DEPOSIT AND CONSULTATION OF DISSERTATION

One copy of your dissertation will be deposited in ORA (Oxford University Research Archive), where it is intended to be freely available online. In order to facilitate this, you are requested to complete and sign the form below.

Please use block capitals

Surname **DE SOUZA**
First names (in full) **SYDELLE**
Faculty board **EDUCATION**
Degree name and pathway **MSC ALSLA**
Title of dissertation **PROCESSING COLLOCATIONS ON THE L2: SEMANTIC TRANSPARENCY & CONGRUENCY**
Supervisor **DR HAMISH CHALMERS**
Subject keywords *Enter your own keywords or phrases to describe your work. This information helps us describe your work in ORA* **SLA, APPLIED PSYCHOLINGUISTICS, VOCABULARY ACQUISITION, MULTI-WORD UNITS, COLLOCATIONS, PHRASEOLOGICAL FRAMEWORK, L2 ENGLISH, L1 PORTUGUESE**
Research methods used *This information helps us describe your work on SOLO for future students e.g., quantitative, interviews, vocabulary test, systematic review, etc.* **QUANTITATIVE, SEMANTIC ACCEPTABILITY JUDGEMENT TASK, LEXTALE, GENERALIZED LINEAR MIXED-EFFECTS MODELS**

Declaration by the candidate as author of the dissertation

1. I understand that I am the owner of this dissertation and that the copyright rests with me unless I specifically transfer it to another person.
2. I understand that the Department requires that I shall deposit one copy of my dissertation in the Oxford University Research Archive ('ORA') where it shall be freely available online for use in accordance with ORA's Terms and Conditions of Use [https://ora.ox.ac.uk/terms_of_use].
3. I understand that this dissertation should not contain material that can be used to personally identify individuals or specific groups of individuals, and that such material should be removed before this dissertation is deposited in ORA.
4. I agree to be bound by the terms of the ORA Grant of Non-exclusive Licence [www.bodleian.ox.ac.uk/ora/deposit-in-ora/deposit-licence] and I warrant that to the best of my knowledge, making my thesis available on the internet will not infringe copyright or any other rights of any other person or party, nor contain defamatory material.
5. I agree that my dissertation shall be available for download in ORA in accordance with paragraphs 2, 3 and 4 above.

Signed [an electronic signature is sufficient]:
Date: **26TH OCTOBER 2021**



Processing Collocations in the L2:

Semantic Transparency & Congruency

Sydelle de Souza

St. Catherine's College

Department of Education

University of Oxford

Dissertation submitted in part-fulfillment of the requirements for the degree of

MSc in Applied Linguistics and Second Language Acquisition

Submitted to ORA: October 29, 2021

Acknowledgements

Starting this ALSLA degree in the midst of a veritable pandemic (let alone under normal circumstances) was not a trivial task and neither was taking on the challenge of a dissertation that required actual data collection. Now that all is said and done, I finally have the time to look in the rear-view mirror, and I am struck by a deep sense of gratitude for the help and guidance I have received from several remarkable people, both within Oxford and without.

First and foremost, I would like to thank my supervisor, **Dr. Hamish Chalmers**, for all the helpful advice, much needed guidance, and understanding throughout this process. The cool conversations about Radiohead were definitely a plus, but, most importantly, your advice to reach out to people was crucial (Spoiler alert: it worked!). To my ALSLA colleagues (and friends), **Victoria**, **Bruce**, **Bao-linh** and **Molly**, thank you for your feedback on the stimuli selection and experiment building. You guys rock! I am very grateful to **Faidra Faitaki** (now Dr, congratulations!) for listening to my dissertation plans and giving me valuable feedback; **Peter Mitchell** for running the Stats Surgery; and **Cathy Scutt**, our librarian, for always making time for me. Despite the pandemic, the **Department of Education** provided me with everything I needed to excel and then some.

At every stage of this journey, there were many well-established academics who took the time out of their busy schedules to help me. I am inspired by you all. I am grateful to **Dr Alina Villalva** for giving me a start in Psycholinguistics. I shall never forget your influence on the trajectory of my incipient career as a researcher. To **Dr Maria Celeste Rodrigues** for advice on constructing the stimuli sets and for always believing in me. I am fortunate to have your support. My fascination with collocations started with Wolter and Yamashita (2015) and I am grateful to **Dr Brent Wolter** for taking the time to share helpful insights on my ideas and data. Your work on collocations served as inspiration and motivation for this dissertation. To **Dr Sophia Skoufaki** for reaching out to me out of the blue and introducing me to LexTALE, for your methodological insights, and for helping out with data collection. You are a godsend! Your kindness and encouragement will always be remembered. **Dr João Veríssimo**, thank you for inviting me to talk about this project at your lab meetings. The opportunity to discuss my work with other established psycholinguists was key to strengthening my narrative, and your statistical advice and insights were sorely needed and much appreciated. Special thanks to **Dr Indaiá Bassani**

for helping me with the Brazilian side of things. Your expertise and advice was much appreciated.

To my professors **Dr Madalena Colaço**, **Dr Carlos Gouveia**, **Dr Paula Luegi**, and **Dr Gerd Hammer** at the University of Lisbon who unhesitatingly helped me with data collection. I am grateful for your continued support, even though I am no longer your student. Thanks to **Dr Carina Pinto**, **Silvana Abalada** at CLUL, all the various departments and research groups at the Universities of Oxford, Essex, and Edinburgh, and **LabLinc** in Brazil, the link to my experiment reached far and wide. Who would've thought that finding participants would be the least of my worries in this venture!

To **Sybil** and **Inês**, my friends from Lisbon, and **Aarati** and **Gabriella**, my friends from Goa. *Muito obrigada!* Sybil, thank you for all your help with data collection and proofreading. I owe you one. Inês, you are the most fantastic human being I know. Thank you for all the lovely conversations, for your constant belief in me, and for proofreading everything I write. I miss you. Aarati, you're the brightest person I know. Thank you for being there. It was beyond helpful to have someone that "got it". The ALSLA struggles were real, but your advice, encouragement, and feedback really helped me through. I owe you one too! Gab, thank you for encouraging me to apply to Oxford, I would not be here if it weren't for you.

To my friends from Ellensburg, WA—**Dr Charles McGehee** and **Albrecht Bartsch**. Even though neither of you were here at Oxford, my memories of this place will always include you. *Ich bin sehr dankbar für Ihre Unterstützung*. Charlie, your advice was always on point, forthright, and much needed. Al, thank you for coming to all my events and supporting me. There are no words that can describe my gratitude.

To **St Catherine's College**, for providing a friendly, supportive, and engaging environment with lots of opportunities to get involved. Thanks to the Graduate Research Expenses Fund for funding the lucky draw for my participants and the College Library for buying all the collocation dictionaries that I needed to build my stimuli sets. My time at Catz will forever be cherished. For the friends I made at the Catz MCR—**Abelardo**, **Francesco**, **Sam**, and **Karl**, I will be eternally grateful. From jumping on board with me on SCAR to joining me for game nights and other fun stuff, I am lucky to count you as friends. Alejandro, all your comments and feedback on various versions of my dissertation were immensely valuable, but most importantly you got me through my statistical turmoil. What would I have done without you? If you were to ask me, everyone doing quantitative research without a stats background needs a friend like you. *¡Muchas gracias!* Gentile Francesco, thank you for painstakingly picking out the flaws in my logic and for sending me memes. There were some dark times and the memes were important to my well-being. *Grazie*. Sam, thanks for your feedback on the experiment, for your upbeat attitude, and for checking in on me regularly. I wish I could have spent more time with you. *Muito obrigada*. Der Karl, your comments and feedback on my manuscript were invaluable, and that tip about the cref saved me so much time and effort. *Vielen Dank*. My memories of Oxford will always include all of you.

There are two types of people in this world—those who build tools and those who use them. I, unabashedly, belong to the latter and I would like to take this opportunity to thank all the builders of the open source world for providing free and open access to invaluable tools without which this dissertation (and a lot of other research) would not have been feasible or even possible. It is my firm belief that contributions such as these should never go unacknowledged. You are truly the proverbial "*giants*" on whose shoulders the rest of us stand.

I would be remiss if I didn't acknowledge my family. Thank you for always believing in me. I wouldn't have gotten this far without your love and support. And last, but by no means the least, I would like to thank **Win**, my editor and proofreader *extraordinaire*. I'll always remember the fun (at least for me!) times that we had thinking up collocations over coffee, lunch, dinner, and every moment in between to the point of semantic satiation. None of this would be possible if it weren't for you. For putting up with me talking about collocations for months on end. For painstakingly poring through every line of this text. For your unconditional love and support for everything that I set out to do, this Oxford journey being no exception. To you, I am indebted.

To those whom I have forgotten to thank, my sincerest apologies. As is evident from these acknowledgements, it takes a village to complete a dissertation and I have had the good fortune of being part of such a great network of kind, insightful, and intelligent people that helped me along the way.

It goes without saying that all mistakes and omissions are my own.

To all who participated in this study

Abstract

Mastering collocations is one of the most challenging aspects of vocabulary acquisition, especially for adult learners of a second language (L2). Research shows that frequency, congruency, and semantic transparency are important factors that affect L2 collocational processing. Frequency is known to interact with both congruency and semantic transparency. However, congruency and semantic transparency have not been explored simultaneously. This study aims to investigate L2 collocational processing in proficient L2 English speakers (L1 Portuguese) by exploring the role of semantic transparency in conjunction with congruency. Specifically, it investigates whether free combinations, congruent collocations, and incongruent collocations differ in terms of processing by incorporating congruency into Howarth's Continuum Model.

A semantic acceptability judgement task was completed by 329 adults (122 Brazilian Portuguese, 108 European Portuguese, and 99 L1 English speakers). The reaction times and error rates were analysed using Generalized Linear Mixed-effects Models. The results showed that the Brazilian Portuguese speakers processed free combinations faster and more accurately than congruent and incongruent collocations. The European Portuguese speakers processed free combinations faster and more accurately than incongruent collocations. Both groups processed congruent collocations faster and more accurately than incongruent collocations. The L1 English group showed no processing advantages for any of the three conditions. Moreover, the L1 group outperformed the L2 groups only on the incongruent condition.

Although there seems to be an effect of semantic transparency in the results of the L2 groups, it could not be isolated from the effects of frequency and congruency. Future research should consider testing incongruent free combinations along with incongruent collocations to properly isolate these effects. With respect to congruency, this study lends support to the growing body of evidence that L2 speakers incur processing costs for incongruent collocations. This highlights the critical role of the L1 in L2 collocational processing. Based on these findings, recommendations are made for incongruent collocations to be explicitly taught in the L2 classroom.

Keywords: Collocations, Semantic Transparency, Congruency, L1 Portuguese, L2 English, Semantic Acceptability Judgement Task, Generalized Linear Mixed-effects Models

Contents

List of Figures vi

List of Tables viii

Acronyms x

1 Introduction 1

1.1 Collocations & the L2 . . . 1

1.2 English as an L2 . . . 3

1.2.1 English in Brazil & Portugal . . . 3

1.3 Aims & Scope . . . 4

1.4 Structure . . . 4

1.5 A Note on Terminology . . . 5

2 Theoretical Background 6

2.1 Defining Collocations . . . 6

2.1.1 The Phraseological Approach . . . 6

2.1.2 The Frequency-based Approach . . . 9

2.1.3 Some Critiques & Considerations . . . 9

2.2 Collocations in SLA Research . . . 10

2.2.1 Evidence from Applied Linguistics and Language Teaching . . . 11

2.2.2 Evidence from Applied Psycholinguistics . . . 13
Limitations . . . 15
2.3 Summary . . . 17

3 Methods **18**

3.1 Research Design . . . 18
3.2 Participants . . . 20
3.2.1 Recruitment . . . 20
3.2.2 Initial Exclusions . . . 20
3.2.3 Socio-demographics . . . 21
Age . . . 21
Language Pathologies . . . 21
Handedness . . . 22
Vision . . . 22
Education . . . 22
3.2.4 L2 Background . . . 23
L2 English Proficiency . . . 23
English Variety . . . 24
3.3 Materials . . . 25
3.3.1 Sketch Engine . . . 25
TenTen Corpora . . . 25
logDice . . . 26
3.3.2 LexTALE . . . 27
3.3.3 Item Development . . . 28
Free Combinations & Collocations . . . 28

Baseline Items . . . 30
3.4 Data Collection . . . 30
3.4.1 Pilot . . . 30
3.4.2 Procedure . . . 31
3.5 Data Analysis . . . 33
3.5.1 Data Trimming . . . 33
3.5.2 Generalized Linear Mixed-Effect Models (GLMMs) . . . 34
3.5.3 Model Specification . . . 36
Specifying Fixed and Random Factors . . . 36
Specifying Probability Distributions and Link Functions . . . 37
Procedure for Dealing with Convergence Issues . . . 39
Statistical Tools . . . 39
3.5.4 Modeling RT Data . . . 40
3.5.5 Modeling ER Data . . . 40
3.6 Ethics . . . 41

4 Results **42**

4.1 Global Results . . . 42
4.2 Model Results . . . 45
4.3 Post-hoc Analyses . . . 46
4.3.1 RQ 1: Free Combinations versus Congruent Collocations . . . 47
4.3.2 RQ 2: Free Combinations versus Incongruent Collocations . . . 48
4.3.3 RQ 3: Congruent Collocations versus Incongruent Collocations . . . 48
4.3.4 Baseline Items . . . 48
4.3.5 RQ 4: L1 versus L2 . . . 52
4.4 Summary . . . 56

5 Discussion **58**

5.1 Processing Collocations in the L2 . . . 58

5.1.1 Free Combinations versus Congruent Collocations . . . 59

5.1.2 Free Combinations versus Incongruent Collocations . . . 60

5.1.3 Congruent Collocations versus Incongruent Collocations . . . 61

5.1.4 Baselines . . . 62

5.2 Processing Collocations in the L1 . . . 62

5.3 L1 versus L2 . . . 63

5.4 Limitations & Future Work . . . 64

5.5 Implications for Teaching & Learning Collocations . . . 65

6 Conclusion **67**

References **69**

Appendices **86**

A Methods **87**

A.1 Participants . . . 87

A.2 Items . . . 89

A.3 Ethics . . . 95

A.3.1 Approval . . . 95

A.3.2 Recruitment Material . . . 96

A.3.3 Participant Recruitment Emails . . . 97

A.3.4 Participant Information Sheet . . . 99

A.3.5 Sociodemographics Questionnaire . . . 101

A.3.6 L2 Background Questionnaire . . . 101

A.3.7 Debrief Form 103

A.3.8 Task Instructions 104

B Results **105**

B.1 Global Results 105

B.2 GLMM Model Outputs 110

B.3 logRT Model 111

List of Figures

2.1 Howarth's Collocational Continuum Model 7

2.2 Congruent and Incongruent Collocations in Howarth's Continuum 17

3.1 Experiment Flow Diagram 32

3.2 Main Task Screen 33

3.3 Distribution of Raw Reaction Time (ms) 37

3.4 Cullen and Frey Graph 38

4.1 Mean RT (ms) of L1 Groups by Condition 43

4.2 Mean Accuracy of L1 Groups by Condition 43

4.3 Acceptability Judgements of L1 Groups by Condition 44

4.4 Overall Mean Accuracy of Baseline Items 44

4.5 Estimated Marginal Means of Condition by L1 (RT Model) 46

4.6 Estimated Marginal Means of Condition by L1 (ER Model) 46

4.7 Estimated Marginal Means of L1 by Condition (RT Model) 52

4.8 Estimated Marginal Means of L1 by Condition (ER Model) 53

5.1 Howarth's Continuum with Congruency 61

A.1 CUREC Approval (received via email) 95

A.2 Recruitment Poster for Brazilian Participants 96

A.3 Recruitment Poster for Portuguese Participants 96

A.4 Recruitment Poster for UK Participants 97

A.5 Recruitment Poster for Other L1 English Participants 97

A.6 Semantic Acceptability Judgement Task Instructions (Screenshot) 104

A.7 LexTALE Instructions (Screenshot) 104

B.1 Mean Error Rate by Item (Incongruent Collocations) 108

B.2 Mean Error Rate by Item (Congruent Collocations) 108

B.3 Mean Error Rate by Item (Free Combinations) 109

B.4 Estimated Marginal Means of Condition by L1 (logRT Model) 112

B.5 Estimated Marginal Means of L1 by Condition (logRT Model) 112

List of Tables

3.1 Mean Age (in years) by L1 Group 21

3.2 Handedness by L1 Group (Self-reported) 22

3.3 Mean Starting Age of English Learning 23

3.4 Mean LexTALE Scores 24

3.5 Self-reported English Variety Spoken 25

3.6 Model Comparison: Inverse Gaussian versus Gamma Distributions 39

4.1 Mean Reaction Time and Mean Accuracy 43

4.2 Mixed Model ANOVA Results for Reaction Times and Error Rates 45

4.3 Pairwise Contrasts of Condition by L1 (RT Model) 50

4.4 Pairwise Contrasts of Condition by L1 (ER Model) 51

4.5 Pairwise Contrasts of L1 by Condition (RT Model) 54

4.6 Pairwise Contrasts of L1 by Condition (ER Model) 55

4.7 Summary of Comparisons by Condition 57

4.8 Summary of L1 Comparisons 57

A.1 Age: Dunn's Test Results 87

A.2 Education Level (Self-reported) 87

A.3 Place of L2 English Learning (Self-reported) 88

A.4 L2 Proficiency (Self-reported) 88

A.5 Average English Proficiency Across the Four Skills (Self-reported) 88

A.6 ENtenten18 Sub-Corpora by Domain 89

A.7 Tukey Results for logDice scores 89

A.8 Free Combinations 90

A.9 Congruent Collocations 91

A.10 Incongruent Collocations 92

A.11 Baseline Items 93

A.12 Baseline Items (contd.) 94

B.1 Descriptive Statistics: RT by L1 Group 105

B.2 Descriptive Statistics: ER by L1 Group 105

B.3 Descriptive Statistics: RT by L1 Group and Condition 106

B.4 Descriptive Statistics: ER by L1 Group and Condition 107

B.5 GLMM Outputs: RT Model, logRT Model, ER Model 110

B.6 Mixed Model ANOVA Results (logRT Model) 111

B.7 Pairwise Contrasts of Condition by L1 for Log-Transformed Reaction Time 113

B.8 Pairwise Contrasts of Condition by L1 for Log-Transformed Reaction Time 114

Acronyms

H_0 null hypothesis. 18

H_1 alternative hypothesis. 18

ADJ+N Adjective+Noun. 11, 14, 64

ANOVA Analysis of Variance. 13, 19, 21, 33, 35, 39

AoA Age of Acquisition. 14, 61

BNC British National Corpus. 24, 26, 29

BP Brazilian Portuguese. 4, 5, 12, 18, 20, 21, 23–25, 29, 42, 47–49, 53, 56, 58–60, 63, 67, 68

CEFR Common European Framework of Reference for Languages. 4, 20, 23, 24, 28

COCA Corpus of Contemporary American English. 12, 24, 26, 29

EFL English as a Foreign Language. 11–13, 63

EMI English as a Medium of Instruction. 23

EMM estimated marginal means. 40, 46, 47, 52, 67

EN English. 20

ENtenten18 English TenTen Corpus, 2018. 24–27, 30

EP European Portuguese. 4, 5, 18, 20, 21, 23–25, 28, 29, 42, 47–49, 53, 56, 58–60, 63, 67

ER error rate. 5, 13, 14, 20, 30, 36, 39, 40, 42, 45, 47, 48, 56, 62, 65

ESL English as a Second Language. 12, 13

GLMM Generalized Linear Mixed-effects Model. 4, 19, 33–36, 38, 39, 45, 67

L1 native or first language. 1–5, 10–15, 18–22, 29–31, 36, 42, 45, 47, 48, 52, 53, 56, 58–65, 67, 68

L2 second or foreign language. 1–6, 10–16, 18–21, 23, 24, 26–28, 30, 31, 47, 48, 52, 53, 56, 58–68

LDT Lexical Decision Task. 13, 14

LMM Linear Mixed-effects Model. 12–16, 33–36

LRT Likelihood-ratio Test. 39

MI Mutual Information. 9, 26, 27

mRT mean reaction time. 20, 22, 42

OoA Order of Acquisition. 14, 61

PTtenten11 Portuguese TenTen Corpus, 2011. 25, 64

QPT Quick Placement Test. 16, 27, 28

RQ research question. 18, 19, 42, 46–48, 52, 53

RT reaction time. 2, 5, 13–15, 20, 30, 31, 33, 34, 36–38, 40, 42, 45, 47, 48, 56, 60, 62, 65

SAJT semantic acceptability judgement task. 4, 15, 19, 30, 31, 37, 42

SE standard error. 47–49, 52, 53

SLA Second Language Acquisition. 1–4, 6, 11, 14, 16, 26, 27, 35, 65, 68

TOEIC Test of English for International Communication. 16

V+N Verb+Noun. 8, 9, 12–15, 28, 30, 64

VIF Variance Inflation Factor. 40

Chapter 1

Introduction

1.1 Collocations & the L2

From *killing time* and *playing dead* to *running baths* and *making beds*, lexical patterns with semi-transparent meanings are ubiquitous in language use. Often referred to as *collocations*, these semi-transparent lexical items are comprised of one word used in its literal sense, and one other word used figuratively. These word combinations are characterized by an arbitrary constraint on substitutions (Cowie, 1998; Howarth, 1998). For example, one can *raise doubts* or *lift bans*, but not **lift doubts* or **raise bans*. Collocations are always syntagmatically related (Nesselhauf, 2004), and occur together in a certain lexical context more frequently than chance would allow (Sinclair, 1991). Alongside idioms, metaphors, phrasal verbs, etc., collocations constitute a significant subset of formulaic language which accounts for more than half of any text, be it written or spoken (Erman & Warren, 2000). Owing to the predominantly formulaic nature of natural human language, they form an important part of vocabulary knowledge, and are therefore indispensable to language users in both the native or first language (L1) and the second or foreign language (L2) (Durrant & Schmitt, 2009; Gyllstad, 2013; Nation, 2001).

In the fields of Second Language Acquisition (SLA), Applied Linguistics, and Language Teaching, the importance of formulaic language (also referred to as *multi-word units*) has been widely acknowledged (Nattinger & DeCarrico, 1992; Wray, 2002). According to Howarth (1998), knowing and using such sequences has a positive impact on a L2 speaker's fluency. A widely cited paper by Pawley and Syder (1983) claims that proper use of multi-word units allows L2 speakers to sound more "native-like" as it grants them idiomaticity, and in so doing, fluency. However, it has long been acknowledged by researchers and teachers alike that even advanced L2 speakers face many challenges in acquiring, understanding, and using formulaic language, and collocations in particular, in the L2 (Nesselhauf, 2004; Siyanova-Chanturia & Omidian, 2019; Wolter, 2019; Wray, 2002).

The last two decades have seen a growing body of SLA research on formulaic language, a vast amount of which focuses on collocations. Classroom and corpus-based evidence points to three influential factors that affect the acquisition of collocations: (i) frequency, i.e., exposure to the L2 input, (ii) congruency, i.e., availability of a direct word-for-word translation of the L2 collocation in the L1, and most importantly (iii) semantic transparency, i.e., the degree to which the meaning of a word combination can be gleaned by analysing the constituent words. The effects of frequency and congruency are fairly straightforward—findings suggest that increased frequency of exposure to the L2 input facilitates collocational learning (e.g., Durrant & Schmitt, 2009; Webb, Newton, & Chang, 2013), as does availability of the collocation in the L1 (e.g., Peters, 2016; Tsai, 2020). However, evidence suggests that it is the semi-transparent nature of collocations that most hinders acquisition—L2 speakers are often unaware of the alternate meanings and hence tend to overestimate their understanding of these lexical units, while remaining oblivious to the failure in comprehension (Laufer, 2011; Laufer & Waldman, 2011; Martinez & Murphy, 2011).

Similarly, there is a growing interest in understanding L2 collocational acquisition and processing within the field of Applied Psycholinguistics. Over the last decade, a slow, but steady accumulation of psycholinguistic evidence mirrors the aforementioned findings that frequency, congruency, and semantic transparency affect L2 collocational acquisition and processing. Evidence from a series of online[1] experiments carried out by Brent Wolter and his colleagues, Henrik Gyllstad, and Junko Yamashita clearly points to the effects of frequency and congruency (Wolter & Gyllstad, 2011, 2013; Wolter & Yamashita, 2015, 2018). High-frequency collocations show processing advantages over low-frequency ones, and congruent collocations (available in the L1 and L2) show processing advantages over incongruent collocations (not available in the L1). In terms of semantic transparency, Gyllstad and Wolter (2016) found that free combinations, which have completely transparent meanings (e.g., *eat cake*), show a processing advantage in both the L1 and L2 over congruent collocations which have semi-transparent meanings. Furthermore, Yamashita (2018) reported that congruent collocations tend to be more transparent than incongruent collocations, and called for future research "to examine congruency and semantic transparency simultaneously" by extending Gyllstad and Wolter's (2016) work to include incongruent collocations.

To the best of my knowledge, this call has not yet been answered and semantic transparency in general has been largely ignored in the psycholinguistic literature.

[1] Experiments that measure reaction time (RT) as the dependent variable.

1.2 English as an L2

Spoken by more than a billion people (EF Education First, 2020), English has become the common language of communication amongst people from different origins, nationalities, and backgrounds, not only in the social realm, but also in the realms of economy, commerce, science, and technology (Crystal, 2003). It is undeniable that the English language has become a veritable global *lingua franca* (Galloway & Rose, 2015), and provides a competitive advantage to those striving to climb the socio-economic ladder (Cronquist & Fiszbein, 2017; EF Education First, 2020). Advances in digital technologies, easy access to the global market, and emerging opportunities in the gig economy have changed the face of the modern labour market (Graham, Hjorth, & Lehdonvirta, 2017). From research labs to global conglomerates, workplaces now depend on high levels of international and intercultural collaboration, which makes English proficiency a critical skill in terms of employability (EF Education First, 2020).

Against this backdrop, it is evident that the playing field is not level. L1 English speakers have a clear competitive edge. Moreover, the rising tide of economic immigration to English-speaking countries such as the UK, Canada, Australia, and the USA is yet to ebb (Triandafyllidou, 2016), and proficiency in English is critical not only for economic gains, but also for successful social assimilation (Bleakley & Chin, 2010). Navigating the intricacies and complexities of a brave new world is not a trivial task, and L2 speakers face many difficulties when operating in a second language (Pavlenko, 2017). Therefore, understanding how a second language is acquired and processed will not only contribute to SLA theory, but could also inform effective classroom strategies to tackle the issues faced by L2 learners.

1.2.1 English in Brazil & Portugal

In Brazil and Portugal, as in many other countries across the globe, proficiency in English is viewed as the price of entry to the perceived economic, technological, and cultural benefits of our globalized world (British Council, 2015; Finardi, 2017; Santos David, 2016). This attitude can be witnessed by the top-down language policies requiring that English be taught in schools at the primary level (Cronquist & Fiszbein, 2017; Santos David, 2016) and the push for internationalization in higher education in both countries (Coelho & Arau Ribeiro, 2018; Guimarães, Mendes, & Hildeblando Júnior, 2021).

In addition, many Portuguese and Brazilian citizens are immigrating for economic reasons to countries where English is the dominant language (Jouët-Pastré & Braga, 2008; Pereira & Azevedo, 2019), leading to an increased demand for English language courses. The linguistic landscape in the big cities like Rio de Janeiro and Lisbon is littered with ad campaigns by language schools catering to this demand. It comes as no surprise then that there is an increasing number of L2 English speakers whose L1 is Portuguese (EF Education First, 2020), a language

with approximately 232.4 million speakers (Eberhard, Simons, & Fennig, 2021).

In SLA research, the impact of the L1 on L2 acquisition has been widely acknowledged and investigated (Gass, Plonsky, & Behney, 2020; Joseph et al., 2018; McManus, 2019). Yet, despite recurring calls by researchers in the field to diversify (e.g., Siyanova-Chanturia & Spina, 2020; Yamashita, 2018), most SLA research on L2 English seems to focus on a select few L1 backgrounds such as Chinese, Japanese, German, Dutch, and Spanish. In comparison, the body of SLA research on L2 English with L1 Portuguese participants is scarce.

1.3 Aims & Scope

The present study aimed to explore the role of semantic transparency and congruency in L2 collocational processing by extending Gyllstad and Wolter (2016). In order to do so, incongruent collocations in conjunction with congruent collocations and free combinations were tested. To suppress the effects of frequency, only high-frequency items were used.

A quantitative approach to data collection and analysis was employed. The target populations were adult L1 Brazilian Portuguese (BP), and European Portuguese (EP) speakers of L2 English with a Common European Framework of Reference for Languages (CEFR) proficiency level of B2 or above, along with a control group of L1 English speakers. The LexTALE was administered to the L2 speakers as a proxy for general English proficiency. Data was collected by means of a semantic acceptability judgement task (SAJT) and analysed using Generalized Linear Mixed-effects Models (GLMMs). The results suggested that although semantic transparency affected collocational processing in the L2, the effect of congruency was more pronounced. Additionally, in contrast with Gyllstad and Wolter (2016), there were discrepancies in the processing tendencies of the L1 group and L2 groups which could be attributed to frequency.

1.4 Structure

This dissertation is structured as follows. Chapter 2 contains a review of the literature that informed the research questions and the methods. Chapter 3 begins by presenting the research questions that this dissertation seeks to answer and then describes and justify the methods used to address them. Chapter 4 presents the results of the statistical analyses carried out on the data. Chapter 5 discusses these results in detail followed by a discussion of the limitations of this study and recommendations for teaching collocations in the L2 classroom. Finally, chapter 6 summarizes the conclusions drawn from this study.

1.5 A Note on Terminology

The use of certain terms must be addressed beforehand in order to facilitate reading. Due to the stigma associated with the terms *native* and *non-native* when referring to language users (see Jenkins, 2007, 2014; Kachru, 1982), these terms will not be used unless absolutely unavoidable, and the terms *L1* and *L2* will take their place. *L1* and *L2* without any qualifiers should henceforth be understood as referring to L1 English and L2 English, respectively, unless stated otherwise (e.g., L1 Portuguese). References such as *Portuguese speakers* or *Italian university students* should be understood as L1 Portuguese speakers of L2 English or L1 Italian university students learning L2 English, unless specified otherwise. The term *Portuguese* includes both European Portuguese (EP) and Brazilian Portuguese (BP). Contradicting Krashen's (1982) view that *learning* is a conscious activity and *acquisition* is a subconscious one, the two terms will be used interchangeably.

To facilitate the interpretation of results, the terms *speed* and *accuracy* will be frequently used to refer to the dependent variables reaction time (RT) and error rate (ER), respectively. It must be noted that *higher speed* and *higher accuracy* indicates lower RT and ER, while *lower speed* and *lower accuracy* indicate higher RT and ER, respectively. Finally, the term *processing cost* refers to higher speed and lower accuracy, while *processing advantage* refers to lower speed and higher accuracy.

Chapter 2

Theoretical Background

This chapter reviews the literature on collocational acquisition and processing. Section 2.1 describes how the term *collocation* is operationalized in this study based on relevant SLA research. Most importantly, it describes Howarth's Collocational Continuum which forms the main theoretical basis of this study (section 2.1.1). Section 2.2 briefly outlines how collocations fit within usage-based theories of language acquisition and reviews key studies that investigate the roles of frequency, congruency, and semantic transparency in L2 collocational processing. Finally, section 2.3 concludes this chapter with a summary of the key theoretical assumptions underlying this study.

2.1 Defining Collocations

One of the most difficult challenges faced by researchers investigating collocations is the lack of consensus on their definition. Despite vociferous calls (see Siyanova-Chanturia & Omidian, 2019; Wolter, 2019; Wood, 2019) for a uniform definition, the term is used varyingly and often vaguely which makes it difficult to compare and consolidate findings (Nesselhauf, 2004; Wray, 2013). However, traditionally, there have been two main approaches to operationalizing the term—the phraseological approach, and the frequency-based approach (Henriksen, 2013).

2.1.1 The Phraseological Approach

As the name suggests, this approach is rooted in phraseology—a sub-field of linguistics that concerns itself with the study of multi-word units in terms of grammatical structure, meaning, and use (Cowie, 1998). These units are distinguished based on: (i) compositionality, i.e., the degree to which the individual constituents that make

up the combination contribute to its overall meaning, and (ii) lexical fixedness, i.e., whether or not the constituent words can be substituted with a synonym (Cowie, 1994). Based on a continuum of least fixed and most compositional to completely fixed and non-compositional, Howarth (1998) proposed a descriptive, typological model of multi-word units with four main categories—free combinations, restricted collocations (henceforth, *collocations*), figurative idioms, and pure idioms (see figure 2.1).

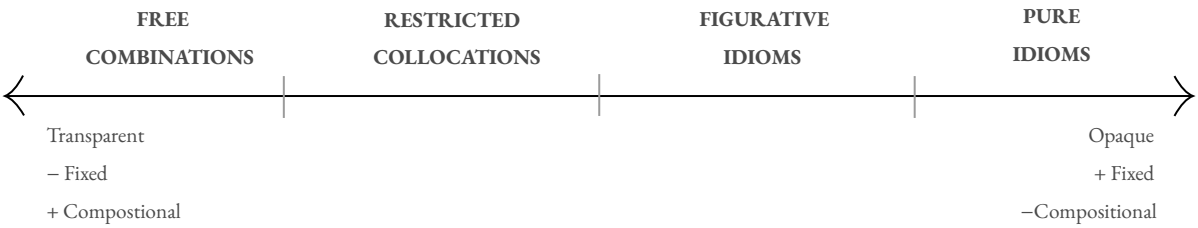


Figure 2.1. Howarth's Collocational Continuum Model

On the transparent end of this continuum lie *free combinations* , wherein both content words are used in their literal sense and substitutions can be made such that neither of their meanings are affected (e.g., blow a trumpet). Then come *collocations* wherein, minimally, one word is figurative, and the other is literal, generating a transparent meaning (e.g., blow a fuse) that has an arbitrary restriction on the substitution of individual words (e.g., *burst a fuse[1]). *Figurative idioms* are those combinations wherein both words can be interpreted entirely figuratively, but also literally (e.g., blow your own trumpet). Finally, on the far end of the continuum lie *pure idioms* that only have non-compositional meanings and allow no substitutions (e.g., blow a gaff[2]).

It must be noted that there are various other descriptive models that have been put forth to classify multi-word units within and without the phraseological tradition (e.g., Cowie, 1994; Erman & Warren, 2000; Mel'čuk, 1998; Nattinger & DeCarrico, 1992). However, only Howarth's model will be considered here[3], as the present study aims to extend Gyllstad and Wolter (2016) which set out to empirically test this model. Therefore, it forms an integral part of the underlying theoretical basis of the present study. Nevertheless, the term "collocation" is still vaguely defined in this model. In order to further constrain this definition, this study invokes Mel'čuk's (1998) definition which resorts to formal linguistics.

According to Mel'čuk (1998), any lexical item X in a given language L can be denoted as follows:

$$X = \langle 'X'; \ /X/; \ S_X \rangle, \tag{2.1}$$

wherein, *'X'* is the meaning of lexical item X, */X/* is the phonetic representation of X, and S_X is the syntactic

[1] The asterisk "*" denotes ungrammaticality.

[2] To reveal a secret (mostly used in British English).

[3] An extensive discussion of other models is beyond the scope of this dissertation, hence the reader is directed to Wray (2002).

data on its co-occurrence with other lexical items.

Let *A* and *B* be lexical items of language *L*, wherein *'A'* is the literal meaning of *A*, *'B'* is the literal meaning of *B*, and *'C'* is the meaning expressed by lexical item *B*, contingent on *A*.

AB is a collocation iff *B* expresses *'C'* contingent on *A*, under one of the following conditions:

1. *'C'* ≠ *'B'*, **and**
    a. *'C'* is empty, such that A selects B to support it in a particular syntactic configuration (e.g., to launch an appeal)

    or

    b. *'C'* is not empty but *B* expresses *'C'* only with *A*, or with a few other similar lexical items (e.g., black coffee)

2. *'C'* = *'B'* , i.e., lexical item *B* has (in the dictionary) the corresponding meaning **and**
    a. *'B'* cannot be expressed with *A* by any possible synonym of *B* (e.g., heavy smoker versus *weighty smoker)

    or

    b. *'B'* includes (an important part of) *'A'*, and is thus bound by *A* (e.g., aquiline nose, the horse neighs)

Therefore, a collocation is denoted as:

$$AB = \langle {}'A \oplus C'; \ /A \oplus B/; \ S_{AB} \rangle. \tag{2.2}$$

In contrast, a free combination which only expresses the literal meanings of *A* and *B* would be denoted as:

$$AB = \langle {}'A \oplus \ B'; \ /A \oplus B/; \ S_{AB} \rangle, \tag{2.3}$$

while a figurative idiom which expresses the meaning *'C'* such that neither the literal meanings of *A*, nor *B* are part of the meaning of *AB* would be expressed as:

$$AB = \langle {}'C'; \ /A \oplus B/; \ S_{AB} \rangle \mid {}'A' \neq \ {}'C' \ \& \ {}'B' \neq \ {}'C'. \tag{2.4}$$

As illustrated, Mel'čuk offers a more precise definition of collocations than Howarth by putting the onus of classification on one constituent, i.e., *B*. In Verb+Noun (V+N) constructions, the differentiating factor between a free combination and a collocation depends on the verb. The noun is always read in the literal sense, and if the

verb expresses its literal meaning, it forms a free combination, but if it expresses a figurative meaning, it forms a collocation.

One final caveat in operationalizing the term collocations adequately for the purposes of this study is that multi-word units, like single words, can have multiple meanings. For example, *blow a fuse*—an example of a collocation provided by Howarth himself—has multiple meanings and fits more than one category on the continuum. If the noun *fuse* is used literally and the verb *blow* figuratively, then it can be rightly classified as a collocation, but *blow a fuse* could also be read as 'to get angry', wherein both words are used figuratively, putting it squarely in the class of figurative idioms. This is problematic when creating items for testing as the alternative meanings could act as confounds. Therefore, the collocations tested in the present study are syntagmatically related V+N items that adhere to condition 1.a of the four conditions laid out by Mel'čuk, with the added constraint that they must have no plausible reading other than that expressed in definition 2.2.

2.1.2 The Frequency-based Approach

The frequency-based approach to defining collocations has its roots in modern Corpus Linguistics (Wray, 2002). This is a quantitative approach that uses statistical measures such as Mutual Information (MI) scores and t-scores to gauge the strength of associations between two words. These measures are used to calculate the probability of two words appearing together in a given body of text.

Collocations are defined as two or more words occurring in close proximity, more frequently than chance would allow (Gries, 2010; Sinclair, 1991), and all multi-word sequences are included under this term (Supasiraprapa, 2018). The semantic nature of the items do not play a role in this approach, i.e., no distinction is made between free combinations, collocations, figurative idioms, and pure idioms.

2.1.3 Some Critiques & Considerations

The phraseological approach is subjective. It relies heavily on semantic analysis without any regard to frequency. Conversely, the frequency-based approach is objective, relying heavily on frequency measures and disregarding the semantic aspects of the word combinations identified. According to Henriksen (2013), the phraseological approach is problematic in that it could leave collocations unidentified, or conversely, it could misidentify word combinations that have appeared together by chance as a collocation. Similarly, the frequency-based approach could also lead to the possible misidentification of collocations (e.g., of the, in the, etc.) while leaving infrequent collocations unidentified (*ibid.*). Additionally, it must be noted that regardless of the differences between the two approaches, the constituent words of a collocation are always syntagmatically related (Nesselhauf, 2004).

The inadequacies of both approaches notwithstanding, Schmitt (2010) suggests that researchers investigating collocations should integrate both approach by identifying probabilistic patterns and analyzing them based on semantic criteria. However, most research on collocations employs the frequency-based definition (*ibid.*). In fact, according to Fioravanti, Senaldi, Lenci, and Siyanova-Chanturia (2021), only two studies investigating L2 collocational acquisition have employed a phraseological framework—Gyllstad and Wolter (2016) and Nesselhauf (2003). Therefore, the present study employs both definitions by first identifying collocations and free combinations according to Mel'čuk's definitions with the help of expert and non-expert native speaker judgements, and then cross-referencing them in a corpus based on statistical scores that indicate the strength of their association, and frequency.

2.2 Collocations in SLA Research

Given the assumption of a strict distinction between the lexicon and syntax, with the lexicon comprising only those items that could not be constructed on the basis of rules (Chomsky, 1969), collocations and other subsets of multi-word units have been largely ignored by the generative approach to linguistics (Wray, 2013). Generativists viewed lexicalized units larger than the word as part of the periphery and not the core, an issue of performance and not competence, and therefore research on collocations languished in the heyday of Chomskian linguistics (Al-Dobaian, 2017).

In contrast, a number of research programs that focus on the socio-pragmatic use of language, i.e., performance, have come to the fore over the last few decades (see Goldberg, 2005; Tomasello, 2001, 2003). Broadly categorized as emergentist, or usage-based theories, these approaches emphasize the role of meaning and use as the basis of language learning and the emergence of language structure (Bybee, 2010; Tomasello, 2009). They postulate that language is acquired through use, and that repeated use of a linguistic unit gives rise to the structure of language itself (Tomasello, 2003). This repetition leads to the entrenchment of that unit in the speakers' memory, allowing for faster access and retrieval (Bybee & Beckner, 2012). Morever, the patterns of language emerge from the learner's cumulative analysis of the distributional characteristics of the linguistic input that they receive, and efficient language processing depends on knowing these patterns (Bybee & Beckner, 2012; Ellis, 2011; Ellis & Ogden, 2017). Simply stated, frequency not only drives language acquisition, but also affords fluency and automaticity in language processing.

Within the usage-based paradigm, language is viewed as a fundamental cognitive ability, such that both adults as well as children are thought to employ the same domain-general cognitive mechanisms when acquiring a language (Evers-Vermeul & Tribushinina, 2017). Therefore, L1 and L2 acquisition are thought to be more similar than different, as they are assumed to employ the same underlying mechanisms. This *Kuhnian* (2012) paradigm

shift from the generative to the usage-based approaches provided a much needed impetus not only for research in SLA, wherein frequency is a well-known factor that affects language learning (Joseph et al., 2018), but also for research on collocations, which are frequently used lexical patterns and therefore, easily fit this view.

The last decade has seen a rise in the number of studies investigating collocational acquisition and processing that have given credence to usage-based theories. Emerging trends suggest that frequency, congruency, and semantic transparency are important factors in L2 collocational acquisition and processing. The following sections will discuss key studies that investigate these factors.

2.2.1 Evidence from Applied Linguistics and Language Teaching

Given that collocations are studied from a usage-based perspective, it is not surprising that a considerable amount of research on collocations exists that has explored the effect of frequency. Durrant and Schmitt (2009) investigated the use of Adjective+Noun (ADJ+N) collocations in written language and found that L2 writers tend to overuse high-frequency collocations while under-using low-frequency ones. In a lab-based study, Durrant and Schmitt (2010) found that repeated encounters resulted in better recall in adult L2 learners, suggesting that they retain probabilistic information about the co-occurrence of words in the input. Another study by Webb et al. (2013) which investigated the effects of repeated exposure on incidental collocational acquisition reported that increased exposure resulted in a positive effect on receptive and productive knowledge. Edmonds and Gudmestad (2021) found that L2 French learners increased their use of low frequency collocations after spending a year abroad in France. Similarly, Siyanova-Chanturia and Spina (2020) reported that advanced learners used more infrequent and opaque collocations than beginners. Although these are just a few examples, the relationship between frequency and collocational acquisition is well-established and acknowledged.

The role of the L1 has long been known to have an effect on L2 acquisition (James, 1980; Lado, 1957), and congruency is another important factor known to affect collocational acquisition. In a classroom-based study, Peters (2016) investigated the influence of congruency on collocational acquisition with forty-one Dutch university students, advanced English as a Foreign Language (EFL) learners, and reported that even in intentional, collocation-focused instructional exercises, incongruent collocations (not available in the L1) were more difficult for learners to acquire than congruent collocations. Similar findings were reported by Nesselhauf (2003). In a learner corpus of thirty-two essays written by German university students, advanced EFL learners, Nesselhauf (2003) reported that almost one quarter of the collocations were judged as erroneous by L1 speakers, and a higher percentage of errors was found in the production of incongruent collocations than congruent ones. Tsai (2020) reported that Chinese EFL learners recalled congruent collocations more accurately than incongruent collocations in a content-base instruction setting. Laufer and Waldman (2011) observed that irrespective of proficiency, L2 learners used fewer collocations than native speakers, and that L1-related errors persisted even

at advanced levels. Interestingly, Leśniewska and Witalisz (2007) reported that advanced Polish EFL learners rejected acceptable congruent collocations in English, possibly due to uncertainty regarding the transferability of their L1 collocational knowledge. Evidently, the L1 can support as well as hinder collocational learning in the L2.

Apart from frequency and congruency, there is also evidence that learners are often unaware of collocations due to their semi-transparent nature. Martinez and Murphy (2011) administered two reading comprehension tasks that differed in terms of whether or not the texts contained multi-word expressions to 101 Brazilian Portuguese (BP) English as a Second Language (ESL) learners. They reported a significant decrease in comprehension for the texts with multi-word units, while the learners' self-reported comprehension scores were higher than their actual scores for the those same texts. The authors concluded that these over-estimations could be attributed to the partially transparent meaning of collocations. Similarly, while investigating the effects of dictionary-use in collocational acquisition with ninety-five Hebrew ESL students, Laufer (2011) observed that although dictionary-use contributed to collocational learning, many learners overestimated their collocational knowledge and reported not having consulted a dictionary. In an exploratory study of fifty-four collocations in the Corpus of Contemporary American English (COCA), Macis and Schmitt (2017) found that a substantial percentage of collocations have both opaque and transparent meanings, and strongly recommend that they be explicitly taught.

Employing the phraseological framework and Howarth's continuum, Fioravanti et al. (2021) investigated L1 Italian speakers' and L2 Italian learners' sensitivity to lexical fixedness and compositionality in free combinations, collocations, and idioms. They administered two tasks to two different groups of participants. In the first task, an acceptability judgement task was administered to twenty L1 and twenty L2 speakers, all university students. Forty free combinations, forty collocations, and forty idioms (all V+N items) were incorporated into 120 sentences. Then another 120 sentences with synonyms of the verbs used in the target set were created. The participants were asked to rate the acceptability of the sentences. The second task employed a multiple choice task wherein the same 120 target stimuli sentences from the first task were used with the verbs removed. Twenty L1 and twenty-two L2) were asked to pick the acceptable verb from the choices available to complete the sentences. The data were analysed using Linear Mixed-effects Models (LMMs). The results showed that L1 speakers were less likely to judge a synonym used in a collocation as acceptable than in a free combination, while the opposite was found for L2 speakers, i.e., they were more open to substituting verbs in a collocation than in a free combination. Both groups perceived idioms as the least flexible. These findings suggest that L2 learners are unaware of the lexical fixedness of collocations, and that idioms are less difficult than collocations for L2 learners to acquire. The authors conclude that their findings support usage-based theories, as well as provide empirical support to Howarth's descriptive categories.

From the above discussion, it is evident that collocations are problematic for L2 learners and that frequency, congruency, and semantic transparency are intervening factors in collocational acquisition. Evidence from psycholinguistic studies mirror these findings.

2.2.2 Evidence from Applied Psycholinguistics

In an eye-tracking study, Siyanova-Chanturia, Conklin, and van Heuven (2011) investigated how binomial phrases (e.g., bride and groom, black and white) that differ in phrasal frequency were processed by twenty-eight L1 and twenty-eight proficient L2 English speakers, all university students. Participants were asked to read sentences containing binomial phrases and their reversed forms (groom and bride), which are syntactically and semantically identical, but differ in terms of frequency—the former being more frequent than the latter. LMMs revealed that both groups were not only sensitive to the phrasal frequency of the binomials, but also to whether the configuration was reversed. They conclude that their results support usage-based theories of language processing, since eye-movements reflect the reader's experience with language.

In order to test the effects of congruency and frequency on L2 collocation learning, Yamashita and Jiang (2010) administered an online phrase-acceptability judgment task to twenty-four ESL and twenty-three EFL learners (L1 Japanese), and twenty L1 speakers using congruent collocations, incongruent collocations, and fillers. RTs and ERs were analysed using an Analysis of Variance (ANOVA). Their results indicated that while the ESL group performed better than the EFL group (faster RTs and lower ERs), both groups were less accurate in the incongruent condition than the congruent one. Yamashita and Jiang concluded that greater L2 exposure increases the potential for collocational acquisition, and that incongruent collocations are harder to learn even for ESL speakers who had increased exposure to English. They further suggest that frequency (L2 exposure) and congruency interactively affect collocational acquisition.

In light of Yamashita and Jiang's findings, Wolter and Gyllstad (2011) used a primed Lexical Decision Task (LDT) and a test of receptive collocational knowledge to further explore the effect of congruency on L2 collocational processing. Thirty-one L1 Swedish advanced (self-reported) learners of English, and thirty-seven L1 English speakers were presented with V+N collocations (congruent, incongruent, baselines, and fillers). The LMM results showed that the L2 group processed and recognized congruent collocations more effectively than the incongruent ones. No congruency effects were observed for L1 speakers. The authors conclude that the L1 exerts considerable influence on the development of collocational knowledge in the L2, thus corroborating Yamashita and Jiang (2010).

Wolter and Gyllstad (2013) then investigated the interaction of frequency and congruency in L2 collocational processing with twenty-five advanced learners of English (L1 Swedish), and twenty-seven L1 English speakers.

They used an acceptability judgement task to present high to medium frequency ADJ+N incongruent, congruent, and non-collocational items, and LMMs to analyse RTs and ERs. Their results indicated a significant effect of frequency for both groups, suggesting that L1 and L2 speakers use similar mechanisms for collocational learning and processing, and thus providing support for usage-based theories. The L1 group responded significantly faster to incongruent collocations than the L2 group. However, they found no significant differences between the two groups for congruent and non-collocational items, indicating that incongruent collocations were problematic for L2 speakers. Interestingly, they also observed significantly faster RTs for high-frequency incongruent collocations than for low-frequency congruent collocations, indicating that frequency trumps congruency. Based on these findings, Wolter and Gyllstad posit that the processing advantage for congruent collocations could be explained by automatic L2 to L1 lexical priming, i.e., the L1 provides "ready-made" collocations that are inserted into the L2 lexicon.

To test this explanation, Wolter and Yamashita (2015) set out to investigate if collocational patterns acceptable in the L1 but not the L2 would still be activated when processing language entirely in the L2. Their sample consisted of twenty-five advanced and twenty-five intermediate L2 English speakers (L1 Japanese), along with twenty-seven L1 speakers. They used a double LDT to present ADJ+N and V+N items in three conditions: (i) word-for-word infelicitous English translations of Japanese collocations (e.g., *far eye), (ii) incongruent collocations, and (iii) baseline items. Regardless of proficiency, they found no priming effect for the translated infelicitous collocations, which indicated no support for the explanation put forth by Wolter and Gyllstad (2013). They also found no processing advantages for the incongruent collocations over baselines. Based on these results, they hypothesized that Age of Acquisition (AoA)/Order of Acquisition (OoA) effects (c.f. Carroll & White, 1973) could perhaps provide a better explanation for the processing advantages offered by congruency.

In a follow up study, Wolter and Yamashita (2018) set out to empirically test if the AoA/OoA accounts could be better suited to explain the congruency effect. They tested the effects of frequency, congruency, and proficiency on twenty-four advanced and twenty-three intermediate learners of English (L1 Japanese), and twenty-seven L1 speakers. They used an online acceptability judgment task with the same conditions as Wolter and Yamashita (2015) with the addition of a congruent condition. The RT and ER data were analysed using LMMs. Results showed that L2 speakers processed congruent collocations significantly faster than incongruent collocations, and that all three groups were sensitive to both word and phrasal frequency. There was no significant difference between L1 and advanced L2 speakers on the congruent condition. Once again, no processing advantages were found for the Japanese-only collocations (infelicitous in English). As for proficiency, they found similarities between the intermediate and advanced groups in both congruent and incongruent items. They concluded that AoA/OoA did seem to provide the best explanation for the congruency effect, but called for more research on the topic. The psycholinguistic evidence reviewed up until this point clearly lends support to the well-established view in SLA and Language Teaching that frequency and congruency play an important role L2

collocational processing. However, there is a dearth of psycholinguistic evidence with respect to the effect of semantic transparency on L2.

To test the effect of semantic transparency in conjunction with frequency, Gyllstad and Wolter (2016) administered a visual SAJT with thirty-eight L1 English speakers and twenty-seven advanced learners of English (L1 Swedish), all university students. The stimuli set consisted of V+N items in two conditions—free combinations (e.g., write a letter) and collocations (e.g., draw a conclusion), based on Howarth's Continnum (see figure 2.1). To avoid the interference of the congruency effect, only congruent collocations were used. The LMM analysis showed that overall there were no differences between the L1 and L2 groups—collocations were processed slower, and with less accuracy than free combinations. Unsurprisingly, they also observed that phrasal frequency predicted RT, providing additional support for the usage-based theories. Concluding that semantic transparency affects the processing of multi-word units, the authors suggest that their results provide empirical evidence for Howarth's theoretical model. They also call for more research using the phraseological framework and for future studies to expand their work by including other categories from Howarth's model.

Finally, Yamashita (2018) set out to investigate if semantic transparency could explain the congruency effect by categorizing the target collocations from the five experiments (Wolter & Gyllstad, 2011, 2013; Wolter & Yamashita, 2015, 2018; Yamashita & Jiang, 2010) described above based on semantic transparency . The results showed that an overwhelming majority of the congruent items used in those studies were more transparent, while those that were incongruent tended to be more opaque. Based on this evidence, Yamashita ventures that semantic transparency could possibly contribute to the congruency effect. However, she does so without denying or diminishing the role of the L1. Highlighting the significance of Gyllstad and Wolter's (2016) findings, Yamashita calls for the need to examine the effects of congruency and transparency simultaneously. On an interesting side-note, Yamashita (2018) found a 90% overlap between Swedish and Japanese collocations that were also congruent with English, indicating that congruent collocations may be more universal than previously thought. The evidence from Gyllstad and Wolter (2016) and Yamashita (2018) highlights the need for further exploration into the role of semantic transparency in L2 collocational processing.

Limitations

Given that L2 learners tend to transfer collocational knowledge from their L1, linguistic distance might be a relevant factor that could affect collocational processing. However, only L1 Japanese and Swedish learners were tested by Wolter and colleagues. Replicating these studies with learners from different L1 backgrounds would provide a more comprehensive understanding of the effect of the L1 on L2 collocational processing. Moreover, the sample population in almost every study consisted solely of university students, which does not allow for generalization of the results to the rest of the population (see Andringa & Godfroid, 2020, for an in-depth

discussion). An effort should be made to recruit a more diverse sample.

Many of the studies reported here have small sample sizes and are therefore underpowered. Yet, they use LMMs, which require sufficient power for the results to be reliable and replicable (Kumle, Võ, & Draschkow, 2021). For example, a quick power analysis using the 'sjstats' package (Lüdecke, 2021) in R for Fioravanti et al.'s (2021) study reveals that a minimum of fifty participants per group, and a total sample size of ninety-nine was required for the study to be sufficiently powered for an LMM analysis. However, their sample size per group is substantially lower than fifty. This is problematic as: (i) underpowered studies have a greater probability of erroneously showing larger effect sizes than high-powered studies (Gelman & Weakliem, 2009), and (ii) the social sciences are rife with publication bias which leads to studies reporting larger effect sizes being more likely to be published (Roettger, 2021). In the case of a type II error (rejecting the null hypothesis when it should be accepted), an underpowered study is likely to show a larger effect size and thus more likely to get published, impacting not only the findings of the study in particular, but the field in general, giving rise to an "inconsistent and misleading" body of literature (Smith, 2012). Therefore, it is important to estimate the sample size required during the research planning stage, before data collection can commence.

Furthermore, as demonstrated by Yamashita (2018), semantic transparency must be taken into account when developing items for psycholinguistic experiments. A frequency-only definition—used in all the studies described here, except for Fioravanti et al. (2021); Gyllstad and Wolter (2016); Nesselhauf (2003)—may not be adequate to model L2 collocational behaviour, as semantic transparency is an important linguistic characteristic of a word combination and is known to affect processing. Another issue when selecting target stimuli is the inclusion of cognates—words that are similar in form and meaning—which are known to boost processing times (Urdaniz & Skoufaki, 2019). For example, Gyllstad and Wolter (2016) use collocations such as *run a **risk*** which in Swedish is *löpa en **risk***. Since this could act as a confounding variable, care must be taken to avoid cognates, or balance them across stimuli sets.

Finally, not all studies accounted for learner proficiency, while some depended on self-reports which are often unreliable (Tomoschuk, Ferreira, & Gollan, 2019), making it difficult to compare findings between studies involving different groups of speakers. Using validated tests, such as the Quick Placement Test (QPT), Test of English for International Communication (TOEIC) or the LexTALE to measure proficiency must become standard procedure in SLA research. Additionally, reliable measures of proficiency could be useful to model collocational processing at different proficiency levels.

2.3 Summary

Although this is by no means an exhaustive review of the literature, the coalescing picture narrows down the underlying factors known to affect collocational acquisition and processing to frequency, congruency, and semantic transparency. Evidence from the classroom as well as the lab has shown that these factors do not work in isolation. Frequency is key to collocational acquisition, given that both congruency and semantic transparency are affected by it. However, to the best of my knowledge, there is no online psycholinguistic study that has investigated semantic transparency and congruency simultaneously.

Therefore, based on the review of the literature presented in this chapter, this study aims to empirically evaluate the effects of semantic transparency in tandem with congruency by extending Gyllstad and Wolter (2016). This will be done by incorporating congruent and incongruent collocations into Howarth's Continuum under the assumption—based on Yamashita's findings—that incongruent collocations are less transparent than congruent ones (see figure 2.2). Paying heed to Schmitt (2010), statistical measures of association will be taken into account for item development. Moreover, to minimize the effect of frequency, only high-frequency collocations will be used.

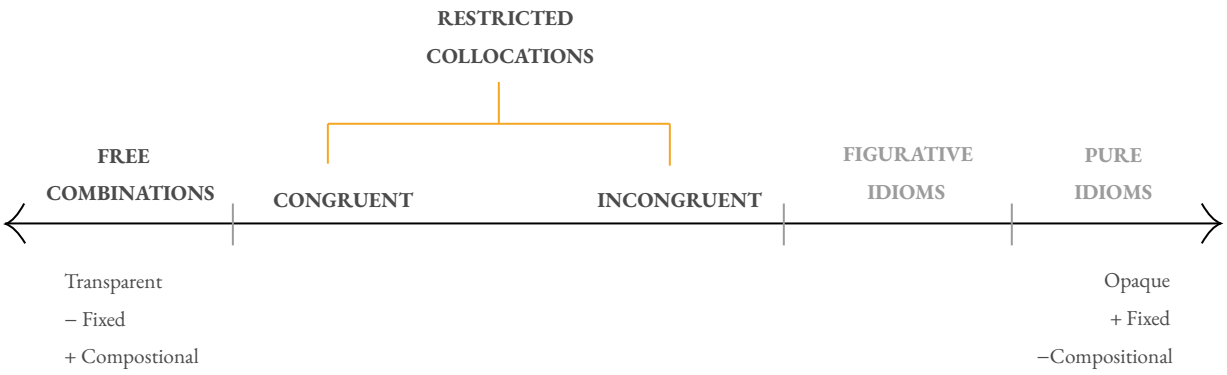


Figure 2.2. Congruent and Incongruent Collocations in Howarth's Continuum

Chapter 3

Methods

This chapter presents and justifies the methods used in this study. First, the research design is outlined in section 3.1. The sample population is described in section 3.2, while section 3.3 describes the materials used in this study. The data collection method and experimental procedure are outlined in section 3.4. In section 3.5, the statistical procedures that were used to analyse the data are described and justified. The chapter concludes with an ethics statement in section 3.6.

3.1 Research Design

A quantitative, cross-sectional design was employed in keeping with the methods used by Gyllstad and Wolter (2016). The cross-sectional aspect of this study entailed a between-participants/within-items factor. Participants belonged either to the control group which consists of L1 English speakers or one of two experimental groups L1 Brazilian Portuguese (BP) and European Portuguese (EP), together referred to as L1 Portuguese. Items belonged to one of four experimental conditions—free combinations, congruent collocations, incongruent collocations, and baseline items. Each item in each condition was seen by every participant in each group that completed the experiment. Given the review of the literature in chapter 2, the following research questions (RQs) and their associated null hypotheses (H_0) and alternative hypotheses (H_1) were put forth:

RQ1: For L2 English speakers (L1 Portuguese), is there a statistically significant difference in processing speed and accuracy between free combinations versus congruent collocations?

H_0: For L2 English speakers (L1 Portuguese), the processing speed and accuracy for free combinations is the same as that of congruent collocations.

H_1: For L2 English speakers (L1 Portuguese), the processing speed and accuracy for free combinations is not the same as that of congruent collocations.

RQ2: For L2 English speakers (L1 Portuguese), is there a statistically significant difference in processing speed and accuracy between free combinations versus incongruent collocations?

H_0: For L2 English speakers (L1 Portuguese), the processing speed and accuracy for free combinations is the same as that of incongruent collocations.

H_1: For L2 English speakers (L1 Portuguese), the processing speed and accuracy for free combinations is not the same as that of incongruent collocations.

RQ3: For L2 English speakers (L1 Portuguese), is there a statistically significant difference in processing speed and accuracy between congruent versus incongruent collocations?

H_0: For L2 English speakers (L1 Portuguese), the processing speed and accuracy for congruent and incongruent collocations is the same.

H_1: For L2 English speakers (L1 Portuguese), the processing speed and accuracy for congruent and incongruent collocations is not the same.

RQ4: Is there a statistically significant difference in processing speed and accuracy between L1 English and L2 English speakers (L1 Portuguese)?

H_0: The processing speed and accuracy for L1 English speakers is the same as that of L2 English speakers (L1 Portuguese).

H_1: The processing speed and accuracy for L1 English speakers is not the same as that of L2 English speakers (L1 Portuguese).

To test these hypotheses, a semantic acceptability judgement task (SAJT) was administered to all participants, following Gyllstad and Wolter (2016). The data were analysed using the GLMM framework with a mixed-model ANOVA. In order to address RQ1 and RQ2, each L2 group was compared by their performance on the free combinations versus the congruent and the incongruent conditions, respectively. For RQ3, each L2 group's performance on the congruent condition was compared to their performance on the incongruent condition. Finally, to address RQ4, L1 English speakers' performance was compared to that of the two L2 groups (L1 Portuguese) for each of the four conditions. The baseline items used in this study played a dual role—they were used as filler items to counter familiarity effects, and also for comparison with the other conditions. It must be noted that the two L2 groups will not be compared to each other as there is no underlying theoretical motivation to do so.

This study involved two main predictor variables—the participants' language group (L1 group), and collocation type (condition)—and two dependent variables—speed (RT), and accuracy (ER).

3.2 Participants

This section describes the recruitment process, the participants' background, and the exclusionary criteria. For the experimental groups, proficient (CEFR B2 level and above) adult L1 BP and European Portuguese (EP) speakers of L2 English were recruited, while for the control group, adult L1 English (EN) speakers were recruited. No consideration was given to the variety of English spoken.

A power analysis was performed to determine the sample size required for a multi-level mixed-effects model using the 'sjstats' package (Lüdecke, 2021) in R. The results showed that for a medium to large effect (.2), a minimum of 48 participants per L1 group and a total sample size of 143 was required.

3.2.1 Recruitment

A total of 749 participants were recruited through email and the snowball effect. For the L2 groups, professors and researchers at various universities in Portugal and Brazil were contacted via email. They were asked to disseminate the link to their students in the classroom, via social media, and mailing lists. L1 English speakers were recruited by sending emails to various departments, research groups, JCRs (Junior Common Room), and MCRs (Middle Common Room) at the University of Oxford, as well as to other universities in the UK. Emails were sent to the researcher's personal contacts who fit the description of the target populations. They were also encouraged to forward the email to their personal contacts and share it on social media. Additionally, at the end of the experiment, participants were encouraged to share a link to the experiment with anyone that might be interested in participating.

After exclusions, the present study had a total sample size of 329 participants. Of these, ninety-nine were L1 English (EN) speakers (48 female, 50 male, 1 non-binary), 122 BP speakers (33 female, 86 male, 3 non-binary), and 108 EP speakers (77 female, 29 male, 2 non-binary). The present study was considered to be sufficiently powered as it surpassed the minimum requirements of the power analysis.

3.2.2 Initial Exclusions

Of the 749 people that were recruited, only 510 completed all nodes in the experiment tree. Those that did not reach the 'end' node were excluded. Fourteen participants were excluded as they reported their L1 to be a language other than EP, BP, or English. Ninety-five participants with mean reaction times (mRTs) faster than

450 ms were eliminated, as it indicated that they were not attending to the task. Similarly, nineteen participants that registered a mean accuracy of less than 50% were eliminated. Although Gyllstad and Wolter (2016) do not eliminate any participants based on accuracy, it was deemed necessary here as the task was not performed in a controlled environment. Other exclusions made on the basis of socio-demographic self-reports and L2 proficiency are described in the following sub-sections.

3.2.3 Socio-demographics

Socio-demographic information was collected by means of a questionnaire. The participants were asked to provide information about their L1, age, level of education, handedness, vision and any known language pathologies (see Appendix A.3.5). The data collected are summarized in the following sub-sections.

Age

In terms of mean age, there was a small but significant difference between all three groups. The L1 English participants were the oldest group with a mean age of 30 years (SD = 12.2), while the mean age of the BP and EP group were 28.5 (SD = 7.17) and 27.6 (SD = 10.5), respectively (table 3.1). All three groups were positively skewed in terms of age, violating the assumption of normality required for an ANOVA. Therefore, a Kruskall-Wallis non-parametric test was used (see Hatch & Lazaraton, 1991) to compare the three groups, and a significant difference was found ($H(2) = 714$, $p < 0.0001$). A post-hoc Dunn's test (see Dinno, 2015) revealed that all groups were significantly different from each other. Hence, all analyses that involved L1 group as a factor were conducted with age as a covariate.

Table 3.1. Mean Age (in years) by L1 Group

	Brazilian Portuguese (BP)	English (EN)	European Portuguese (EP)
Mean	28.5	30.0	27.7
SD	7.2	12.2	10.5

Language Pathologies

The participants were asked if they had been diagnosed with dyslexia or any other language pathologies. Five participants responded with a "yes" and eight responded with an "unsure". A Kruskall-Wallis test showed sig-

nificant differences in mRTs between all groups ($H(2) = 778.26$, $p < 0.001$). The follow-up Dunn's test showed that all groups were significantly different. Given that participants with language pathologies such as dyslexia show processing delays in visual semantic judgement tasks (see Rüsseler, Becker, Johannes, & Münte, 2007), these thirteen participants were excluded.

Handedness

In terms of handedness, 269 of the 329 participants reported themselves as being right-handed (BP = 110, EN = 61, EP = 98), twenty-six were left-handed (BP = 6, EN = 13, EP = 7), and thirty-four were ambidextrous (BP = 6, EN = 25, EP = 3) (table 3.2). The number of ambidextrous self-reports is abnormally high especially given that only 1 in 100 people (1%) are ambidextrous (Rodriguez et al., 2010). These participants were not eliminated, as it would entail losing approximately 25% of the L1 English group. However, given that the task had equal amounts of "yes" and "no" responses, no group had a performance advantage due to handedness. Hence, it was not included as a covariate. As an aside, this illustrates the tenuous and unreliable nature of self-reported data, if untrue (see Tomoschuk et al., 2019).

Table 3.2. Handedness by L1 Group (Self-reported)

Handedness	Brazilian Portuguese (BP)	English (EN)	European Portuguese (EP)
Ambidextrous	6	25	3
Left	6	13	7
Right	110	61	98

Vision

This study is fundamentally based on the participants' visual word recognition abilities. Therefore, twenty-five participants (BP = 12, EN = 7, EP = 6) who reported visual impairments other than corrected vision were eliminated to avoid vision-related processing lags acting as a possible confound.

Education

All participants reported having attained, or being on course to complete, a degree in higher education. No exclusions were made based on education level. See Appendix A.2 for more details.

3.2.4 L2 Background

Information about the participants' L2 background was collected by means of a questionnaire. There was a significant difference in the mean starting age of English learning between the BP (11.2 years old) and the EP (7.55 years old) groups ($t = 75.941$, $df = 19784$, $p < 0.001$).

Table 3.3. Mean Starting Age of English Learning

	Brazilian Portuguese (BP)	European Portuguese (EP)
Mean	11.18	7.54
SD	4.02	2.85

Eighty-three participants (BP = 31, EP = 52) reported having studied in an English as a Medium of Instruction (EMI) institution, while forty-three (BP = 23, EP = 20) reported having studied abroad in a majority English-speaking country. Additionally, fifty-three (BP = 30, EP = 23) participants reported having lived in an English-speaking country for an extended period of time.

The participants were also asked where they first began learning English. A majority of the EP speakers (78%) reported that they first learned English in school, while BP speakers reported first learning English in language institutions (37%). Only 18% of the BP speakers and 16% of the EP speakers reported first learning English at home. For more details, see Appendix A.3. These questions were asked solely for the sake of describing the sample, and no exclusions were made based on any of their responses, nor were they included in any statistical analyses.

L2 English Proficiency

The LexTALE test (Lemhöfer & Broersma, 2012) was administered to gauge general English proficiency. The mean score was 84.56% for the BP group and 82.34% for the EP group. A Welch's two-sample t-test showed a significant difference between the two groups ($t = 16.63$, $df = 19{,}742$, $p < 0.001$). Ninety-two BP and sixty-five EP speakers scored more than 80% and were classified as advanced (CEFR C1-C2). Thirty BP and forty-three EP speakers scored between 60-80%, and were classified as upper-intermediate (CEFR B2).

Table 3.4. Mean LexTALE Scores

Lextale	Brazilian Portuguese (BP)	European Portuguese (EP)
Mean	84.56	82.34
SD	9.10	10.02

Since the European Union stipulates that L2 speakers attain a CEFR level of B2 or above to gain access to higher education, the same was deemed to be an appropriate minimum proficiency level for the purposes of this study. Therefore, seventeen participants who scored less than 60% were excluded.

Additionally, L2 participants were asked to report their CEFR levels (A1-C2) and self-assess their proficiency in the four skills—reading, listening, speaking and writing (see Appendices A.4 and A.5). Owing to inconsistencies between the self-reports and the LexTALE scores, only the latter were used to exclude participants with low proficiency.

English Variety

In order to establish justification for the use of the English TenTen Corpus, 2018 (ENtenten18) in lieu of the British National Corpus (BNC) or the COCA (see discussion in section 3.3.1), the participants were asked to indicate the variety of English that they spoke. Unsurprisingly, from a geopolitical standpoint, more BP (58%) than EP speakers (11%) reported speaking American English, while more EP (34%) than BP speakers (7%) reported speaking British English (table 3.5). A substantial number of participants reported speaking both[1] American and British English (BP = 31%, EP = 48%). This clearly illustrates the need for referencing a more diverse corpus, or at the very least, referencing both the BNC and the COCA, to account for the language to which learners are exposed in experimental design.

[1] Additionally, one of the participants in this study noted in the comments section, "In theory, British English, but American television/movies/music had a lot more influence in my learning path".

Table 3.5. Self-reported English Variety Spoken

	Brazilian Portuguese (BP)	European Portuguese (EP)
British English	8	37
American English	71	11
Both	38	52
Don't Know	5	7
Other	0	1
Total	122	108

3.3 Materials

3.3.1 Sketch Engine

Sketch Engine (Jakubíček, Kilgarriff, Kovář, Rychlý, & Suchomel, 2013) is an online text analysis tool for performing word co-occurrence analyses, extracting terms, generating frequency lists, and other lexicographic functions. The Word Sketch feature was used to identify the items used in this study. Word Sketch finds the collocates of the queried word in a given corpus, and provides a one-page summary of the grammatical and collocational behavior exhibited by that word. The collocations included in the resulting 'word sketch' are organized according to grammatical relations (e.g., subject, object, modifier) and displayed along with their logDice and frequency scores.

TenTen Corpora

For item development, Sketch Engine's Portuguese TenTen Corpus, 2011 (PTtenten11) and English TenTen Corpus, 2018 (ENtenten18) were used as reference corpora. The PTtenten11 corpus contains 3,896,392,719 words (4,622,750,491 tokens) from both varieties—EP and BP—crawled from the internet between 2011 and 2012. An extensive search found no other Portuguese language corpus with both varieties as well as phrasal frequencies. The ENtenten18 is a dynamic web corpus, containing 21,926,740,748 words (25,825,512,637 tokens). The texts were crawled between 2016 and 2018. In addition to its massive size and dynamic nature, it does not restrict itself to a single variety—it incorporates sub-corpora of various Englishes such as Indian, Canadian, and Australian English (see Appendix A.6).

In contrast, the BNC, a corpus of British English, and the COCA, a corpus of American English, are static corpora that only contain texts from their specific varieties, and are widely used in Applied Linguistics and SLA. However, these corpora may not reflect the English to which L2 learners are exposed. Although a learner might be taught British English in school, they watch Hollywood movies, play video games, use social media, etc. outside the classroom. Access to the internet allows for exposure to other varieties of English such as Canadian English, Indian English, etc. It is well-known in the field that learning does not take place solely in the classroom, especially in today's digital age with global mobility and enhanced access to foreign languages online (Lamb, 2017). In the same way that it would be inappropriate to study child language using an adult language corpus, it is no longer appropriate to use mono-varietal corpora to gauge the language to which L2 learners are exposed.

Therefore, the present study used the ENtenten18 for item development. However, every item was checked against the BNC and the COCA to ensure that they occur in both major varieties of English.

logDice

Instead of widely used association scores such as the Mutual Information (MI) score (Church & Hanks, 1990), and t-scores, Sketch Engine provides logDice values as a measure of association. The MI score—used by Wolter and Yamashita (2018), and Gyllstad and Wolter (2016)—expresses the extent to which words co-occur in comparison with the number of times they appear separately in relation to corpus size. It is expressed as follows:

$$M.I.\ Score = log_2 \frac{2 f_{xy} N}{f_x f_y},$$

wherein, 'f_{xy}' is the number of co-occurrences of words x and y, 'f_x' is the number of occurrences of word x, 'f_y' is the number of occurrences of word y, and 'N' is the size of the corpus.

The MI score is strongly affected by frequency and corpus size, wherein low-frequency words tend to have a higher MI score (indicating stronger association), which could be misleading (Rychlý, 2008). Since it takes corpus size into account, it makes between-corpora comparisons impossible (*ibid.*). Additionally, the larger the corpus, the more skewed the MI scores. This makes them hard to interpret. Therefore, Rychlý (2008) suggests using the Dice score (Dice, 1945) which calculates association without accounting for corpus size. It is expressed as follows:

$$Dice\ (D) = \frac{2 f_{xy}}{f_x + f_y}.$$

However, the values of the Dice score (D) are usually very small numbers. Therefore, Rychlý (2008) proposes the logDice, which is easier to interpret and is expressed as follows:

$$logDice = 14 + log_2\ D.$$

LogDice expresses the strength of the association between two words that occur together. A logDice value of 14 (theoretical maximum) indicates that the two words always occur together in a given corpus, while a score of zero means that they never occur together. A negative score indicates dissociation—the words are likely to be unrelated. In essence, the closer the value is to 14, the stronger the association.

Unlike MI scores, logDice values are not affected by corpus size, allow for comparison between corpora, and are easy to interpret. Although MI scores have been standard in research involving collocations in SLA, many recent studies have switched to logDice instead (e.g., Edmonds & Gudmestad, 2021; Frankenberg-Garcia, 2018; Frankenberg-Garcia, Lew, Roberts, Rees, & Sharma, 2019; Gablasova, Brezina, & McEnery, 2017; Gilquin, 2017; Öksüz, Brezina, & Rebuschat, 2021). Given the above discussion, and the fact that the ENtenten18 is a behemoth of a corpus, the logDice values were used to identify the items in this study.

3.3.2 LexTALE

Although self-ratings of language proficiency are prevalent in SLA research, the validity of such ratings is unclear (Tomoschuk et al., 2019). Therefore, the LexTALE, a valid and standardized test of vocabulary knowledge for medium to highly proficient speakers of L2 English (Lemhöfer & Broersma, 2012), was used as a proxy for general English proficiency. This test was designed with cognitive researchers in mind and has become increasingly popular in psycholinguistic circles (e.g., Poort & Rodd, 2019; Urdaniz & Skoufaki, 2019) as it is quick (3-5 minutes), easy to administer, and free. Additionally, it is considered to be a fair indicator of general English proficiency as there is a significant correlation between LexTALE scores and: (i) the QPT, and (ii) the Test of English for International Communication (TOEIC) scores (Lemhöfer & Broersma, 2012).

The LexTALE is a simple un-speeded visual lexical decision task with sixty trials—forty words and twenty non-words. The trials were administered with the same instructions, and in the same order as recommended by Lemhöfer and Broersma (2012), with the exception that a 2,000 ms timeout was added. This was done following Poort and Rodd (2019) since the main task, which was performed first, was speeded.

Vocabulary size is measured by taking into account the percentage of correct responses, adjusted for the unequal proportion of words and non-words in the test by averaging the percentages of correct responses for both item types. This measure is called % $correct_{av}$ (averaged % correct). The formula to calculate LexTALE scores is as follows:

$$\%\, correct_{\text{av}} = \frac{(number\ of\ words\ correct/40 \times 100) + (number\ of\ non-words\ correct/20 \times 100)}{2}$$

According to Lemhöfer and Broersma (2012), LexTALE scores below 60% correspond to lower-intermediate (B1) on the CEFR proficiency scale. Scores from 60% to 80% correspond to the upper-intermediate (B2) proficiency level, while those above 80% correspond to advanced (C1 and C2) proficiency level[2]. These correspondences were based on the equivalences between the LexTALE and QPT score bands, in comparison with those of the QPT score bands and the CEFR levels.

3.3.3 Item Development

All test items were developed with L1 Portuguese speakers and Howarth's model in mind. In Portuguese, the adjective follows the noun (Villalva & Silvestre, 2014), unlike in English wherein the adjective precedes the noun (e.g., *camisa*$_N$ *vermelha*$_{ADJ}$, 'red$_{\text{ADJ}}$ shirt$_{\text{N}}$'). Therefore, only Verb+Noun (V+N) collocations were used here to avoid potential complications arising from word order differences between the two languages.

Free Combinations & Collocations

A preliminary list of potential V+N collocations was first populated with the help of two learner dictionaries: (i) The BBI Combinatory Dictionary of English (2010), and (ii) Collins COBUILD Advanced Learner's Dictionary (2018). Definite articles were avoided by resorting to nouns in their plural forms (e.g., 'share dreams' instead of 'share a dream'). Collocations such as 'run a risk' (*'run risks') were rejected to avoid issues arising due to syntactic violations. Additionally, verbs that could also be interpreted as other parts of speech were avoided (e.g., '**open** accounts', '**store** data'), as were collocations that could have reasonable multiple interpretations (e.g., '**draw** blood', '**catch** fire'). However, some word combinations such as *foot bills* and *kick habits* were included as the alternative interpretations were deemed to be implausible. Words with different spellings in American and British English (e.g., *color* versus *colour*) were also avoided.

After the list was created, all the items were translated word-for-word into Portuguese by the researcher who is a proficient speaker of EP (CEFR C2). This translated list was then verified and categorized by an L1 EP speaker[3]. The translated items were divided into two groups: (i) those that were semantically meaningful collocations in EP with the exact same semantic sense as their English counterparts, and (ii) those that were not meaningful at all in EP. Any items that did not fit these categories were eliminated. Simultaneously, a list of free combinations

[2] These values were adjusted for the gaps and overlaps in the ranges presented in Lemhöfer and Broersma (2012).

[3] Third-year student of general and Portuguese linguistics at the University of Lisbon, advanced L2 English speaker.

in English was populated in compliance with the constraints laid out for the collocational items. All free combinations were congruent in Portuguese. For example, *shave beards* was rejected as its Portuguese equivalent is *fazer a barba* (*'to make/do the beard') or *barbear* ('to shave') which are not direct word-for-word translations.

This list was then checked for cognates[4], irrespective of their etymological relation (e.g., *carro*, 'car', *inundação*, 'inundation') . If an item contained a cognate (e.g., ***satis**fazer pedidos*, '**satis**fy requests'), it was rejected. This filtered list was then verified by an expert L1 BP speaker[5]. If the items were not directly translatable to BP, or if their translations violated any of the above conditions, they were rejected. For example, 'break rules' is a congruent collocation in EP (*quebrar regras*), but in BP, the equivalent collocation is *'driblar regras'* (*dribble rules). Therefore 'break rules' was rejected. Lastly, only items that were unacceptable in both EP and BP were added to the list of incongruent collocations.

Once all the items met the initial inclusionary criteria, Sketch Engine was used to verify their logDice scores in both English and Portuguese in their respective tenten corpora. Collocations with no occurrences in the resulting 'word sketch', or those with low logDice scores were rejected. The English collocations were then checked against the BNC and the COCA to ensure that they were used in both British and American English. Only those that were found in both corpora were accepted.

To ensure that all individual words would be familiar to the L2 participants and to suppress the effects of differences in frequency, the individual words in each collocation were checked for frequency in Nation's BNC/-COCA Word frequency lists using the Compleat Web VP version 2.5 (Cobb, 2021). The aim was to employ only those words that belonged to the K-1 band (most frequent word families), but this proved impossible given the preceding set of constraints on item development. Out of ninety words, seventeen belonged to the K-2 band, and five to the K-3.

Finally, a list of 20 free combinations, 20 congruent collocations, and 20 incongruent collocations (all high-frequency items) was presented to three L1 English speakers (1 linguistics professor, 1 linguistics student, and 1 non-expert), and three L1 EP speakers, all advanced L2 English speakers (1 linguistics professor, 1 linguistics student, and 1 non-expert). They were asked to assess: (i) whether or not the verb was used in its literal sense, (ii) whether or not the noun was used in its literal sense, and (iii) if the word combination had more than one meaning. Additionally, the L1 EP speakers were asked whether or not the English collocation had a literal equivalent in their L1. There was a 100% agreement within both groups on all sixty items.

[4] According to Daulton (2008) cognates can be defined from a synchronic perspective as words that are similar in form and meaning, or from a diachronic perspective, as etymologically related words similar in meaning and form. The synchronic definition is more relevant to this study as language learners perceive cross-linguistic similarities that affect word recognition (see Petrescu, Helms-Park, & Dronjic, 2017; Urdaniz & Skoufaki, 2019).

[5] Linguistics researcher and professor at the Federal University of São Paulo, specializes in Portuguese morphosyntax and verbal semantics.

These items were further controlled for length (total number of letters in the item) and syllables (total number of syllables in the item). Fifteen items per condition—forty-five in total—were then chosen for the semantic acceptability judgement task (SAJT) (see Appendix A.2).

Baseline Items

The baseline items used in this study were semantically meaningless V+N combinations included not only for comparison, but also to avoid the confounding effects of familiarity. To match the number of target items, forty-five baseline items were created by recombining the words from the other three conditions. This ensured that there would be no difference in lexical frequency of individual words between the baseline and target items (see Wolter & Yamashita, 2015, 2018). Each baseline item was checked against the ENtenten18 corpus to ensure novelty. Theoretically, these items should have had zero occurrences in the corpus, but given the sheer size of the ENtenten18 (25,825,512,637 tokens), some baseline items were found when queried. However, the corpus size, their infinitesimally small frequencies, and negative logDice values indicate that these occurrences were due to chance and therefore, these items were deemed not to be meaningful word combinations.

Finally, further checks were conducted for differences between the four experimental conditions. While there were significant differences for collocation length ($p < 0.05$) and the results of Tukey's post-hoc test revealed significant differences between all conditions, none were found for number of syllables ($p > 0.05$). Therefore, collocation length was added as a covariate to all models. Similarly, phrasal frequency was compared (without the baselines, as they do not exist in theory) and a significant difference was found between all groups ($p < 0.05$). Therefore, a separate analysis involving phrasal frequency as a covariate was conducted and it was not found to be a significant predictor for RTs and ERs.

As a result, ninety items in total were used in this study—45 target items and 45 baseline items.

3.4 Data Collection

3.4.1 Pilot

A pilot was run with seven L1 speakers and eight L2 speakers of varying L1 backgrounds to ensure that all transitions between nodes occurred smoothly, and that all the branches filtered participants correctly. Two L2 participants reported that the font size on the instructions page was too small. This could not be rectified in time for data collection, as it appeared to be a problem that arose from using both HTML and markdown in the Gorilla™ experiment builder. No other problems were encountered during the piloting process, and the experiment was deemed fit for actual data collection to commence.

3.4.2 Procedure

The experiment was designed, configured and deployed using Gorilla™ (see figure 3.1).

The experiment could only be accessed by means of a computer with a keyboard, i.e., desktop PC or a laptop. Touch screen devices such as mobile phones and tablets were restricted as they are known to not accurately register RTs. There were no restrictions placed on the participants' browser specifications, nor on location and connection speed.

First, a consent and participant information sheet was presented, wherein a brief explanation about the study was provided and it was made clear that participation was voluntary and anonymous. Once consent was given, a routine socio-demographics questionnaire concerning language background, age, laterality, vision, etc. was administered (see Appendix section A.3.5). Upon completion of this questionnaire, the main task, i.e., the SAJT, commenced.

Participants were asked whether or not the word combinations that appeared on the screen were acceptable in English. They were instructed to press the "A" key if they felt an item was "acceptable", and the "L" key if they felt an item was "unacceptable". They were also instructed that there would be a training round, and that they should answer as quickly and accurately as possible. For screenshots of the instructions see Appendix A.3.8

The stimulus presentation procedure is shown in section 3.4.2. A stimulus (target or baseline item) appeared on the screen and remained for a total duration of 4,500 ms. If no decision was taken within 4,000 ms, a warning was presented to the participant indicating that they were responding too slowly. The warning remained on screen for 500 ms, during which the participant could still respond. A fixation cross would appear on the screen for a duration of 500 ms between each item.

The test began with a short training set of six items with feedback, so that participants could familiarize themselves with the task. After the training set was complete, they were asked to press the spacebar key to start the task. The task consisted of ninety items that were presented in an individualized random order. Participants were also allowed a short break of 15,000 ms after 45 items were presented. Most participants completed the SAJT within 3–5 minutes.

Once the main task was completed, the L1 speakers then proceeded to a debrief questionnaire, while the L2 speakers were presented with an L2 background questionnaire (see Appendix A.3.6). The LexTALE was then administered to assess the L2 participants' proficiency, after which the debrief questionnaire was presented (see Appendix A.3.7). This questionnaire asked the participants if they were distracted or if they faced any connectivity issues during the test. Several participants reported encountering distractions and/or internet connection lags. As Gorilla pre-buffers the experiment by downloading it onto the participants' device (Sauter, Draschkow,

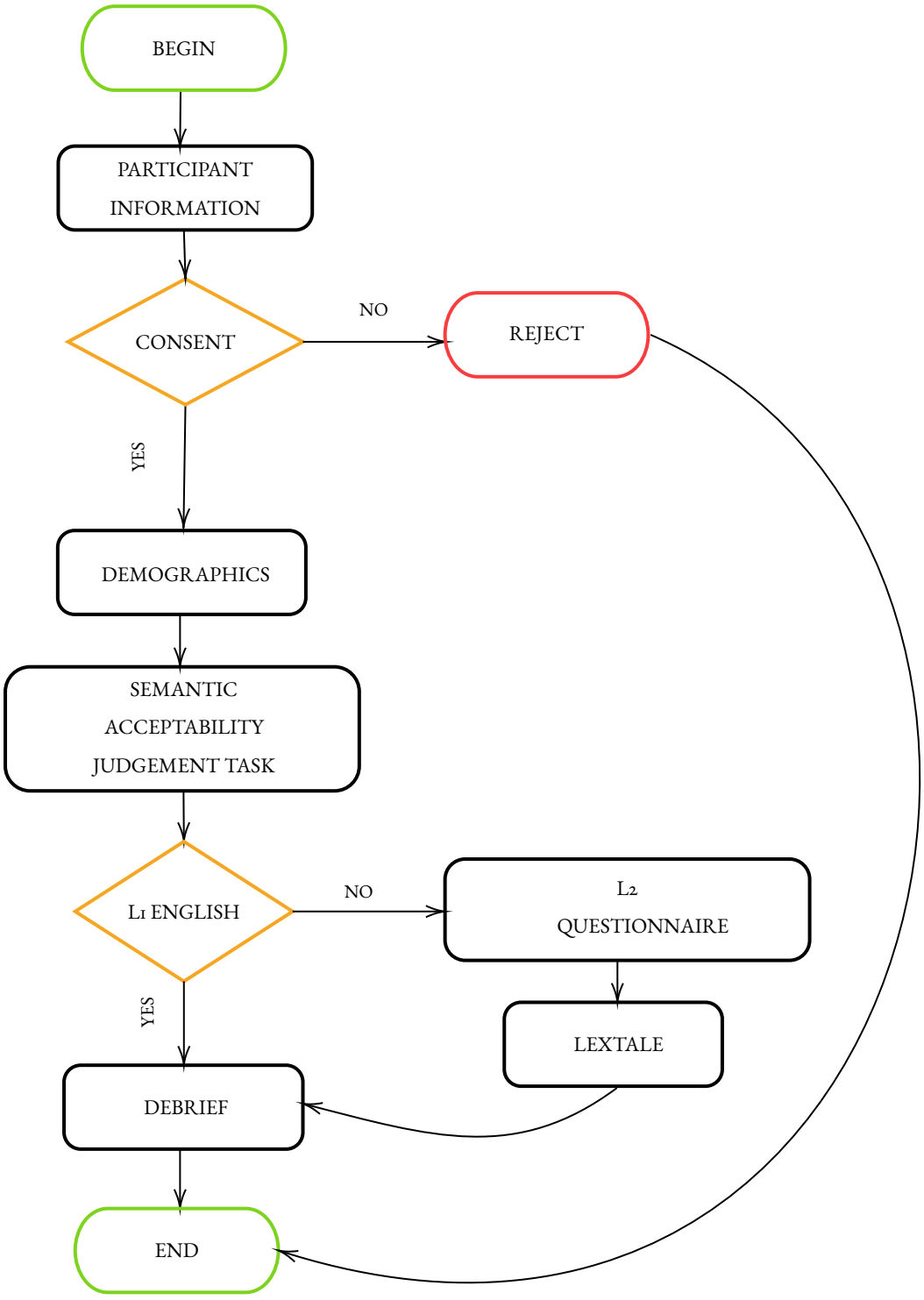
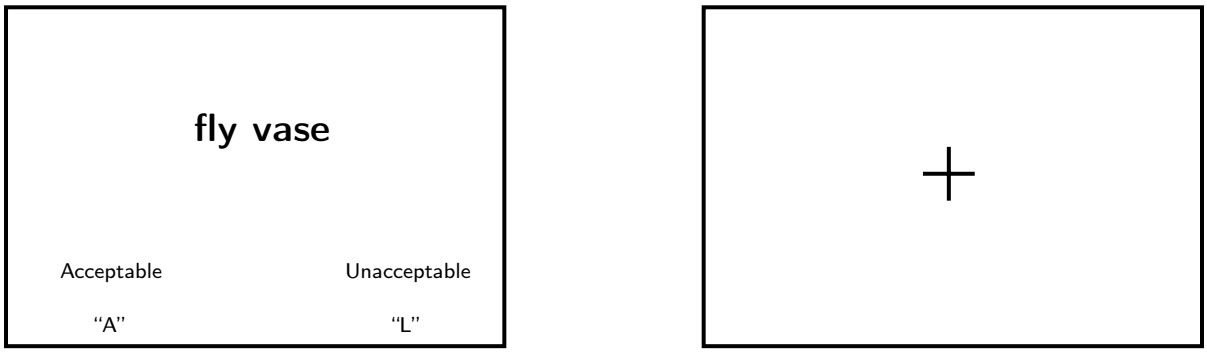


Figure 3.1. Experiment Flow Diagram



Stimulus
(4500 ms)

Fixation Cross
(500 ms)

Figure 3.2. Main Task Screen

& Mack, 2020), connectivity issues were assumed to not be problematic. However, distractions are known to affect processing and therefore were included as a covariate.

All participants that successfully completed all nodes of the experiment were given the chance to participate in a lucky draw to win one of six gift cards worth £25 (in their currency) by submitting their email. The lucky draw was carried out using Microsoft Excel's random function and the six winners were notified via email.

3.5 Data Analysis

This section describes the GLMM framework and justifies its use in this study, after which the model specification procedure is presented.

3.5.1 Data Trimming

In order to get the data to fit a normal distribution, an assumption for many statistical tests such as ANOVAs and LMMs, researchers often trim outliers by using the IQR method (see Ratcliff, 1993). This leads to a bias in the dataset leaning towards the faster RTs and is not truly reflective of the sample. Therefore, following suggestions from Baayen and Milin (2010), minimal *a priori* trimming was carried out, as normality is not an underlying assumption of a GLMM which was used to carry out the statistical analyses in this study (see discussion below). All RTs faster than 450 ms (540 of 29,610 observations; 1.82%), and all observations that timed-out at 4,000 ms (268 of 29,610 observations; 0.9%) were eliminated, culminating in a dataset with 28,802 observations.

3.5.2 Generalized Linear Mixed-Effect Models (GLMMs)

LMMs are able to take into account the non-independence of observations, i.e., when one participant provides multiple measurements (Barr, Levy, Scheepers, & Tily, 2013). Therefore, they are strongly favored by cognitive researchers and recommended by high-impact journals in Psycholinguistics (Lo & Andrews, 2015). However, in this study, GLMMs (which also account for non-independence of observations) were used *in lieu* of LMMs primarily based on three important considerations—theoretical, empirical, and philosophical.

The fundamental theoretical assumption underlying this study—and most research in cognitive psychology (Townsend, 1992)—is that the differences in RTs observed over experimental conditions are assumed to directly reflect the differences in time taken to perform the mental task being measured. Simply stated, they are based on raw RT values. LMMs are constrained by the assumption of normality—the mathematical assumption that the dependent variable follows a normal (Gaussian) distribution (Gill & Torres, 2019). Given that RT data are positively skewed, they violate this assumption, thereby affecting the estimate of the means and distorting the outcome of the LMM. To counter this, most researchers apply a non-linear transformation, such as a log or inverse transformation, to the RT data (see Baayen, 2008) in an attempt to make their data approximate the normal distribution. Analyses are then conducted on the transformed data. However, this does not allow for the results to be interpreted in terms of the original metric. i.e., the raw RTs. Although some researchers "back-transform" the estimates of the LMM to raw RTs, the significance results of the transformed data simply do not provide any information regarding the significance of the non-transformed data, and are therefore uninformative (see Berry, DeMeritt, & Esarey, 2010, for an in depth discussion). Simply put, if the empirical predictions were made based on theoretical assumptions dependent on the raw RTs and not on the transformed metric, there is a risk of failing to answer the research question (Lo & Andrews, 2015).

From an empirical standpoint, transformation distorts the ratio scale properties of RT data (Stevens, 1946), which could mislead the researcher. To illustrate, Lo and Andrews (2015) provide an example from a semantic priming task. Two age groups—one older and one younger—exhibited differences in RTs (e.g., 600 and 700 ms for the younger adults, and 780 and 910 ms for the older adults) when the prime and target were semantically related (e.g., nurse–doctor) versus when they were unrelated (e.g., plane–doctor). However, when the RTs are transformed using natural logs, the differences between these two groups vanish (e.g., log(700 ms) - log(600 ms) = 0.15415; log(910 ms) - log(780 ms) = 0.15415). Based on the transformed values, the researcher may conclude that there is no interaction between age and priming. Evidently, uncritical analyses of log-transformed RT data with LMMs are problematic as it can lead to incorrect conclusions being drawn.

Philosophically speaking, the epistemic simplicity of a model is a virtue that should not be dismissed (Rochefort-Maranda, 2016). Models are a way of describing or representing the real world, based on assumptions that are known to be "approximately true (and exactly false)" (Gabaix & Laibson, 2011). As scientists, we are constantly

analysing and formulating models in order to understand the relationship between our variables of interest. A model that makes it easier to understand this relationship is therefore valuable not only to the SLA researcher, but to all stakeholders involved. Using the raw RT values instead of transformed values allows for easier interpretation of the results, making the model epistemically simpler. Although this comes at the price of predictive precision (Rochefort-Maranda, 2016), which is a valuable attribute of any good model (Gabaix & Laibson, 2011), it is mitigated by the research context of this study, wherein the aim is not to make predictions, but rather to understand the nature of the relationship between the dependent and independent variables.

In the same way that research in Psycholinguistics has moved away from using ANOVAs due to problems associated with reliance on the mean (see Speelman & McGann, 2013), it is now time to critically consider the problems associated with LMMs and their underlying assumption of normality which forces the researcher to make decisions based on mathematical rather than theoretical criteria. GLMMs provide an adequate solution as they allow for the separation of the mathematical and theoretical components of the model. While the model itself satisfies the mathematical criteria of normalized, homoscedastic residuals required for linear regression, the researcher specifies the following components based on their research question, data and experimental design (Gill & Torres, 2019; Lo & Andrews, 2015):

- **Outcome Variable**: The dependent variable being measured by the researcher (e.g., test scores, reaction time, accuracy, etc.)
- **Fixed Factors**: The predictor variables (covariates) responsible for systematic variation in the dependent variable (e.g., age, sex, item type, etc.).
- **Random Factors**: The sampling structure of the experimental design that contributes to the random variability in the dependent variable (e.g., participant, item, etc.).
- **Distribution**: The probability distribution that best describes the plausible processes underlying the observations (e.g., Gaussian, Beta, Poisson, etc.).
- **Link Function**: The mathematical function that characterizes the relationship between the fixed factors and the dependent variable (e.g., log, identity, etc.).

This allows the researcher to conduct a computationally powerful, analytical, and more robust statistical analysis. Moreover, by eliminating the need to transform the dependent variable, the results become easier to interpret.

3.5.3 Model Specification

The present study followed Barr et al.'s (2013) recommendation to specify the "maximal random effects structure as justified by the design". Although these recommendations were made for specifying LMMs, the same procedure can be applied to specify GLMMs (Singmann & Kellen, 2019). Therefore, two maximal models were first specified, one with reaction time (RT) and the other with error rate (ER) as the dependent variable.

Specifying Fixed and Random Factors

Participant (`1|participant`) and item (`1|item`) were included as crossed random effects to account for individual differences, along with the main predictor variables, i.e., L1 group, condition, and the interaction between the two (`l1*condition`). Random slopes were included to allow for individual and group variations—the effect of experimental condition was included as a random slope by participant (`1+condition|participant`), and the effect of L1 group was included as a random slope by item (`1+l1|item`).

Four covariates were specified in the maximal models. As there were significant differences in age between L1 groups, and significant differences in collocation length (`length`) between experimental conditions, *age* and *collocation length* were added. Given that these covariates were continuous, they were centered on the mean. Additionally, *trial order* (`trial`) was included to account for task familiarity, and *distraction* was added to account for environmental factors that could affect attention, and thereby RTs. With the exception of probability distribution and link function, all the above components were common to both models.

The basic R syntax for the maximal models were as follows:

```
rt ~ l1*condition + trial + length + distraction + age
+ (1|item) + (1|participant) + (1+condition|participant) + (1+l1|item)
```

```
er ~ l1*condition + trial + length + distraction + age
+ (1|item) + (1|participant) + (1+condition|participant) + (1+l1|item)
```

It must be noted that Gyllstad and Wolter (2016) used a backward step-wise regression, which involves running a maximal model and then reducing it by eliminating the covariate with the lowest absolute t-value (or any other metric specified by the researcher), until only significant covariates remain. Given that the sample size of the present study was greater than the number of fixed factors, an analysis in this vein could have been pursued. However, as step-wise regression is dependent on sampling error, the results generated are idiosyncratic and not easily replicable in other datasets (Menard, 2002). Moreover, for research in the social sciences and education,

Lewis (2007) suggests that hierarchical regression is the "appropriate tool" due to the covariates often being correlated with each other. In contrast with backward step-wise regression, the covariates in a hierarchical regression are added to the model in a bottom-up fashion, and their inclusion is determined by the researcher before the analysis is conducted (*ibid.*). This is valuable as it allows for models to be built with a grounding in theory and research, and not through arbitrary mathematical criteria. Although this procedure was not utilized in this study, Lewis's recommendations were followed, and only those covariates (fixed factors) that were theoretically or practically motivated were included in the models.

Specifying Probability Distributions and Link Functions

In order to run the maximal models, it was necessary to first specify the probability distribution. RT data are generally unimodal and positively skewed, which is characteristic of chronometric measurements, as they have a minimum but no maximum value (Ratcliff, 1993). A number of distributions such as the ex-Gaussian, ex-Wald, Shifted Wald, Gamma, and Inverse Gaussian can be applied to approximate such data (*ibid.*). However, there is no consensus in the field as to the most optimal distribution (*ibid.*). Therefore, to approximate an appropriate distribution, an *ad hoc* procedure was used.

For the RT analysis, only "correct" responses[6] were analysed, following Gyllstad and Wolter (2016). All "incorrect" responses (4,677 of 28,802; 1.6%) were eliminated. This resulted in a dataset comprising 24,125 observations. The observed RTs had a unimodal distribution with a positive skew, same as the untrimmed dataset (see figure 3.3).

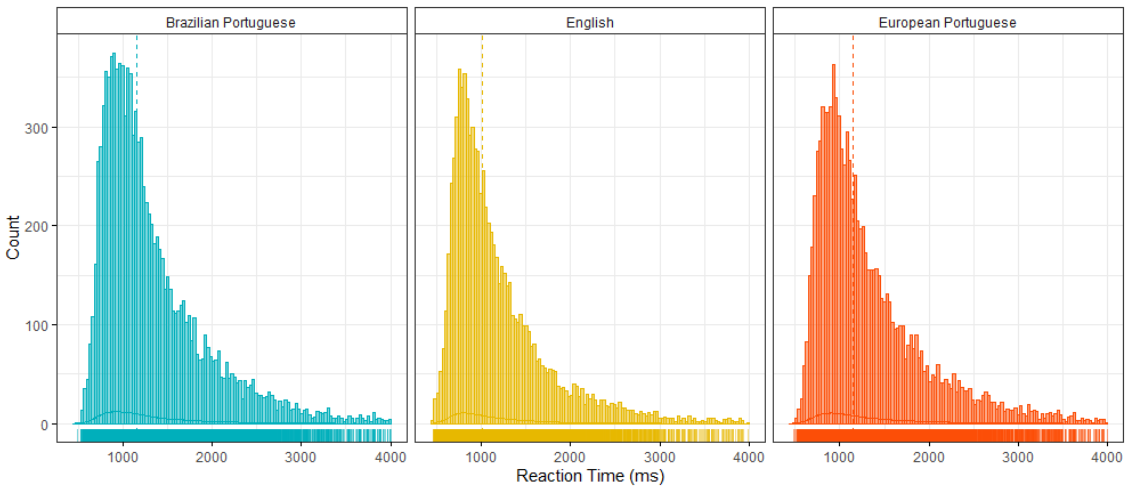


Figure 3.3. Distribution of Raw Reaction Time (ms)

(Dotted line indicates group mean)

[6]It must be noted that this was a SAJT, wherein the participants were asked to judge if they felt that the word combinations on the screen were acceptable in English or not, which means that there is no real "correct" or "incorrect" answer.

Apart from a visual check for normality, an Anderson-Darling test[7] was performed. The significant result (A = 1,086, p < 0.0001) confirmed that the data were indeed not normally distributed.

A Cullen and Frey graph (see figure 3.4) was produced using the R package, 'fitdistrplus' (Delignette-Muller & Dutang, 2015). The graph showed that the observations lay within the range of the Beta distribution, close to the Gamma distribution line.

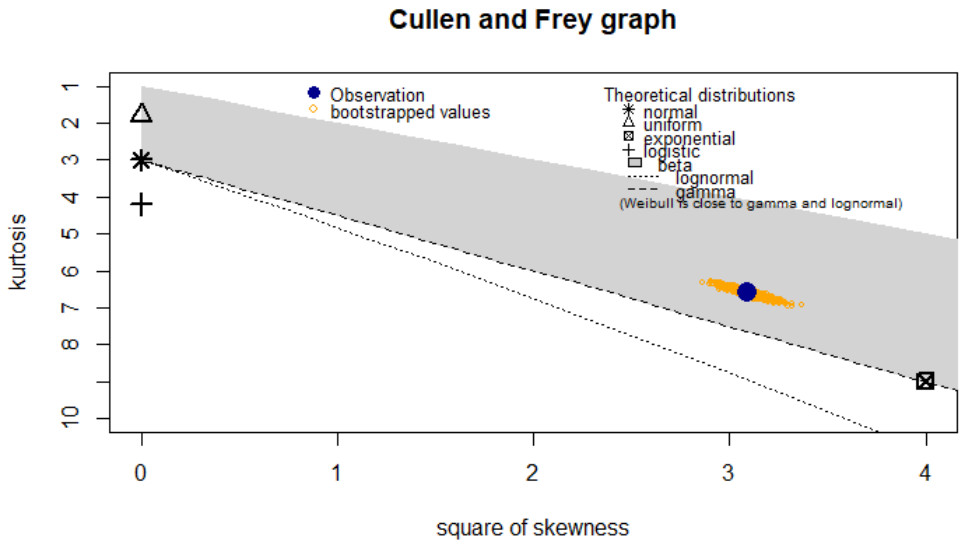


Figure 3.4. Cullen and Frey Graph

Since the Beta distribution only comprises values between 0 and 1, the data were initially assumed to approximate a Gamma distribution. The default 'stats' package (R Core Team, 2019) also implements the Inverse Gaussian distribution which can be alternately specified for use in a GLMM. Therefore, the maximal RT model was first run with the Gamma distribution, and then with the Inverse Gaussian. The two models were compared, and the latter proved to be a slightly better fit as demonstrated by the lower Akaike Information Criterion (AIC) values (see table 3.6). Therefore, the Inverse Gaussian distribution was specified in all models with RT as the dependent variable. Although this procedure leaves room for error, the GLMM framework is said to be relatively robust against "moderate misspecification" of the probability distribution[8] (McCulloch & Neuhaus, 2011).

[7] The widely used Shapiro-Wilks test for normality could not be used as the dataset was too large.

[8] It must be noted that these distributions are merely theoretical models of real-world data that allow for approximations and simplification of large amounts of data into a few parameters. As such, no theoretical distribution fits large sets of real-world data perfectly.

Model	df	logLikelihood	AIC	delta	weight
Inverse Gaussian	31	−174, 024.50	348, 111.10	0	1
Gamma	31	−175, 597.70	351, 257.40	3, 146.30	0

Table 3.6. Model Comparison: Inverse Gaussian versus Gamma Distributions

In contrast, specifying the probability distribution of the ER model was straightforward. All 28,802 observations were analysed in this model. Given that it is a binary variable, the binomial distribution—a discrete probability distribution—was specified (see Lo & Andrews, 2015). The 'logit' link function was used, following Poort and Rodd (2019).

Procedure for Dealing with Convergence Issues

Models with complex random-effects structures often do not converge (Barr et al., 2013; Singmann & Kellen, 2019). Therefore, in the case of convergence issues, the models in the present study would be simplified by first suppressing the correlations (e.g., `condition||participant`), then the random intercepts, and finally random slopes until convergence was reached. The same would be done in the case of a singular fit warning—a critical warning that indicates over-fitting of the model and reduced statistical power (Matuschek, Kliegl, Vasishth, Baayen, & Bates, 2017).

Statistical Tools

Although the 'lme4' package (Bates, Mächler, Bolker, & Walker, 2015) is widely used in psychological research, the present study used the `mixed()` function in the 'afex' package (Singmann, Bolker, Westfall, Aust, & Ben-Shachar, 2021) to run the models in R. This function uses the 'lme4' package to calculate the estimates, and then calls an ANOVA to automatically calculate p-values for the main effects (predictor variables) and interactions. In contrast, use of the standard 'lme4' package only provides p-values of the coefficients of the linear model, but not of the effects of the predictor variables themselves (Singmann et al., 2021). Additionally, 'afex' provides two methods of calculating the p-values for GLMMs—Likelihood-ratio Test (LRT), and Parametric Bootstrapping. The LRT method was chosen as it is computationally less expensive than the latter, which is strongly recommended for small sample sizes and not a concern in the case of the present study. A check for multicollinearity was performed using the `check_collinearity()` function from the 'performance' pack-

age (Lüdecke, Ben-Shachar, Patil, Waggoner, & Makowski, 2021) and any fixed factor with a Variance Inflation Factor (VIF) greater than five was eliminated from the model. Post-hoc analyses were carried out by computing the estimated marginal means (EMM), also known as least-squares means, using the package 'emmeans' (Lenth, 2021), which allows for pairwise comparisons and contrasts.

3.5.4 Modeling RT Data

Once all the components were specified, the maximal RT model was run. Although convergence was reached, the singularity check failed. Therefore, the model was simplified as per the procedure laid out in section 3.5.3, until no more singular fit warnings were encountered. The reduced model retained all the covariates and intercepts of the maximal model. The correlation of the random slope by participant was suppressed, and the random slope by item was removed altogether. A check for multicollinearity showed a moderate collinearity for trial (VIF = 6.20). Additionally, the effect of trial was not significant (χ^2 (1) = 0.45, p = 0.504), therefore it was eliminated.

The R syntax specified for the final RT model was as follows:

```
rt ~ l1*condition + length + distractions + age
+ (1|item) + (1|participant) + (condition||participant)
```

3.5.5 Modeling ER Data

Due to computational constraints, the analysis of the maximal ER model could not be completed. Therefore, this model was also simplified, based on the same procedure as the RT model. The most complex model that converged without any singularity warnings was the same as the final model specified for RTs. A multicollinearity check once again showed a moderate correlation for trial order (VIF = 6.87), which was eliminated and the model was re-run.

The R syntax specified for the final ER model was as follows:

```
er ~ l1*condition + length + distractions + age
+ (1|item) + (1|participant) + (condition||participant)
```

The results of these analyses will be discussed in the following chapter.

3.6 Ethics

The present study was approved by the University of Oxford's Central University Research and Ethics Committee (CUREC) (see Appendix A.3.1). Data collection commenced upon receipt of approval. Information outlining the nature of the study, its objectives, and procedures was provided to all potential participants. It was made clear that participation was voluntary, anonymous, and that participants had the right to withdraw at any time during the task by closing their browser tab/window.

Chapter 4

Results

This chapter presents the findings of the present study. The three L1 groups are compared based on their performance on the semantic acceptability judgement task (SAJT) in terms of the dependent variables, i.e., speed (RTs), and accuracy (ERs). Section 4.1 briefly outlines the global results and descriptive statistics. Then, section 4.2 presents the results of the two models and the post-hoc analyses contextualized by the research questions (RQs) outlined in section 3.1. Lastly, the findings are summarized in section 4.4.

4.1 Global Results

The mean reaction time (mRT) and mean accuracy (mean ER) for each L1 group was first calculated (see table 4.1). The mRT was 1,187.34 ms (SD = 561.82) for the English group, 1,328.09 ms (SD = 589.03) for the BP group, and 1,328.66 ms (SD = 607.80) for the EP group. In terms of mean accuracy, the BP group scored 83%, the English group scored 86% , and the EP group scored 83%. The ER values were transformed to percentages in order to facilitate comparison.

Figure 4.1 illustrates the average performance of each group by condition (item type), while figure 4.2 illustrates the mean accuracy of each group by condition. For a detailed report of the descriptive statistics, see table B.3 for RT, and table B.4 for ER.

Table 4.1. Mean Reaction Time and Mean Accuracy

	Reaction Time		Error Rate	
L1	Mean	SD	Mean	SD
Brazilian Portuguese (BP)	1,328.09	589.03	0.83	0.37
English (EN)	1,187.36	561.82	0.86	0.35
European Portuguese (EP)	1,328.66	607.80	0.83	0.38

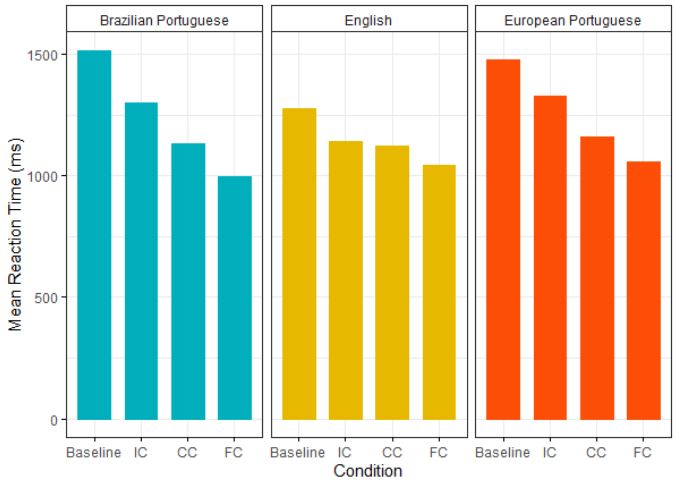


Figure 4.1. Mean RT (ms) of L1 Groups by Condition

IC: Incongruent Collocations, CC: Congruent Collocations, FC: Free Combination

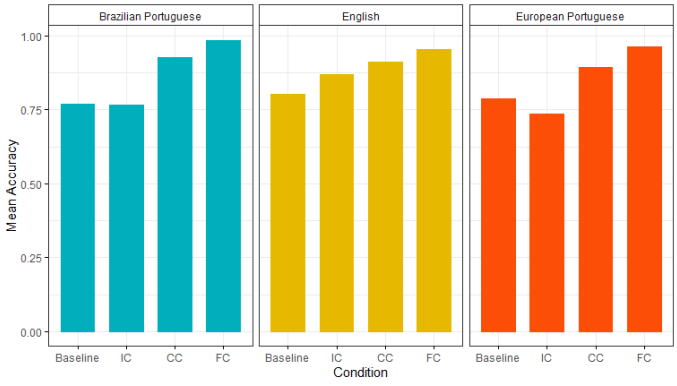


Figure 4.2. Mean Accuracy of L1 Groups by Condition

IC: Incongruent Collocations, CC: Congruent Collocations, FC: Free Combination

Given that this was an acceptability judgement task, it was deemed important to get an overview of the participants' judgements (see figure 4.3). All baseline items are unacceptable, while all items from the remaining three conditions are acceptable. However, several baseline items were found to be acceptable by all groups. Therefore, the mean accuracy of each item was calculated, and four items—*lose goals*, *build secrets*, *raise bans*, and *meet peace*—were found to have a mean accuracy of less than 60% (see figure 4.4). This would be problematic if only one of the groups were to find these items acceptable. However, as this was not the case, these items were not excluded from the analysis.

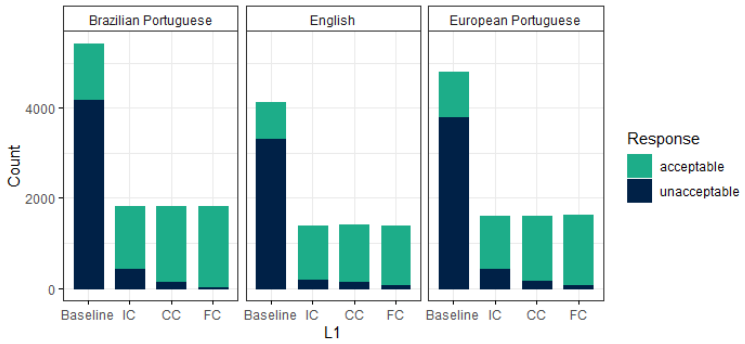


Figure 4.3. Acceptability Judgements of L1 Groups by Condition

(*Note: CC = Congruent Collocations, IC = Incongruent Collocations, FC = Free Combinations*)

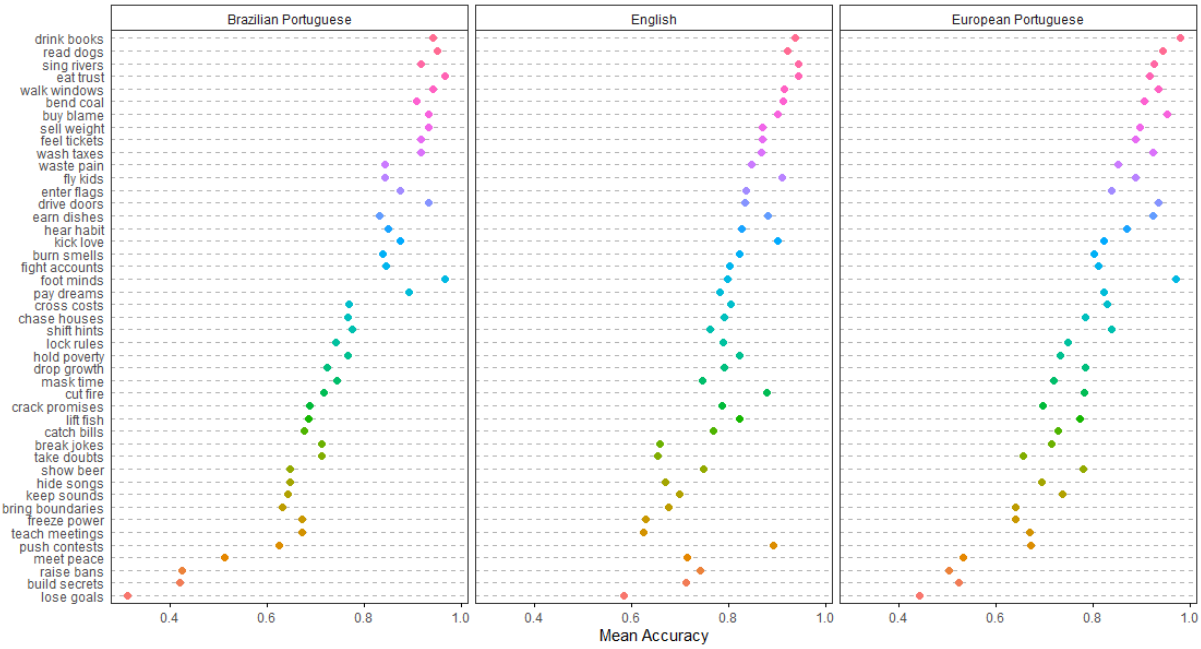


Figure 4.4. Overall Mean Accuracy of Baseline Items

For visualizations of the other conditions by accuracy, see Appendix section B.1

4.2 Model Results

The results of the final RT and ER models are presented in this section (table 4.2). For the standard 'lme4' outputs of the GLMMs, see Appendix 4.2.

In terms of the main predictor variables, the effect of L1 was not significant for speed [χ^2 (2) = 0.91, p = 0.634], but was significant for accuracy [χ^2 (2) = 16.67, p < 0.001], while the effect of condition (item type) was significant for both [RT: χ^2 (3) = 39.99, p < 0.001; ER: χ^2 (3) = 41.46, p < 0.001]. Additionally, the main effect of the interaction between L1 and condition was also significant for both speed [χ^2 (6) = 58.82, p < 0.001] and accuracy [χ^2 (6) = 100.35, p < 0.001]. Taken together, these results indicated that the three L1 groups performed in a very similar manner with respect to overall RTs, but not for ERs, and that their performance varied based on condition.

As for the covariates, the effect of age was significant for speed [χ^2 (1) = 15.84, p < 0.001], but not for accuracy [χ^2 (1) = 0.93, p = 0.334]. These results were expected, as research indicates not only that RTs increase with an increase in age (Woods, Wyma, Yund, Herron, & Reed, 2015), but also that older adults are likely to be more cautious and tend to favour accuracy rather than speed (Starns & Ratcliff, 2010). Moreover, distractions were significant in both models [RT: χ^2 (1) = 8.72, p = 0.003; ER: χ^2 (1) = 15.82, p < 0.001]. This was also expected, since distractions are known to affect RTs (Du & Clark, 2020), as well as accuracy (LaPointe, Heald, Stierwalt, Kemker, & Maurice, 2007). No significant effects were found for collocation length in either model [RT: χ^2 (1) = 2.12, p = 0.146; ER: χ^2 (1) = 0.00, p = 0.978].

Table 4.2. Mixed Model ANOVA Results for Reaction Times and Error Rates

		Reaction Time			Error Rate		
	Effect	df	Chisq	P-value	df	Chisq	P-value
1	L1	2	0.91	.634	2	16.67 ***	<.001
2	Condition	3	39.99 ***	<.001	3	41.46 ***	<.001
3	Distractions	1	8.72 **	.003	1	15.82 ***	<.001
4	Collocation Length	1	2.12	.146	1	0.00	.978
5	Age	1	15.84 ***	<.001	1	0.93	.334
6	L1*Condition	6	58.82 ***	<.001	6	100.35 ***	<.001

Significance codes: '***' p < 0.001; '**' p < 0.01; '*' p < 0.05

4.3 Post-hoc Analyses

In order to address the RQs, as well as to understand the results more clearly, post-hoc analyses were carried out by computing the estimated marginal means (EMM). The trends in the EMMs can be visualized in figure 4.5 and figure 4.6. The overall results show that free combinations are processed faster and with fewer errors than the collocations and baselines, and congruent collocations are processed faster and with fewer errors than incongruent collocations.

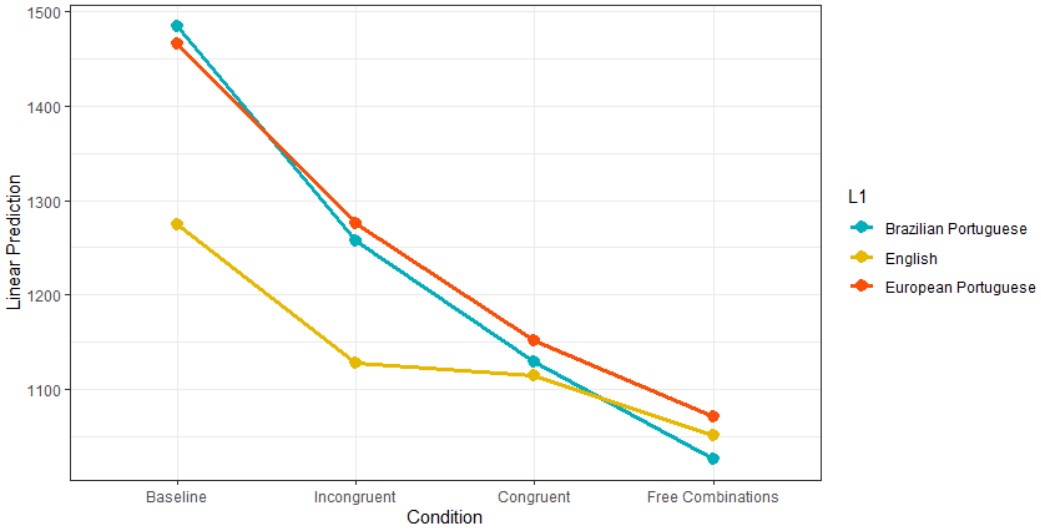


Figure 4.5. Estimated Marginal Means of Condition by L1 (RT Model)

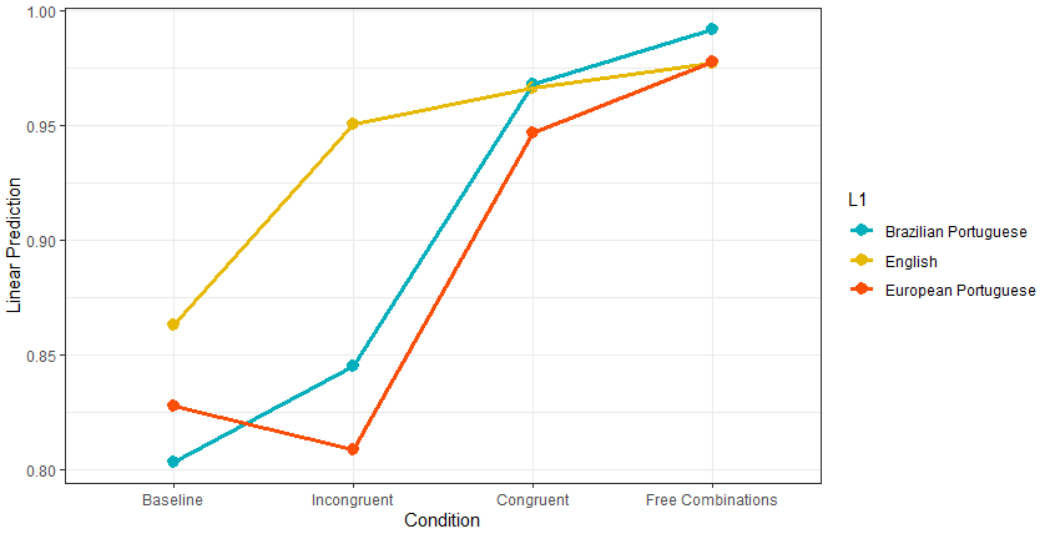


Figure 4.6. Estimated Marginal Means of Condition by L1 (ER Model)

Pairwise contrasts were conducted for the interaction terms (`l1*condition`) and condition to answer RQ 1,

RQ 2, and RQ 3 (see tables 4.3 and 4.4). For RQ 4, pairwise contrasts were conducted for the interaction terms and L1 (see tables 4.5 and 4.6). All contrasts were carried out for both models and all reported p-values were adjusted according to the Tukey method (Singmann et al., 2021).

The estimates of the RT model were presented in terms of reaction time in milliseconds. The pairwise contrasts were carried out by calculating the difference in the EMMs of the RTs of the conditions being compared. This difference is denoted by the delta symbol (Δ). Negative Δ values indicated faster reaction times, i.e., a facilitative effect for the first condition being compared over the second. Positive Δ values indicated slower reaction times, i.e., inhibitory effects for the first condition being compared over the second.

The estimates of the ER model were presented in terms of odds ratios which express the strength of associations from logistic regression when a binary outcome is being predicted, in this case, accuracy. It must be noted that odds ratios should not be equated to probabilities (Norton, Dowd, & Maciejewski, 2018). In all cases, when comparing two estimates, an odds ratio greater than one indicated that the odds of being correct on the first condition are higher than that of the second. Conversely, an odds ratio of less than one indicated that the odds of being correct on the first condition are lower than that of the second condition. However, to facilitate interpretation, all discussions that involved an odds ratio value less than 1 were reversed to express the ratio in terms of higher accuracy (lower ERs). For example, an odds ratio value of 2 indicates that the odds of being correct on the first condition are twice that of the second condition.

4.3.1 RQ 1: Free Combinations versus Congruent Collocations

RQ 1 asked whether there would be a statistically significant difference in processing time and accuracy for L1 Portuguese speakers of L2 English for free combinations versus congruent collocations, and it was hypothesized that there would be a significant difference between the two conditions. A statistically significant difference was found, but only for the BP group.

The results show that free combinations were processed significantly faster than congruent collocations by the BP group [$p = 0.0054$, $\Delta = -103.22$ ms, standard error (SE) = 31.31], but not by the EP group [$p = 0.0597$, $\Delta = -80.75$ ms, SE = 32.28]. Similar results were found for accuracy. The EP group showed no significant differences ($p = 0.2606$) between free combinations and congruent collocations. However, the BP group showed a small, but significant difference with an odds ratio of 4.00 (SE = 2.10, $p = 0.0404$) favoring the free combinations.

A significant difference was found for speed, and accuracy for the BP group accuracy, but not for the EP group.

4.3.2 RQ 2: Free Combinations versus Incongruent Collocations

RQ 2 asked whether there would be a statistically significant difference in processing time and accuracy for L1 Portuguese speakers of L2 English for free combinations versus incongruent collocations, and it was hypothesized that there would be a significant difference between the two conditions. A statistically significant difference was found.

Both groups processed free combinations significantly faster than incongruent collocations (BP: $p < 0.0001$, $\Delta = -231.18$ ms, SE = 32.95 ; EP: $p < 0.0001$, $\Delta = -204.98$ ms, SE = 34.04), and with significantly lower ERs. The odds ratio of the BP group was 22.26 ($p < 0.0001$, SE = 11.61), while that of the EP group was 10.46 ($p < 0.0001$, SE = 5.22), both indicating that free combinations were processed more accurately than incongruent collocations.

The L2 speakers responded faster and more accurately to the free combinations than the incongruent collocations. Therefore, the null hypothesis was rejected.

4.3.3 RQ 3: Congruent Collocations versus Incongruent Collocations

RQ 3 asked whether there would be a statistically significant difference in processing time and accuracy for L1 Portuguese speakers of L2 English for congruent versus incongruent collocations, and it was hypothesized that there would be a significant difference between the two conditions. A statistically significant difference was found.

The pairwise comparisons of the two congruency conditions show that congruent collocations were processed significantly faster than incongruent ones by both groups (BP: $p < 0.0001$, $\Delta = -127.95$ ms, SE = 29.64 ; EP: $p = 0.0002$, $\Delta = -124.23$ ms, SE = 30.35). In terms of accuracy, the odds ratio of the BP group was 5.56 ($p < 0.0002$, SE = 2.67), while that of the EP group was 4.20 ($p = 0.0142$, SE = 2.01), both indicating that congruent collocations were processed more accurately than incongruent collocations.

The L2 speakers responded faster and more accurately to the congruent collocations than the incongruent collocations. Therefore, the null hypothesis was rejected.

4.3.4 Baseline Items

In terms of RT, significant differences ($p < 0.0001$) between the baselines and the other three conditions were found for both L2 groups. The differences in estimates indicated that free combinations (BP: $p < 0.0001$ $\Delta = -458.29$ ms, SE = 30.45 ; EP: $p < 0.0001$ $\Delta = -395.83$ ms, SE = 31.61), congruent collocations (BP: $p < 0.0001$, Δ

= −355.07 ms, SE = 25.60 ; EP: $p < 0.0001$, $\Delta = -315.09$ ms, SE = 26.19), and incongruent collocations (BP: $p < 0.0001$, $\Delta = -227.12$ ms, SE = 27.00 ; EP: $p < 0.0001$, $\Delta = -190.86$ ms, SE = 27.71) were all processed faster than the baseline items.

However, for accuracy, only free combinations and congruent collocations showed significantly more accurate responses than baselines. The odds ratio of the BP group was 29.70 ($p < 0.0001$, SE = 13.34), while that of the EP group was 9.22 ($p < 0.0001$, SE = 3.92) for free combinations versus baselines. The odds ratio of the BP group was 7.42 ($p < 0.0001$, SE = 3.09), while that of the EP group was 3.71 ($p < 0.0090$, SE = 1.54) for congruent collocations versus baselines.

No significant differences between incongruent collocations and baselines were found for either group.

Table 4.3. Pairwise Contrasts of Condition by L1 (RT Model)

L1	Contrast	Estimate (Δ)	SE	Z-ratio	P-value
Brazilian Portuguese	FC – CC	−103.22	31.31	−3.297	0.0054**
	FC – IC	−231.18	32.95	−7.017	< 0.0001***
	CC – IC	−127.95	29.64	−4.316	0.0001***
	FC – BL	−458.29	30.45	−15.051	< 0.0001***
	CC – BL	−355.07	25.60	−13.872	< 0.0001***
	IC – BL	−227.12	27.00	−8.411	< 0.0001***
English	FC – CC	−62.69	33.00	−1.900	0.2281
	FC – IC	−77.13	34.12	−2.261	0.1073
	CC – IC	−14.44	29.79	−0.485	0.9625
	FC – BL	−223.97	32.54	−6.884	< 0.0001***
	CC – BL	−161.28	26.30	−6.132	< 0.0001***
	IC – BL	−146.84	27.02	−5.435	< 0.0001***
European Portuguese	FC – CC	−80.75	32.28	−2.501	0.0597
	FC – IC	−204.98	34.04	−6.022	< 0.0001***
	CC – IC	−124.23	30.35	−4.093	0.0002***
	FC – BL	−395.83	31.61	−12.521	< 0.0001***
	CC – BL	−315.09	26.19	−12.031	< 0.0001***
	IC – BL	−190.86	27.71	−6.887	< 0.0001***

FC: Free combinations; CC: Congruent Collocations; IC: Incongruent Collocations; BL: Baseline Items

Degrees of Freedom (df) = Inf

Results are averaged over the levels of: Distractions

P-value adjustment: Tukey method for comparing a family of 4 estimates

Significance codes: '***' p < 0.001; '**' p < 0.01; '*' p < 0.05

Table 4.4. Pairwise Contrasts of Condition by L1 (ER Model)

L1	Contrast	Odds Ratio	SE	Z-ratio	P-value
Brazilian Portuguese	FC/CC	4.00	2.10	2.648	0.0404**
	FC/IC	22.26	11.61	5.948	< 0.0001***
	CC/IC	5.56	2.67	3.572	0.0020**
	FC/BL	29.70	13.34	7.551	< 0.0001***
	CC/BL	7.42	3.09	4.814	< 0.0001***
	IC/BL	1.33	0.55	0.698	0.8978
English	FC/CC	1.49	0.75	0.793	0.8577
	FC/IC	2.22	1.12	1.578	0.3910
	CC/IC	1.49	0.73	0.818	0.8460
	FC/BL	6.72	2.87	4.466	< 0.0001***
	CC/BL	4.52	1.92	3.542	0.0022**
	IC/BL	3.03	1.31	2.577	0.0489**
European Portuguese	FC/CC	2.49	1.24	1.827	0.2606
	FC/IC	10.46	5.22	4.699	< 0.0001***
	CC/IC	4.20	2.01	3.003	0.0142*
	FC/BL	9.22	3.92	5.231	< 0.0001***
	CC/BL	3.71	1.54	3.145	0.0090**
	IC/BL	0.88	0.37	−0.301	0.9905

FC: Free combinations; CC: Congruent Collocations; IC: Incongruent Collocations; BL: Baseline Items

Degrees of Freedom (df) = Inf; Null = 1.00

Results are averaged over the levels of: Distractions

P-value adjustment: Tukey method for comparing a family of 4 estimates

Tests are performed on the log odds ratio scale

Significance codes: '***' p < 0.001; '**' p < 0.01; '*' p < 0.05

4.3.5 RQ 4: L1 versus L2

RQ 4 asked whether there would be a statistically significant difference in processing time and accuracy between the L1 and L2 groups. The alternate hypothesis was that there would be a difference between the groups. Significant differences were found mainly in the incongruent and baseline conditions.

In terms of the post-hoc analyses conducted for the first three RQs, the overall performance of L1 group differed from the L2 groups. For both speed and accuracy, there were no significant differences between free combinations versus congruent collocations (RT: $p = 0.2281$; ER: $p = 0.8577$), free combinations versus incongruent collocations (RT: $p = 0.1073$; ER: $p = 0.3910$), and incongruent versus congruent collocations (RT: $p = 0.9625$; ER: $p = 0.8460$). However, their responses were significantly faster ($p < 0.0001$) for free combinations ($p < 0.0001$, $\Delta = -223.97$; SE = 32.54), congruent collocations ($p < 0.0001$, $\Delta = -161.28$; SE = 26.30), and incongruent collocations ($p < 0.0001$, $\Delta = -146.84$; SE = 27.02) when compared to baselines. In contrast, there were varying degrees of significant differences for accuracy. Free combinations were 6.72 times ($p < 0.0001$, SE = 2.87) more likely to elicit an accurate response than baselines, while congruent collocations were 4.52 times ($p = 0.0022$, SE = 1.92), and incongruent collocations were 3.03 times ($p = 0.0489$, SE = 1.31) more likely to elicit an accurate response than baselines.

The trends in the EMMs calculated for the L1 groups can be visualized in figure 4.7 for speed and in figure 4.8 for accuracy.

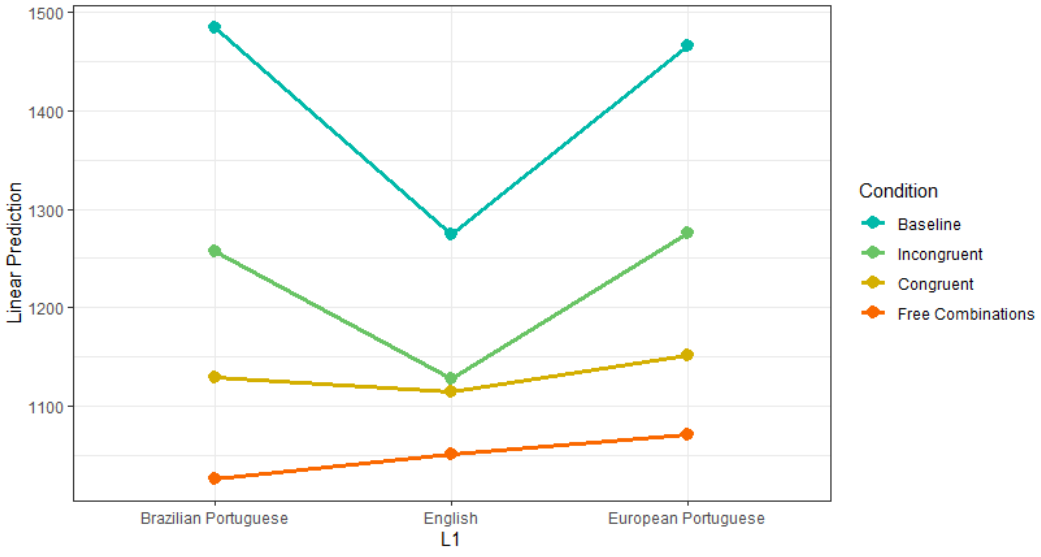


Figure 4.7. Estimated Marginal Means of L1 by Condition (RT Model)

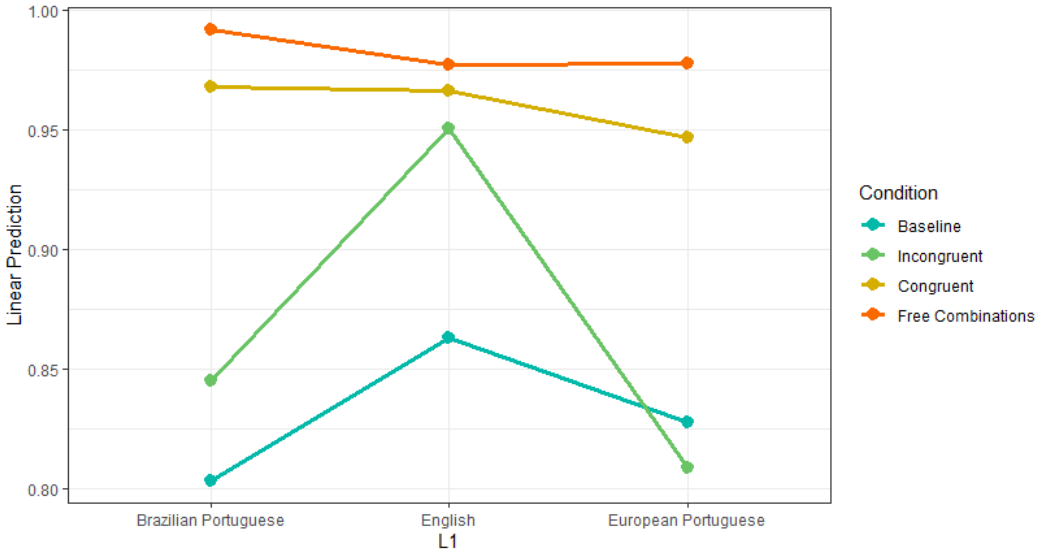


Figure 4.8. Estimated Marginal Means of L1 by Condition (ER Model)

A pairwise contrast of the interaction term (`l1*condition`) by L1 was carried out for both models (see tables 4.5 and 4.6). Significant differences between the L1 and L2 groups were found for incongruent collocations in terms of speed and accuracy. On average, the L1 group was 129.21 ms faster than the BP group ($p < 0.0001$, SE = 24.95), and 147.15 ms faster than the EP group ($p < 0.0001$, SE = 25.84). The odds ratios in favor of the L1 group are 3.45 versus the BP group ($p < 0.0001$, SE = 0.06), and 4.55 versus the EP group ($p < 0.0001$, SE = 0.05).

No significant differences between the L1 and L2 groups were found for congruent collocations ($p > 0.05$) for both speed and accuracy. However, while no differences were found for speed in terms of free combinations ($p > 0.05$), a significant difference was found for accuracy, but only between the L1 and the BP group which was 2.86 times ($p = 0.0004$, SE = 0.79) more accurate.

In terms of baselines, the L1 speakers outperformed both the L2 groups in terms of speed (BP–EN: $p < 0.0001$, $\Delta = -209.49$, SE = 25.18; EP–EN: $p < 0.0001$, $\Delta = -191.17$, SE = 25.84), but not in terms of accuracy ($p > 0.05$). These results suggested that the L2 speakers were slower than the L1 group, but not any less accurate when processing baseline items.

These results only provide a partial answer to RQ 4.

Table 4.5. Pairwise Contrasts of L1 by Condition (RT Model)

Condition	Contrast	Estimate (Δ)	SE	Z-ratio	P-value
Free Combinations	BP – EN	−24.84	28.18	−0.881	0.6521
	EP – EN	19.31	29.14	0.663	0.7852
	BP – EP	−44.14	27.24	−1.621	0.2368
Congruent Collocations	BP – EN	15.69	23.21	0.676	0.7774
	EP – EN	37.36	24.01	1.556	0.2648
	BP – EP	−21.67	22.62	−0.958	0.6036
Incongruent Collocations	BP – EN	129.21	24.95	5.179	< 0.0001***
	EP – EN	147.15	25.84	5.694	< 0.0001***
	BP – EP	−17.94	25.16	−0.713	0.7558
Baseline Items	BP – EN	209.49	25.18	8.321	< 0.0001***
	EP – EN	191.17	25.84	7.398	< 0.0001***
	BP – EP	18.32	24.72	0.741	0.7391

BP: Brazilian Portuguese; EP: European Portuguese; EN: English

Degrees of Freedom (df) = Inf

Results are averaged over the levels of: Distractions

P-value adjustment: Tukey method for comparing a family of 3 estimates

Significance codes: '***' p < 0.001; '**' p < 0.01; '*' p < 0.05

Table 4.6. Pairwise Contrasts of L1 by Condition (ER Model)

Condition	Contrast	Odds Ratio	SE	Z-ratio	P-value
Free Combinations	BP / EN	2.86	0.79	3.820	0.0004***
	EP / EN	1.04	0.24	0.183	0.9817
	EP / BP	0.36	0.10	-3.716	0.0006***
Congruent Collocations	BP / EN	1.06	0.23	0.287	0.9557
	EP / EN	0.62	0.14	-2.168	0.0768
	EP / BP	0.59	0.12	-2.609	0.0247**
Incongruent Collocations	BP / EN	0.29	0.06	-5.708	< 0.0001***
	EP / EN	0.22	0.05	-6.706	< 0.0001***
	EP / BP	0.78	0.16	-1.263	0.4162
Baseline Items	BP / EN	0.65	0.12	-2.300	0.0559
	EP / EN	0.76	0.15	-1.409	0.3361
	EP / BP	1.17	0.21	0.890	0.6467

BP: Brazilian Portuguese; EP: European Portuguese; EN: English

Degrees of Freedom (df) = Inf; Null = 1.00

Results are averaged over the levels of: Distractions

P-value adjustment: Tukey method for comparing a family of 3 estimates

Tests are performed on the log odds ratio scale

Significance codes: '***' p < 0.001; '**' p < 0.01; '*' p < 0.05

4.4 Summary

The findings suggest that both L2 groups process free combinations and congruent collocations faster and more accurately than incongruent collocations. However, discrepancies arise when free combinations are compared to congruent collocations. The BP group processed free combinations significantly faster and more accurately than congruent collocations. No significant differences were found for the EP group. In terms of speed, both L2 groups were faster on all the conditions in comparison to the baselines. In terms of accuracy, their ER was lower when responding to free combinations and congruent collocations in comparison to baseline items, but no significant difference was found between incongruent collocations and baselines. Table 4.7 provides a simplified overview of the findings.

The L2 groups' performance was on par with the L1 group in the congruent condition but the L1 group performed better in the incongruent and baseline conditions, for both speed and accuracy. For free combinations, no significant differences were found between the L1 and L2 groups in terms of speed, but, surprisingly, the BP group significantly outperformed the L1 English speakers in terms of accuracy. No such differences were found for the EP group. A simplified summary of these findings can be found in table 4.8.

Overall, the results suggest that there could be an effect of semantic transparency on processing speed for the L2 speakers, but not for the L1 speakers. L2 speakers process free combinations faster than congruent collocations, which in turn are processed faster than incongruent collocations. The L1 speakers show no such difference. The effect of congruency, although visibly present in the RTs, is more noticeable in the ERs, especially highlighted by the fact that the L2 groups performed similarly on the incongruent and baseline conditions. These results only partially support the findings reported by Gyllstad and Wolter (2016). The following chapter will discuss these findings in more detail.

Table 4.7. Summary of Comparisons by Condition

DV	Brazilian Portuguese (BP)	English (EN)	European Portuguese (EP)
RT	FC > CC FC > IC CC > IC	FC = CC FC = IC CC = IC	FC = CC FC > IC CC > IC
RT	FC > BL CC > BL IC > BL	FC > BL CC > BL IC > BL	FC > BL CC > BL IC > BL
ER	FC > CC FC > IC CC > IC	FC = CC FC = IC CC = IC	FC = CC FC > IC CC > IC
ER	FC > BL CC > BL IC = BL	FC > BL CC > BL IC > BL	FC > BL CC > BL IC = BL

FC: Free combinations; CC: Congruent Collocations; IC: Incongruent Collocations; BL: Baseline Items

">" = Significantly faster than/more accurate than; "=" = No significant differences found

Table 4.8. Summary of L1 Comparisons

DV	Free Combinations	Congruent Collocations	Incongruent Collocations	Baseline Items
RT	EN = BP EN = EP	EN = BP EN = EP	EN > BP EN > EP	EN > BP EN > EP
ER	EN < BP EN = EP	EN = BP EN = EP	EN > BP EN > EP	EN = BP EN = EP

BP: Brazilian Portuguese; EN: English; EP: European Portuguese

">":Significantly faster than/more accurate than; "<": Significantly slower than/less accurate than; "=": No significant differences found

Chapter 5

Discussion

This chapter discusses and interprets the results described in chapter 4. It is divided into five sections. Section 5.1 discusses the results of RQ 1, RQ 2, and RQ 3 within the context of Gyllstad and Wolter's (2016) findings, as well as other related studies. In section 5.2, the results of the L1 groups performance are briefly discussed. Section 5.3 then discusses the results of RQ 4, by comparing the L2 groups' performance to that of the L1 group. The limitations of the present study are detailed in section 5.4 and suggestions for future studies are also provided. Finally, section 5.5 makes recommendations for teaching collocations in the L2 classroom.

5.1 Processing Collocations in the L2

The aim of this study was to investigate the effects of semantic transparency and congruency on collocational processing in L1 Portuguese speakers of L2 English. The data have been analysed and the findings will be discussed through the adapted theoretical framework of Howarth's Collocational Continuum Model (see figure 2.2), and within the context of Gyllstad and Wolter's findings. Wherever pertinent, the findings from other studies will be drawn on, in order to contextualize these results, potentially adding another piece to the puzzle that is L2 collocational processing.

To recap, the results of both L2 groups indicated that congruent collocations as well as free combinations showed processing advantages over incongruent collocations. However, only the BP group showed processing advantages for free combinations over congruent collocations. This same advantage was not observed for the EP group. Moreover, the L1 group showed no processing advantages for any of the three main conditions. These findings pertinent to RQ 1, RQ 2, and RQ 3 can be summarized by the following statements:

$$BP: \textit{Free Combinations} > \textit{Congruent Collocations} > \textit{Incongruent Collocations} \quad (5.1)$$

$$EN: \textit{Free Combinations} = \textit{Congruent Collocations} = \textit{Incongruent Collocations} \quad (5.2)$$

$$EP: \textit{Free Combinations} = \textit{Congruent Collocations} > \textit{Incongruent Collocations} \quad (5.3)$$

">": faster than/more accurately than; "=": no significant difference

5.1.1 Free Combinations versus Congruent Collocations

Gyllstad and Wolter (2016) reported processing advantages (faster speed and higher accuracy) for free combinations over congruent collocations in both L1 and L2 (L1 Swedish) groups in their study. They posit that their results lend support to Howarth's descriptive typologies at least for free combinations versus congruent collocations. In comparison, the findings of the present study are not as clear cut. As can be seen from statements 5.1, 5.2, and 5.3, the BP group showed processing advantages for free combinations over congruent collocations, but the EP and L1 groups did not. This reveals two discrepancies when compared to Gyllstad and Wolter (2016): (i) the lack of a processing advantage for free combinations over congruent collocations for the EP and L1 groups, and (ii) the dissimilarity between the BP and the other groups.

The first discrepancy could probably be explained by frequency. All the items used in this study were highly frequent in both English and Portuguese. Wolter and Gyllstad (2013) found that higher frequency incongruent collocations are processed faster than low-frequency congruent collocations suggesting that frequency can undermine the effect of congruency. Seen in this context, the findings of the present study suggest that frequency can also undermine the effect of semantic transparency. This is directly in line with the usage-based approaches to language acquisition.

In order to address the second discrepancy, there are three details that must be noted: (i) despite being significantly more proficient than the EP group, the BP speakers' performance was not similar to that of L1 group, (ii) the L2 groups were on par with one another on every other contrast, and (iii) the BP group were more accurate (but not faster) than both the EP group and, surprisingly, the L1 group on the free combinations. Taken together, these details indicate that the locus of the discrepancy in the BP speakers' performance lies in the free combinations. This result was unexpected. By the law of large numbers, the BP group's performance should have been 'average', so to speak. However, their performance was significantly better than even the L1 speakers, which points to the discrepancy being potentially driven by the nature of the BP sample itself.

Sampling validity is a key issue in the social and behavioral sciences. The procedure used to recruit participants can introduce different types of biases into the results of a study, reducing the external validity of the sample (Emerson, 2015; McEwan, 2020). Given the methods used for data collection (see section 3.4), the samples in this study cannot be considered truly random. The BP group's data was mainly collected through forwarded emails and social media posts. In contrast, much of the EP group's data was collected by professors at various Portuguese universities sharing the link to the experiment in the classroom. The BP participants actively and freely chose to participate in the experiment, while the EP speakers possibly felt an obligation to participate. These differences in recruitment and in inclinations to participate may have lead to a selection bias in the sample which could be reflected in the performance of the two groups and thereby the findings. Other possible explanations for this discrepancy could be that the BP group were simply faster at processing language in general, or that there could have been a difference in the criteria applied for judging an item as acceptable (B. Wolter, personal communication, 21/08/2021).

As an interesting side note, in an attempt to find an explanation for the BP group's anomalous performance on the free combinations, a model with log-transformed RTs was run. The L2 groups were on par with one another on all contrasts in this model. Furthermore, the BP group outperforming the L1 group in terms of speed was also completely obscured in these results (see Appendix section B.3). This clearly underlines the problems discussed in section 3.5.2 that arise from transforming chronometric data. If the present study were to have log-transformed the RTs, these discrepancies between the groups would not have come to light, and therefore would not have instigated the critical reflection surrounding the implications of the methodological choices of this study.

5.1.2 Free Combinations versus Incongruent Collocations

This study extended Gyllstad and Wolter (2016) by testing incongruent collocations. As can be seen from statements 5.1 and 5.3, free combinations show a processing advantage over incongruent collocations for both L2 groups. This is not surprising given that free combinations are transparent, while collocations are opaque. However, the processing advantages hypothesized for free combinations over congruent collocations were not clearly observed in the findings. Therefore, the processing advantages seen here cannot be directly attributed to semantic transparency alone.

The free combinations used in this study are not only completely transparent, but also congruent in Portuguese, which unintentionally resulted in a comparison of congruent free combinations to incongruent collocations. Although the findings suggest that there is an effect of semantic transparency, wherein free combinations have a processing advantage over incongruent collocations, the present study was unable to fully determine the extent of that effect, as the conditions being compared also differed in terms of congruency. Follow-up studies should

compare incongruent collocations to incongruent free combinations (see figure 5.1). If incongruent free combinations were found to be processed faster and more accurately than incongruent collocations by advanced L2 speakers, then the processing advantages could be more directly attributable to semantic transparency.

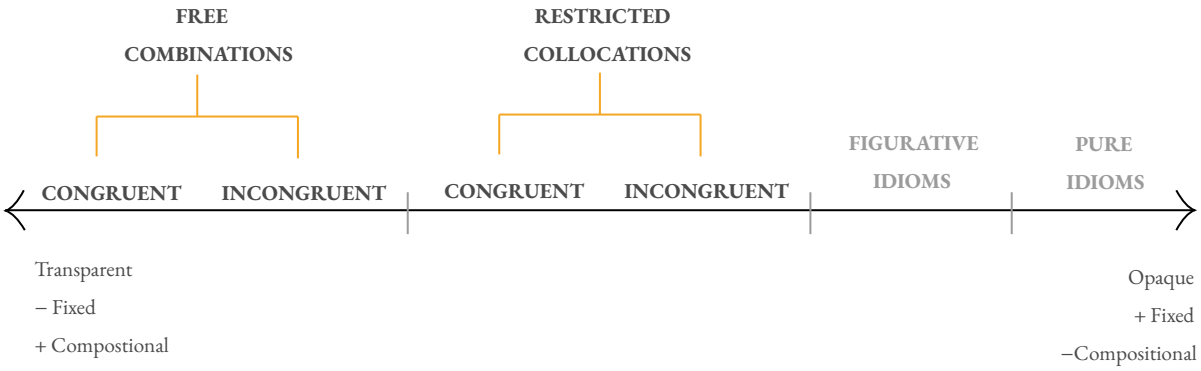


Figure 5.1. Howarth's Continuum with Congruency

5.1.3 Congruent Collocations versus Incongruent Collocations

Both L2 groups show a clear processing advantage for congruent over incongruent collocations. These findings are consistent with the growing body of evidence discussed in chapter 2 (e.g., Wolter & Gyllstad, 2011, 2013; Yamashita & Jiang, 2010) which points to the facilitative role of the L1 when recognizing collocations in the L2. Similar evidence has also been found for congruent versus incongruent idioms (see Carrol & Conklin, 2014; Cieślicka & Heredia, 2017; Titone, Columbus, Whitford, Mercier, & Libben, 2015,).

In an attempt to explain the effect of congruency on L2 collocational processing, Wolter and Gyllstad (2013) ventured that it could be due to the effects of AoA/OoA—the order in which words are acquired has an influence on how fast their representations are activated, independently of frequency (Carroll & White, 1973). Wolter and Yamashita (2015) and Wolter and Yamashita (2018) also posit the same. Numerous studies with single-word units have shown that words which are learned earlier are easier to access than words that are learned later (e.g., Izura et al., 2011; Stewart & Ellis, 2008). The effects were observed to manifest even between high-frequency/late-acquired words and low-frequency/early-acquired words (Stadthagen-Gonzalez, Bowers, & Damian, 2004). Arnon, Mccauley, and Christiansen (2017) found that speakers are sensitive to the OoA of multi-word units. They argue that multi-word units acquired early leave traces in the adult lexicon and are important building blocks of language, just like single-word units, thereby challenging the long-held ontological distinction between single-word and multi-word units.

The findings of the present study seem to lend support to the AoA/OoA hypotheses as an explanation for

the effect of congruency, given that only high frequency collocations were used in both the congruent and the incongruent conditions, and that no significant effect of frequency was found for RTs and ERs. The congruent collocations were conceptually encountered earlier in the L1 by L2 speakers. Therefore, they were probably able to access them faster than the equally frequent incongruent collocations that were encountered later (or quite possibly never). However, Wolter and Gyllstad (2013) report that high frequency incongruent collocations are processed faster in the L2 than low-frequency congruent collocations. This contradicts the findings reported by Stadthagen-Gonzalez et al. (2004) for single words and indicates that frequency and congruency may interact differently with multi-word units in comparison to single words—an avenue for future research.

5.1.4 Baselines

Up until now, the discussion has revolved around the three main conditions, and little has been said about the baseline items, which elicited interesting results in the L2 groups. A processing cost (slower RTs and higher ERs) for baselines over congruent collocations and free combinations was observed. Despite being high-frequency items, incongruent collocations were processed faster, but not more accurately than the baselines. These findings can be plausibly explained by the possibility of incongruent collocations being just as "nonsensical" as the baseline items to an L2 speaker when encountered without context, as in this experiment. To illustrate, compare the incongruent collocation *foot bills* to the baseline item *sing rivers*. If an L2 learner is unaware of the usage of the verb *to foot*, then they are likely to erroneously judge the item as unacceptable just as they would *sing rivers*.

This finding regarding the baselines and incongruent collocations is only partially in keeping with Wolter and Yamashita (2015) which found no processing advantages for Japanese speakers, and Wolter and Gyllstad (2011, 2013) which reported processing advantages for incongruent collocations over baselines for Swedish speakers. Wolter and Yamashita (2015) posit that the discrepancy can be attributed to the greater linguistic distance between Japanese and English versus Swedish and English, and greater exposure to English in Sweden versus Japan. Like Swedish, Portuguese uses the same orthography, word order, and shares a large number of cognates with English. However, exposure to L2 input in day-to-day life is not as widespread in Portugal, and even less so in Brazil, when compared to Sweden. When taken in this context, the findings of the present study suggest that the Portuguese speakers performance lies in between that of the Swedish and Japanese speakers, further underlining the roles of frequency and congruency in L2 acquisition.

5.2 Processing Collocations in the L1

The L1 speakers showed no processing advantages for free combinations when compared to either of the collocational conditions. Although the intention was to use the L1 group as a control against which the L2 groups could

be compared, it is important to address these findings as they suggest that Howarth's descriptive typology has no psychological reality for collocational processing in the L1. In contrast, the L1 group in Gyllstad and Wolter (2016) showed processing advantages for free combinations over congruent collocations, supporting Howarth's typology.

This discrepancy could be explained by the fact that the present study used only high-frequency items with strong association scores. Most of the constituent words came from the K-1 and K-2 levels of Paul Nation's BNC-COCA list. Usage-based models such as those put forth by Ellis (2008) posit that frequent exposure leads to chunking, i.e., holistic storage, retrieval, and access in short-term and long-term memory. Studies investigating compositionality have shown that familiarity, which comes from frequent exposure, plays an important role in determining processing advantages for compositional and non-compositional idioms in comparison to matched novel phrases (e.g., Tabossi, Fanari, & Wolf, 2009). According to Ellis, Simpson-Vlach, and Maynard (2008), multi-word units are encountered and used more frequently by L1 speakers and are therefore processed faster than matched novel language. The free combinations used in the present study were compositional, and both the collocations were semi-compositional, but all three sets of items were highly frequent in English and therefore presumably highly familiar to the L1 speakers. Therefore, if any effects of semantic transparency were to have manifested, they were inadvertently drowned out by the overwhelming effect of frequency.

5.3 L1 versus L2

In terms of RQ 4, it was hypothesized that there would be differences between the L1 and L2 groups for all four experimental conditions. This was based on the assumption that the BP and EP speakers were EFL learners and therefore their behavior would be more in line with Wolter and Yamashita's (2015; 2018) Japanese speakers than Gyllstad and Wolter's (2016) Swedish speakers. However, the L1 and L2 groups were on par with one another for congruent collocations and free combinations, but the L1 group outperformed the L2 group on the incongruent and baseline conditions. When considered in tandem with the findings that L2 speakers show processing advantages for both free combinations and congruent collocations over incongruent collocations, the impact of the L1 on L2 collocational processing becomes even more apparent. Incongruent collocations are a serious hurdle for even proficient L2 speakers.

In terms of Howarth's continuum, these findings suggest that when frequency is controlled, there may be a tipping point at which semantic transparency begins to affect L2 collocational processing. However, this is merely a speculation as this study was unable to fully isolate the effects of semantic transparency and congruency.

5.4 Limitations & Future Work

One of the main limitations of this study was that the data were collected online and therefore the environment in which the participants' performed the task could not be controlled. Given that a wide net was cast in order to ensure that the sample size was large enough for valid statistical analyses, this was inevitable. Some participants were administered the task in the classroom, while others happened upon a link to the study on social media and might have performed the task at home or in a café. This is problematic since environmental distractions could affect processing. An effort was made to address this by adding distractions as a covariate to all the models, and as can be seen from the results, there was a significant effect for distractions in all the analyses.

Another limitation that has already been mentioned is the unreliability of self-reported data. Apart from the curious number of participants claiming to be ambidextrous in the English group, another issue that arose was the definition of L1, or what it means to be a native speaker. The socio-demographic questionnaire used in this study asked the participants to indicate their L1 with the following options: English, Portuguese (European), Portuguese (Brazil), Other (Please specify). This provided no opportunity for simultaneous bilinguals to indicate their other language(s), thereby creating an incomplete picture of the participants' language profile. Moreover, in a private conversation with the researcher, one participant stated that they reported being an L1 English speaker, since English was their "dominant language" and as they "[thought] in English". However, the participant in question had moved to the UK at the age of five and only began learning English thereafter. Evidently, the operationalization of terms such as *L1* or *native speaker* is problematic and has become increasingly so in today's globalized world. In terms of this study, it is not entirely clear how many L1 English speakers are truly so in the sense that English was the first language that they acquired as a child. This is bound to have ramifications on the conclusions drawn, as studies have repeatedly shown that the L1 plays a significant role in L2 acquisition. Future work should gather more detailed information on L1 background and language dominance in order to gain a deeper understanding of linguistic processing in both the L1 and L2

Only V+N collocations were used in this study. ADJ+N collocations were not included so as to avoid interference due to word order, as the adjective is placed after the noun in Portuguese (as in other Romance languages) (Villalva & Silvestre, 2014). Investigating the effects of differences in word order on collocational processing could lead to further advancements in our understanding of language processing at the morphosyntactic interface in the L2. Additionally, other syntactic categories of collocations such as Noun+Noun, Adverb+Verb, and Adverb+Adjective should also be investigated.

A major oversight in the item development process was a failure to cross-check whether the baselines existed in Portuguese. Although this could have been problematic, an *a posteriori* check showed that most of the baseline items were non-existent and some had very few occurrences in the PTtenten11. Future work should preemptively

ensure that baseline items do not correspond to meaningful units in the L2, as they could act as confounds.

Given that not all participants could be compensated for their time, the decision was made to keep the experiment as short as possible. Therefore, the LexTALE was not administered to the L1 group. In retrospect, this was an important measure that should have been recorded. Proficiency could have been important predictor of RTs and ERs and therefore should have been specified in the model. Moreover, these measures would provide an important metric that could quantitatively gauge the proficiency of the L2 speakers by comparing their vocabulary knowledge scores against that of the L1 speakers. Therefore, a reliable and quantifiable measure of language proficiency is an absolute must for any study in SLA. This would also allow for a more accessible comparison of results between different studies with disparate samples.

Given the pervasiveness of collocations in natural language, there are two other areas of linguistic research that could benefit from studying collocations—Ageing, and First Language Acquisition. In the present study, the diversity of the sample population in terms of age was reflected in processing speed, but not accuracy, which is in keeping with the general observations of real-time processing in the ageing literature (see Lo & Andrews, 2015; Woods et al., 2015). Exploration of age-based differences in collocational processing could provide further insights into the effects of ageing on language processing. In first language acquisition, usage-based approaches have repeatedly underlined the importance of multi-word units and chunking (see Goldberg, 2005; Tomasello, 2003). Yet, most studies on collocations focus on adult speakers. Recently, Jiang, Jiang, and Siyanova-Chanturia (2020) investigated collocational acquisition in children. They found that children, like adults, read high-frequency phrases faster than low-frequency phrases. Follow-up studies investigating semantic transparency and congruency (in bilingual children) could help broaden our understanding of L1 acquisition.

5.5 Implications for Teaching & Learning Collocations

According to Laufer (2005), lexical items that pose a challenge for L2 learners will be poorly learned, and there is already a resounding call for explicit teaching of collocations in the L2 classroom (Nation, 2001; Tsai, 2020; Vedder & Benigno, 2016). The findings of the present study indicate that even high-frequency incongruent collocations are problematic for advanced L2 English speakers. Therefore, explicit teaching of incongruent collocations is recommended especially in classrooms with fairly homogeneous L1 backgrounds, wherein the teacher can leverage off of the students' shared L1.

Moreover, students should be encouraged to use collocation dictionaries, as well as online resources such as Sketch Engine's SkeLL (Baisa & Suchomel, 2014) and ColloCaid (Frankenberg-Garcia et al., 2019) all of which have been shown to improve vocabulary gains (Frankenberg-Garcia et al., 2019; Hirata & Hirata, 2019; Laufer, 2011). Translation activities wherein explicit contrasts between L1 and L2 collocations are the key focus were

also found to improve collocation learning (Laufer & Girsai, 2008).

Finally, similar to single words, incidental learning of collocations is known to increase with frequency of encounters (Webb et al., 2013), while typographic enhancements of collocations in a text are also known to improve collocational gains (Choi, 2017; Sonbul & Schmitt, 2013). Nation's (2011) extensive reading component along with typographically enhanced collocations in the texts would be a good addition to the L2 classroom.

Chapter 6

Conclusion

This dissertation aimed to investigate the role of semantic transparency and congruency in L2 collocational processing. A semantic acceptability judgement task was administered to test free combinations, congruent collocations, and incongruent collocations alongside an equal number of baseline items. The data of 122 Brazilian Portuguese (BP) speakers, 108 European Portuguese (EP) speakers, and 99 L1 English speakers were analysed using the Generalized Linear Mixed-effects Model (GLMM) framework. Post-hoc analyses were carried out on the estimated marginal means (EMM) calculated by the models.

In terms of congruency, the results of the L2 groups revealed that the L1 facilitates collocational processing in the L2—congruent collocations have a processing advantage over incongruent collocations. This is further highlighted in the comparison between the L1 and L2 groups. The L2 speakers were on par with the L1 speakers on the free combinations and congruent collocations, but were at a disadvantage on the incongruent collocations. Furthermore, the L2 groups were not significantly better at judging incongruent collocations than baselines. Finally, the incongruent collocations used in this study were highly frequent, yet they negatively affected the speed and accuracy even in proficient L2 speakers. This was in keeping with the results from numerous other studies in the field (e.g., Wolter & Gyllstad, 2011, 2013) and clearly underlined the challenge faced by L2 speakers in processing incongruent collocations in the L2. Based on these findings, recommendations were made for the explicit teaching of incongruent collocations in the L2 classroom.

In terms of semantic transparency, the results of this study only provide partial support for Howarth's Collocational Continuum. In contrast with Gyllstad and Wolter (2016), the EP and L1 groups showed no processing advantages for free combinations over congruent collocations. This partial lack of processing advantages could be explained by the effect of frequency since all items used in the study comprised high-frequency constituent words, as well as strong association scores and high phrasal frequency. Since the idea that frequent exposure to input plays a role in L2 acquisition was well established in the field under the usage-based paradigm (Bybee,

2010; Ellis, 2008; Tomasello, 2001), it is hardly surprising that frequency could override the effects of semantic transparency. However, even though the BP speakers showed a processing advantage for free combinations over congruent collocations, it was not clearly attributable to semantic transparency.

Furthermore, free combinations showed a clear processing advantage over incongruent collocations in both L2 groups. Although this may be attributable to the effect of semantic transparency, it could not be isolated from the effect of congruency since the free combinations were also congruent in Portuguese. Future research should compare incongruent free combinations to incongruent collocations. If a processing advantage is seen for incongruent free combinations, the findings could be seen as lending support to Howarth's Collocational Continuum Model. Even though the findings of the present study are limited, they do highlight the need for more research on the topic.

From *stealing hearts* and *drawing ire* to *chasing dreams* and *catching fire*, collocations with their semi-transparent meanings have been ignored for far too long in the field of linguistics. They are ubiquitous in natural language and critical to fluency in both the L1 and L2, making research on collocations an important contribution to linguistic theory in SLA and beyond.

References

Al-Dobaian, A. (2017). Collocations in Generative Theory. *Journal of Arts, Literature, Humanities and Social Sciences*(14).

2.2

Andringa, S., & Godfroid, A. (2020). Sampling Bias and the Problem of Generalizability in Applied Linguistics. *Annual Review of Applied Linguistics*, *40*(2020), 134–142. doi: 10.1017/S0267190520000033

2.2.2

Arnon, I., Mccauley, S. M., & Christiansen, M. H. (2017). Digging up the building blocks of language: Age-of-acquisition effects for multiword phrases. *Journal of Memory and Language*, *92*, 265–280. Retrieved from http://dx.doi.org/10.1016/j.jml.2016.07.004 doi: 10.1016/j.jml.2016.07.004

5.1.3

Baayen, R. H. (2008). *Analyzing Linguistic Data : A practical introduction to statistics using R*. Cambridge: Cambridge University Press.

3.5.2

Baayen, R. H., & Milin, P. (2010). Analyzing reaction times. *International Journal of Psychological Research*, *3*(2), 12–28. Retrieved from https://revistas.usb.edu.co/index.php/IJPR/article/view/807 doi: 10.21500/20112084.807

3.5.1

Baisa, V., & Suchomel, V. (2014). SkELL : Web Interface for English Language Learning. In A. Horák & P. Rychlý (Eds.), *Proceedings of recent advances in slavonic natural language processing* (pp. 63–70). NLP Consulting.

5.5

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278. Retrieved from http://dx.doi.org/10.1016/j.jml.2012.11.001 doi: 10.1016/j.jml.2012.11.001

3.5.2, 3.5.3, 3.5.3

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal*

of Statistical Software, *67*(1), 1–48. doi: 10.18637/jss.v067.i01

3.5.3

Berry, W. D., DeMeritt, J. H. R., & Esarey, J. (2010). Testing for Interaction in Binary Logit and Probit Models: Is a Product Term Essential? *American Journal of Political Science*, *54*(1), 248–266. Retrieved from https://onlinelibrary.wiley.com/doi/10.1111/j.1540-5907.2009.00429.x doi: 10.1111/j.1540-5907.2009.00429.x

3.5.2

Bleakley, H., & Chin, A. (2010). Age at arrival, English proficiency, and social assimilation among us immigrants. *American Economic Journal: Applied Economics*, *2*(1), 165–192. doi: 10.1257/app.2.1.165

1.2

British Council. (2015). Learning English in Brazil. , 0–36. Retrieved from https://www.britishcouncil.org.br/sites/default/files/learning_english_in_brazil.pdf

1.2.1

Bybee, J. (2010). Language as a complex adaptive system: the interaction of cognition, culture and use. In *Language, usage and cognition* (pp. 194–221). Cambridge: Cambridge University Press. Retrieved from https://www.cambridge.org/core/books/language-usage-and-cognition/language-as-a-complex-adaptive-system-the-interaction-of-cognition-culture-and-use/B87A65C94D81F22999260B633E24C6C4 doi: 10.1017/CBO9780511750526.011

2.2, 6

Bybee, J., & Beckner, C. (2012). Usage-Based Theory. In *The oxford handbook of linguistic analysis.* Oxford: Oxford University Press. doi: 10.1093/oxfordhb/9780199544004.013.0032

2.2

Carrol, G., & Conklin, K. (2014). Getting your wires crossed: Evidence for fast processing of L1 idioms in an L2. *Bilingualism: Language and Cognition*, *17*(4), 784–797. doi: 10.1017/S1366728913000795

5.1.3

Carroll, J. B., & White, M. N. (1973). Word Frequency and Age of Acquisition as Determiners of Picture-Naming Latency. *Quarterly Journal of Experimental Psychology*, *25*(1), 85–95. doi: 10.1080/14640747308400325

2.2.2, 5.1.3

Choi, S. (2017). Processing and learning of enhanced English collocations: An eye movement study. *Language Teaching Research*, *21*(3), 403–426. doi: 10.1177/1362168816653271

5.5

Chomsky, N. (1969). *Aspects of the Theory of Syntax*. Cambridge, MA: MIT Press.

2.2

Church, K. W., & Hanks, P. (1990). Word association norms, mutual information, and lexicography. *Computational Linguistics*.

3.3.1

Cieślicka, A. B., & Heredia, R. R. (2017). How to Save Your Skin when Processing L2 Idioms: An Eye Movement Analysis of Idiom Transparency and Cross-language Similarity among Bilinguals. *Iranian Journal of Language Teaching Research*, *5*(June), 81–107.

5.1.3

Cobb, T. (2021). *Compleat Web VP*. Retrieved from https://www.lextutor.ca/vp/comp/

3.3.3

Coelho, M., & Arau Ribeiro, M. D. C. (2018). Internationalization strategies in Portuguese Higher Education Institutions – time to move on and to move beyond. In *Epic series in language and linguistics* (Vol. 3, pp. 33–25). doi: 10.29007/dcmw

1.2.1

Collins. (2018). *Collins COBUILD advanced learner's dictionary*. (9th ed.). Glasgow, [Scotland.

3.3.3

Cowie, A. P. (1994). *Phraseology* (R. E. Asher & J. M. Y. Simpson, Eds.). Oxford: Pergamon.

2.1.1, 2.1.1

Cowie, A. P. (1998). *Phraseology : theory, analysis, and applications*. Oxford: Clarendon Press.

1.1, 2.1.1

Cronquist, K., & Fiszbein, A. (2017). English Language Learning in Latin America. *The Dialogue: Leadership for the Americas*(September), 1–81.

1.2, 1.2.1

Crystal, D. (2003). Why a global language? In *English as a global language* (2nd ed., pp. 1–28). Cambridge: Cambridge University Press. doi: 10.1017/cbo9781139196970.003

1.2

Daulton, F. E. (2008). *Japan's built-in lexicon of English-based loanwords*. Clevedon: Multilingual Matters.

4

Delignette-Muller, M. L., & Dutang, C. (2015). fitdistrplus: An R package for fitting distributions. *Journal of Statistical Software*, *64*(4), 1–34. Retrieved from https://www.jstatsoft.org/v64/i04/

3.5.3

Dice, L. R. (1945). Measures of the Amount of Ecologic Association Between Species. *Ecology*, *26*(3), 297–302. Retrieved from https://esajournals.onlinelibrary.wiley.com/doi/abs/10.2307/1932409 doi: 10.2307/1932409

3.3.1

Dinno, A. (2015). Nonparametric Pairwise Multiple Comparisons in Independent Groups using Dunn's Test. *The Stata Journal: Promoting Communications On Statistics And Stata*, *15*(1), 292–300.

3.2.3

Du, Y., & Clark, J. E. (2020). Beyond the mean reaction time: Trial-by-trial reaction time reveals the distraction effect on perceptual-motor sequence learning. *Cognition*, *202*(April), 104287. Retrieved from https://doi.org/10.1016/j.cognition.2020.104287 doi: 10.1016/j.cognition.2020.104287

4.2

Durrant, P., & Schmitt, N. (2009). To what extent do native and non-native writers make use of collocations? *IRAL - International Review of Applied Linguistics in Language Teaching*, *47*(2), 157–177. doi: 10.1515/iral.2009.007

1.1, 2.2.1

Durrant, P., & Schmitt, N. (2010). Adult learners' retention of collocations from exposure. *Second Language Research*, *26*(2), 163–188. doi: 10.1177/0267658309349431

2.2.1

Eberhard, D. M., Simons, G. F., & Fennig, C. D. (2021). *How many languages are there in the world?* (24th ed.). SIL International. Retrieved 2021-07-01, from https://www.ethnologue.com/guides/how-many-languages

1.2.1

Edmonds, A., & Gudmestad, A. (2021). Collocational development during a stay abroad. *Languages*, *6*(1), 1–17. doi: 10.3390/languages6010012

2.2.1, 3.3.1

EF Education First. (2020). English Profiency Index. *EF EPI*, 1–27.

1.2, 1.2.1

Ellis, N. (2008). Constructions, Chunking, and Connectionism: The Emergence of Second Language Structure. In *The handbook of second language acquisition* (pp. 63–103). Oxford, UK: Blackwell Publishing Ltd. Retrieved from http://doi.wiley.com/10.1002/9780470756492.ch4 doi: 10.1002/9780470756492.ch4

5.2, 6

Ellis, N. (2011). Frequency-based accounts of SLA. *Handbook of second language acquisition*, 193–201.

2.2

Ellis, N., & Ogden, D. (2017). Thinking About Multiword Constructions: Usage-Based Approaches to Acquisition and Processing. *Topics in Cognitive Science*, *9*(3), 604–620. doi: 10.1111/tops.12256

2.2

Ellis, N., Simpson-Vlach, R., & Maynard, C. (2008). Formulaic Language in Native and Second Language

Speakers: Psycholinguistics, Corpus Linguistics, and TESOL. *TESOL Quarterly*, *42*(3), 61–78. doi: 10.1515/CLLT.2009.003

5.2

Emerson, R. W. (2015). Convenience Sampling, Random Sampling, and Snowball Sampling: How Does Sampling Affect the Validity of Research? *Journal of Visual Impairment & Blindness*, *109*(2), 164–168.

5.1.1

Erman, B., & Warren, B. (2000). The idiom principle and the open choice principle. *Text*, *20*(1), 29–62.

1.1, 2.1.1

Evers-Vermeul, J., & Tribushinina, E. (2017). Language acquisition and language teaching in the usage-based framework. In J. Evers-Vermeul & E. Tribushinina (Eds.), *Usage-based approaches to language acquisition and language teaching* (pp. 1–10). Berlin, Boston: De Gruyter Mouton. Retrieved from https://doi.org/10.1515/9781501505492 doi: doi:10.1515/9781501505492

2.2

Finardi, K. R. (2017). English as a global language in Brazil: A local contribution. In T. Gimenez, M. S. El Kadri, & L. C. S. Calvo (Eds.), *English as a lingua franca in teacher education: A brazilian perspective.* De Gruyter Mouton.

1.2.1

Fioravanti, I., Senaldi, M. S. G., Lenci, A., & Siyanova-Chanturia, A. (2021). Lexical fixedness and compositionality in L1 speakers' and L2 learners' intuitions about word combinations: Evidence from Italian. *Second Language Research*, *37*(2), 291–322. doi: 10.1177/0267658320941560

2.1.3, 2.2.1, 2.2.2

Frankenberg-Garcia, A. (2018). Investigating the collocations available to EAP writers. *Journal of English for Academic Purposes*, *35*, 93–104. Retrieved from https://doi.org/10.1016/j.jeap.2018.07.003 doi: 10.1016/j.jeap.2018.07.003

3.3.1

Frankenberg-Garcia, A., Lew, R., Roberts, J. C., Rees, G. P., & Sharma, N. (2019). Developing a writing assistant to help EAP writers with collocations in real time. *ReCALL*, *31*(01), 23–39. Retrieved from https://www.cambridge.org/core/product/identifier/S0958344018000150/type/journal_article doi: 10.1017/S0958344018000150

3.3.1, 5.5

Gabaix, X., & Laibson, D. (2011). The Seven Properties of Good Models. In A. Caplin & A. Schotter (Eds.), *The foundations of positive and normative economics* (Online ed., pp. 292–299). Oxford, UK: Oxford University Press. Retrieved from https://oxford.universitypressscholarship.com/view/10.1093/acprof:oso/9780195328318.001.0001/acprof-9780195328318-chapter-12 doi:

10.1093/acprof:oso/9780195328318.003.0012

3.5.2

Gablasova, D., Brezina, V., & McEnery, T. (2017). Collocations in Corpus-Based Language Learning Research: Identifying, Comparing, and Interpreting the Evidence. *Language Learning*, *67*(June), 155–179. doi: 10.1111/lang.12225

3.3.1

Galloway, N., & Rose, H. (2015). *Introducing Global Englishes*. London: Routledge.

1.2

Gass, S. M., Plonsky, L., & Behney, J. (2020). *Second language acquisition: An introductory course* (Fifth Edition ed.). London: Routledge.

1.2.1

Gelman, A., & Weakliem, D. (2009). Of beauty, sex and power: Too little attention has been paid to the statistical challenges in estimating small effects. *American Scientist*, *97*(4), 310–316. doi: 10.1511/2009.79.310

2.2.2

Gill, J., & Torres, M. (2019). *Generalized linear models : A unified approach* (2nd ed.). Los Angeles: SAGE Publications.

3.5.2

Gilquin, G. (2017). Applied cognitive linguistics and second/ foreign language varieties: Towards an explanatory account. In J. Evers-Vermeul & E. Tribushinina (Eds.), *Usage-based approaches to language acquisition and language teaching* (chap. 2). Berlin; Boston: De Gruyter Mouton.

3.3.1

Goldberg, A. (2005). *Constructions at Work: The Nature of Generalization in Language*. Oxford ; New York: Oxford University Press. doi: 10.1093/acprof:oso/9780199268511.001.0001

2.2, 5.4

Graham, M., Hjorth, I., & Lehdonvirta, V. (2017). Digital labour and development: Impacts of global digital labour platforms and the gig economy on worker livelihoods. *Transfer*, *23*(2), 135–162. doi: 10.1177/1024258916687250

1.2

Gries, S. (2010). Useful statistics for corpus linguistics. In A. Sánchez & M. Almela (Eds.), *A mosaic of corpus linguistics: Selected approaches* (pp. 269–291).

2.1.2

Guimarães, F. F., Mendes, A. R. M., & Hildeblando Júnior, C. A. (2021). Student perceptions of university internationalisation in Brazil – a case study. *Fórum Linguístico*, *18*(1), 5701–5718. doi: 10.5007/1984-8412

.2021.e70371

1.2.1

Gyllstad, H. (2013). Looking at L2 vocabulary knowledge dimensions from an assessment perspective – challenges and potential solutions. In C. Bardel, C. Lindqvist, & B. Laufer (Eds.), *L2 vocabulary acquisition, knowledge and use: New perspectives on assessment and corpus analysis* (pp. 11–28). Eurosla.

1.1

Gyllstad, H., & Wolter, B. (2016). Collocational Processing in Light of the Phraseological Continuum Model: Does Semantic Transparency Matter ? *Language Learning*, *66*(2), 296–323. doi: 10.1111/lang.12143

1.1, 1.3, 2.1.1, 2.1.3, 2.2.2, 2.2.2, 2.3, 3.1, 3.2.2, 3.3.1, 3.5.3, 3.5.3, 4.4, 5, 5.1, 5.1.1, 5.1.2, 5.2, 5.3, 6

Hatch, E., & Lazaraton, A. (1991). *The Research Manual: Design and statistics for applied linguistics*. Boston, MA: Heinle & Heinle. doi: 10.2307/3586777

3.2.3

Henriksen, B. (2013). Research on L2 learners' collocational competence and development – a progress report. In C. Bardel, C. Lindqvist, & B. Laufer (Eds.), *L2 vocabulary acquisition, knowledge and use: New perspectives on assessment and corpus analysis* (pp. 211–230). Eurosla.

2.1, 2.1.3

Hirata, Y., & Hirata, Y. (2019). Applying 'Sketch Engine for Language Learning' in the Japanese English classroom. *Journal of Computing in Higher Education*, *31*(2), 233–248. Retrieved from https://doi.org/10.1007/s12528-019-09208-z doi: 10.1007/s12528-019-09208-z

5.5

Howarth, P. (1998). Phraseology and second language proficiency. *Applied Linguistics*, *19*(1), 24–44. doi: 10.1093/applin/19.1.24

1.1, 2.1.1, 2.1.1, 2.2.1, 2.2.2, 2.3, 5.1.1, 5.2, 5.3

Ilson, R. F. (2010). *The BBI Combinatory Dictionary of English*. John Benjamins Publishing Company. doi: https://doi.org/10.1075/z.bbi

3.3.3

Izura, C., Pérez, M. A., Agallou, E., Wright, V. C., Marín, J., Stadthagen-González, H., & Ellis, A. W. (2011). Age / order of acquisition effects and the cumulative learning of foreign words: A word training study. *Journal of Memory and Language*, *64*(1), 32–58. Retrieved from http://dx.doi.org/10.1016/j.jml.2010.09.002 doi: 10.1016/j.jml.2010.09.002

5.1.3

Jakubíček, M., Kilgarriff, A., Kovář, V., Rychlý, P., & Suchomel, V. (2013). The TenTen Corpus Family. *7th International Corpus Linguistics Conference*, 125–127. Retrieved from https://www.sketchengine.co.uk/wp-content/uploads/The_TenTen_Corpus_2013

.pdf%5Cnhttp://ucrel.lancs.ac.uk/cl2013/doc/CL2013-ABSTRACT-BOOK.pdf

3.3.1

James, C. (1980). *Contrastive Analysis*. London: Longman.

2.2.1

Jenkins, J. (2007). *English as a Lingua Franca: Attitude and identity*. Oxford: Oxford University Press.

1.5

Jenkins, J. (2014). *English as a Lingua Franca in the International University: The politics of academic English language policy*. London.

1.5

Jiang, S., Jiang, X., & Siyanova-Chanturia, A. (2020). The processing of multiword expressions in children and adults: An eye-tracking study of Chinese. *Applied Psycholinguistics*, *41*(4), 901–931. doi: 10.1017/S0142716420000296

5.4

Joseph, B. D., Janda, R. D., Horn, L. R., Ward, G., Ball, M. J., Perkins, M. R., … Rourke, E. O. (2018). *The Handbook of Advanced Proficiency in Second Language Acquisition*. doi: 10.1002/9781119261650

1.2.1, 2.2

Jouët-Pastré, C., & Braga, L. (2008). *Becoming Brazuca: Brazilian Immigration to the US* (David Rockefeller Center for Latin American Studies ed.). Cambridge: Harvard University Press.

1.2.1

Kachru, B. B. (1982). *The other tongue: English across cultures*. Urbana: University of Illinois Press.

1.5

Krashen, S. D. (1982). *Principles and Practice in Second Language Acquisition*. Oxford: Pergamon.

1.5

Kuhn, T. S. (2012). *The Structure of Scientific Revoolutions* (50th Anniv ed.). Chicago: Chicago University Press.

2.2

Kumle, L., Võ, M. L., & Draschkow, D. (2021). Estimating power in (generalized) linear mixed models: An open introduction and tutorial in R. *Behavior Research Methods*. doi: 10.3758/s13428-021-01546-0

2.2.2

Lado, R. (1957). *Linguistics across cultures: Applied linguistics for language teachers*. Ann Arbor: University of Michigan Press.

2.2.1

Lamb, M. (2017). The motivational dimension of language teaching. *Language Teaching*, *50*(3), 301–346. doi: 10.1017/S0261444817000088

3.3.1

LaPointe, L. L., Heald, G. R., Stierwalt, J. A., Kemker, B. E., & Maurice, T. (2007). Effects of auditory distraction on cognitive processing of young adults. *Journal of Attention Disorders*, *10*(4), 398–409. doi: 10.1177/1087054706293221

4.2

Laufer, B. (2005). Focus on Form in Second Language Vocabulary Learning. *EUROSLA Yearbook*, *5*(April), 223–250. doi: 10.1075/eurosla.5.11lau

5.5

Laufer, B. (2011). The contribution of dictionary use to the production and retention of collocations in a second language. *International Journal of Lexicography*, *24*(1), 29–49. doi: 10.1093/ijl/ecq039

1.1, 2.2.1, 5.5

Laufer, B., & Girsai, N. (2008). Form-focused instruction in second language vocabulary learning: A case for contrastive analysis and translation. *Applied Linguistics*, *29*(4), 694–716. doi: 10.1093/applin/amn018

5.5

Laufer, B., & Waldman, T. (2011). Verb-Noun Collocations in Second Language Writing: A Corpus Analysis of Learners' English. *Language Learning*, *61*(2), 647–672. Retrieved from https://doi.org/10.1111/j.1467-9922.2010.00621.x doi: https://doi.org/10.1111/j.1467-9922.2010.00621.x

1.1, 2.2.1

Lemhöfer, K., & Broersma, M. (2012). Introducing LexTALE: A quick and valid Lexical Test for Advanced Learners of English. *Behavior Research Methods*, *44*(2), 325–343. Retrieved from www.lextale.com.http://link.springer.com/10.3758/s13428-011-0146-0 doi: 10.3758/s13428-011-0146-0

3.2.4, 3.3.2, 2

Lenth, R. V. (2021). emmeans: Estimated marginal means, aka least-squares means [Computer software manual]. Retrieved from https://CRAN.R-project.org/package=emmeans (R package version 1.6.2-1)

3.5.3

Leśniewska, J., & Witalisz, E. (2007). Cross-linguistic influence and acceptability judgments of L2 and L1 collocations. *EUROSLA Yearbook*, *7*, 27–48. doi: 10.1075/eurosla.7.04les

2.2.1

Lewis, M. (2007). Stepwise versus Hierarchal Regression: Pros and Cons. In *Annual meeting of the southwest educational research association*. San Antonio, TX. Retrieved from https://eric.ed.gov/?id=ED534385

3.5.3

Lo, S., & Andrews, S. (2015). To transform or not to transform: using generalized linear mixed models to

analyse reaction time data. *Frontiers in Psychology*, *6*(August), 1–16. Retrieved from http://journal.frontiersin.org/Article/10.3389/fpsyg.2015.01171/abstract doi: 10.3389/fpsyg.2015.01171

3.5.2, 3.5.3, 5.4

Lüdecke, D. (2021). sjstats: Statistical functions for regression models (version 0.18.1) [Computer software manual]. Retrieved from https://CRAN.R-project.org/package=sjstats doi: 10.5281/zenodo.1284472

2.2.2, 3.2

Lüdecke, D., Ben-Shachar, M. S., Patil, I., Waggoner, P., & Makowski, D. (2021). performance: An R package for assessment, comparison and testing of statistical models. *Journal of Open Source Software*, *6*(60), 3139. doi: 10.21105/joss.03139

3.5.3

Macis, M., & Schmitt, N. (2017). The figurative and polysemous nature of collocations and their place in ELT. *ELT Journal*, *71*(1), 50–59. doi: 10.1093/elt/ccw044

2.2.1

Martinez, R., & Murphy, V. A. (2011). Effect of frequency and idiomaticity on second language reading comprehension. *TESOL Quarterly*, *45*(2), 267–290. doi: 10.5054/tq.2011.247708

1.1, 2.2.1

Matuschek, H., Kliegl, R., Vasishth, S., Baayen, H., & Bates, D. (2017). Balancing Type I error and power in linear mixed models. *Journal of Memory and Language*, *94*, 305–315. Retrieved from http://dx.doi.org/10.1016/j.jml.2017.01.001 doi: 10.1016/j.jml.2017.01.001

3.5.3

McCulloch, C. E., & Neuhaus, J. M. (2011). Misspecifying the shape of a random effects distribution: Why getting it wrong may not matter. *Statistical Science*, *26*(3), 388–402. doi: 10.1214/11-STS361

3.5.3

McEwan, B. (2020). Sampling and validity. *Annals Of The International Communication Association*, *44*(3), pp235—-247.

5.1.1

McManus, K. (2019). Awareness of L1 form-meaning mappings can reduce crosslinguistic effects in L2 grammatical learning. *Language Awareness*, *28*(2), 114–138. doi: 10.1080/09658416.2019.1620756

1.2.1

Mel'čuk, I. a. (1998). Collocations and lexical functions. *Phraseology. Theory, analysis, and applications*(1), 23–53. Retrieved from http://olst.ling.umontreal.ca/pdf/CowieLFs.pdf

2.1.1, 2.1.1, 2.1.3

Menard, S. W. (2002). *Applied Logistic Regression Analysis* (2nd ed. ed.). Thousand Oaks, Calif.: Sage Publications.

3.5.3

Nation, P. (2001). The goals of vocabulary learning. In *Learning vocabulary in another language* (2nd ed., pp. 6–25). Cambridge: Cambridge University Press. doi: 10.1016/s0889-4906(02)00014-5

1.1, 5.5

Nation, P. (2011). Research into practice: Vocabulary. *Language Teaching*, *44*(4), 529–539. doi: 10.1017/s0261444811000267

5.5

Nattinger, J. R., & DeCarrico, J. S. (1992). *Lexical phrases and language teaching*. Oxford: Oxford University Press.

1.1, 2.1.1

Nesselhauf, N. (2003). The Use of Collocations by Advanced Learners of English and Some Implications for Teaching. *Applied Linguistics*, *24*(2), 223–242.

2.1.3, 2.2.1, 2.2.2

Nesselhauf, N. (2004). *Collocations in a Learner Corpus*. Philadelphia: John Benjamins Publishing Company. Retrieved from http://ebookcentral.proquest.com/lib/oxford/detail.action?docID=622326

1.1, 2.1, 2.1.3

Norton, E., Dowd, B., & Maciejewski, M. (2018). Odds Ratios—Current Best Practice and Use. *JAMA*, *320*(1), 84–85. Retrieved from http://jama.jamanetwork.com/article.aspx?doi=10.1001/jama.2017.3091 doi: 10.1001/jama.2017.3091

4.3

Öksüz, D., Brezina, V., & Rebuschat, P. (2021). Collocational Processing in L1 and L2: The Effects of Word Frequency, Collocational Frequency, and Association. *Language Learning*, *71*(1), 55–98. doi: 10.1111/lang.12427

3.3.1

Pavlenko, A. (2017). Do you wish to waive your rights? Affect and decision-making in multilingual speakers. *Current Opinion in Psychology*, *17*, 74–78. Retrieved from http://dx.doi.org/10.1016/j.copsyc.2017.06.005 doi: 10.1016/j.copsyc.2017.06.005

1.2

Pawley, A., & Syder, F. H. (1983). Two puzzles for linguistic theory: Nativelike selection and nativelike fluency. *Language and Communication*, 191–226. Retrieved from https://www.researchgate.net/publication/240365989_Two_puzzles_for_linguistic

_theory_Nativelike_selection_and_nativelike_fluency doi: 10.4324/9781315836027-12

1.1

Pereira, C., & Azevedo, J. (2019). *New and Old Routes of Portuguese Emigration. Uncertain Futures at the Periphery of Europe* (Online ed.). Springer International Publishing. Retrieved from https://link.springer.com/book/10.1007/978-3-030-15134-8 doi: 10.1007/978-3-030-15134-8

1.2.1

Peters, E. (2016). The learning burden of collocations: The role of interlexical and intralexical factors. *Language Teaching Research*, *20*(1), 113–138. doi: 10.1177/1362168814568131

1.1, 2.2.1

Petrescu, M. C., Helms-Park, R., & Dronjic, V. (2017). The impact of frequency and register on cognate facilitation: Comparing Romanian and Vietnamese speakers on the Vocabulary Levels Test. *English for Specific Purposes*, *47*, 15–25. Retrieved from http://dx.doi.org/10.1016/j.esp.2017.03.001https://linkinghub.elsevier.com/retrieve/pii/S0889490617301035 doi: 10.1016/j.esp.2017.03.001

4

Poort, E. D., & Rodd, J. M. (2019). Towards a distributed connectionist account of cognates and interlingual homographs: evidence from semantic relatedness tasks. *PeerJ*, *7*, e6725. Retrieved from https://peerj.com/articles/6725 doi: 10.7717/peerj.6725

3.3.2, 3.5.3

R Core Team. (2019). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria. Retrieved from https://www.R-project.org/

3.5.3

Ratcliff, R. (1993). Methods for Dealing with Reaction Time Outliers. *Psychological Bulletin*, *114*(3), 510. Retrieved from http://search.proquest.com/docview/38457340/

3.5.1, 3.5.3

Rochefort-Maranda, G. (2016). Simplicity and model selection. *European Journal for Philosophy of Science*, *6*(2), 261–279. Retrieved from http://dx.doi.org/10.1007/s13194-016-0137-1 doi: 10.1007/s13194-016-0137-1

3.5.2

Rodriguez, A., Kaakinen, M., Moilanen, I., Taanila, A., McGough, J. J., Loo, S., & Järvelin, M. R. (2010). Mixed-handedness is linked to mental health problems in children and adolescents. *Pediatrics*, *125*(2). doi: 10.1542/peds.2009-1165

3.2.3

Roettger, T. B. (2021). Preregistration in experimental linguistics: applications, challenges, and limita-

tions. *Linguistics*. Retrieved from https://www.degruyter.com/document/doi/10.1515/ling-2019-0048 doi: 10.1515/ling-2019-0048

2.2.2

Rüsseler, J., Becker, P., Johannes, S., & Münte, T. F. (2007). Semantic, syntactic, and phonological processing of written words in adult developmental dyslexic readers: an event-related brain potential study. *BMC Neuroscience*, *8*(1), 52. Retrieved from http://bmcneurosci.biomedcentral.com/articles/10.1186/1471-2202-8-52 doi: 10.1186/1471-2202-8-52

3.2.3

Rychlý, P. (2008). A lexicographer-friendly association score. In *Raslan 2008 - recent advances in slavonic natural language processing: 2nd workshop on recent advances in slavonic natural language processing, proceedings* (pp. 6–9). Retrieved from https://nlp.fi.muni.cz/raslan/2008/papers/13.pdf

3.3.1

Santos David, R. (2016). O Ensino-Aprendizagem De Língua Inglesa Em Escolas Públicas: O Real E O Ideal. *Revista Ciências Humanas*, *9*(2), 24–31. doi: 10.32813/2179-1120.2016.v9.n2.a323

1.2.1

Sauter, M., Draschkow, D., & Mack, W. (2020). Building, Hosting and Recruiting: A Brief Introduction to Running Behavioral Experiments Online. *Brain Sciences*, *10*(4), 251. doi: 10.3390/brainsci10040251

3.4.2

Schmitt, N. (2010). *Researching vocabulary: a vocabulary research manual*. Basingstoke: Palgrave Macmillan.

2.1.3, 2.3

Sinclair, J. (1991). *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.

1.1, 2.1.2

Singmann, H., Bolker, B., Westfall, J., Aust, F., & Ben-Shachar, M. S. (2021). afex: Analysis of factorial experiments [Computer software manual]. Retrieved from https://CRAN.R-project.org/package=afex (R package version 0.28-1)

3.5.3, 4.3

Singmann, H., & Kellen, D. (2019). An introduction to mixed models for experimental psychology. *New Methods in Cognitive Psychology*, 4–31. doi: 10.4324/9780429318405-2

3.5.3, 3.5.3

Siyanova-Chanturia, A., Conklin, K., & van Heuven, W. J. (2011). Seeing a Phrase " Time and Again" Matters: The Role of Phrasal Frequency in the Processing of Multiword Sequences. *Journal of Experimental Psychology: Learning Memory and Cognition*, *37*(3), 776–784. doi: 10.1037/a0022531

2.2.2

Siyanova-Chanturia, A., & Omidian, T. (2019). Key issues in researching multiword items. In S. A. Webb (Ed.),

The routledge handbook of vocabulary studies (pp. 529–544). Routledge. doi: 10.4324/9780429291586-33

1.1, 2.1

Siyanova-Chanturia, A., & Spina, S. (2020). Multi-Word Expressions in Second Language Writing: A Large-Scale Longitudinal Learner Corpus Study. *Language Learning*, *70*(2), 420–463. doi: 10.1111/lang.12383

1.2.1, 2.2.1

Smith, M. (2012). *Detrimental Effects of Underpowered or Overpowered Studies.* Retrieved 2021-08-09, from https://web.ma.utexas.edu/users/mks/statmistakes/UnderOverPower.html

2.2.2

Sonbul, S., & Schmitt, N. (2013). Explicit and Implicit Lexical Knowledge: Acquisition of Collocations Under Different Input Conditions. *Language Learning*, *63*(1), 121–159. doi: 10.1111/j.1467-9922.2012.00730.x

5.5

Speelman, C. P., & McGann, M. (2013). How Mean is the Mean? *Frontiers in Psychology*, *4*(JUL). Retrieved from http://journal.frontiersin.org/article/10.3389/fpsyg.2013.00451/abstract doi: 10.3389/fpsyg.2013.00451

3.5.2

Stadthagen-Gonzalez, H., Bowers, J. S., & Damian, M. F. (2004). Age-of-acquisition effects in visual word recognition : evidence from expert vocabularies. *Cognition*, *93*, B11–B26. doi: 10.1016/j.cognition.2003.10.009

5.1.3

Starns, J. J., & Ratcliff, R. (2010). The Effects of Aging on the Speed-Accuracy Compromise: Boundary Optimality in the Diffusion Model. *Psychology and Aging*, *25*(2), 377–390.

4.2

Stevens, S. S. (1946). On the Theory of Scales of Measurement. *Science*, *103*(2684), 677–680. Retrieved from https://science.sciencemag.org/content/103/2684/677 doi: 10.1126/science.103.2684.677

3.5.2

Stewart, N., & Ellis, A. W. (2008). Order of acquisition in learning perceptual categories: A laboratory analogue of the age-of-acquisition effect ? *Psychonomic Bulletin*

Review, *15*(1), 70–74. doi: doi.10.3758/PBR.15.1.70

5.1.3

Supasiraprapa, S. (2018). Second Language Collocation Acquisition: Challenges for Learners and Pedagogical Insights from Empirical Research. *Journal of Asia TEFL*, *15*(3), 797–804. doi: http://dx.doi.org/10.18823/asiatefl.2018.15.3.15.797

2.1.2

Tabossi, P., Fanari, R., & Wolf, K. (2009). Why are idioms recognized fast? *Memory & Cognition*, *37*(4), 529–540. Retrieved from https://doi.org/10.3758/MC.37.4.529

5.2

Titone, D., Columbus, G., Whitford, V., Mercier, J., & Libben, M. (2015). Contrasting Bilingual and Monolingual Idiom Processing. In A. B. Cieślicka & R. R. Heredia (Eds.), *Bilingual figurative language processing* (pp. 171–207). Cambridge: Cambridge University Press. Retrieved from https://www.cambridge.org/core/books/bilingual-figurative-language-processing/contrasting-bilingual-and-monolingual-idiom-processing/9EC3DC75952D181F6BB6C03CF603AC81 doi: 10.1017/CBO9781139342100.011

5.1.3

Tomasello, M. (2001). First steps toward a usage-based theory of language acquisition. *Cognitive Linguistics*, *11*(1-2), 61–82. doi: 10.1515/cogl.2001.012

2.2, 6

Tomasello, M. (2003). *Constructing a language: A usage-based theory of language acquisition.* Cambridge, MA, US: Harvard University Press.

2.2, 5.4

Tomasello, M. (2009). The usage-based theory of language acquisition. In E. L. Bavin (Ed.), *The cambridge handbook of child language* (pp. 69–88). Cambridge: Cambridge University Press.

2.2

Tomoschuk, B., Ferreira, V. S., & Gollan, T. H. (2019). When a seven is not a seven: Self-ratings of bilingual language proficiency differ between and within language populations. *Bilingualism: Language and Cognition*, *22*(3), 516–536. doi: 10.1017/S1366728918000421

2.2.2, 3.2.3, 3.3.2

Townsend, J. T. (1992). *On the proper scales of reaction time* (H.-G. Geissler, S. W. Link, & J. T. Townsend, Eds.). Hillsdale, N.J, London: L. Erlbaum Associates. Retrieved from https://catalogue.nla.gov.au/Record/1757300

3.5.2

Triandafyllidou, A. (2016). *Routledge Handbook of Immigration and Refugee Studies*. Abingdon, Oxon; New York, NY: Routledge.

1.2

Tsai, M.-h. (2020). The effects of explicit instruction on L2 learners' acquisition of verb – noun collocations. *Language Teaching Research*, *24*(2), 138–162. doi: 10.1177/1362168818795188

1.1, 2.2.1, 5.5

Urdaniz, R. P., & Skoufaki, S. (2019). *Spanish L1 EFL learners' recognition knowledge of English academic*

vocabulary: The role of cognateness, word frequency and length. doi: 10.1515/applirev-2018-0109

2.2.2, 3.3.2, 4

Vedder, I., & Benigno, V. (2016). Lexical richness and collocational competence in second-language writing. *International Review of Applied Linguistics in Language Teaching*, *54*(1), 23–42. Retrieved from https://ezproxy-prd.bodleian.ox.ac.uk:2200/view/journals/iral/54/1/article-p23.xml doi: 10.1515/iral-2016-0015

5.5

Villalva, A., & Silvestre, J. P. (2014). *Introdução ao estudo do léxico*. Petrópolis: Editora Vozes Ltda.

3.3.3, 5.4

Webb, S., Newton, J., & Chang, A. (2013). Incidental Learning of Collocation. *Language Learning*, *63*(1), 91–120. doi: 10.1111/j.1467-9922.2012.00729.x

1.1, 2.2.1, 5.5

Wolter, B. (2019). Key Issues in Teaching Multiword Items. In S. A. Webb (Ed.), *The routledge handbook of vocabulary studies* (pp. 493–510). Routledge. Retrieved from https://www.taylorfrancis.com/books/9781000005561/chapters/10.4324/9780429291586-31 doi: 10.4324/9780429291586-31

1.1, 2.1

Wolter, B., & Gyllstad, H. (2011). Collocational Links in the L2 Mental Lexicon and the Influence of L1 Intralexical Knowledge. *Applied Linguistics*, *32*(4), 430–449. Retrieved from https://academic.oup.com/applij/article/32/4/430/235690 doi: 10.1093/applin/amr011

1.1, 2.2.2, 5.1.3, 5.1.4, 6

Wolter, B., & Gyllstad, H. (2013). Frequency of Input and L2 Collocational Processing: A Comparison of Congruent and Incongruent Collocations. *Studies in Second Language Acquisition*, *35*, 451–482. Retrieved from https://doi.org/10.1017/S0272263113000107 doi: 10.1017/S0272263113000107

1.1, 2.2.2, 5.1.1, 5.1.3, 5.1.4, 6

Wolter, B., & Yamashita, J. (2015). Processing collocations in a second language: A case of first language activation? *Applied Psycholinguistics*, *36*(5), 1193–1221. Retrieved from https://www.cambridge.org/core. doi: 10.1017/S0142716414000113

(document), 1.1, 2.2.2, 3.3.3, 5.1.3, 5.1.4, 5.3

Wolter, B., & Yamashita, J. (2018). Word Frequency, Collocational Frequency, L1 Congruency, and Proficiency in L2 Collocational Processing: What accounts for L2 performance? *Studies in Second Language Acquisition*, *40*(2), 395–416. doi: 10.1017/S0272263117000237

1.1, 2.2.2, 3.3.1, 3.3.3, 5.1.3, 5.3

Wood, D. (2019). Classifying and identifying formulaic language. In S. A. Webb (Ed.), *The routledge handbook*

of vocabulary studies (pp. 30–45). Routledge. doi: 10.4324/9780429291586-3

2.1

Woods, D. L., Wyma, J. M., Yund, E. W., Herron, T. J., & Reed, B. (2015). Age-related slowing of response selection and production in a visual choice reaction time task. *Frontiers in Human Neuroscience*, *9*, 193. Retrieved from https://www.frontiersin.org/article/10.3389/fnhum.2015.00193 doi: 10.3389/fnhum.2015.00193

4.2, 5.4

Wray, A. (2002). *Formulaic Language and the Lexicon* (Vol. 80). Cambridge: Cambridge University Press. doi: 10.1353/lan.2004.0209

1.1, 3, 2.1.2

Wray, A. (2013). Formulaic language. *Language Teaching*, *46*(3), 316–334. doi: 10.1017/S0261444813000013

2.1, 2.2

Yamashita, J. (2018). Possibility of semantic involvement in the L1-L2 congruency effect in the processing of L2 collocations. *Journal of Second Language Studies*, *1*(1), 60–78. doi: 10.1075/jsls.17024.yam

1.1, 1.2.1, 2.2.2, 2.2.2, 2.3

Yamashita, J., & Jiang, N. (2010). L1 influence on the acquisition of L2 collocations: Japanese ESL users and EFL learners acquiring English collocations. *TESOL Quarterly*, *44*(4), 647–668. doi: 10.5054/tq.2010.235998

2.2.2, 5.1.3

Appendices

Appendix A

Methods

A.1 Participants

Table A.1. Age: Dunn's Test Results

	Comparison	Z	P-Unadj.	P-Adj.
1	BP - EN	8.460	0	0
2	BP - EP	26.403	0	0
3	EN - EP	16.846	0	0

Table A.2. Education Level (Self-reported)

	Education Level	Brazilian Portuguese (BP)	English (EN)	European Portuguese (EP)
1	High School (or equivalent)	15	13	7
2	Bachelor's Degree (or equivalent)	73	45	64
3	Master's Degree (or equivalent)	20	30	19
4	Ph.D. (or equivalent)	14	11	18
	Total	122	99	108

Table A.3. Place of L2 English Learning (Self-reported)

Place	Brazilian Portuguese (BP)	European Portuguese (EP)
Home	22	18
Language School	42	3
School	45	85
University	1	0
Online	6	0
Other	6	2
Total	122	108

Table A.4. L2 Proficiency (Self-reported)

CEFR Level	Brazilian Portuguese (BP)	European Portuguese (EP)
A1-A2 (Beginner)	7	0
B1 (Lower-Intermediate)	13	7
B2 (Upper-Intermediate)	24	27
C1-C2 (Advanced)	90	81
Total	122	108

Table A.5. Average English Proficiency Across the Four Skills (Self-reported)

Skills	Brazilian Portuguese (BP)	European Portuguese (EP)
Reading	4.66	4.61
Writing	3.81	4.12
Listening	4.27	4.45
Speaking	3.69	4.14

Table A.6. ENtenten18 Sub-Corpora by Domain

Domain	Tokens	Percentage (%)
Australia (.au)	699,931,081	2.71
Canada (.ca)	514,362,319	1.992
EU (.eu)	79,375,718	0.307
Indian (.in)	5,489,170	0.021
Irish (.ie)	76,969,678	0.298
New Zealand (.nz)	138,028,382	0.534
UK (.uk)	1,731,454,108	6.704
US (.us)	240,119,290	0.93

A.2 Items

Stimuli

Table A.7. Tukey Results for logDice scores

Contrast	Estimate	Low	High	P-Adj.
CC−BL	-0.731	-1.678	0.216	0.188
IC−BL	-0.012	-0.959	0.935	1.000
FC−BL	0.743	-0.204	1.690	0.176
IC−CC	0.719	-0.441	1.879	0.370
FC−CC	1.474	0.314	2.634	0.007***
FC−IC	0.755	-0.405	1.915	0.328

Table A.8. Free Combinations

Tag	Item	Length	Syllables	Frequency	logDice	MI
fc1	buy tickets	10	3	65,311	9.300	7.640
fc2	sell books	9	2	24,958	7.370	3.630
fc3	burn coal	8	2	8,334	8.960	8.070
fc4	pay taxes	8	3	94,237	10.360	6.060
fc5	sing songs	9	2	66,963	11.500	8.960
fc6	wash dishes	10	3	7,441	9.690	7.890
fc7	build houses	11	3	72,980	9.100	5.220
fc8	cross rivers	11	3	11,389	9.390	6.970
fc9	eat fish	7	2	20,725	8.460	6.250
fc10	feel pain	8	2	21,536	8.990	4.090
fc11	lock doors	9	2	17,657	9.720	7.440
fc12	drink beer	9	2	10,631	10.040	8.070
fc13	break windows	12	3	9,753	8.140	3.850
fc14	hear sounds	10	2	21,598	9.160	5.310
fc15	teach kids	9	2	26,187	8.480	5.780

Table A.9. Congruent Collocations

Tag	Item	Length	Syllables	Frequency	logDice	MI
cc1	waste time	9	2	63,654	8.540	5.990
cc2	lose weight	10	2	74,300	10.170	6.730
cc3	take power	9	2	19,929	5.560	0.520
cc4	bring peace	10	2	20,605	7.550	4.850
cc5	earn trust	9	2	5,380	7.590	5.910
cc6	keep promises	12	4	16,360	7.490	4.990
cc7	raise doubts	11	2	6,802	6.620	5
cc8	catch fire	9	2	29,614	9.260	6.740
cc9	hide secrets	11	3	2,892	7.860	6.230
cc10	mask smells	10	2	2,310	6.570	6.580
cc11	cut costs	8	2	29,210	8.840	4.960
cc12	lift bans	8	2	9,939	9.910	8.880
cc13	show love	8	2	17,175	7.300	2.880
cc14	freeze accounts	14	3	1,846	6.640	4.460
cc15	read minds	9	2	12,877	6.650	2.710

Table A.10. Incongruent Collocations

Tag	Item	Length	Syllables	Frequency	logDice	MI
e01	enter contests	13	3	7,870	7.440	6.210
e02	push boundaries	14	3	21,371	10.130	8
e03	drive growth	11	2	38,996	8.660	5.670
e04	walk dogs	9	2	12,059	8.960	4.700
e05	foot bills	9	2	7,659	8.850	10.750
e06	hold meetings	12	3	174,292	10.330	6.870
e07	crack jokes	11	2	3,225	9.530	8.820
e08	meet goals	9	2	57,372	9.020	5.340
e09	shift blame	10	2	2,354	8.130	6.560
e010	kick habits	9	3	2,505	8.350	6.520
e011	drop hints	9	2	2,161	7.520	5.950
e012	chase dreams	11	2	5,327	8.110	7.750
e013	fly flags	8	2	8,197	8.620	7.080
e014	fight poverty	12	4	5,538	7.590	6.100
e015	bend rules	9	2	2,541	6.100	4.830

Table A.11. Baseline Items

Tag	Item	Length	Syllables	Frequency
bl1	bend coal	8	2	2
bl2	break jokes	10	2	7
bl3	bring boundaries	15	3	9
bl4	build secrets	12	2	21
bl5	burn smells	10	2	19
bl6	buy blame	8	2	7
bl7	catch bills	10	2	4
bl8	chase houses	11	3	8
bl9	crack promises	13	4	0
bl10	cross costs	10	2	28
bl11	cut fire	7	2	85
bl12	drink books	10	2	34
bl13	drive doors	10	2	30
bl14	drop growth	10	2	35
bl15	earn dishes	10	3	0
bl16	eat trust	8	2	8
bl17	enter flags	10	3	2
bl18	feel tickets	11	3	3
bl19	fight accounts	13	3	4
bl20	fly kids	7	2	47

Table A.12. Baseline Items (contd.)

Tag	Item	Length	Syllables	Frequency
bl21	foot minds	9	2	0
bl22	freeze power	11	3	30
bl23	hear habits	9	3	2
bl24	hide songs	9	2	9
bl25	hold poverty	11	4	14
bl26	keep sounds	10	2	43
bl27	kick love	8	2	15
bl28	lift fish	8	2	49
bl29	lock rules	9	2	36
bl30	lose goals	9	2	53
bl31	mask time	8	2	45
bl32	meet peace	9	2	42
bl33	pay dreams	9	2	2
bl34	push contests	12	3	7
bl35	raise bans	9	2	1
bl36	read dogs	8	2	71
bl37	sell weight	10	2	39
bl38	shift hints	10	2	7
bl39	show beer	8	2	54
bl40	sing rivers	10	3	5
bl41	take doubts	10	2	12
bl42	teach meetings	13	3	8
bl43	walk windows	11	3	5
bl44	wash taxes	9	3	0
bl45	waste pain	9	2	19

A.3 Ethics

A.3.1 Approval

Re: MSc Applied Linguistics for Second Language Acquisition/Applied Linguistics for Language Teaching CUREC application for Sydelle de Souza

Sandra Mathers <sandra.mathers@education.ox.ac.uk>
Mon 15/03/2021 10:57
To: Student CUREC <student.curec@education.ox.ac.uk>; Sydelle De Souza <sydelle.desouza@stcatz.ox.ac.uk>; Hamish Chalmers <hamish.chalmers@education.ox.ac.uk>

Dear Hamish and Sydelle

Processing Collocations in the L2: Congruency & Semantic Transparency - approval reference ED-CIA-21-126

The above application has been considered on behalf of the Departmental Research Ethics Committee (DREC) in accordance with the procedures laid down by the University for ethical approval of all research involving human participants. I am pleased to inform you that, on the basis of the information provided to DREC, the proposed research has been judged as meeting appropriate ethical standards, and accordingly, approval has been granted.

Please continue to follow all current guidance issued by CUREC during the pandemic, notably COVID-19: CUREC guidance on research involving human participants, https://researchsupport.admin.ox.ac.uk/governance/ethics/coronavirus

If relevant please also check the CUREC website for their best practice research guides, https://researchsupport.admin.ox.ac.uk/governance/ethics/resources/bpg

Should there be any subsequent changes to the project which raise ethical issues not covered in the original application you should submit details to research.office@education.ox.ac.uk for consideration.

Good luck with your research study.

With kind regards,

Dr. Sandra Mathers
Senior Researcher
Department of Education
University of Oxford

Figure A.1. CUREC Approval (received via email)

A.3.2 Recruitment Material



Figure A.2. Recruitment Poster for Brazilian Participants



Figure A.3. Recruitment Poster for Portuguese Participants



Figure A.4. Recruitment Poster for UK Participants



Figure A.5. Recruitment Poster for Other L1 English Participants

A.3.3 Participant Recruitment Emails

Email 1

Dear XXX,

To whom it may concern,

I am a student at the Department of Education at the University of Oxford, pursuing an MSc in Applied Linguistics and Second Language Acquisition. For my dissertation, I am conducting a psycholinguistic study on how L2 learners of English process collocations with a focus on L1 Portuguese speakers. Therefore, I am writing to ask you if it would be possible to undertake data collection in your classroom/department, as your stu-

dents fit the criteria of my study. The study has been approved by the Education Department Research Ethics Committee (Reference No. ED-CIA-21-126). The data collection will take place online via Gorilla and the tasks will take approximately 20 minutes to finish. Participation is voluntary and the research data will be anonymous and confidential. Upon publication, a plain language summary of the results will be made available to those who are interested. Anyone that agrees to participate and completes all the tasks will be eligible to enter their email address into a lucky draw to win 6 gift cards of XXX amount. I would be glad to talk to your students about my research after the data has been collected. If you have any further questions, please do not hesitate to contact me (sydelle.desouza@stcatz.ox.ac.uk) or my supervisor, Dr Hamish Chalmers (hamish.chalmers@education.ox.ac.uk). We would be glad to hear from you. Thank you for your time. I look forward to hearing from you.

Regards,

Sydelle de Souza

Email 2

To whom it may concern,

I am reading for an MSc in Applied Linguistics and Second Language Acquisition at the Department of Education, University of Oxford and for my dissertation I am studying how L1 Portuguese speakers process word combinations in L2 English. Psycholinguistic data is sparse in the collocational acquisition literature and my study aims to contribute to filling this gap. Given the pandemic, I'm doing an online experiment. This means that I need more participants than would usually suffice. So, I was wondering if you could help me reach out to potential participants through your mailing list or by other means such as your Instagram or FB page.

Below, you will find a visual and the details of my study along with the link to access it. I would be very grateful if you could help me out. Additionally, I am the Academic Officer at St Catherine's College, so if researchers at your department need participants for their studies, or if you'd like to advertise your department's events, I can do the same with the mailing lists to which I have access. Thank you so much for your time.

Regards,

Sydelle de Souza

A.3.4 Participant Information Sheet

Processing Collocations in the L2: Congruency & Semantic Transparency

CUREC Approval Reference: ED-CIA-21-126

General Information

The aim of this study is to contribute to the understanding of how people process word combinations. We appreciate your interest in participating in this online task. You have been invited to participate as you are above the age of 18 and a native English or Portuguese speaker. Please read through this information before agreeing to participate (if you wish to) by ticking the 'yes' box below. You may ask any questions before deciding to take part by contacting the researcher (details below). The Principal Researcher is Sydelle de Souza, who is attached to the Department of Education at the University of Oxford. This project is being completed under the supervision of Dr Hamish Chalmers.

You will be asked to fill out a survey about your linguistic background, dexterity, and vision, after which you will be asked to perform a simple word recognition task. In this task you will be shown a combination of words in English (for example, FRY ONIONS). You will then be asked to say whether this combination of words is 'allowed' or not in English. The whole process should take about 15 minutes. No background knowledge is required. There will be instructions to guide you throughout this process.

Your responses will help us to understand how word combinations are recognized and processed visually by language users. Only the research team will be able to see your answers. However, if necessary and in line with University policy, the data may be made available to responsible members of the University of Oxford with oversight for data management or as representatives of the university in its capacity as data controller. On completion and publication of the study we will make fully anonymized data available to other researchers on request, assuming we believe such requests to be reasonable. Data will be de-identified before it is shared with other researchers or results are made public.

Do I have to take part?

No. Please note that participation is voluntary. If you do decide to take part, you may withdraw at any point for any reason before submitting your answers by closing the browser. However, only the participants who complete all three activities will be eligible to enter the lucky draw.

Will I be compensated for my time?

Unfortunately, we are not able to pay all participants, but we do have a lucky draw set up for 6 vouchers of EUR 25 each. If you wish to participate, you can leave your email at the end of the experiment and if you win, you will be contacted directly via that email address.

How will my data be used?

We will not collect any data that could directly identify you, as routine. For example, we will not ask you for your name, and we will not collect your IP address . However, if you choose to enter the lucky draw, we will ask you for your email address. We will not share your email address with anyone. All email addresses will be deleted as soon as the lucky draw is completed, and the winners have been contacted. Your responses on the tests will be labelled by a code number and not by your name. They will be stored in a password-protected electronic file. When we write up the study we will not include your name or any information that could identify you. Research data will be stored for three years after publication or public release. This project will be written up for an M.Sc. degree in Applied Linguistics and Second Language Acquisition.

Who will have access to my data?

Your data will be collected through Gorilla™, please see their privacy notice here Gorilla Due Diligence. After data collection is complete, the data will be downloaded to the researcher's computer.

Who has reviewed this study?

This project has been reviewed by, and received ethics clearance through, the University of Oxford Central University Research Ethics Committee ED-CIA-21-126.

Who do I contact if I have a concern or complaint?

If you have a concern about any aspect of this study, please contact Sydelle de Souza (sydelle.desouza@stcatz.ox.ac.uk) or their supervisor, Dr Hamish Chalmers (hamish.chalmers@education.ox.ac.uk), and we will do our best to answer your query. We will acknowledge your concern within 10 working days and give you an indication of how it will be dealt with.

If you remain unhappy or wish to make a formal complaint, please contact the Chair of the Research Ethics Committee at the University of Oxford who will seek to resolve the matter as soon as possible:

Social Sciences & Humanities Interdivisional Research Ethics Committee
Email: ethics@socsci.ox.ac.uk
Address: Research Services, University of Oxford, Wellington Square, Oxford OX1 2JD

OR

Education Departmental Research Ethics Committee (DREC)

Email: student.curec@education.ox.ac.uk

Address: Department of Education, University of Oxford, 15 Norham Gardens, Oxford, OX2 6PY

Please note that you may only participate in this survey if you are 18 years of age or over.

☐ I certify that I am 18 years or over

If you have read the information above and agree to participate with the understanding that the data (including any personal data) you submit will be processed accordingly, please check the relevant box below to get started.

☐ Yes, I agree to take part

A.3.5 Sociodemographics Questionnaire

Please answer the following questions about yourself.

Age: _____

Gender: ☐ Female ☐ Male ☐ Non-binary ☐ Other (please specify)

Handedness: ☐ Left ☐ Right ☐ Ambidextrous (both)

Do you have any visual impairments, aside from wearing contacts and/or glasses?

☐ Yes ☐ No

Have you ever been diagnosed with dyslexia or any other language pathologies?

☐ Yes ☐ No ☐ Unsure

What is your highest level of education? If you are a student, please select your current level of study.

☐ High School (or equivalent) ☐ Bachelor's Degree (or equivalent) ☐ Master's Degree (or equivalent) ☐ Ph.D. (or equivalent) ☐ Other (please specify)

What is your first (native) language?

☐ English ☐ Portuguese (European) ☐ Portuguese (Brazil) ☐ Portuguese (Other) ☐ Other (please specify)

A.3.6 L2 Background Questionnaire

Please answer the following questions about your English learning abilities and background.

Age when you started learning English: ____

Where did you start learning English?

☐ Home ☐ School ☐ University ☐ Online (e.g. Duolingo) ☐ Language School ☐ Other (please specify)

Which variety of English did/do you learn?

☐ American English ☐ British English ☐ Both ☐ Don't know ☐ Other (please specify)

On a scale of 1 (beginner) to 5 (proficient), how would you rate your English proficiency in the following areas:

Reading: 1☐ 2☐ 3☐ 4☐ 5☐

Listening: 1☐ 2☐ 3☐ 4☐ 5☐

Writing: 1☐ 2☐ 3☐ 4☐ 5☐

Speaking: 1☐ 2☐ 3☐ 4☐ 5☐

What is your CEFR level in English?

☐ A1-A2 (Beginner) ☐ B1 (Intermediate) ☐ B2 (Intermediate) ☐ C1(Proficient) ☐ C2 (Proficient)

Did/Do you study at a school, college or university where English is/was the medium of instruction?

☐ Yes ☐ No

Have you lived (or do you live) or spent an extended period of time (more than 3 months)) in a majority English-speaking country?

☐ Yes ☐ No

If yes, for how long?

☐ Less than 6 months ☐ 6 months to 1 year ☐ More than 1 year ☐ 1 to 3 years ☐ More than 3 years

Have you ever learned English in a study abroad context?

☐ Yes ☐ No

Excluding English and your first language, please list the other languages in which you are fluent enough to read a children's story. If you do not speak any other languages, then please type "none" to continue.

_____

A.3.7 Debrief Form

Did you experience any distractions while completing the tasks? ☐ Yes ☐ No

Did you experience any internet connectivity issues during the tasks? ☐ Yes ☐ No

Were there times when the task was slow or did not respond when you clicked or pressed a button? ☐ Yes ☐ No

Would you like to leave a comment about your experience? _____

Thank You For Your Participation!

We greatly appreciate your interest and participation in our research project.

If you have a concern about any aspect of this study, please contact the principal researcher, Sydelle de Souza (sydelle.desouza@stcatz.ox.ac.uk), or Dr Hamish Chalmers (hamish.chalmers@education.ox.ac.uk), and we will do our best to answer your query. We will acknowledge your concern within 10 working days and give you an indication of how it will be dealt with.

If you remain unhappy or wish to make a formal complaint, please contact the Chair of the Research Ethics Committee at the University of Oxford who will seek to resolve the matter as soon as possible:

Social Sciences & Humanities Interdivisional Research Ethics Committee
Email: ethics@socsci.ox.ac.uk;
Address: Research Services, University of Oxford, Wellington Square, Oxford OX1 2JD

OR

Education Departmental Research Ethics Committee (DREC)
Email: student.curec@education.ox.ac.uk;
Address: Department of Education, University of Oxford, 15 Norham Gardens, Oxford, OX2 6P

If you would like to receive a plain language summary of the results of this study, please leave your email below:

_____

LUCKY DRAW

If you would like a chance to win gift vouchers worth €25, please leave your email below:

_____

A.3.8 Task Instructions

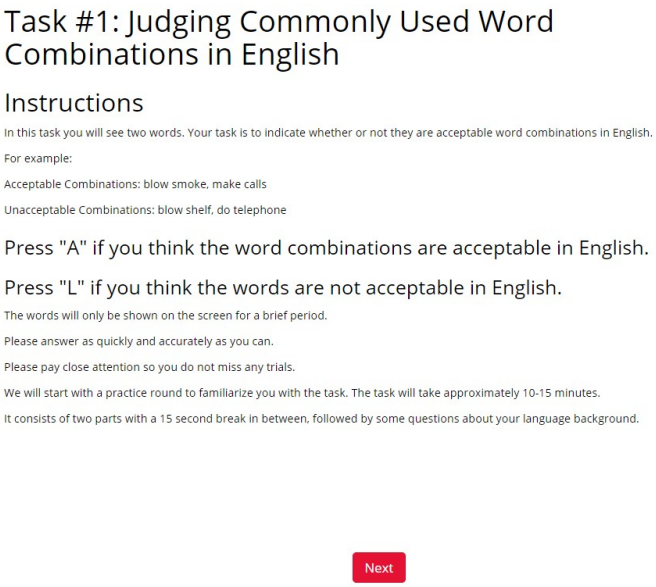


Figure A.6. Semantic Acceptability Judgement Task Instructions (Screenshot)

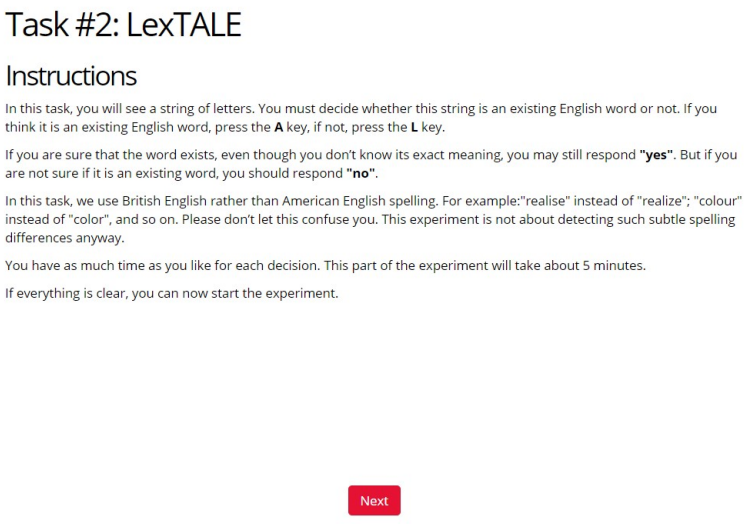


Figure A.7. LexTALE Instructions (Screenshot)

Appendix B

Results

B.1 Global Results

Table B.1. Descriptive Statistics: RT by L1 Group

	L1	Mean	SD	Median	IQR	Min	Max
1	BP	1328.09	589.03	1151.51	661	487	3999.5
2	EN	1187.34	561.82	1015.44	591.48	459.51	3996.59
3	EP	1328.66	607.80	1143.5	678.14	483.51	3997.32

Table B.2. Descriptive Statistics: ER by L1 Group

	L1	Mean	SD	Median	IQR	Min	Max
1	BP	0.83	0.37	1.00	0.00	0	1
2	EN	0.86	0.35	1.00	0.00	0	1
3	EP	0.83	0.38	1.00	0.00	0	1

Table B.3. Descriptive Statistics: RT by L1 Group and Condition

Condition	L1	Mean	SD	Median	IQR	Min	Max
Baseline Items	BP	1514.94	622.12	1335.28	751.60	487	3999.5
	EN	1274.39	570.77	1104	624.40	459.51	3996.59
	EP	1475.81	634.85	1285.37	767.85	504.39	3997.32
Incongruent Collocations	BP	1297.12	543.2	1136.65	599.87	570.70	3989.88
	EN	1142.2	563.98	952.33	553.52	462.93	3938.45
	EP	1329.24	589.78	1146.03	674.00	483.51	3963.05
Congruent Collocations	BP	1133.65	485.76	986.40	441.59	540	3914.30
	EN	1120.99	549.99	936.89	559.57	471.51	3924.97
	EP	1160.77	522.76	999.1	504.77	500.85	3947.66
Free Combinations	BP	997.3	372.35	896.68	307.92	528.89	3739
	EN	1041.05	496.74	884.85	462.85	466.30	3929.07
	EP	1058.28	471.05	912	401.31	534.91	3875

Table B.4. Descriptive Statistics: ER by L1 Group and Condition

Condition	L1	Mean	SD	Median	IQR	Min	Max
Baseline Items	BP	0.77	0.42	1.00	0.00	0	1
	EN	0.80	0.40	1.00	0.00	0	1
	EP	0.79	0.41	1.00	0.00	0	1
Incongruent Collocations	BP	0.77	0.42	1.00	0.00	0	1
	EN	0.87	0.34	1.00	0.00	0	1
	EP	0.74	0.44	1.00	1.00	0	1
Congruent Collocations	BP	0.93	0.26	1.00	0.00	0	1
	EN	0.91	0.28	1.00	0.00	0	1
	EP	0.89	0.31	1.00	0.00	0	1
Free Combinations	BP	0.99	0.12	1.00	0.00	0	1
	EN	0.95	0.21	1.00	0.00	0	1
	EP	0.96	0.19	1.00	0.00	0	1

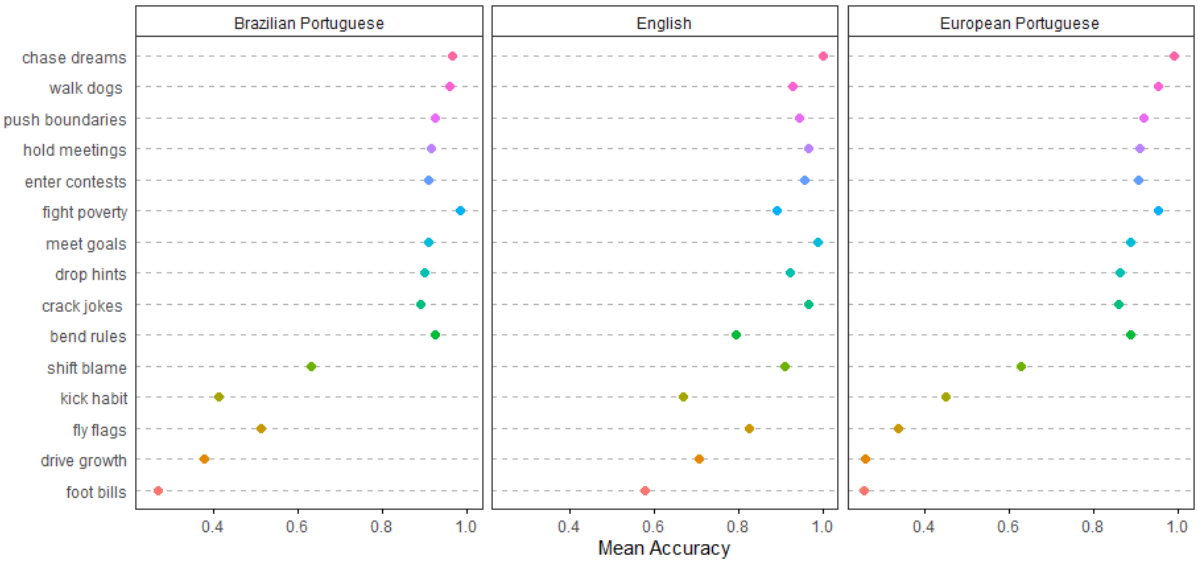


Figure B.1. Mean Error Rate by Item (Incongruent Collocations)

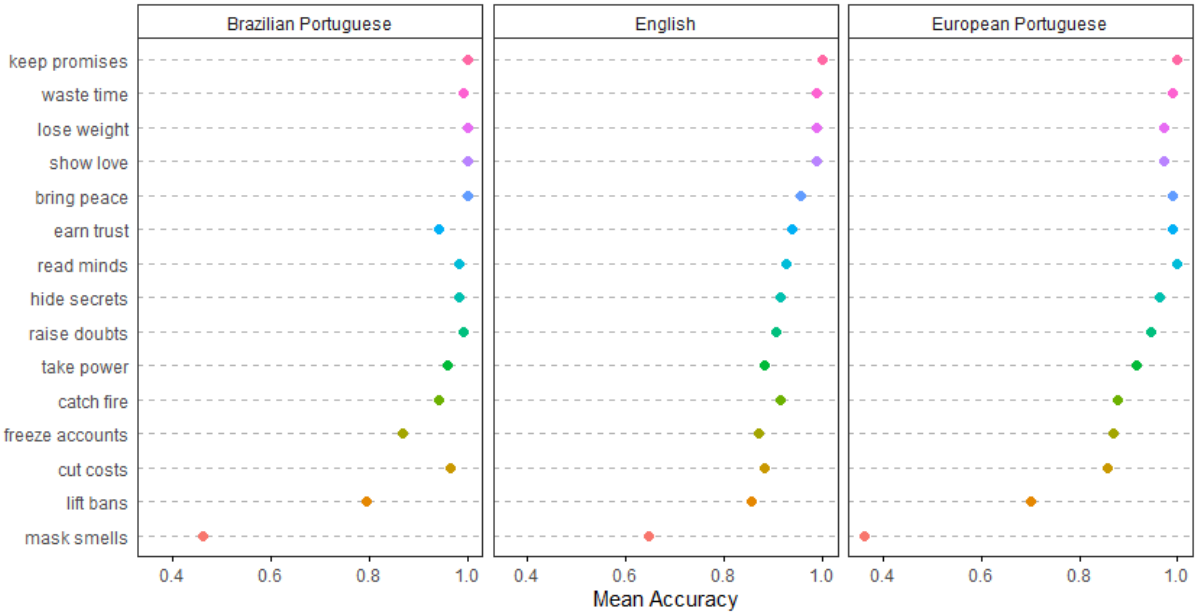


Figure B.2. Mean Error Rate by Item (Congruent Collocations)

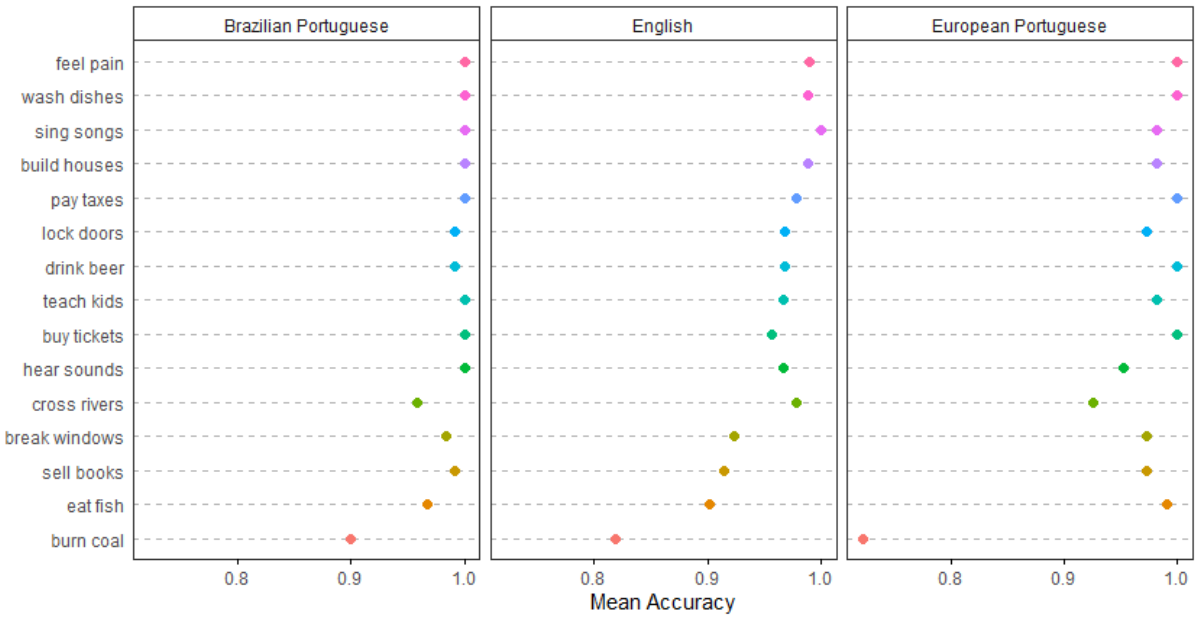


Figure B.3. Mean Error Rate by Item (Free Combinations)

B.2 GLMM Model Outputs

Table B.5. GLMM Outputs: RT Model, logRT Model, ER Model

	Dependent variable:		
	RT	logRT	ER
	(1)	(2)	(3)
BP	210.237*** (27.326)	0.177*** (0.017)	−0.434** (0.189)
EP	191.684*** (28.047)	0.158*** (0.017)	−0.274 (0.194)
IC	−145.099*** (29.751)	−0.134*** (0.022)	1.110*** (0.431)
CC	−165.678*** (28.545)	−0.149*** (0.021)	1.508*** (0.426)
FC	−237.500*** (29.025)	−0.206*** (0.021)	1.905*** (0.427)
Distractions (Yes)	91.245*** (18.819)	0.069*** (0.014)	−0.543*** (0.140)
Collocation Length	28.702*** (7.726)	0.026*** (0.006)	0.002 (0.136)
Age	58.699*** (7.586)	0.055*** (0.006)	−0.056 (0.058)
BP:IC	−86.374*** (29.624)	−0.038* (0.021)	−0.821*** (0.272)
EP:IC	−47.608 (30.528)	−0.005 (0.021)	−1.235*** (0.279)
BP:CC	−197.592*** (27.149)	−0.141*** (0.019)	0.496* (0.265)
EP:CC	−157.992*** (27.926)	−0.107*** (0.020)	−0.198 (0.266)
BP:FC	−227.940*** (27.897)	−0.175*** (0.020)	1.486*** (0.303)
EP:FC	−168.143*** (28.723)	−0.121*** (0.020)	0.316 (0.268)
Constant	1,234.231*** (23.020)	7.054*** (0.015)	2.112*** (0.237)
Observations	24,125	24,125	28,802
Log Likelihood	−174,128.200	−3,920.891	−9,250.976
Akaike Inf. Crit.	348,314.300	7,899.782	18,557.950
Bayesian Inf. Crit.	348,549.000	8,134.421	18,789.460

FC: Free combinations; CC: Congruent Collocations; IC: Incongruent Collocations; BL: Baseline Items

*p<0.1; **p<0.05; ***p<0.01

Standard Errors in parentheses

Factor levels: Baselines, English (EN)

B.3 logRT Model

The R syntax specified for the logRT model:

```
logRT ~l1*condition + length + distractions + age +
(1|item) + (1|participant) + (condition||participant)
```

Table B.6. Mixed Model ANOVA Results (logRT Model)

	Effect	df	Chisq	P-value
1	L1	2	9.09 *	.011
2	Condition	3	30.76 ***	<.001
3	Distractions	1	6.21 *	.013
4	Collocation Length	1	2.26	.133
5	Age	1	15.16 ***	<.001
6	L1*Condition	6	57.30 ***	<.001

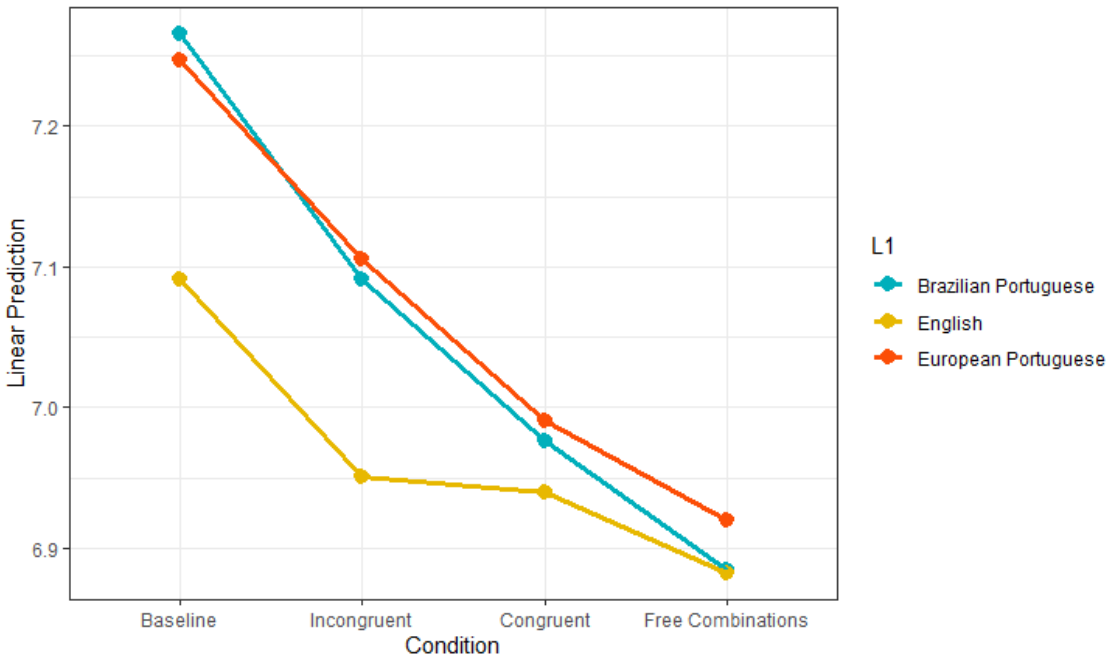


Figure B.4. Estimated Marginal Means of Condition by L1 (logRT Model)

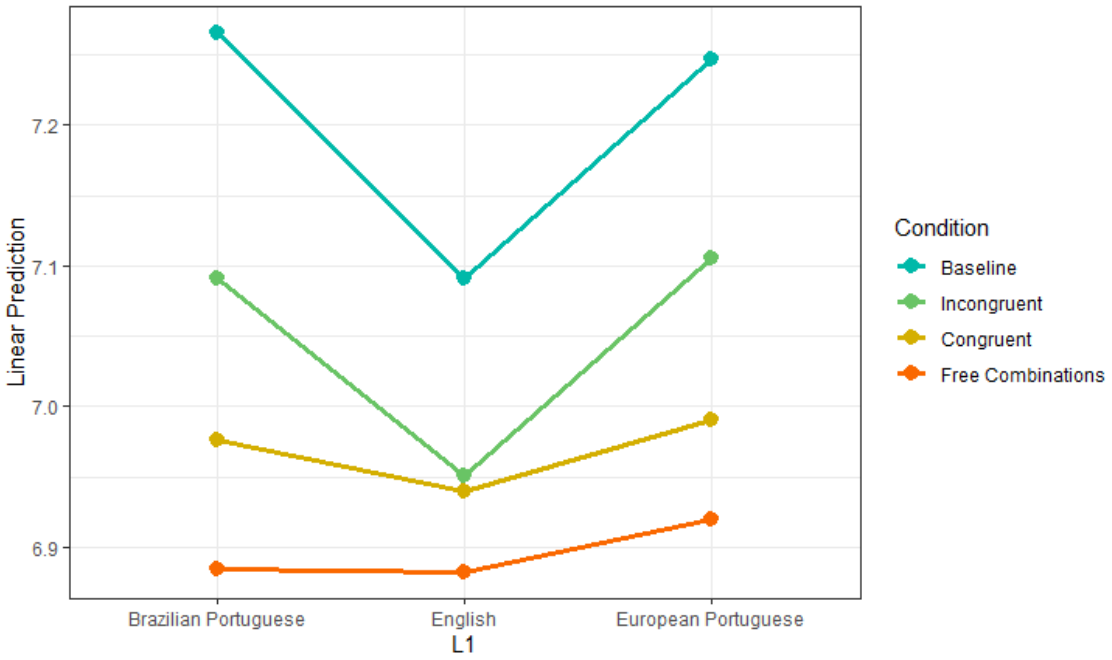


Figure B.5. Estimated Marginal Means of L1 by Condition (logRT Model)

Table B.7. Pairwise Contrasts of Condition by L1 for Log-Transformed Reaction Time

L1	Contrast	Estimate (Δ)	SE	Z-ratio	P-value
Brazilian Portuguese	FC–CC	−0.09	0.02	−3.846	0.0007***
	FC–IC	−0.21	0.02	−8.355	<.0001***
	CC–IC	−0.12	0.02	−5.130	<.0001***
	FC–BL	−0.38	0.02	−17.253	<.0001***
	CC–BL	−0.29	0.02	−15.335	<.0001***
	IC–BL	−0.17	0.02	−8.906	<.0001***
English	FC–CC	−0.06	0.03	−2.267	0.1057
	FC–IC	−0.07	0.03	−2.649	0.0402**
	CC–IC	−0.01	0.02	−0.500	0.9591
	FC–BL	−0.21	0.02	−8.795	<.0001***
	CC–BL	−0.15	0.02	−7.643	<.0001***
	IC–BL	−0.14	0.02	−6.910	<.0001***
European Portuguese	FC–CC	−0.07	0.02	−2.910	0.0189**
	FC–IC	−0.19	0.03	−7.278	<.0001***
	CC–IC	−0.11	0.02	−4.966	<.0001***
	FC–BL	−0.33	0.02	−14.337	<.0001***
	CC–BL	−0.26	0.02	−13.240	<.0001***
	IC–BL	−0.14	0.02	−7.053	<.0001***

FC: Free combinations; CC: Congruent Collocations; IC: Incongruent Collocations; BL: Baseline Items

Degrees of Freedom (df) = Inf

Results are averaged over the levels of: Distractions

P-value adjustment: Tukey method for comparing a family of 4 estimates

Significance codes: '***' < 0.001; '**' < 0.01; '*' < 0.05

Table B.8. Pairwise Contrasts of Condition by L1 for Log-Transformed Reaction Time

Condition	Contrast	Estimate (Δ)	SE	Z-ratio	P-value
Baseline Items	BP−EN	0.18	0.02	10.133	<.0001***
	EP−EN	0.16	0.02	8.818	<.0001***
	EP−BP	−0.02	0.02	−1.116	0.5044
Incongruent Collocations	BP−EN	0.14	0.02	7.584	<.0001***
	EP−EN	0.15	0.02	8.055	<.0001***
	EP−BP	0.01	0.02	0.760	0.7276
Congruent Collocations	BP−EN	0.04	0.02	2.087	0.0925
	EP−EN	0.05	0.02	2.828	0.0130**
	EP−BP	0.01	0.02	0.860	0.6655
Free Combinations	BP−EN	0.00	0.02	0.103	0.9942
	EP−EN	0.04	0.02	1.725	0.1959
	EP−BP	0.03	0.02	1.732	0.1933

BP = Brazilian Portuguese; EP = European Portuguese; EN = English

Degrees of Freedom (df) = Inf

Results are averaged over the levels of: Distractions

P-value adjustment: Tukey method for comparing a family of 3 estimates

Significance codes: '***' < 0.001; '**' < 0.01; '*' < 0.05