6), 565–581. <https://doi.org/10.1080/01690960143000137>
- Luce, P. A., & Pisoni, D. B. (1998). Recognizing spoken words: the neighborhood activation model. *Ear and Hearing*, 19(1), 1–36. <https://doi.org/10.1097/00003446-199802000-00001>

- Luce, P. A., Pisoni, D. B., & Goldinger, S.D. (1990). Similarity neighborhoods of spoken words. In G. T. M. Altman (Eds.), *Cognitive models of speech processing: Psycholinguistic and computational perspectives* (pp. 105–121), MIT Press.
- Macchi, L., Casalis, S., & Schelstraete, M. A. (2019). Phonological and orthographic reading routes in French-speaking children with severe developmental language disorder. *Journal of Communication Disorders*, 81, 1-14. <https://doi.org/10.1016/j.jcomdis.2019.05.002>
- Magnuson, J. S., Dixon, J. A., Tanenhaus, M. K., & Aslin, R. N. (2007). The dynamics of lexical competition during spoken word recognition. *Cognitive Science*, 31(1), 133–156. <https://doi.org/10.1080/03640210709336987>
- Malmkær, K. (1991). *The linguistics encyclopedia*. Routledge.
- Mandera, P., Keuleers, E., & Brysbaert, M. (2020). Recognition times for 62 thousand English words: Data from the English Crowdsourcing Project. *Behavior Research Methods*, 52(2), 741–760. <https://doi.org/10.3758/s13428-019-01272-8>
- Mangen, A., & Balsvik, L. (2016). Pen or keyboard in beginning writing instruction? Some perspectives from embodied cognition. *Trends in Neuroscience and Education*, 5(3), 99–106. <https://doi.org/10.1016/j.tine.2016.06.003>
- Marian, V., Bartolotti, J., Chabal, S., & Shook, A. (2012). CLEARPOND: cross-linguistic easy-access resource for phonological and orthographic neighborhood densities. *PloS One*, 7(8). <https://doi.org/10.1371/journal.pone.0043230>
- Martin, George R. R. (2013). *A Dance With Dragons*. Bantam Books.
- Martin, K. I. (2017). The impact of L1 writing system on ESL knowledge of vowel and consonant spellings. *Reading and Writing: An Interdisciplinary Journal*, 30(2), 279–298. <https://doi.org/10.1007/s11145-016-9673-5>

- Martin R. C. (2021). The Critical Role of Semantic Working Memory in Language Comprehension and Production. *Current Directions in Psychological Science*, 30(4), 283–291. <https://doi.org/10.1177/0963721421995178>
- Masri, R. A., Grünert, U., & Martin, P. R. (2020). Analysis of Parvocellular and Magnocellular Visual Pathways in Human Retina. *The Journal of Neuroscience: the Official Journal of the Society for Neuroscience*, 40(42), 8132–8148. <https://doi.org/10.1523/JNEUROSCI.1671-20.2020>
- McClelland, J. L., & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception: I. An account of basic findings. *Psychological Review*, 88(5), 375–407. <https://doi.org/10.1037/0033-295X.88.5.375>
- McClelland, J. L., McNaughton, B. L., & O'Reilly, R. C. (1995). Why there are complementary learning systems in the hippocampus and neocortex: insights from the successes and failures of connectionist models of learning and memory. *Psychological Review*, 102(3), 419–457. <https://doi.org/10.1037/0033-295X.102.3.419>
- Mehler, J., Peña, M., Nespors, M., & Bonatti, L. (2006). The "soul" of language does not use statistics: reflections on vowels and consonants. *Cortex*, 42(6), 846–854. [https://doi.org/10.1016/s0010-9452\(08\)70427-1](https://doi.org/10.1016/s0010-9452(08)70427-1)
- Michon, M., López, V., & Aboitiz, F. (2019). Origin and evolution of human speech: Emergence from a trimodal auditory, visual and vocal network. *Progress in Brain Research*, 250, 345–371. <https://doi.org/10.1016/bs.pbr.2019.01.005>
- Moats, L. C. (2006). How spelling supports reading and why it is more regular and predictable than you may think. *American Educator*, 12–43.
- Mooshammer, C., Tiede, M., Shattuck-Hufnagel, S., & Goldstein, L. (2019). Towards the Quantification of Peggy Babcock: Speech Errors and Their Position within the Word. *Phonetica*, 76(5), 363–396. <https://doi.org/10.1159/000494140>

- Munson, B., Kurtz, B. A., & Windsor, J. (2005). The influence of vocabulary size, phonotactic probability, and wordlikeness on nonword repetitions of children with and without specific language impairment. *Journal of Speech, Language, and Hearing Research, 48*(5), 1033–1047. [https://doi.org/10.1044/1092-4388\(2005/072\)](https://doi.org/10.1044/1092-4388(2005/072))
- Nagao, S., & Kitazawa, H. (2008). Role of the cerebellum in the acquisition and consolidation of motor memory. *Shinkei Kenkyu no Shinpo, 60*(7), 783-790.
- Nassaji, H., & Geva, E. (1999). The contribution of phonological and orthographic processing skills to adult ESL reading: Evidence from native speakers of Farsi. *Applied Psycholinguistics, 20*(2), 241–267. <https://doi.org/10.1017/S0142716499002040>
- Nation, K. (1997). Children's sensitivity to rime unit frequency when spelling words and nonwords. *Reading and Writing: An Interdisciplinary Journal, 9*(5-6), 321–338. <https://doi.org/10.1023/A:1007938810898>
- Nation, I. S. P., & Beglar, D. (2007). A vocabulary size test. *The Language Teacher, 31*(7), 9-13.
- New, B., Araújo, V., & Nazzi, T. (2008). Differential processing of consonants and vowels in lexical access through reading. *Psychological science, 19*(12), 1223–1227. <https://doi.org/10.1111/j.1467-9280.2008.02228.x>
- Nosofsky, R. M. (1986). Attention, similarity, and the identification–categorization relationship. *Journal of Experimental Psychology: General, 115*(1), 39–57. <https://doi.org/10.1037/0096-3445.115.1.39>
- Nuerk, H. C., Rey, A., Graf, R., & Jacobs, A. M. (2000). Phonographic sublexical units in visual word recognition. *Current Psychology Letters: Behaviour, Brain & Cognition, 2*, 25–36. <https://doi.org/10.4000/cpl.241>

- Nusbaum, H. C., Pisoni, D. B., & Davis, C. K. (1984). Sizing up the Hoosier mental lexicon: Measuring the familiarity of 20,000 words. *Research on Speech Perception Progress Report*, 10(10), 357–376.
- Orbán, G., Fiser, J., Aslin, R. N., & Lengyel, M. (2008). Bayesian learning of visual chunks by human observers. *Proceedings of the National Academy of Sciences of the United States of America*, 105(7), 2745–2750. <https://doi.org/10.1073/pnas.0708424105>
- Orban, G. A., van Essen, D., & Vanduffel, W. (2004). Comparative mapping of higher visual areas in monkeys and humans. *Trends in Cognitive Sciences*, 8(7), 315–324 <https://doi.org/10.1016/j.tics.2004.05.009>
- Ouellette, G. (2010). Orthographic learning in learning to spell: the roles of semantics and type of practice. *Journal of Experimental Child Psychology*, 107(1), 50–58. <https://doi.org/10.1016/j.jecp.2010.04.009>
- O'Connor, R. E., & Forster, K. I. (1981). Criterion bias and search sequence bias in word recognition. *Memory & Cognition*, 9(1), 78–92. <https://doi.org/10.3758/BF03196953>
- Paap, K. R., Newsome, S. L., & Noel, R. W. (1984). Word shape's in poor shape for the race to the lexicon. *Journal of Experimental Psychology. Human Perception and Performance*, 10(3), 413–428. <https://doi.org/10.1037//0096-1523.10.3.413>
- Pacton, S., Fayol, M., & Perruchet, P. (2005). Children's implicit learning of graphotactic and morphological regularities. *Child Development*, 76(2), 324–339. <https://doi.org/10.1111/j.1467-8624.2005.00848.x>
- Pacton, S., Sobaco, A., Fayol, M., & Treiman, R. (2013). How does graphotactic knowledge influence children's learning of new spellings?. *Frontiers in Psychology*, 4, 1-10. <https://doi.org/10.3389/fpsyg.2013.00701>

- Parlato-Oliveira, E., Saint-Georges, C., Cohen, D., Pellerin, H., Pereira, I. M., Fouillet, C., Chetouani, M., Dommergues, M., & Viaux-Savelon, S. (2021). "Motherese" Prosody in Fetal-Directed Speech: An Exploratory Study Using Automatic Social Signal Processing. *Frontiers in Psychology*, 12, 1-10. <https://doi.org/10.3389/fpsyg.2021.646170>
- Patterson, K., & Coltheart, V. (1987). Phonological processes in reading: A tutorial review. In M. Coltheart (Ed.), *Attention and performance 12: The psychology of reading* (pp. 421–447). Lawrence Erlbaum Associates, Inc.
- Peereman, R., & Content, A. (1997). Orthographic and phonological neighborhoods in naming: Not all neighbors are equally influential in orthographic space. *Journal of Memory and Language*, 37(3), 382–410. <https://doi.org/10.1006/jmla.1997.2516>
- Perea, M. (1998). Orthographic neighbours are not all equal: Evidence using an identification technique. *Language and Cognitive Processes*, 13(1), 77–90. <https://doi.org/10.1080/016909698386609>
- Perea, M. (2015). Neighborhood effects in visual word recognition and reading. In A. Pollatsek & R. Treiman (Eds.), *The Oxford handbook of reading* (pp. 76–87). Oxford University Press. <https://doi.org/10.1093/oxfmordhb/9780199324576.013.7>
- Perea, M., & Carreiras, M. (1998). Effects of syllable frequency and syllable neighborhood frequency in visual word recognition. *Journal of Experimental Psychology: Human Perception and Performance*, 24(1), 134–144. <https://doi.org/10.1037/0096-1523.24.1.134>
- Perea, M., & Fraga, I. (2006). Transposed-letter and laterality effects in lexical decision. *Brain and Language*, 97(1), 102–109. <https://doi.org/10.1016/j.bandl.2005.08.004>
- Perea, M., & Lupker, S. J. (2004). Can CANISO activate CASINO! Transposed-letter similarity effects with nonadjacent letter positions. *Journal of Memory and Language*, 51(2), 231–246. <https://doi.org/10.1016/j.jml.2004.05.005>

- Perea, M., & Rosa, E. (2002). Does "whole-word shape" play a role in visual word recognition?. *Perception & Psychophysics*, 64(5), 785–794. <https://doi.org/10.3758/bf03194745>
- Perea, M., Acha, J., & Carreiras, M. (2009). Eye movements when reading text messaging (txt msgng). *Quarterly Journal of Experimental Psychology*, 62(8), 1560–1567. <https://doi.org/10.1080/17470210902783653>
- Perea, M., Duñabeitia, J. A., & Carreiras, M. (2008). R34D1NG W0RD5 W1TH NUMB3R5. *Journal of Experimental Psychology. Human Perception and Performance*, 34(1), 237–241. <https://doi.org/10.1037/0096-1523.34.1.237>
- Perea, M., Rosa, E., & Gómez, C. (2005). The frequency effect for pseudowords in the lexical decision task. *Perception & Psychophysics*, 67(2), 301–314. <https://doi.org/10.3758/bf03206493>
- Perea, M., Marcet, A., Baciero, A., & Gómez, P. (2023). Reading about a RELO-VUTION. *Psychological Research*, 87(4), 1306–1321. <https://doi.org/10.1007/s00426-022-01720-9>
- Perry, C., Ziegler, J. C., & Coltheart, M. (2002). A dissociation between orthographic awareness and spelling production. *Applied Psycholinguistics*, 23(1), 43–73. <https://doi.org/10.1017/S0142716402000036>
- Perry, C., Ziegler, J. C., & Zorzi, M. (2010). Beyond single syllables: large-scale modeling of reading aloud with the Connectionist Dual Process (CDP++) model. *Cognitive Psychology*, 61(2), 106–151. <https://doi.org/10.1016/j.cogpsych.2010.04.001>
- Peterson, D. J., & Berryhill, M. E. (2013). The Gestalt principle of similarity benefits visual working memory. *Psychonomic Bulletin & Review*, 20(6), 1282–1289. <https://doi.org/10.3758/s13423-013-0460-x>

- Peterson, R. L., Pennington, B. F., & Olson, R. K. (2013). Subtypes of developmental dyslexia: testing the predictions of the dual-route and connectionist frameworks. *Cognition*, *126*(1), 20–38. <https://doi.org/10.1016/j.cognition.2012.08.007>
- Pierrehumbert, J. B. (1994). Syllable structure and word structure: A study of triconsonantal clusters in English. In P. Keating (Ed.), *Papers in laboratory phonology III: Phonological structure and phonetic form* (pp. 168-188). Cambridge University Press.
- Playfoot, D., Billington, J., & Tree, J. J. (2018). Reading and visual word recognition ability in semantic dementia is not predicted by semantic performance. *Neuropsychologia*, *111*, 292–306. <https://doi.org/10.1016/j.neuropsychologia.2018.02.011>
- Polka, L., & Werker, J. F. (1994). Developmental changes in perception of nonnative vowel contrasts. *Journal of experimental psychology. Human Perception and Performance*, *20*(2), 421–435. <https://doi.org/10.1037//0096-1523.20.2.421>
- Pollo, T. C., Kessler, B., & Treiman, R. (2005). Vowels, syllables, and letter names: differences between young children's spelling in English and Portuguese. *Journal of Experimental Child Psychology*, *92*(2), 161–181. <https://doi.org/10.1016/j.jecp.2005.01.006>
- Pollo, T. C., Treiman, R., & Kessler, B. (2008). *Three perspectives on spelling development* (1st ed.) (pp. 175-189). Psychology Press.
- Pontecorvo, C., & Zucchermaglio, C. (1988). Modes of differentiation in children's writing construction. *European Journal of Psychology of Education*, *3*(4), 371-384. <https://doi.org/10.1007/BF03172661>
- Quijada Van der Berghe, C., Bäumlér, L., & Weiland, V. (2023). La realización de <ll> y <y> en el corpus Fonología del Español Contemporáneo (FEC): Categorización

- acústica del continuum entre oclusión y elisión. *Boletín de Filología*, LVIII(1), 485-527. <https://dx.doi.org/10.4067/S0718-93032023000100485>
- Rayner, K., & Pollatsek, A. (1989). *The psychology of reading*. Prentice Hall.
- Rayner, K., White, S. J., Johnson, R. L., & Liversedge, S. P. (2006). Reading words with jumbled letters: there is a cost. *Psychological Science*, 17(3), 192–193. <https://doi.org/10.1111/j.1467-9280.2006.01684.x>
- Rapp, D. N., & Samuel, A. G. (2002). A reason to rhyme: phonological and semantic influences on lexical access. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 28(3), 564–571. <https://doi.org/10.1037//0278-7393.28.3.564>
- Read, C. (1986). *Children's creative spelling*. Routledge & Kegan Paul.
- Regev, T. I., Kim, H. S., Chen, X., Affourtit, J., Schipper, A. E., Bergen, L., Mahowald, K., & Fedorenko, E. (2024). High-level language brain regions process sublexical regularities. *Cerebral Cortex*, 34(3), 1-18. <https://doi.org/10.1093/cercor/bhae077>
- Reynolds, M., & Besner, D. (2006). Reading aloud is not automatic: processing capacity is required to generate a phonological code from print. *Journal of Experimental Psychology: Human perception and performance*, 32(6), 1303–1323. <https://doi.org/10.1037/0096-1523.32.6.1303>
- Ring, J., Leoné, F., & Dijkstra, T. (2023a). L2 Vocabulary Learning Benefits from Skill-Based Learner Models. *Proceedings of the 15th International Conference on Computer Supported Education*, 1, 322-329. <https://doi.org/10.5220/0011981800003470>
- Ring, J., Leoné, F., & Dijkstra, T. (2023b). Native and foreign language orthotactic probability and neighborhood density in word learning. *The Mental Lexicon*, 18(6). <https://doi.org/10.1075/ml.22006.rin>

- Roux, F. E., Dufor, O., Giussani, C., Wamain, Y., Draper, L., Longcamp, M., & Démonet, J. F. (2009). The graphemic/motor frontal area Exner's area revisited. *Annals of Neurology*, 66(4), 537–545. <https://doi.org/10.1002/ana.21804>
- Sakaguchi, K., Duh, K., & Post, M. (2016). Robust Word Recognition via Semi-Character Recurrent Neural Network. *Proceedings of the AAAI Conference on Artificial Intelligence*, 31, 3281-3287. <https://doi.org/10.1609/aaai.v31i1.10970>
- Santana Marrero, J. (2016). Seseo, ceceo y distinción en el sociolecto alto de la ciudad de Sevilla: Nuevos datos a partir de los materiales de PRESEEA. *Boletín de Filología de La Universidad de Chile*, 51(2), 255–280. <https://doi.org/10.4067/S0718-93032016000200010>
- Schnider, A. (2007). *The Confabulating Mind. How the brain creates reality* (2nd ed.) (pp. 174-184). Oxford University Press.
- Schwartz, B. L., & Metcalfe, J. (2011). Tip-of-the-tongue (TOT) states: retrieval, behavior, and experience. *Memory & Cognition*, 39(5), 737–749. <https://doi.org/10.3758/s13421-010-0066-8>
- Seidenberg, M. S., & McClelland, J. L. (1989). A distributed, developmental model of word recognition and naming. *Psychological Review*, 96(4), 523–568. <https://doi.org/10.1037/0033-295x.96.4.523>
- Seitz, F. I. (2023). Relative Attention Across Features Predicts that Common Features Increase Similarity. *Proceedings of the 21st International Conference on Cognitive Modelling*, 217-222. <https://doi.org/10.31234/osf.io/a3d8s>
- Sendmeier, W. (1987). Auditive Judgements of Word Similarity. *STUF - Language Typology and Universals*, 40(1-6), 538-547. <https://doi.org/10.1524/stuf.1987.40.14.538>

- Shahar-Yames, D., & Share, D. L. (2008). Spelling as a self-teaching mechanism in orthographic learning. *Journal of Research in Reading, 31*(1), 22–39. <https://doi.org/10.1111/j.1467-9817.2007.00359.x>
- Share, D. L., & Blum, P. (2005). Syllable splitting in literate and preliterate Hebrew speakers: onsets and rimes or bodies and codas?. *Journal of Experimental Child Psychology, 92*(2), 182–202. <https://doi.org/10.1016/j.jecp.2005.05.003>
- Shepard, R. N. (1962). The analysis of proximities: Multidimensional scaling with an unknown distance function. Part 1. *Psychometrika, 27*(2), 125–140. <https://doi.org/10.1007/BF02289630>
- Silva, C., & Alves-Martins, M. (2002). Phonological skills and writing of presyllabic children. *Reading Research Quarterly, 37*(4), 466–483. <https://doi.org/10.1598/RRQ.37.4.6>
- Skottun B. C. (2014). The magnocellular system versus the dorsal stream. *Frontiers in Human Neuroscience, 8*, 1-2. <https://doi.org/10.3389/fnhum.2014.00786>
- Skottun B. C. (2015). The need to differentiate the magnocellular system from the dorsal stream in connection with dyslexia. *Brain and Cognition, 95*, 62–66. <https://doi.org/10.1016/j.bandc.2015.01.001>
- Snow, C. E., Burns, M. S., & Griffin, P. (1998). *Preventing reading difficulties in young children*. National Academy Press.
- Soares, A. P., Perea, M., & Comesaña, M. (2014). Tracking the emergence of the consonant bias in visual-word recognition: evidence with developing readers. *PloS One, 9*(2), 1-9. <https://doi.org/10.1371/journal.pone.0088580>
- Sobaco, A., Treiman, R., Peereman, R., Borchardt, G., & Pacton, S. (2015). The influence of graphotactic knowledge on adults' learning of spelling. *Memory & Cognition, 43*(4), 593–604. <https://doi.org/10.3758/s13421-014-0494-y>

- Spencer, M., Kaschak, M. P., Jones, J. L., & Lonigan, C. J. (2015). Statistical Learning is Related to Early Literacy-Related Skills. *Reading and Writing, 28*(4), 467–490. <https://doi.org/10.1007/s11145-014-9533-0>
- Stanback, M. L. (1992). Syllable and rime patterns for teaching reading: Analysis of a frequency-based vocabulary of 17,602 words. *Annals of Dyslexia, 42*(1), 196–221. <https://doi.org/10.1007/BF02654946>
- Stella, M. (2018). Cohort and Rhyme Priming Emerge from the Multiplex Network Structure of the Mental Lexicon. *Complexity, 2018*(1), 1-14. <https://doi.org/10.1155/2018/6438702>
- Stella, G., & Biancardi, A. (1990). Accesso alla lingua scritta e sistema verbale: Una integrazione complessa. *Etá Evolutiva, 35*, 38-49.
- Stine-Morrow, E. A. L., McCall, G. S., Manavbasi, I., Ng, S., Llano, D. A., & Barbey, A. K. (2022). The Effects of Sustained Literacy Engagement on Cognition and Sentence Processing Among Older Adults. *Frontiers in Psychology, 13*, 1-15. <https://doi.org/10.3389/fpsyg.2022.923795>
- Storkel, H. L. (2001). Learning new words: phonotactic probability in language development. *Journal of Speech, Language, and Hearing Research, 44*(6), 1321–1337. [https://doi.org/10.1044/1092-4388\(2001/103\)](https://doi.org/10.1044/1092-4388(2001/103))
- Stokes, S. F. (2010). Neighborhood density and word frequency predict vocabulary size in toddlers. *Journal of Speech, Language, and Hearing Research, 53*(3), 670–683. [https://doi.org/10.1044/1092-4388\(2009/08-0254\)](https://doi.org/10.1044/1092-4388(2009/08-0254))
- Strauß, A., Wu, T., McQueen, J. M., Scharenborg, O., & Hintz, F. (2022). The differential roles of lexical and sublexical processing during spoken-word recognition in clear and in noise. *Cortex, 151*, 70–88. <https://doi.org/10.1016/j.cortex.2022.02.011>

- Suárez, L., Tan, S. H., Yap, M. J., & Goh, W. D. (2011). Observing neighborhood effects without neighbors. *Psychonomic Bulletin & Review*, 18(3), 605–611. <https://doi.org/10.3758/s13423-011-0078-9>
- Sundara, M., Zhou, Z. L., Breiss, C., Katsuda, H., & Steffman, J. (2022). Infants' developing sensitivity to native language phonotactics: A meta-analysis. *Cognition*, 221(3). <https://doi.org/10.1016/j.cognition.2021.104993>
- Takashima, A., Bakker, I., van Hell, J. G., Janzen, G., & McQueen, J. M. (2017). Interaction between episodic and semantic memory networks in the acquisition and consolidation of novel spoken words. *Brain and Language*, 167, 44–60. <https://doi.org/10.1016/j.bandl.2016.05.009>
- Tang, H., Singer, J., Ison, M. J., Pivazyan, G., Romaine, M., Frias, R., Meller, E., Boulin, A., Carroll, J., Perron, V., Dowcett, S., Arellano, M., & Kreiman, G. (2016). Predicting episodic memory formation for movie events. *Scientific Reports*, 6, 1-14. <https://doi.org/10.1038/srep30175>
- Taraban, R., & McClelland, J. L. (1987). Conspiracy effects in word pronunciation. *Journal of Memory and Language*, 26(6), 608–631. [https://doi.org/10.1016/0749-596X\(87\)90105-7](https://doi.org/10.1016/0749-596X(87)90105-7)
- Tootell, R. B. H., & Nasr, S. (2017). Columnar Segregation of Magnocellular and Parvocellular Streams in Human Extrastriate Cortex. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 37(33), 8014–8032. <https://doi.org/10.1523/JNEUROSCI.0690-17.2017>
- Treiman, R. (1992). The role of intrasyllabic units in learning to read and spell. In P. B. Gough, L. Ehri, & R. Treiman (Eds.), *Reading acquisition* (pp. 65-106). Hillsdale, NJ: Erlbaum.

- Treiman, R. (1994). Use of consonant letter names in beginning spelling. *Developmental Psychology, 30*(4), 567–580. <https://doi.org/10.1037/0012-1649.30.4.567>
- Treiman, R., & Broderick, V. (1998). What's in a name: Children's knowledge about the letters in their own names. *Journal of Experimental Child Psychology, 70*(2), 97–116. <https://doi.org/10.1006/jecp.1998.2448>
- Treiman, R., Kessler, B., & Bourassa, D. (2001). Children's own names influence their spelling. *Applied Psycholinguistics, 22*(4), 555–570. <https://doi.org/10.1017/S0142716401004040>
- Treiman, R., & Chafetz, J. (1987). Are there onset- and rime-like units in printed words? In Coltheart, M. (Ed.), *Attention and performance 12: The psychology of reading* (pp. 281–298). Lawrence Erlbaum Associates, Inc.
- Treiman, R., & Kessler, B. (2006). Spelling as statistical learning: Using consonantal context to spell vowels. *Journal of Educational Psychology, 98*(3), 642–652. <https://doi.org/10.1037/0022-0663.98.3.642>
- Treiman, R., & Kessler, B. (2014). *How children learn to write words*. Oxford University Press.
- Treiman, R., & Wolter, S. (2018). Phonological and graphotactic influences on spellers' decisions about consonant doubling. *Memory & Cognition, 46*(4), 614–624. <https://doi.org/10.3758/s13421-018-0793-9>
- Treiman, R., Kessler, B., & Bick, S. (2002). Context sensitivity in the spelling of English vowels. *Journal of Memory and Language, 47*(3), 448–468. [https://doi.org/10.1016/S0749-596X\(02\)00010-4](https://doi.org/10.1016/S0749-596X(02)00010-4)
- Treiman, R., Kessler, B., Knewasser, S., Tincoff, R., & Bowman, M. (1996). English speakers' sensitivity to phonotactic patterns. In M. B. Broe & J. B. Pierrehumbert

(Eds.), *Papers in Laboratory Phonology V: Acquisition and the lexicon* (pp. 269-282). Cambridge University Press.

Treiman, R., Mullennix, J., Bijeljac-Babic, R., & Richmond-Welty, E. D. (1995). The special role of rimes in the description, use, and acquisition of English orthography. *Journal of Experimental Psychology. General*, 124(2), 107–136. <https://doi.org/10.1037//0096-3445.124.2.107>

Tribushinina, E., Niemann, G., & Meuwissen, J. (2023). Explicit cognate instruction facilitates vocabulary learning by foreign language learners with developmental language disorder. *Child Language Teaching and Therapy*, 39(3), 248-265. <https://doi.org/10.1177/02656590231202177>

Tulving, E. (1983). *Elements of Episodic Memory*. Oxford University Press.

Turker, S., & Hartwigsen, G. (2021). Exploring the neurobiology of reading through non-invasive brain stimulation: A review. *Cortex*, 141, 497–521. <https://doi.org/10.1016/j.cortex.2021.05.001>

Tversky, A. (1977). Features of similarity. *Psychological Review*, 84(4), 327–352. <https://doi.org/10.1037/0033-295X.84.4.327>

Ulicheva, A., & Coltheart, M. (2015). How word-beginnings constrain the pronunciations of word-ends in the reading aloud of English: the phenomena of head- and onset-conditioning. *PeerJ*, 3. <https://doi.org/10.7717/peerj.1482>

Ullman M. T. (2001). A neurocognitive perspective on language: the declarative/procedural model. *Nature reviews. Neuroscience*, 2(10), 717–726. <https://doi.org/10.1038/35094573>

Ullman M. T. (2004). Contributions of memory circuits to language: the declarative/procedural model. *Cognition*, 92(1-2), 231–270. <https://doi.org/10.1016/j.cognition.2003.10.008>

- Ullman, M. T., & Lovelett, J. T. (2018). Implications of the declarative/procedural model for improving second language learning: The role of memory enhancement techniques. *Second Language Research*, 34(1), 39-65. <https://doi.org/10.1177/0267658316675195>
- Ullman, M. T., Earle, F. S., Walenski, M., & Janacsek, K. (2020). The Neurocognition of Developmental Disorders of Language. *Annual review of psychology*, 71, 389–417. <https://doi.org/10.1146/annurev-psych-122216-011555>
- Van Assche, E., & Grainger, J. (2006). A study of relative-position priming with superset primes. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, 32(2), 399–415. <https://doi.org/10.1037/0278-7393.32.2.399>
- Van der Wouden, T. (1990). Celex: Building a multifunctional polytheoretical lexical data base. *Proceedings of BudaLex*, 88, 363–373.
- Van Heuven, W.J.B., Mandera, P., Keuleers, E., & Brysbaert, M. (2014). Subtlex-UK: A new and improved word frequency database for British English. *Quarterly Journal of Experimental Psychology*, 67(6), 1176–1190. <https://doi.org/10.1080/17470218.2013.850521>
- Van Polanen, V., & Davare, M. (2015). Interactions between dorsal and ventral streams for controlling skilled grasp. *Neuropsychologia*, 79(Pt B), 186–191. <https://doi.org/10.1016/j.neuropsychologia.2015.07.010>
- Varnhagen, C. K., Boechler, P. M., & Steffler, D. J. (1999). Phonological and orthographic influences on children's vowel spelling. *Scientific Studies of Reading*, 3(4), 363–379. https://doi.org/10.1207/s1532799xssr0304_3
- Venezky, R. L. (1970). *The structure of English orthography*. Mouton. <https://doi.org/10.1515/9783110804478>

- Vergara-Martínez, M., Perea, M., Gómez, P., & Swaab, T. Y. (2013). ERP correlates of letter identity and letter position are modulated by lexical frequency. *Brain and Language*, 125(1), 11–27. <https://doi.org/10.1016/j.bandl.2012.12.009>
- Vernon, S. A., & Ferreiro, E. (1999). Writing development: A neglected variable in the consideration of phonological awareness. *Harvard Educational Review*, 69(4), 395–415. <https://doi.org/10.17763/haer.69.4.p411667586738x0w>
- Vitevitch, M. S., & Luce, P. A. (1998). When words compete: Levels of processing in perception of spoken words. *Psychological Science*, 9(4), 325–329. <https://doi.org/10.1111/1467-9280.00064>
- Vitevitch, M. S., & Luce, P. A. (1999). Probabilistic phonotactics and neighborhood activation in spoken word recognition. *Journal of Memory and Language*, 40(3), 374–408. <https://doi.org/10.1006/jmla.1998.2618>
- Vitevitch, M. S.; Luce, P. A. (2016). Phonological Neighborhood Effects in Spoken Word Perception and Production. *Annual Review of Linguistics*, 2(1), 75–94. <https://doi.org/10.1146/annurev-linguistics-030514-124832>
- Vitevitch, M., Luce, P., Pisoni, D., & Auer, E. (1999). Phonotactics, Neighborhood Activation, and Lexical Access for Spoken Words. *Brain and Language*, 68(1-2), 306-311. <https://doi.org/10.1006/brln.1999.2116>.
- Vitevitch, M. S., Luce, P. A., Charles-Luce, J., & Kemmerer, D. (1997). Phonotactics and syllable stress: implications for the processing of spoken nonsense words. *Language and Speech*, 40(Pt 1), 47–62. <https://doi.org/10.1177/002383099704000103>
- Vyshedskiy, A., Mahapatra, S., & Dunn, R. (2017). Linguistically deprived children: Meta-analysis of published research underlines the importance of early syntactic language use for normal brain development. *Research Ideas and Outcomes*, 3, 1-39. <https://doi.org/10.3897/rio.3.e20696>

- Wang, M., & Cheng, C. (2008). Subsyllabic unit preference in young Chinese children. *Applied Psycholinguistics*, 29(2), 291-314. <https://doi.org/10.1017/S0142716407080137>
- Wertheimer, M. (1950a) Gestalt theory. In Ellis, W. D. (Ed.), *A sourcebook of Gestalt psychology* (pp. 1-11). Humanities Press.
- Wertheimer, M. (1950b). Untersuchungen zur Lehre von der Gestalt: II [Principles of perceptual organization]. In Ellis, W. D. (Ed.), *A sourcebook of Gestalt psychology*. (pp. 71–81). Humanities Press.
- White, S. J., Johnson, R. L., Liversedge, S. P., & Rayner, K. (2008). Eye movements when reading transposed text: the importance of word-beginning letters. *Journal of Experimental Psychology. Human Perception and Performance*, 34(5), 1261–1276. <https://doi.org/10.1037/0096-1523.34.5.1261>
- Wilkins, A., Cleave, R., Grayson, N., & Wilson, L. (2009). Typography for children may be inappropriately designed. *Journal of Research in Reading*, 32(4), 402-412. <https://doi.org/10.1111/j.1467-9817.2009.01402.x>
- Williams, C. C., Perea, M., Pollatsek, A., & Rayner, K. (2006). Previewing the neighborhood: the role of orthographic neighbors as parafoveal previews in reading. *Journal of Experimental Psychology. Human Perception and Performance*, 32(4), 1072–1082. <https://doi.org/10.1037/0096-1523.32.4.1072>
- Wong, A. C., Bukach, C. M., Yuen, C., Yang, L., Leung, S., & Greenspon, E. (2011). Holistic processing of words modulated by reading experience. *PloS One*, 6(6), 1-7. <https://doi.org/10.1371/journal.pone.0020753>
- Wright, D. M., & Ehri, L. (2007). Beginners remember orthography when they learn to read words: The case of double letters. *Applied Psycholinguistics*, 28(1), 115–133. <https://doi.org/10.1017/S0142716406070068>

- Wurm, L. H., & Samuel, A. G. (1997). Lexical inhibition and attentional allocation during speech perception: Evidence from phoneme monitoring. *Journal of Memory and Language*, 36(2), 165–187. <https://doi.org/10.1006/jmla.1996.2482>
- Yalnizyan-Carson, A., & Richards, B. A. (2022). Forgetting Enhances Episodic Control With Structured Memories. *Frontiers in Computational Neuroscience*, 16, 1-16. <https://doi.org/10.3389/fncom.2022.757244>
- Yarkoni, T., Balota, D., & Yap, M. (2008). Moving beyond Coltheart's N: a new measure of orthographic similarity. *Psychonomic Bulletin & Review*, 15(5), 971–979. <https://doi.org/10.3758/PBR.15.5.971>
- Yeari, M. (2017). The role of working memory in inference generation during reading comprehension: Retention, (re)activation, or suppression of verbal information?. *Learning and Individual Differences*, 56, 1-12. <https://doi.org/10.1016/j.lindif.2017.04.002>
- Yoon, H.-K., Bolger, D. J., Kwon, O-S., & Perfetti, C. A. (2002). Subsyllabic units in reading. In L. Verhoeven, C. Elbro, & P. Reitsma (Eds.), *Precursors of functional literacy (Vol. 11)* (pp. 139–163). Benjamins. <https://doi.org/10.1075/swll.11.12yoo>
- Zechmeister, E. B. (1972). Orthographic distinctiveness as a variable in word recognition. *The American Journal of Psychology*, 85(3), 425–430. <https://doi.org/10.2307/1420843>
- Zhang, S., Hudson, A., Ji, X. R., Joshi, R. M., Zamora, J., Gómez-Velázquez, F. R., & González-Garrido, A. A. (2021). Spelling acquisition in Spanish: Using error analyses to examine individual differences in phonological and orthographic processing. *Scientific Studies of Reading*, 25(1), 64–83. <https://doi.org/10.1080/10888438.2020.1754834>

- Ziegler, J. C., & Goswami, U. (2005). Reading acquisition, developmental dyslexia, and skilled reading across languages: a psycholinguistic grain size theory. *Psychological Bulletin*, 131(1), 3–29. <https://doi.org/10.1037/0033-2909.131.1.3>
- Ziegler, J. C., & Perry, C. (1998). No more problems in Coltheart's neighborhood: resolving neighborhood conflicts in the lexical decision task. *Cognition*, 68(2), B53–B62. [https://doi.org/10.1016/s0010-0277\(98\)00047-x](https://doi.org/10.1016/s0010-0277(98)00047-x)
- Ziegler, J. C., Stone, G. O., & Jacobs, A. M. (1997). What is the pronunciation for -ough and the spelling for /u/? A database for computing feedforward and feedback consistency in English. *Behavior Research Methods, Instruments & Computers*, 29(4), 600–618. <https://doi.org/10.3758/BF03210615>

APPENDIX

CUESTIONARIO

Nombre/Apellido de Participante: _____

Fecha de Realización: _____

Nivel de Vocabulario: _____

(Obtener de <https://my.vocabularysize.com/session/evstxx20>)

DECLARACIÓN DE CONSENTIMIENTO INFORMADO

Sr./Sra., de años de edad.

Manifiesto que he hecho las preguntas que me surgieron sobre el proyecto y que he recibido información suficiente sobre el mismo. He sido también informado/a de que mis datos personales serán protegidos. Comprendo que mi participación es totalmente voluntaria, que puedo retirarme del testeo cuando quiera, sin tener que dar explicaciones y sin que esto repercuta de ninguna manera.

Presto libremente mi conformidad para participar en el Proyecto de Investigación titulado “*Influence of Lexical and Sublexical Typicality Knowledge on the Neighbor Generation Task and Wordlikeness Judgements in English as L2*”.

Dicho todo esto, otorgo mi consentimiento para cubrir los objetivos especificados en el proyecto.

Mendoza, a _____ del mes de _____ del
202__

TEST 1

Se le mostrará una secuencia típica de la lengua inglesa. La tarea consiste en escribir todas las palabras monosilábicas que conozca o recuerde que finalicen con dicha secuencia.

Ejemplo

X) AVE

have, gave, cave, pave.

No se admiten palabras que modifiquen la vocal o la consonante final
(p.ej., *heave* o *paved*)

a) OO

_____	_____	_____	_____
_____	_____	_____	_____
_____	_____	_____	_____
_____	_____	_____	_____

b) UT

_____	_____	_____	_____
_____	_____	_____	_____
_____	_____	_____	_____
_____	_____	_____	_____

c) AB

_____	_____	_____	_____
_____	_____	_____	_____
_____	_____	_____	_____
_____	_____	_____	_____

d) OOK

_____	_____	_____	_____
_____	_____	_____	_____
_____	_____	_____	_____
_____	_____	_____	_____

TEST 2

Se le mostrará una *pseudopalabra* (palabra inventada) en cada ejercicio. Deberá evaluar con una escala del 1 al 7 cuán probable es que dichas *pseudopalabras* sean consideradas como inglesas, es decir, juzgar cuán “inglesas” son.

Un juicio de **1** implica que es “**imposible**” que dicha *pseudopalabra* sea inglesa, un juicio de **4** significa que hay una “**probabilidad moderada**”, y un juicio de **7** significa que hay una “**muy alta probabilidad**” de que sea inglesa. Los juicios de **2** y **3** implican que es “**improbable**”, mientras que un juicio de **5** o **6** señala que es “**probable**” que sea considerada inglesa.

a) KNADE

1 2 3 4 5 6 7

b) CLIRTH

1 2 3 4 5 6 7

c) PRATCH

1 2 3 4 5 6 7

d) PLESK

1 2 3 4 5 6 7

e) DRITH

1 2 3 4 5 6 7

f) FLIEL

1 2 3 4 5 6 7

g) TRULF

1 2 3 4 5 6 7

h) TWEAL

1 2 3 4 5 6 7

i) BLYRRH

1 2 3 4 5 6 7

j) GRELVE

1 2 3 4 5 6 7

TEST 3

Los últimos dos ejercicios muestran 5 *pseudopalabras* cada uno. El participante deberá ordenarlas según cuál sea la de mayor probabilidad a menor probabilidad de que sea considerada inglesa.

a) DRULLE - CHAITH - PRELL - QUILST - BRELT

+ “*Inglesa*”

- “*Inglesa*”

b) FLUGE - KNORD - BLAUB - TROOD - CLUDE

+ “*Inglesa*”

- “*Inglesa*”

EN CASO DE HABER VISTO ALGUNA DE LAS PSEUDOPALABRAS AQUÍ PRESENTADAS, POR FAVOR ESCRÍBALA(S) AQUÍ.

