

**Prosódia e
Bilinguismo**

**Prosodia y
Bilingüismo**

**Prosody and
Bilingualism**

Amanda Post da Silveira
Pablo Arantes
Organizadores

Amanda Post da Silveira
Pablo Arantes
(Organizadores)

Amanda Post da Silveira
(Versões português-inglês e traduções inglês e espanhol-português)

Prosódia e Bilinguismo

Prosodia y Bilingüismo

Prosody and Bilingualism

Araraquara
Letraria
2022

FICHA CATALOGRÁFICA

Dados Internacionais de Catalogação na Publicação (CIP) (Câmara Brasileira do Livro, SP, Brasil)

Prosódia e bilinguismo = Prosodia y bilingüismo = Prosody and bilingualism [livro eletrônico] / Amanda Post da Silveira, Pablo Arantes (organizadores). -- Araraquara, SP : Letraria, 2022.

PDF.

Vários autores.

Ed. trilingue: português/espanhol/inglês.

ISBN 978-65-5434-005-2

1. Análise prosódica (Linguística) 2. Bilinguismo I. Silveira, Amanda Post da. II. Arantes, Pablo.

III. Título: Prosodia y bilingüismo

IV. Título: Prosody and bilingualism

22-125665

CDD-410

Índices para catálogo sistemático:

1. Linguística 410

Eliete Marques da Silva - Bibliotecária - CRB-8/9380

Conselho editorial

Maria de Fátima de Almeida Baia (UESB)

Ubiratã Kickhöfel Alves (UFRGS)

| SUMÁRIO

Sobre os autores About the authors	6 7
Introdução Introduction	10 13
Prosody	
Chapter 1 - Algunos aspectos metodológicos en el estudio del acento léxico Joaquim Llisterri	17
Alguns aspectos metodológicos no estudo do acento lexical Joaquim Llisterri	46
Chapter 2 - Towards a comprehensive analysis of Franconian tonal accent Björn Köhnlein	76
Rumo a uma análise abrangente do acento tonal do francônio Björn Köhnlein	98
Bilingualism	
Chapter 3 - O usuário de L2 sob perspectiva Bilíngue Ricardo Augusto de Souza	119
Chapter 4 - Implications of a usage-based approach to contact linguistics Ad Backus	132
Implicações de um enfoque baseado no uso para a linguística de contato Ad Backus	149
Prosody and bilingualism	
Chapter 5 - Resolving the prosody paradox Vincent J. van Heuven	168
Resolvendo o paradoxo da prosódia Vincent J. van Heuven	205
Chapter 6 – O acento primário em modelos psicolinguísticos de produção de palavra: propostas para L1 e L2 Amanda Post da Silveira	242
Chapter 7 - Análise quantitativa da entoação de perguntas interrogativas pronominais: evidências de aprendizagem e atrito na produção de falantes de espanhol L2 brasileiros Cristiane Conceição Silva Pablo Arantes	257

| Sobre os autores

Vincent van Heuven é Professor Emérito de Linguística Experimental e Fonética na Universidade de Leiden (Holanda, desde 2014) e na Universidade da Panônia (Hungria, desde 2019). Ele tem cátedras honorárias na Universidade de Nankai (Tianjin, China), Universidade de Groningen (Holanda) e na Frisian Academy (Leeuwarden, Holanda). Ele passou um ano de pós-graduação na Escócia (Dept. Linguistics, Edinburgh University) e um ano de pós-doutorado nos EUA (Dept. Linguistics, UC Los Angeles). Foi diretor do Holland Institute of Linguistics (1999-2001) e do Leiden University Centre for Linguistics (2001-2006). Seus interesses de pesquisa são em prosódia (acento, tom, entonação), reconhecimento de palavras, inteligibilidade entre idiomas, aquisição de segunda língua e fonética forense. É membro vitalício da Royal Netherlands Academy of Arts and Sciences. Para mais informações, veja: <https://www.universiteitleiden.nl/en/staffmembers/vincent-van-heuven#tab-1>; <https://www.researchgate.net/profile/Vincent-Van-Heuven>

Joaquim Llisterri é Professor de Linguística Geral na Universidade Autônoma de Barcelona. Sua pesquisa se concentrou na incorporação do conhecimento fonético às tecnologias da fala, na criação e anotação de corpus oral, na interferência fonética na aquisição de segundas línguas e na descrição prosódica do espanhol.

Ricardo Augusto de Souza é Professor Titular de Língua Inglesa da Faculdade de Letras da UFMG. Como docente do curso de graduação em Letras: Inglês da UFMG, ocupa-se primariamente do ensino da descrição da organização linguística do inglês. Como pesquisador, concentra seus esforços nos seguintes temas: (1) aprendizagem e uso de construções e seus efeitos sobre o processamento da linguagem e sobre as habilidades linguísticas bilíngues; (2) apreensões dos sentidos de bilinguismo e seus efeitos sobre a docência em propostas de educação bilíngue; (3) constituição institucional e epistemológica da psicolinguística brasileira; (4) exploração de diálogos entre teorias linguísticas baseadas no uso e as formações do inconsciente na teoria psicanalítica. Atua como docente do quadro permanente do Programa de Pós-Graduação em Estudos Linguísticos da UFMG e é bolsista de Produtividade em Pesquisa (nível 2) do CNPq.

Ad Backus é Professor Titular de Linguística e Sociolinguística na Universidade de Tilburg, na Holanda. Ele também é o editor do *International Journal of Bilingualism*. Seu trabalho é principalmente sobre contato linguístico, de vários ângulos teóricos e empíricos, com maior atenção empírica para o contato turco-holandês na Holanda. Em trabalhos recentes, ele tentou articular uma abordagem baseada no uso para questões de variação linguística e mudança linguística.

Amanda Post da Silveira recebeu o título de doutora pelo Instituto Donders para o Cérebro, Cognição e Comportamento da Universidade Radboud Nijmegen nos Países Baixos. Fez parte de seu doutorado no Centro para Linguística da Universidade de Leiden. Atua nos seguintes temas: Psicolinguística

Experimental, Fonética Acústica Experimental, Aquisição de Segundas Línguas, Fonética e Fonologia da Língua Inglesa, Linguística de Corpus e Linguística Aplicada Crítica. É atualmente Professora Adjunta de Língua Inglesa e Ensino da Universidade Federal de Jataí (UFJ).

Cristiane Conceição Silva é Professora Adjunta no Departamento de Letras e Literaturas Estrangeiras da Universidade Federal de Santa Catarina (UFSC). Doutora em Linguística (UNICAMP). Mestre em Linguística pela USP e especialização em Fonética e Fonologia do Espanhol realizada em Madri pelo Centro Superior de Investigaciones Científicas e Universidad Menéndez Pelayo. Atua no âmbito da fonética experimental e desenvolve pesquisa sobre a prosódia do espanhol L2.

Pablo Arantes é Doutor (2010) e Bacharel (2003) em Linguística pela Universidade Estadual de Campinas. Interesses de pesquisa principais: (1) descrição fonético-acústica detalhada de fenômenos linguísticos, especialmente os prosódicos, (2) desenvolvimento de ferramentas computacionais e estatísticas necessárias para a atividade descritiva e (3) aplicações de conhecimento linguístico para a fonética forense e (4) aspectos fonéticos do bilinguismo. Coordena o Laboratório de Fonética da Universidade Federal de São Carlos.

Björn Köhnelin é Professor Associado do Departamento de Linguística da Ohio State University. Sua especialidade principal é a fonologia – particularmente interações de estrutura métrica, tom (entonacional) e estrutura segmentar – e suas interfaces com fonética e morfologia, tanto de uma perspectiva tipológica sincrônica quanto diacrônica.

| **About the authors**

Vincent van Heuven is an emeritus professor of Experimental Linguistics and Phonetics at Leiden University (Netherlands, since 2014) and at the University of Pannonia (Hungary, since 2019). He has honorary professorships at Nankai University (Tianjin, China), Groningen University (Netherlands), and the Frisian Academy (Leeuwarden, Netherlands). He spent a postgraduate year in Scotland (Dept. Linguistics, Edinburgh University) and a postdoctoral year in the USA (Dept. Linguistics, UC Los Angeles). He was the director of the Holland Institute of Linguistics (1999-2001) and the Leiden University Centre for Linguistics (2001-2006). His research interests are prosody (stress, tone, intonation), word recognition, cross-lingual intelligibility, second language acquisition, and forensic phonetics. He is a life member of the Royal Netherlands Academy of Arts and Sciences. For more information, see <https://www.universiteitleiden.nl/en/staffmembers/vincent-van-heuven#tab-1>; <https://www.researchgate.net/profile/Vincent-Van-Heuven>

Joaquim Llisterri is full-time lecturer in General Linguistics at the Universitat Autònoma de Barcelona. His research has been devoted to the applications of experimental phonetics to speech technology, the creation and annotation of spoken corpora, the study of phonetic transfer in second language acquisition, and the prosodic description of Spanish.

Ricardo Augusto de Souza is a Full Professor of English at the Faculty of Arts of UFMG. As a professor of the undergraduate course in Letters: English at UFMG, he is primarily concerned with teaching the description of the linguistic organization of English. As a researcher, he focuses his efforts on the following topics: (1) learning and use of constructions and their effects on language processing and bilingual language skills; (2) apprehensions of the meanings of bilingualism and its effects on teaching in proposals for bilingual education; (3) institutional and epistemological constitution of Brazilian psycholinguistics; (4) exploration of dialogues between usage-based linguistic theories and unconscious formations in psychoanalytic theory. He works as a permanent professor at the Postgraduate Program in Linguistic Studies at UFMG and is a Research Productivity Scholar (level 2) from CNPq.

Ad Backus is a Full Professor of Linguistics and Sociolinguistics at Tilburg University in the Netherlands. He is also the editor of the *International Journal of Bilingualism*. His work is mostly on language contact, from various theoretical and empirical angles, with most empirical attention going to Turkish-Dutch contact in the Netherlands. In recent work, he has attempted to articulate a usage-based approach to issues of language variation and language change.

Amanda Post da Silveira holds a PhD from the Donders Institute for Brain, Cognition and Behavior at Radboud Nijmegen University in the Netherlands. She did part of her doctorate at the Centre for Linguistics at the University of Leiden. She works on the following subjects: Experimental Psycholinguistics, Experimental Acoustic Phonetics, Second Language Acquisition, English Phonetics and Phonology, Corpus Linguistics, and Critical Applied Linguistics. She is currently an Assistant Professor of English Language and Teaching at the Federal University of Jataí (UFJ).

Cristiane Conceição Silva is an Assistant Professor at the Department of Foreign Letters and Literature at the Federal University of Santa Catarina (UFSC). She holds a PhD in Linguistics from the State University of Campinas, a master in Linguistics from USP, and a specialization in Phonetics and Phonology of Spanish held in Madrid by Centro Superior de Investigaciones Científicas and Universidad Menéndez Pelayo. She works in the field of experimental phonetics and develops research on the prosody of L2 Spanish.

Pablo Arantes holds a PhD (2010) and a Bachelor's (2003) in Linguistics from the State University of Campinas. His main research interests include (1) detailed phonetic-acoustic description of linguistic phenomena, especially prosodic ones, (2) development of computational and statistical tools necessary for descriptive activity and (3) applications of linguistic knowledge to forensic phonetics, and (4) phonetic aspects of bilingualism. He coordinates the Phonetics Laboratory at the Federal University of São Carlos.

Björn Köhnlein is an Associate Professor in the Department of Linguistics at the Ohio State University. His core expertise is phonology – particularly interactions of metrical structure, (intonational) tone, and segmental structure – and its interfaces with phonetics and morphology, from both a synchronic and a diachronic typological perspective.

| Introdução

O presente livro é uma coletânea de capítulos produzidos por um grupo de especialistas nacionais e internacionais sobre os tópicos de prosódia, bilinguismo e sobre a interação entre prosódia e bilinguismo. A intenção deste volume é a de estabelecer uma ponte entre esses dois campos de pesquisa, que é uma relação promissora, embora pouco explorada tanto no cenário nacional quanto internacional de pesquisas sobre a linguagem.

O livro teve origem em um evento acadêmico, organizado por Amanda Post da Silveira e por Pablo Arantes, intitulado “Simpósio Prosódia e Bilinguismo”, que aconteceu de forma remota no período de maio a setembro de 2020 no contexto da pandemia.

A excepcionalidade da pandemia e a disseminação do trabalho remoto, que levou muitos eventos a adotarem um formato virtual, permitiram a organização de um evento que pode ser realizado com um custo muito baixo e ao mesmo tempo contar com a participação de pesquisadores de renome internacional, além de permitir a participação de um público ouvinte amplo do ponto de vista da procedência geográfica.

As falas deste simpósio foram tão interessantes e instrutivas, em uma área tão carente de eventos específicos, quanto os trabalhos em prosódia e bilinguismo, que intentamos, com a proposta deste livro, registrar de forma permanente e em textos mais longos os trabalhos apresentados, pelo caráter formador que podem ter.

É importante chamar a atenção dos leitores para o fato de que o termo bilinguismo, que aparece no título tanto do Simpósio quanto do presente livro, não é o único que aparece na literatura. Há expressões usadas em paralelo, como aquisição/desenvolvimento de língua estrangeira, aquisição/desenvolvimento de línguas não nativas e multilinguismo e nem todos eles podem ser considerados sinônimos. Tal heterogeneidade tem a ver com diferenças teóricas e metodológicas na definição do fenômeno. Essa diversidade de concepções e terminologias aparecem nas contribuições feitas pelos pesquisadores convidados para o simpósio e para produzirem os capítulos deste livro. No capítulo de autoria de van Heuven, por exemplo, cuja atuação primária concentra-se na área da fonética, o autor usa a expressão aquisição de língua estrangeira. Outras vezes, é simplesmente usado o termo bilinguismo em textos de viés psicolinguístico (como nos capítulos de Ricardo Augusto de Sousa e Amanda Post da Silveira), uma vez que a Psicolinguística tipicamente define o grau de bilinguismo dentro da realidade de cada comunidade investigada. Os organizadores do evento e do livro optaram pelo uso do termo bilinguismo conscientes da multiplicidade teórica e metodológica da área e conscientes, também, de que as discussões e contribuições feitas pelos pesquisadores convidados podem iluminar e enriquecer o debate, embora nem os autores nem os organizadores se pretendam a unificar a nomenclatura do campo.

Outro ponto particular e original deste livro é o de que ele traz contribuições em inglês e espanhol, que foram traduzidas para o português, para possibilitar a difusão dos trabalhos para o público estudantil e acadêmico brasileiro, mesmo para aqueles que não têm o domínio de leitura nessas línguas estrangeiras. As traduções dos textos em inglês e espanhol para o português são apresentadas logo em seguida aos textos originais nas línguas estrangeiras.

Um dos propósitos tanto do Simpósio Prosódia e Bilinguismo quanto do livro, que leva o mesmo nome, é o de reunir pesquisadores da comunidade internacional e brasileira interessados no entrecruzamento entre estudos sobre bilinguismo e prosódia e sobre a interação entre esses dois. Algumas das contribuições tratam primordialmente de um ou de outro dos dois tópicos e alguns exploram de maneira mais direta a relação entre os dois, o que pode ser explicado pelo fato de esse estudo conjunto ainda ser uma relativa novidade no cenário mundial. Um segundo propósito do evento e do livro foi o de abarcar diferentes perspectivas e temas no âmbito dessa relação: pesquisa experimental, pesquisa sócio-interacional e reflexão sobre modelos de análise teórica. Dado que a experimentação em linguística ainda é um modo de pesquisa em consolidação no ambiente da linguística brasileira, o evento também abriu espaço para contribuições de caráter metodológico, com o objetivo de acrescentar para a formação do público assistente, que majoritariamente constituiu-se de estudantes de graduação e pós-graduação.

Sobre as contribuições deste livro, os capítulos foram agrupados em três partes: na primeira, são reunidos os capítulos que tratam primordialmente do tema da prosódia; na segunda, são reunidas as contribuições relacionadas ao tema do bilinguismo; na terceira parte, são apresentados os capítulos que desenvolvem em igual medida os dois tópicos, a prosódia e o bilinguismo.

A primeira parte é aberta com o capítulo de Joaquim Llisterri (Universidade Autònoma de Barcelona), que apresenta uma revisão bastante abrangente da literatura a respeito de estratégias e técnicas de pesquisa adotadas para o estudo da produção dos parâmetros acústicos do acento lexical. Traz considerações sobre aspectos relevantes da atividade de experimentação, como desenho experimental, definição da composição e número de participantes na amostra, o que torna este capítulo um verdadeiro guia para aqueles que se interessam em investigar o acento lexical com excelência metodológica.

O segundo capítulo, de autoria de Björn Köhnlein (Ohio State University), apresenta uma análise da oposição entre dois tons lexicais no francônio, uma língua multi-regional falada em partes da Alemanha, Bélgica e Holanda. O texto descreve os fatos distribucionais dos tons, considerações sobre sua implementação melódica e duracional e algumas funções morfológicas que podem exercer. O capítulo termina com uma análise de caráter fonológico que propõe uma representação fonológica dos tons lexicais francônios, baseada em princípios autosegmentais.

Em seguida, a seção sobre bilinguismo começa com o capítulo em que Ricardo Augusto Souza (Universidade Federal de Minas Gerais) faz uma interessante discussão sobre bilinguismo, debatendo noções de transferência linguística e déficit representacional, muitas vezes atribuídas ao processamento e à representação da linguagem por bilíngues. Aponta maneiras mais adequadas de tratar o objeto de investigação do fenômeno do bilinguismo pelo viés da psicolinguística. É uma revisão de literatura bastante didática e necessária para a sequência deste livro.

Finalizando esta seção, o quarto capítulo, de autoria de Ad Backus, traz uma introdução bastante ampla e completa sobre o tema do contato linguístico, ilustrando sua discussão teórica principalmente com exemplos dos seus vários anos de pesquisa a respeito das línguas de herança na Holanda. Propõe no capítulo que a investigação do fenômeno de contato linguístico seja feita não só pelo viés sociolinguístico, ou psicolinguístico, mas por ambos.

Na última seção, sobre prosódia e bilinguismo, o quinto capítulo, produzido por Vincent van Heuven (Emeritus Leiden Centre for Linguistics), faz uma apresentação daquilo que caracteriza o nível prosódico em contraposição ao segmental e se encaminha para uma discussão que traz conceitualizações, explicações e exemplos bastante didáticos sobre o papel do nível segmental e da prosódia na percepção, inteligibilidade e compreensão monolíngues e bilíngues, apresentando exemplos de trabalhos internacionais icônicos no estudo da prosódia e promovendo um debate interessante sobre a importância (ou não) desta para inteligibilidade e compreensão de fala de modo geral.

No sexto capítulo, Amanda Post da Silveira (Universidade Federal de Jataí) discute importantes modelos psicolinguísticos de produção de fala de vertente conexionista, fazendo comparações destes modelos quanto à implementação do acento de palavra, visto que a maioria desses modelos deu conta somente de palavras monossilábicas. Além da revisão importante e didática desses modelos de produção, é feita a crítica da ausência ou pouca relevância dada à atribuição do acento de palavra em L1 e, ainda menos frequentemente, no processamento bilíngue.

Por fim, o sétimo capítulo, escrito por Cristiane Conceição Silva (Universidade Federal de Santa Catarina) e Pablo Arantes (Universidade Federal de São Carlos), apresenta uma análise quantitativa de contornos entoacionais de falantes bilíngues brasileiros de espanhol como L2 que mostra evidência da ocorrência de aprendizagem, no caso dos contornos em L2, e também de atrito linguístico, no caso dos contornos de L1 produzidos pelos bilíngues.

Encerra-se esta introdução com um indispensável agradecimento à comissão editorial, cujas observações e comentários contribuíram para o aperfeiçoamento tanto desta introdução quanto da organização deste livro. Tais melhorias devem-se, em grande parte, às excelentes sugestões do professor Ubiratã Kirchhoff Alves, a quem os organizadores dedicam especial agradecimento.

| Introduction

This book is a collection of chapters produced by a group of national and international experts on the topics of prosody, bilingualism, and the interaction between prosody and bilingualism. The intention of this volume is to establish a bridge between these two fields of research, which is a promising relationship, although little explored both in the national and international scenario of research on language.

The book originated from an academic event, organized by Amanda Post da Silveira and Pablo Arantes, entitled “International Symposium on Prosody and Bilingualism”, which occurred remotely from May to September 2020 in the context of the pandemic.

The exceptionality of the pandemic and the spread of remote work, which led many events to adopt a virtual format, favored the organization of the event that could be held at a very low cost and at the same time count on the participation of internationally renowned researchers, in addition to allowing for participants from all over the world to attend it.

The lectures of the symposium were so interesting and instructive, in an area as lacking in specific events, as the works in prosody and bilingualism, that we intend, with the proposal of this book, to register permanently and in longer texts the works presented, because of the formative potential they may have.

It is important to draw readers’ attention to the fact that the term bilingualism, which appears in the title of both the Symposium and the present book, is not the only one that appears in the literature. There are expressions used in parallel, such as foreign language acquisition/development, acquisition/development of non-native languages and multilingualism, and not all of them can be considered synonyms. Such heterogeneity has to do with theoretical and methodological differences in the definition of the phenomenon. This diversity of concepts and terminologies appears in the contributions made by the researchers invited to the symposium and to produce chapters for this book. In the chapter by van Heuven, for example, whose primary focus is on phonetics, the author uses the expression foreign language acquisition. Other times, the term is simply used as bilingualism in texts with a psycholinguistic bias (as in the chapters by Ricardo Augusto de Sousa and Amanda Post da Silveira), since Psycholinguistics typically defines the degree of bilingualism within the reality of each investigated community. The organizers of the event and the book chose to use the term bilingualism, aware of the theoretical and methodological multiplicity of the area and also aware that the discussions and contributions made by the invited researchers can illuminate and enrich the debate about the term, although neither the authors nor the organizers intended to unify the nomenclature of the field.

Another particular and original point of this book is that it brings contributions in English and Spanish, which have been translated into Portuguese, to enable the dissemination of the works to the Brazilian student and academic public, even to those who do not have the knowledge of reading in these foreign languages. The translations of the texts in English and Spanish into Portuguese are presented immediately after the original texts in the foreign languages.

One of the purposes of both the International Symposium on Prosody and Bilingualism and this book, which bears the same name, is to bring together researchers from the international and Brazilian communities interested in the intersection between studies on bilingualism and prosody and the interaction between these two. Some of the contributions primarily deal with one or the other of the two topics and some more directly explore the relationship between the two, a fact that can be explained by the fact that this joint study is still a relative novelty on the world stage. A second purpose of the event and the book was to encompass different perspectives and themes within the scope of this relationship: experimental research, socio-interactional research and the review of models of theoretical analysis. Given that experimentation in linguistics is still a form of research in consolidation in the Brazilian linguistics environment, the event also made room for methodological contributions, with the objective of adding to the training of the attending public, which consisted mostly of undergraduate and graduate students .

Regarding the contributions of this book, the chapters were grouped into three parts: in the first, the chapters that dealt primarily with the theme of prosody were gathered; in the second, contributions related to the theme of bilingualism were gathered; the third part presents the chapters that develop the two topics in equal measure, prosody and bilingualism.

The first part opens with the chapter by Joaquim Llisterri (Universidade Autónoma de Barcelona), which presents a very comprehensive review of the literature on strategies and research techniques adopted to study the production of acoustic parameters of lexical stress. It brings considerations about relevant aspects of the experimentation activity, such as experimental design, definition of the composition and number of participants in the sample, which makes this chapter a true guide for those interested in investigating lexical stress with methodological excellence.

The second chapter, authored by Björn Köhnlein (Ohio State University), presents an analysis of the opposition between two lexical tones in Franconian, a multi-regional language spoken in parts of Germany, Belgium and the Netherlands. The text describes the distributional facts of tones, considerations about their melodic and durational implementation and some morphological functions they may perform. The chapter ends with a phonological analysis that proposes a phonological representation of Franconian lexical tones based on autosegmental principles.

Then, the section on bilingualism begins with the chapter by Ricardo Augusto Souza (Federal University of Minas Gerais) making an interesting discussion on bilingualism, debating notions of

linguistic transfer and representational deficit, often attributed to the processing and representation of language by bilinguals. It points out more adequate ways of treating the object of investigation of the phenomenon of bilingualism through the bias of psycholinguistics. It is a very didactic and necessary literature review for the sequel to this book.

At the end of this section, the fourth chapter, authored by Ad Backus, provides a very broad and complete introduction on the subject of linguistic contact, illustrating its theoretical discussion mainly with examples from his several years of research on heritage languages in the Netherlands. It proposes in the chapter that the investigation of the phenomenon of linguistic contact should be carried out not only from a sociolinguistic or psycholinguistic perspective, but from both.

In the last section, on prosody and bilingualism, the fifth chapter, produced by Vincent van Heuven (Emeritus Leiden Center for Linguistics), presents what characterizes the prosodic level as opposed to the segmental level and moves towards a discussion that brings conceptualizations, explanations and very didactic examples on the role of segmental level and prosody in monolingual and bilingual perception, intelligibility and understanding, presenting examples of iconic international works in the study of prosody and promoting an interesting debate about the importance (or not) of this for intelligibility and understanding speech in general.

In the sixth chapter, Amanda Post da Silveira (Federal University of Jataí) discusses important connectionist models of speech production, making comparisons of these models regarding the implementation of word stress, since most of these models only accounted for monosyllabic words. In addition to the important and didactic review of these production models, a critique is made of the absence or little relevance given to the attribution of word stress in L1 and, even less frequently, in bilingual processing.

Finally, the seventh chapter, written by Cristiane Conceição Silva (Federal University of Santa Catarina) and Pablo Arantes (Federal University of São Carlos), presents a quantitative analysis of the intonation contours of Brazilian bilingual speakers of Spanish as L2, which shows evidence of the occurrence of learning, in the case of contours in L2, and also of linguistic friction, in the case of contours of L1 produced by bilinguals.

We close this introduction with indispensable acknowledgments to the editorial board, whose observations and comments have contributed to the improvement of both this introduction and the organization of this book. Such improvements are largely due to the excellent suggestions of Professor Ubiratã Kirchhoff Alves, to whom the organizers dedicate special thanks.



Prosody

Chapter 1

**Algunos aspectos
metodológicos en el
estudio del acento
léxico**

Joaquim Llisterri

El estudio experimental del acento léxico

El acento léxico (en la terminología anglosajona *lexical stress*, *word stress* o, en algunos casos, simplemente *stress*) suele caracterizarse como una prominencia mediante la que se crea un contraste entre una unidad acentuada y las unidades inacentuadas adyacentes (CRYSTAL, 1980; TRASK, 1996). Se tiende a considerar que la unidad sobre la que recae el acento es la sílaba, tal como se pone de manifiesto, por ejemplo, en el hecho de que en el Alfabeto Fonético Internacional la marca que indica la presencia de un acento no se sitúa sobre la vocal, sino al inicio de la sílaba tónica (INTERNATIONAL PHONETIC ASSOCIATION, 1999, p. 15, 22).

Sin entrar en detalles sobre la definición de las unidades prosódicas, puede decirse que la palabra es el dominio en el que opera el acento léxico, por lo que en las lenguas que tradicionalmente se conocen como ‘de acento libre’ (GARDE, 1968, p. 5-6) la posición del acento léxico en la palabra en algunos casos da lugar a diferencias en el significado, de modo que el acento adquiere un valor contrastivo, tal como se muestra en los ejemplos de (1a) para el español y de (1b) para el portugués.

(1) a. *límite* [ˈlimite] (sustantivo) / *limite* [liˈmite] (1.^a y 3.^a persona del singular del presente de subjuntivo del verbo *limitar*) / *limité* [limiˈte] (1.^a persona del singular del pretérito indefinido del verbo *limitar*) (MARTÍNEZ CELDRÁN et al., 2003, p. 257)

b. *sábia* [ˈsabiɐ] (adjetivo) / *sabía* [saˈbiɐ] (3.^a persona del singular del pretérito imperfecto de indicativo del verbo *saber*) / *sabiá* [sabiˈa] (sustantivo) (Barbosa y Albano, 2004, p. 230)

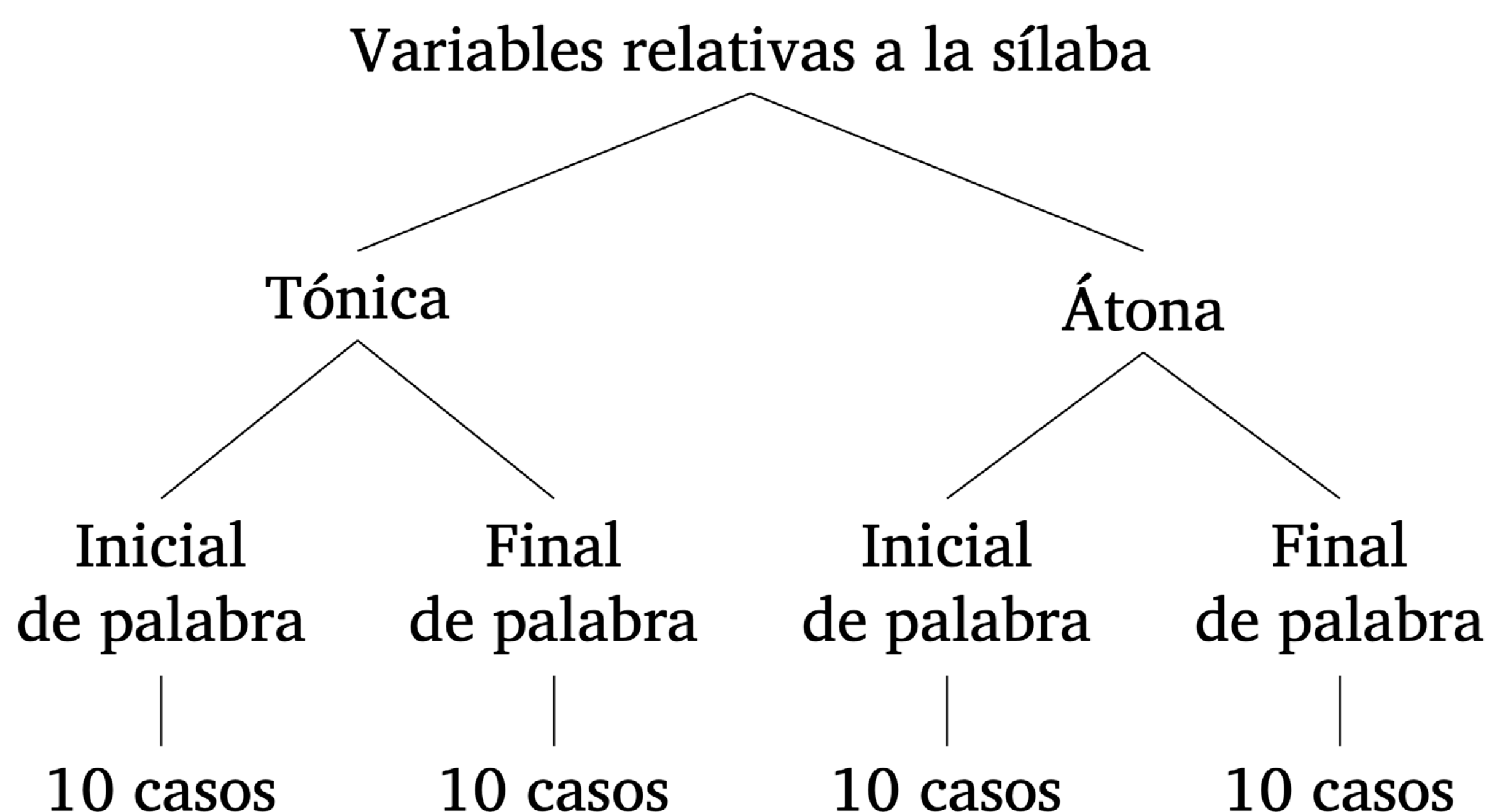
Buena parte de los estudios fonéticos experimentales sobre el acento léxico se han ocupado de establecer sus correlatos acústicos o perceptivos (VAN HEUVEN, 2018; VAN HEUVEN; TURK, 2020) y, en menor medida, articulatorios. Por ‘correlatos’ se entiende aquellas propiedades de naturaleza fonética que se relacionan con la producción por parte del hablante o con la percepción por parte del oyente de la prominencia que realza la sílaba tónica frente a las átonas. Llevar a cabo una investigación sobre estos correlatos requiere tomar una serie de decisiones, algunas de las cuales se examinan a continuación. El lector interesado en profundizar en los aspectos metodológicos encontrará información relevante en el capítulo 5 («Experimentação em fonética acústica») de Barbosa y Madureira (2015), así como en los apartados dedicados al diseño de los experimentos en las numerosas publicaciones que han abordado el estudio del acento léxico.

Para analizar los correlatos acústicos del acento léxico es preciso plantearse cuestiones relativas a los hablantes, al corpus sobre el que se va a trabajar, a los procedimientos y las herramientas empleadas para la adquisición y el análisis de los datos, y al tratamiento que se dará a los mismos. En lo que se refiere a los hablantes, resulta especialmente relevante su número y su perfil lingüístico, además de factores de interés sociolingüístico como el género, la edad o el nivel sociocultural.

Constituir el corpus de trabajo implica tomar decisiones sobre los estilos de habla que se considerarán y sobre el método más adecuado para obtener muestras procedentes de diferentes estilos, así como sobre las variables que se tendrán en cuenta en el diseño del corpus y el número de casos del que se dispondrá para el análisis.

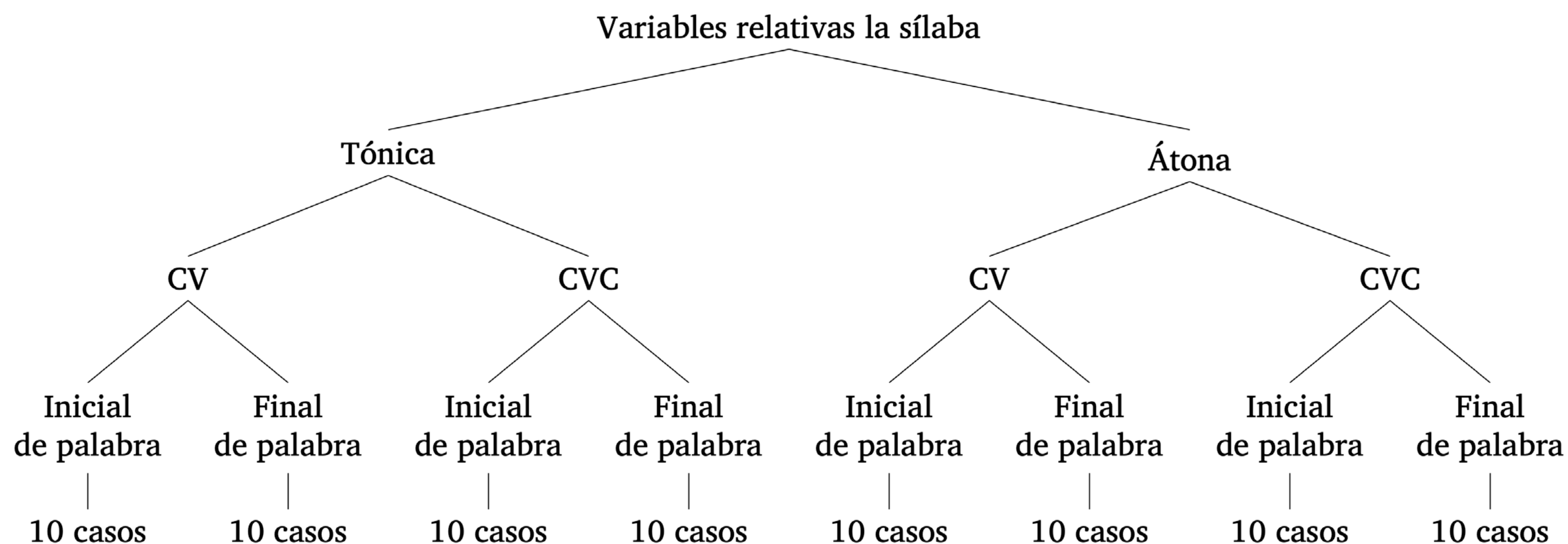
Cabe precisar que todos estos aspectos, que se explican más detalladamente en el apartado 2, se encuentran, en mayor o menor medida, relacionados entre sí. Por ejemplo, en un experimento en el que se investigue el efecto en los correlatos acústicos del acento de la posición de la sílaba dentro de la palabra y se decida emplear únicamente palabras bisílabas, deben incluirse sílabas tónicas y sílabas átonas tanto en posición inicial como en posición final de palabra, como se muestra en la figura 1. Si se quiere contar con 10 casos de cada posible combinación de las dos variables (tonicidad y posición en la palabra), el corpus constará de 40 ítems.

Figura 1: Diseño experimental de un estudio con dos variables: tonicidad de la sílaba y posición de la sílaba en la palabra



Al añadir una nueva variable como, por ejemplo, la estructura silábica, limitada a sílabas del tipo consonante-vocal (CV) y consonante-vocal-consonante (CVC), el número de casos por analizar se eleva a 80, como puede comprobarse en la figura 2.

Figura 2: Diseño experimental de un estudio con tres variables: tonicidad de la sílaba, estructura silábica y posición de la sílaba en la palabra



Si, además, el mismo experimento se llevara a cabo comparando dos estilos de habla, el número de ítems llegaría a 160 (80 para cada estilo). Suponiendo que, en lugar de 10 casos de cada combinación posible de las variables, el investigador considerara necesario contar con 30 casos, un diseño con dos variables implicaría analizar 120 ítems, uno con tres variables supondría un corpus de 240 ítems y, si se introduce una cuarta variable, la cifra se eleva a 480 ítems para cada locutor. Así, en un estudio en el que participen 10 hablantes se analizan 4 800 realizaciones.

Con este ejemplo ficticio, y tal vez excesivamente simple, se pretende poner de relieve que a medida que se incrementa el número de variables contempladas en un diseño experimental, aumenta también el tamaño del corpus y, con ello, el volumen de datos que deberá tratarse. Puede argumentarse que, en la actualidad, esto no supone un problema dada la posibilidad de llevar a cabo los análisis mediante procedimientos como la segmentación y el etiquetado automáticos y usando programas que permiten obtener, también de modo automático, los valores de los parámetros acústicos. No obstante, conviene no olvidar que, por una parte, no se dispone de herramientas de segmentación y etiquetado para todas las lenguas y que, por otra, los procesos automáticos suelen requerir una revisión manual. En el trabajo de Garrido (2018) se discuten con más detalle las ventajas y los inconvenientes del tratamiento automático de corpus para estudios prosódicos.

El estudio de los correlatos acústicos del acento léxico

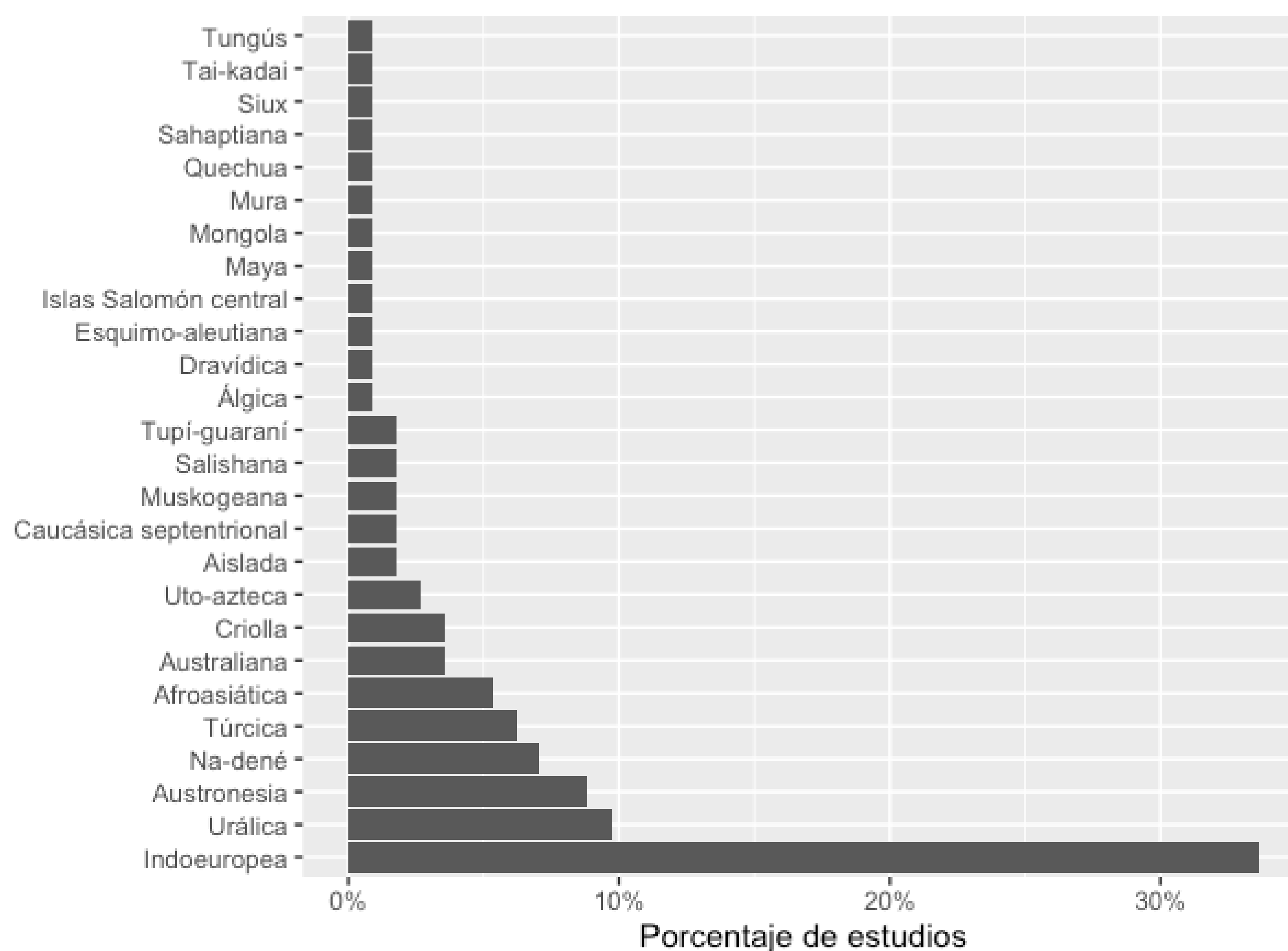
Como se ha explicado, diseñar un experimento para establecer los correlatos acústicos del acento léxico supone tomar una serie de decisiones metodológicas. A continuación, se exponen algunas de las que se relacionan con los hablantes (apartado 2.1) y con el corpus empleado (apartado 2.2).

Sin embargo, antes de entrar en cuestiones específicas, conviene tal vez empezar por una visión general de los principales aspectos metodológicos que inciden en el estudio experimental del acento

léxico desde la perspectiva del análisis acústico. Para ello, resulta sumamente útil el trabajo de Roettger y Gordon (2017) en el que se examinan 86 publicaciones sobre los correlatos acústicos del acento léxico en 76 lenguas pertenecientes a 26 familias lingüísticas. El artículo se complementa con una base de datos en línea (GORDON; ROETTGER, 2018) que contiene la información extraída de las publicaciones revisadas y con un conjunto de *scripts* en R (R Core Team, 2021) para reproducir los cálculos ejecutados por los autores.

Una primera pregunta que cabe formularse —al menos, desde la perspectiva de un lingüista— es en qué lenguas se basan los conocimientos con los que se cuenta en la actualidad sobre los parámetros acústicos que señalan la presencia del acento léxico. En la figura 3, elaborada a partir de los datos de Gordon y Roettger (2018), puede observarse que, como cabría esperar, aproximadamente un tercio de los estudios incluidos en la revisión de Roettger y Gordon (2017) se han realizado sobre lenguas que pertenecen a la familia indoeuropea.

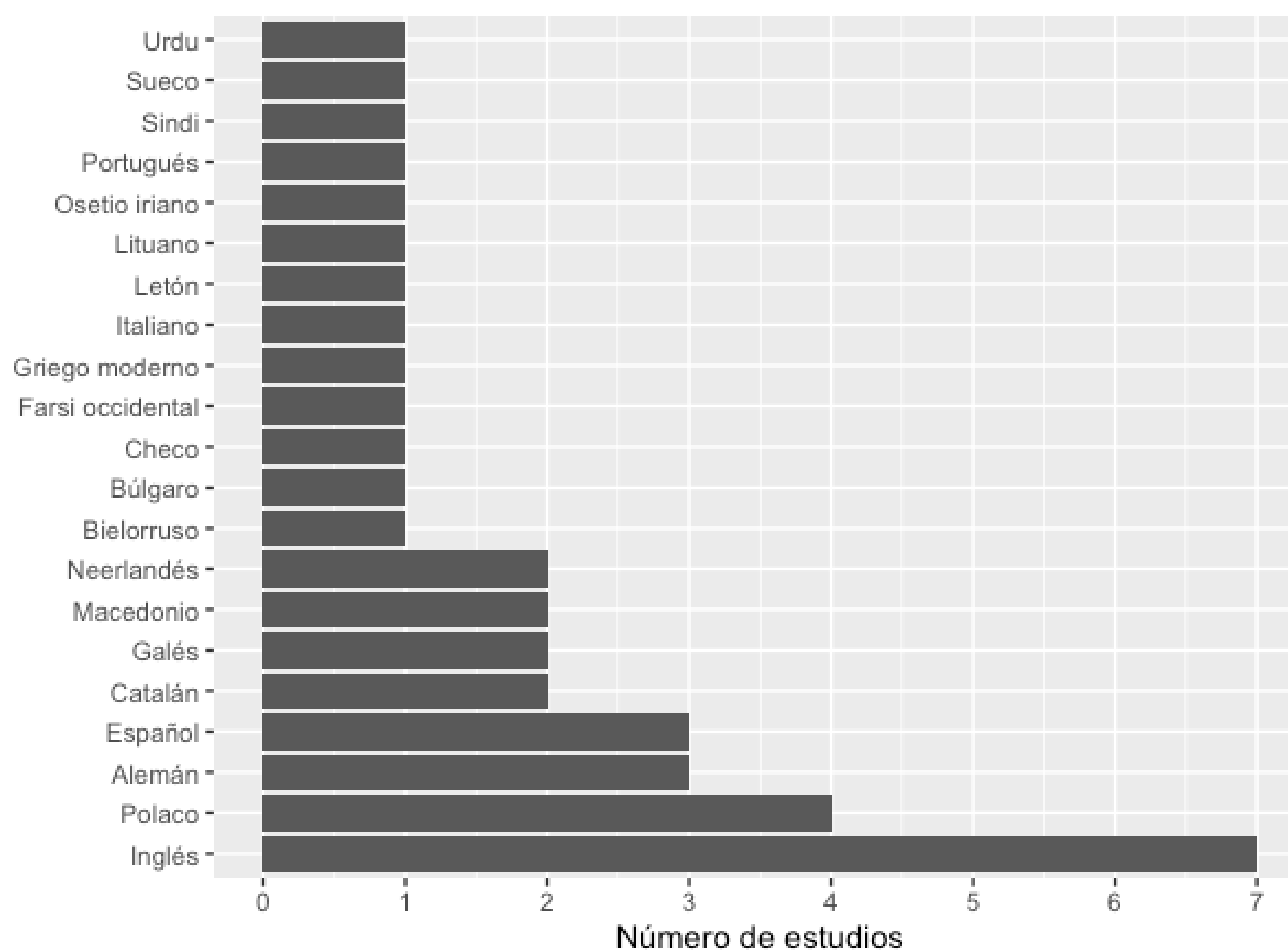
Figura 3: Familias lingüísticas representadas en los estudios sobre los correlatos acústicos del acento léxico analizados por Roettger y Gordon (2017)



Nota. Gráfico elaborado a partir de la información recogida en la base de datos del proyecto *Studies on acoustic correlates of word stress*, de M. K. Gordon y T. B. Roettger, 18 de abril de 2018, OSF (<https://doi.org/10.17605/OSF.IO/9R2CD>).

Tampoco resulta sorprendente que en los trabajos considerados por Roettger y Gordon (2017) el inglés sea la lengua a la que se dedica un mayor número de estudios, como se puede comprobar en la figura 4.

Figura 4: Lenguas indoeuropeas representadas en los estudios sobre los correlatos acústicos del acento léxico analizados por Roettger y Gordon (2017)



Nota. Gráfico elaborado a partir de la información recogida en la base de datos del proyecto *Studies on acoustic correlates of word stress*, de M. K. Gordon y T. B. Roettger, 18 de abril de 2018, OSF (<https://doi.org/10.17605/OSF.IO/9R2CD>).

Para el español, Roettger y Gordon (2017) seleccionaron tres publicaciones escritas en inglés, dos de las cuales se ocupan también de otras lenguas distintas del español (ORTEGA-LLEBARIA, 2006; ORTEGA-LLEBARIA; PRIETO, 2011; VOGEL *et al.*, 2016), tal como sucede en la única publicación elegida sobre el portugués, una contribución en inglés en la que el portugués se compara con el sueco (BARBOSA *et al.*, 2013a).

A la vista de los datos, puede pensarse que al analizar los métodos empleados en las investigaciones sobre el acento léxico (ROETTGER; GORDON, 2017) y al valorar los resultados obtenidos (GORDON; ROETTGER, 2017), existe un cierto sesgo a favor de los trabajos sobre el inglés y de los publicados en este idioma cuando se trata de aportaciones relativas a otras lenguas.

Hablantes

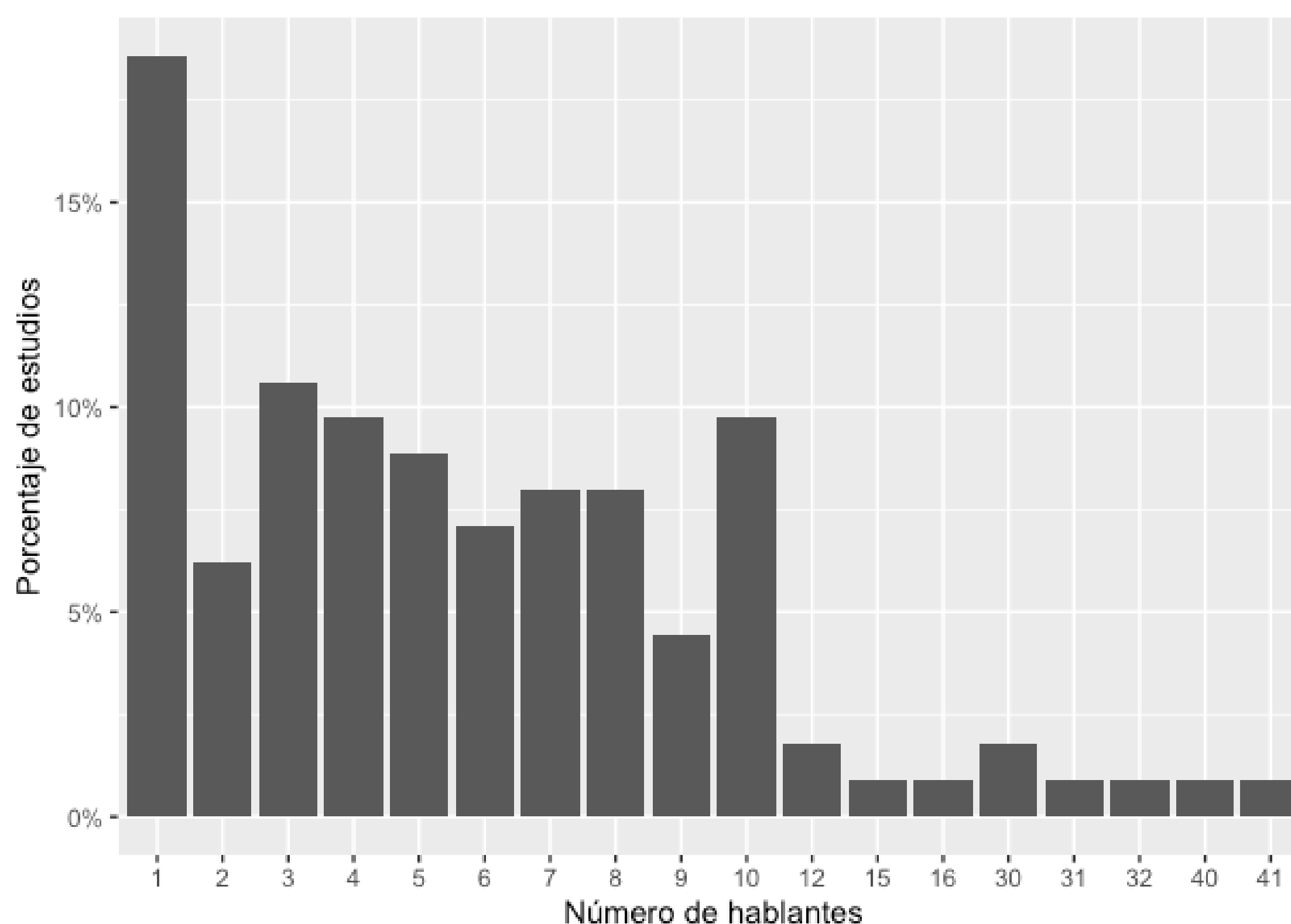
La selección de los hablantes que proporcionarán los datos para un experimento no debe considerarse en modo alguno una tarea trivial. Aunque pueda verse condicionada por factores de naturaleza práctica como la disponibilidad para formar parte del estudio, la distancia geográfica o la financiación con la que se cuenta si la participación es remunerada, cabe tener en cuenta también

factores sobre los que el investigador tiene un cierto grado de control como el número de hablantes, que se trata en el apartado 2.1.1 y, muy especialmente, su perfil lingüístico, abordado en el apartado 2.1.2.

Número de hablantes

Entre los aspectos metodológicos examinados por Roettger y Gordon (2017) se cuenta el número de hablantes en los que se basan los datos recogidos en las publicaciones incluidas en su revisión. Como señalan los autores, y como puede también apreciarse en la figura 5, “The majority of studies employ data from between one and ten speakers but vary widely within this range. On the other hand, the mode (20 studies) is a single speaker” (ROETTGER; GORDON, 2017, § 3.2).

Figura 5: Número de hablantes en los estudios sobre los correlatos acústicos del acento léxico analizados por Roettger y Gordon (2017)



Nota. Gráfico elaborado a partir de la información recogida en la base de datos del proyecto *Studies on acoustic correlates of word stress*, de M. K. Gordon y T. B. Roettger, 18 de abril de 2018, OSF (<https://doi.org/10.17605/OSF.IO/9R2CD>).

A modo de ejemplo, el trabajo sobre el portugués seleccionado por Roettger y Gordon (2017) se basa en 10 hablantes nativos de portugués brasileño y en otros tantos hablantes nativos de sueco (BARBOSA et al., 2013a), mientras que, en el caso de las tres publicaciones sobre el español elegidas por estos mismos autores, en una de ellas se incluyen cinco hablantes de español peninsular (ORTEGA-LLEBARIA, 2006), en otra con 10 hablantes de español peninsular y 10 de catalán (ORTEGA-LLEBARIA; PRIETO, 2011) y, en la tercera, se analizan datos procedentes de nueve hablantes de español peninsular, 10 de griego, 8 de húngaro y 9 de turco (VOGEL et al., 2016).

A la hora de tomar decisiones sobre el número de locutores, es pertinente recordar la recomendación de Ladefoged (2003) para los trabajos de campo: «Ideally, you want about half a dozen speakers of each sex. If you can eventually find 12 or even 20 members of each sex, so much the better» (p. 14). Como puede constatarse en la figura 5, pocas publicaciones superan este mínimo de 12 hablantes propuesto por Ladefoged.

Perfil lingüístico de los hablantes

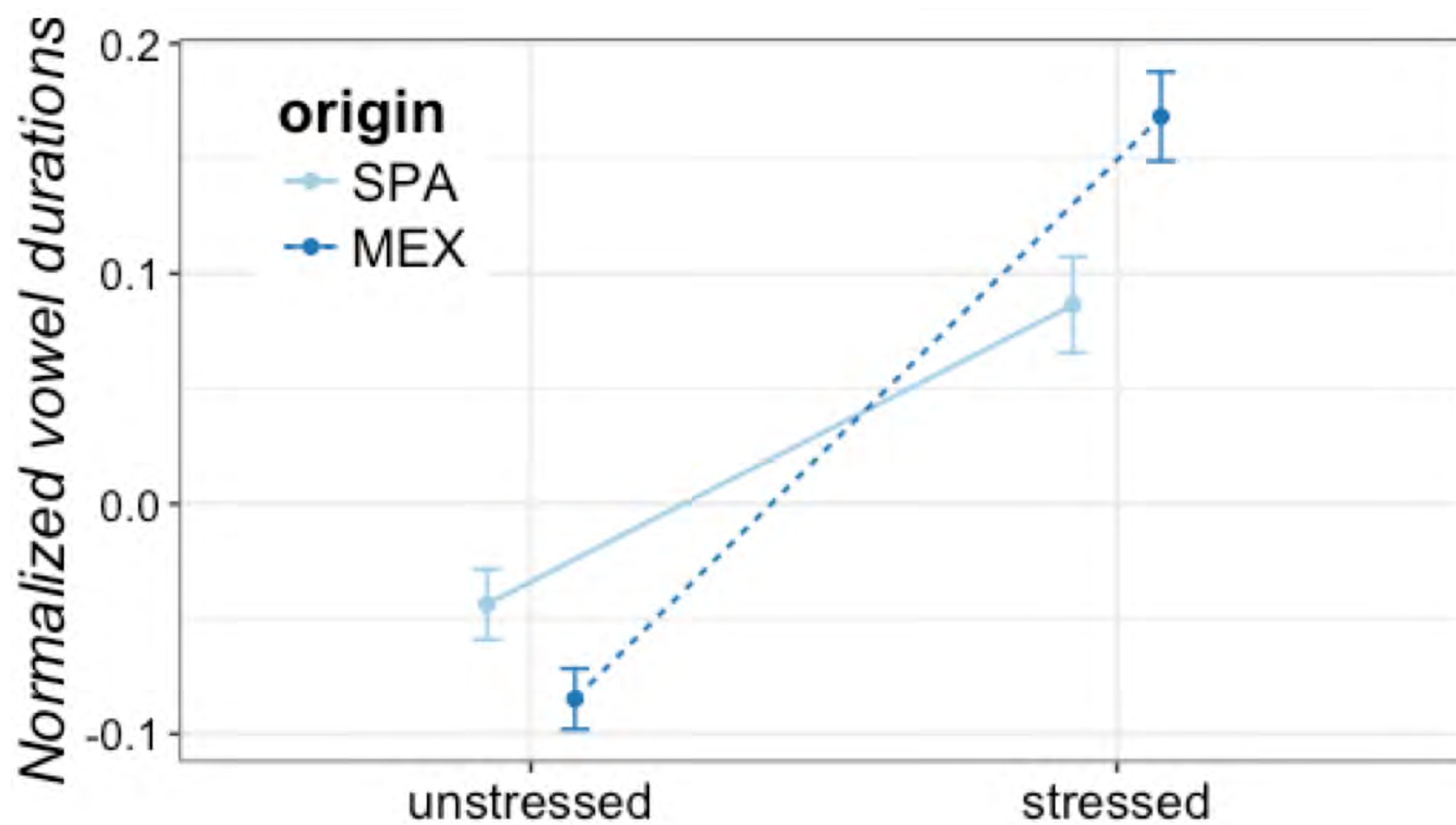
En los estudios fonéticos que tienen como objetivo describir los correlatos acústicos del acento léxico, se recurre, como es lógico, a hablantes monolingües de la lengua por la que se interesa el investigador. Así, por ejemplo, en el trabajo de Vogel et al. (2016), incluido entre los seleccionados por Roettger y Gordon (2017), se encuentra la siguiente descripción de los participantes en el experimento:

All of the speakers who participated in our study were university students between the ages of 18 and 30, with no reported language / hearing issues. Moreover, they were monolingual to the extent possible, though it is understood that university students will have studied at least one other language as part of their education. For each language, ten speakers from similar standard dialect backgrounds were recorded at their local universities: Hungarian – Debrecen, Spanish – Madrid, Turkish – Istanbul, and Greek – Athens (VOGEL et al., 2016, p. 132).

En esta cita se hace patente, en primer lugar, que en función de la edad y del nivel de educación, es muy posible que los hablantes elegidos para un determinado estudio posean conocimientos de otras lenguas. Decidir en qué medida una persona capaz de usar más de un idioma puede describirse como bilingüe o plurilingüe depende, en realidad, de la definición de bilingüismo o de plurilingüismo que adopte el investigador, un asunto complejo sobre el que se volverá más adelante. Un segundo aspecto que merece la pena destacar es la variedad geográfica empleada por los hablantes, pues algunos de los datos disponibles hacen pensar que la manifestación acústica del acento léxico no tiene que ser necesariamente homogénea en todo un dominio lingüístico.

En la figura 6 se reproducen los resultados obtenidos por Santiago y Mairano (2018) sobre el efecto del acento léxico en la duración de las vocales en un grupo de 12 hablantes de español estudiantes universitarios en la Ciudad de México y en otro de 10 hispanohablantes que cursaban estudios universitarios en Madrid. Como puede observarse, la diferencia de duración entre vocales átonas y tónicas es más marcada en los hablantes de español de México que en los locutores peninsulares; en concreto, «Spanish speakers reduce unstressed vowels by 7 % with respect to their stressed counterparts, whereas Mexican speakers reduce them by 13 %» (SANTIAGO; MAIRANO, 2018, p. 455).

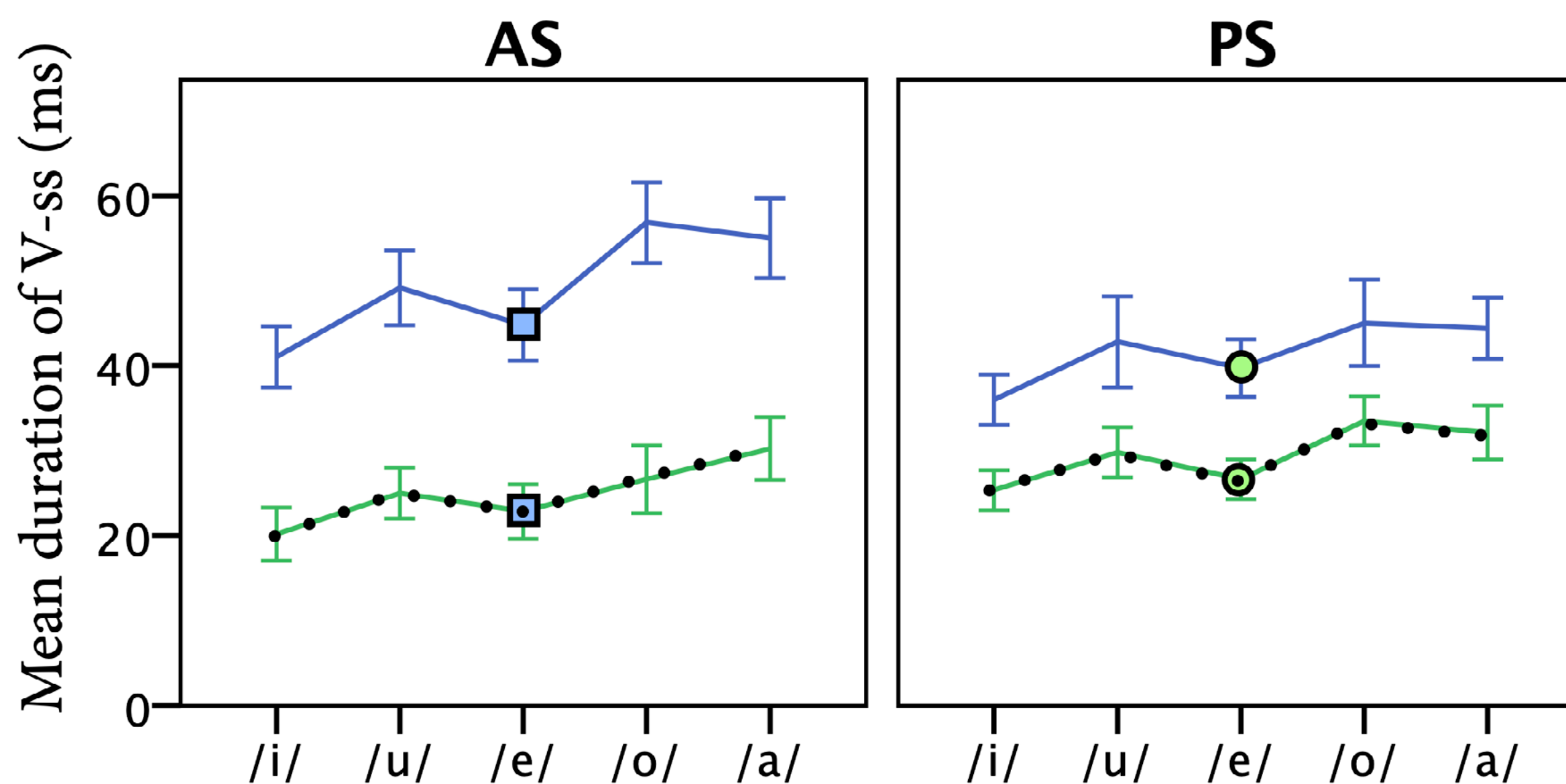
Figura 6: Diferencias de duración entre vocales tónicas y átonas en español de México y en español peninsular



Nota. SPA = Español peninsular; MEX = Español de México. Figura extraída de «The role of lexical stress on vowel duration and vowel space in two varieties of Spanish», de F. Santiago y P. Mairano, 2018, en K. Klessa, J. Bachan, A. Wagner, M. Karpiński y D. Śledziński (Eds.), *9th International Conference on Speech Prosody 2018*, p. 455, ISCA Archive (<https://doi.org/10.21437/SpeechProsody.2018-92>).

Un fenómeno similar se encuentra al comparar la duración de las vocales en función del acento léxico en cinco hablantes limeños de español y cinco hispanohablantes nacidos o educados en Buenos Aires, estudiados por Kim (2011). En la figura 7 se aprecia que la diferencia de duración entre vocales tónicas y átonas es menor en la variedad peruana que en la argentina.

Figura 7: Diferencias de duración entre vocales tónicas y átonas en español de Argentina y en español de Perú



Nota. V-ss = Parte estable de la vocal; AS = Hablantes de español de Argentina; PS = Hablantes de español de Perú. La línea continua corresponde a las vocales tónicas y la línea de puntos, a las átonas. Figura extraída de *The phonetics of stress manifestation: Segmental variation, syllable constituency and rhythm*, de M. Kim, 2011, Tesis doctoral, Stony Brook University, p. 121, SUNY Digital Repository (<http://hdl.handle.net/1951/56031>).

Estos dos ejemplos sugieren que la variante geográfica constituye una variable pertinente en lo que se refiere a las características de los hablantes monolingües, por lo que no tomarla en consideración puede llegar a afectar la homogeneidad del grupo de locutores analizado en un experimento.

Para estudiar los correlatos acústicos del acento léxico se ha recurrido también a locutores bilingües. Como se ha propuesto, este término puede abarcar perfiles lingüísticos muy diversos; por ejemplo, Hoffmann (1991, p. 16-17) recoge hasta 15 casos de uso de dos lenguas en las que podría hablarse de bilingüismo según la manera en que se defina el término. También a modo de ejemplo, en la descripción de los participantes en el trabajo de Ortega-Llebaria y Prieto (2011, p. 80), uno de los que Roettger y Gordon (2017) incluyeron en su revisión, se lee lo siguiente:

Ten native speakers of Castilian Spanish (7 females and 3 males) and ten native speakers of Central Catalan (10 females) participated in the experiment. Their ages ranged from 20 to 32 years old. All subjects had earned university degrees and spoke an educated variety of their native tongue which they used regularly with their parents and siblings. The ten Catalan subjects had learnt Spanish as a second language in school at age five or later while only five of the Spanish subjects also knew Catalan as a second language, which they learnt in secondary school, that is, at age ten or later.

De los 15 hablantes que pueden definirse como bilingües, 10 han aprendido la segunda lengua a partir de los cinco años y cinco empezaron a aprenderla a partir de los 10, lo que introduce una cierta asimetría entre los dos grupos. En estas situaciones resulta útil recurrir al concepto de ‘dominancia lingüística’ (TREFFERS-DALLER, 2019), que, con la información disponible, permitiría caracterizar a 10 hablantes como bilingües catalán-español con el catalán como lengua dominante y a cinco hablantes como bilingües español-catalán cuya lengua dominante es el español.

Cuando se trabaja con bilingües conviene tener en cuenta, además de la dominancia de los hablantes, las diferencias entre las lenguas que pueden llegar a incidir en el fenómeno estudiado. Así, si se analiza el efecto del acento léxico en los valores de frecuencia de los formantes de las vocales, el hecho de que en catalán central se dé un proceso de reducción vocálica, ausente en español, podría llegar a afectar las realizaciones átonas en español de un bilingüe con una fuerte dominancia catalana (BADIA; MARGARIT, 1975, p. 150), ya que en catalán central /e/, /ɛ/ y /a/ se realizan como [ə] en posición átona y /o/ y /ɔ/ se realizan como [u] en dicha posición (JULIÀ, 2005, p. 160).

En conjunto, el término ‘hablante bilingüe’ engloba un amplio abanico de perfiles e historias lingüísticas, de modo que la dominancia de los locutores, junto a otros factores, es una variable que cabe considerar si se pretende contar con un grupo relativamente homogéneo de participantes en un experimento. Por esta razón se han desarrollado diversas herramientas para caracterizar con un cierto detalle el perfil de un hablante bilingüe. Una de ellas es el *Bilingual Language Profile* (BLP), un cuestionario de autoevaluación consistente en 19 preguntas que puede responderse en papel o

en línea —con una puntuación automática en este último caso— y que cuenta con versiones para 15 pares de lenguas (BIRDSONG et al., 2012; GERTKEN et al., 2014).

Una tercera categoría de hablantes con los que se ha contado para el análisis de los correlatos acústicos del acento léxico corresponde a la de las personas que aprenden o adquieren una segunda lengua o una lengua extranjera (L2/LE). Por ejemplo, en el trabajo de Schwab (2014) sobre la producción del acento léxico en francés hablado por estudiantes cuya primera lengua es el español, se describe a los participantes del siguiente modo:

[...] eight native Spanish advanced learners of French (B2–C1) (3 males and 5 females), living in Geneva, with at least two years spent in a French-speaking country. Spanish speakers were from the Iberian Peninsula (Aragon, La Rioja, Galicia, La Rioja, Extremadura, and Andalusia) and French was their second foreign language after English (SCHWAB, 2014, p. 80).

Como puede comprobarse, aunque frecuentemente se haga referencia a una L2, en investigaciones sobre estudiantes que aprenden una lengua distinta de la primera y esa lengua no es el inglés, es muy probable que la L2 sea, en realidad, una tercera lengua (L3).

Resulta también interesante observar las características de los participantes en un análisis de la producción del acento en francés y en español por parte de estudiantes anglófonos canadienses que proporcionan Colantoni et al. (2014) y que se reproduce en la figura 8.

Figura 8: Características de los participantes en un estudio sobre la adquisición del acento en francés y en español por parte de estudiantes anglófonos canadienses

	French		Spanish	
	Learners	Native Speakers	Learners	Native Speakers
n	9	4	8	3
Gender (M;F)	2;7	1;3	4;4	2;1
Mean (range) Age at Acquisition onset	6 (5-12)	-	17 (11-26)	-
Mean (range) Immersion (months)	23 (19-40)	-	25 (18-29)	-
Immersion (Months)	19 (0-150)	-	12 (0-66)	-

Nota. Tabla extraída de «Learning to realize prosodic prominence in French and Spanish», de L. Colantoni, O. Marasco, J. Steele y S. Sunara, 2014, en R. T. Miller, K. I. Martin, C. M. Eddington, A. Henery, N. Marcos, A. M. Tseng, A. Tuninetti y D. Walter (Eds.), *Selected Proceedings of the 2012 Second Language Research Forum: Building bridges between disciplines*, p. 18, Cascadilla Proceedings Project (<http://www.lingref.com/cpp/slrf/2012/abstract3082.html>).

La especificidad del sistema educativo canadiense explica las diferencias en la edad a la que los estudiantes de francés y los de español empezaron a aprender estas lenguas, así como el distinto tiempo de inmersión (Colantoni et al., 2014, p. 18), dos factores que hacen que los dos grupos de participantes en el experimento no sean exactamente idénticos.

Al igual que sucede con los hablantes bilingües, el perfil lingüístico de las personas que aprenden o adquieren una L2/LE puede resultar muy diverso y verse influenciado por diferentes factores. Entre los más relevantes destacan la edad a la que se empieza a aprender la lengua, el tiempo de estudio o de contacto con la L2/LE, el contexto en el que se ha aprendido la lengua y las variantes geográficas de la L2/LE a las que el estudiante ha estado expuesto. Por ello, suele ser conveniente realizar una evaluación del nivel de competencia lingüística —e, idealmente, de la competencia fonética— de los participantes en un experimento.

Una primera posibilidad es considerar simplemente que el nivel de conocimiento de la L2/LE viene dado por el número de años de estudio de la lengua o por la etapa educativa en la que se encuentra el estudiante; esto conlleva asumir que, por ejemplo, todos los alumnos de primer curso de un grado en español de una determinada universidad en un país que no sea hispanohablante poseen el mismo nivel de competencia lo que, evidentemente, no siempre corresponde del todo a la realidad.

Una segunda opción, como en el trabajo ya citado de Schwab (2014) y en muchos otros, consiste en determinar el grado de competencia con relación a un sistema de niveles globalmente aceptado. En el contexto europeo, es habitual emplear el Marco Común Europeo de Referencia (MCER), en el que se establecen tres etapas, cada una de ellas dividida en dos niveles: usuario básico —A1 (Acceso) y A2 (Plataforma)—, usuario independiente —B1 (Umbral) y B2 (Avanzado)— y usuario competente —C1 (Dominio operativo eficaz) y C2 (Maestría)—. Para alcanzar la certificación correspondiente a cada nivel es preciso superar una prueba estandarizada (CONSEJO DE EUROPA, 2002). Sin embargo, como acertadamente señalan Norris y Ortega (2003), este tipo de pruebas no se concibieron inicialmente para responder a las necesidades de los investigadores en el campo de la adquisición de segundas lenguas:

For example, Shohamy (1994) observed that tests intended for educational decision making, such as the Test of English as a Foreign Language (TOEFL), are frequently utilized by SLA researchers as measures of learning or acquisition, even though such tests were designed as indicators of global academic language abilities. Likewise, holistic proficiency measures, such as the ACTFL (1986) *Guidelines* and related procedures, may be used as a basis for assigning learners to instructional research conditions, even though the scores on such measures may have nothing to do with the particular L2 forms or abilities being investigated (NORRIS; ORTEGA, 2003, p. 739).

En consecuencia, puede suceder que el nivel reflejado en estas evaluaciones no coincida con el conocimiento real del componente lingüístico —el componente fónico, en el caso del acento léxico— por el que se interesa el investigador.

Una tercera vía sería la que se explica en el estudio de Colantoni et al. (2014) anteriormente mencionado: «Learners' proficiency is based on their own self-reported evaluation confirmed by the authors during data analysis» (COLANTONI et al., 2014, p. 18, n. 4). En este caso, el grado de conocimiento de la L2/LE se determina mediante un cuestionario de autoevaluación, o bien desarrollado por los autores de cada estudio, o bien recurriendo a instrumentos ya existentes. Cuando se emplea esta técnica, es el propio participante en el experimento quien proporciona la información inicial sobre cómo percibe su nivel de competencia y de uso de la L2/LE.

Entre los cuestionarios de autoevaluación de la competencia lingüística específicamente orientados a hablantes que conocen más de una lengua están el *Language History Questionnaire* (LHQ) y el *Language Experience and Proficiency Questionnaire* (LEAP-Q). El primero de ellos, descrito en su última versión en el trabajo de Li et al. (2020), consta de 24 preguntas y cuenta con traducciones a 14 lenguas; si se usa el cuestionario en línea (The Brain, Language, and Computation (BLC) Lab, s. f.), la puntuación se calcula automáticamente. El LEAP-Q, del que se dispone de cuestionarios adaptados en 26 idiomas, incluye nueve preguntas generales y siete preguntas específicas para cada una de las lenguas evaluadas (KAUSHANSKAYA et al., 2020); puede responderse en un documento en Word con la posibilidad de exportar los datos a Excel y se encuentra también en la red (Bilingualism and Psycholinguistics Research Group, s. f.).

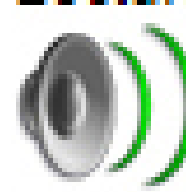

Los tres procedimientos que se acaban de describir permiten caracterizar, cada uno con sus ventajas y sus limitaciones, el conocimiento general de la L2/LE, pero, en muchas ocasiones, no todos los miembros de un grupo de participantes en un experimento con el mismo supuesto grado de competencia —por encontrarse en la misma etapa del sistema educativo, porque cuentan con el mismo certificado de nivel, o porque han respondido de forma similar a un cuestionario de autoevaluación— poseen necesariamente una competencia fónica homogénea. Por ello, en estudios fonéticos como los que atañen al acento léxico, podría ser aconsejable realizar también una valoración del grado de acento extranjero.

En este caso, la dimensión que se evalúa es la que en inglés se conoce como *accentedness*. Para Munro y Derwing (1995), dos de los autores que indudablemente han ejercido una mayor influencia en la caracterización del acento no nativo, «*Accentedness* refers to how strong the talker's foreign accent is perceived to be» (MUNRO; DERWING, 1995, p. 291). Las respuestas de los evaluadores pueden recogerse mediante una escala conocida como VAS (*Visual Analogue Scale*) en la que, como se muestra en la figura 9, el grado de acento se establece deslizando un cursor sobre una línea continua.

Figura 9: Evaluación del grado de acento extranjero mediante una escala VAS (Visual Analogue Scale)

15. Ahora escucha con atención las grabaciones siguientes y evalúa cómo de fuerte es el acento, si es que lo tiene. *

NO TIENE ACENTO TIENE MUCHISIMO ACENTO

Lectura 	<input type="range"/>
Narración 	<input type="range"/>

Nota. Figura extraída de «The contribution of prosody to foreign accent: A study of Spanish as a foreign language», de C. C. Silva y P. Barbosa, 2017, *Loquens*, 4(2), artículo e041, p. 6 (<http://dx.doi.org/10.3989/loquens.2017.041>).

También cabe utilizar escalas Likert, en las que la puntuación se divide en categorías discretas, como puede apreciarse en la figura 10, procedente del estudio de Avello (2018), en el que se recurrió a una escala de cinco puntos, aunque es frecuente también emplear valoraciones del 1 al 7 o del 1 al 9 (ISBELL, 2018, p. 89). Es interesante mencionar que Praat, un programa de dominio público con una amplia base de usuarios (BOERSMA; WEENINK, 2021), permite diseñar experimentos empleando este tipo de escalas.

Figura 10: Evaluación del grado de acento extranjero mediante una escala Likert

File Query Help
1 / 78

Rate the degree of FOREIGN-ACCENT in each sentence.
Please MAKE USE OF ALL THE NUMBERS in the scale.

1 (native)	2	3	4	5 (heavy foreign accent)
------------	---	---	---	--------------------------

[Click here to play it again](#)

Nota. Figura extraída de «Assessing learners' changes in foreign accent during Study Abroad», de P. Avello, 2018, en C. Pérez Vidal, S. López Serrano, J. Ament y D. J. Thomas-Wilhelm (Eds.), *Learning context effects: Study abroad, formal instruction and international immersion classrooms*, p. 142, Language Science Press (<https://doi.org/10.5281/zenodo.1300630>).

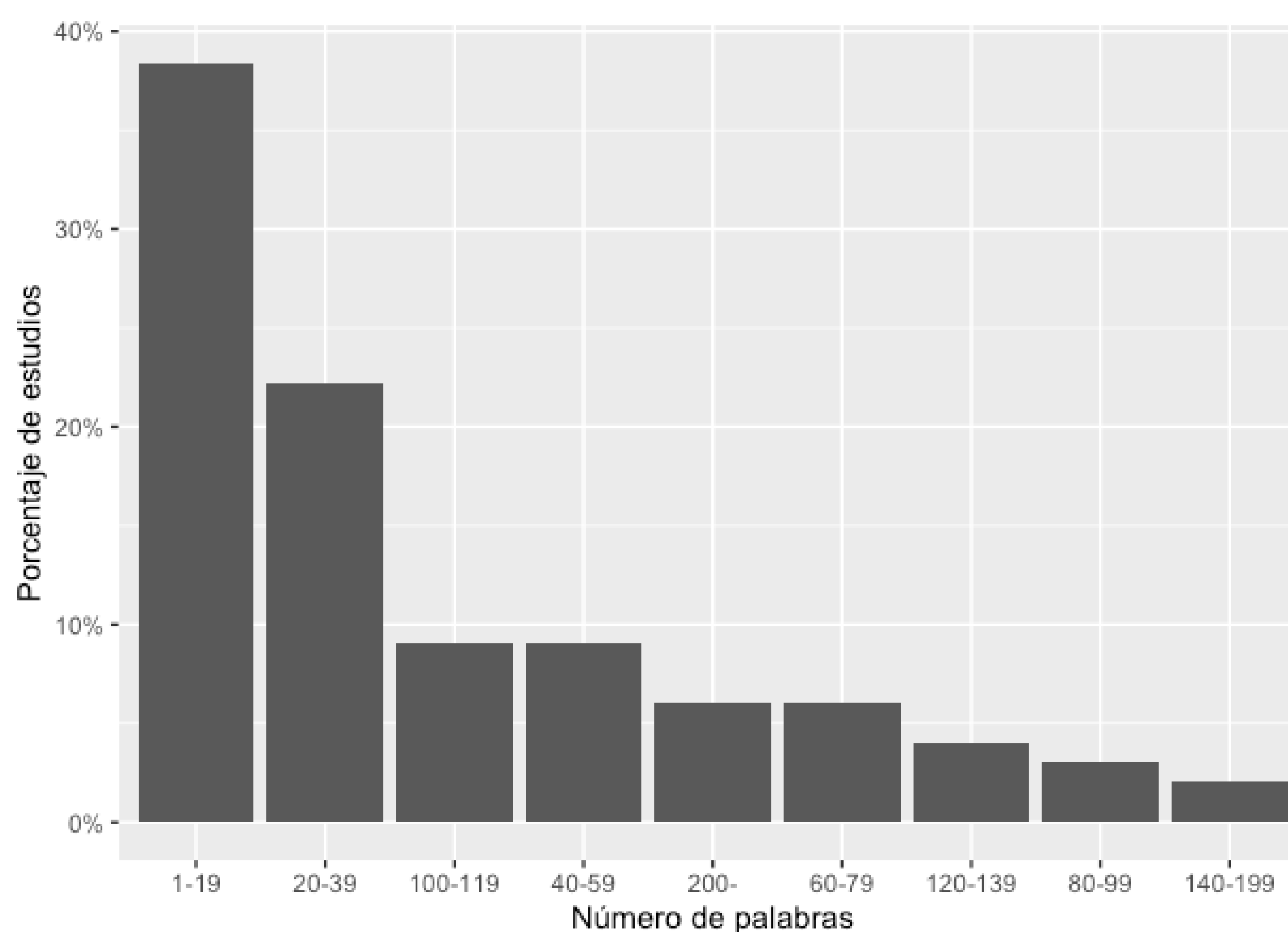
Corpus

Un segundo ámbito en el que se plantean problemas de naturaleza metodológica cuando se estudian experimentalmente los correlatos acústicos del acento léxico se relaciona con el diseño del corpus que se empleará para obtener la información necesaria. En primer lugar, se presentan algunos datos sobre el número de casos con el que se cuenta en los análisis (apartado 2.2.1) y, a continuación, se discuten algunos aspectos que atañen al estilo de habla considerado en las investigaciones (apartado 2.2.2).

Número de casos

En la revisión de las publicaciones sobre los correlatos acústicos del acento léxico llevada a cabo por Roettger y Gordon (2017), se aborda específicamente el número de casos analizados en los estudios seleccionados por los autores. Los datos sobre el número de palabras diferentes presentes en el corpus empleado en cada trabajo que se recogen en la figura 11 ponen de manifiesto que «the majority of studies in our corpus use between 1 and 40 different lexical items. Depending on how a lexical item was defined (word type or word form) and the corpus (lab speech vs. spontaneous speech), some studies look at many hundreds of words» (ROETTGER; GORDON, 2017, § 3.2).

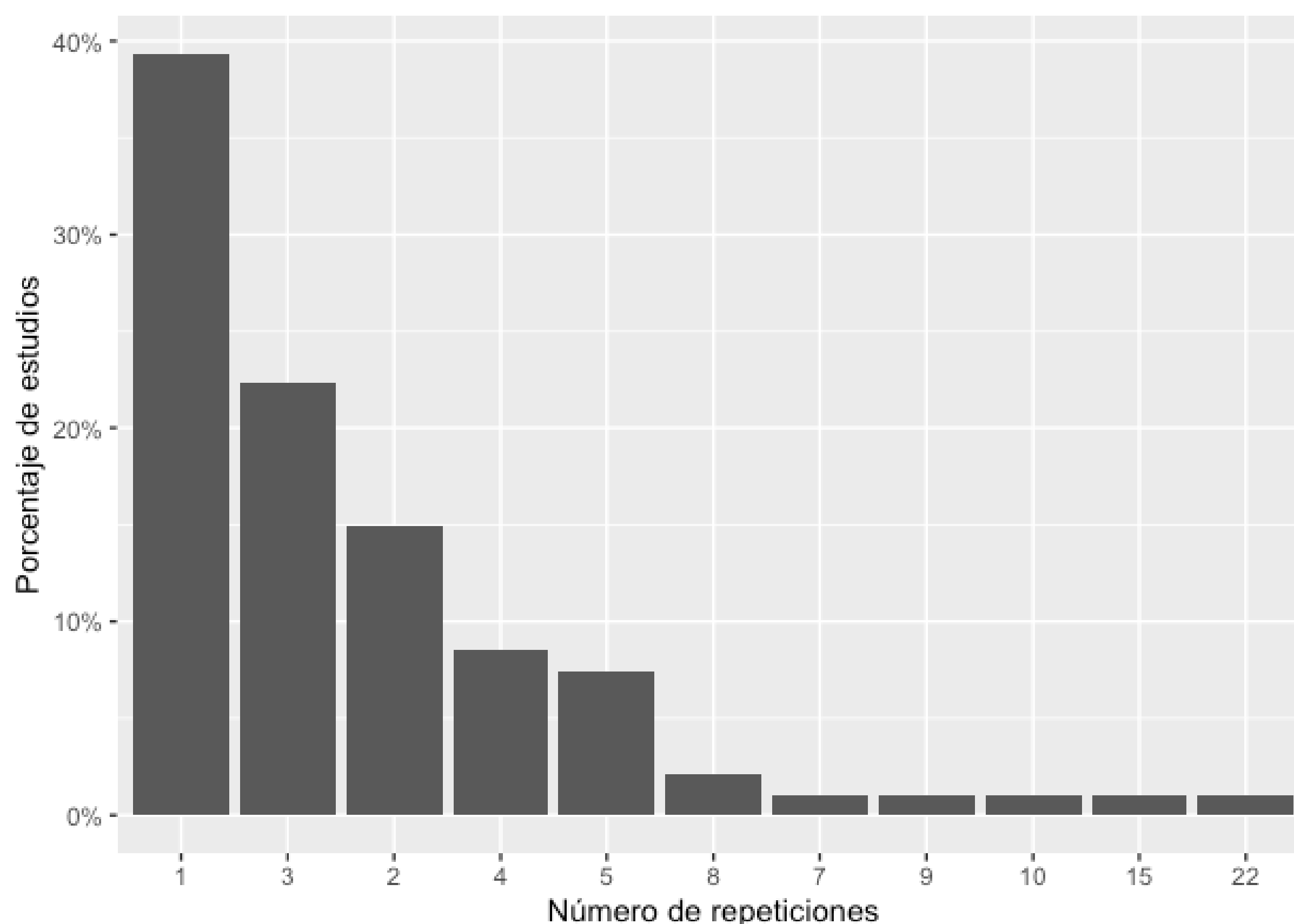
Figura 11: Número de palabras en los corpus empleados en los estudios sobre los correlatos acústicos del acento léxico analizados por Roettger y Gordon (2017)



Nota. Gráfico elaborado a partir de la información recogida en la base de datos del proyecto *Studies on acoustic correlates of word stress*, de M. K. Gordon y T. B. Roettger, 18 de abril de 2018, OSF (<https://doi.org/10.17605/OSF.IO/9R2CD>).

Aunque una manera habitual de contar con un mayor volumen de datos consiste en obtener diversas repeticiones de la misma palabra, en el análisis llevado a cabo por Roettger y Gordon (2017) se observa que «the majority of studies use between 1 and 6 repetitions with a single repetition being most common» (ROETTGER; GORDON, 2017, § 3.2), tal como se muestra en la figura 12.

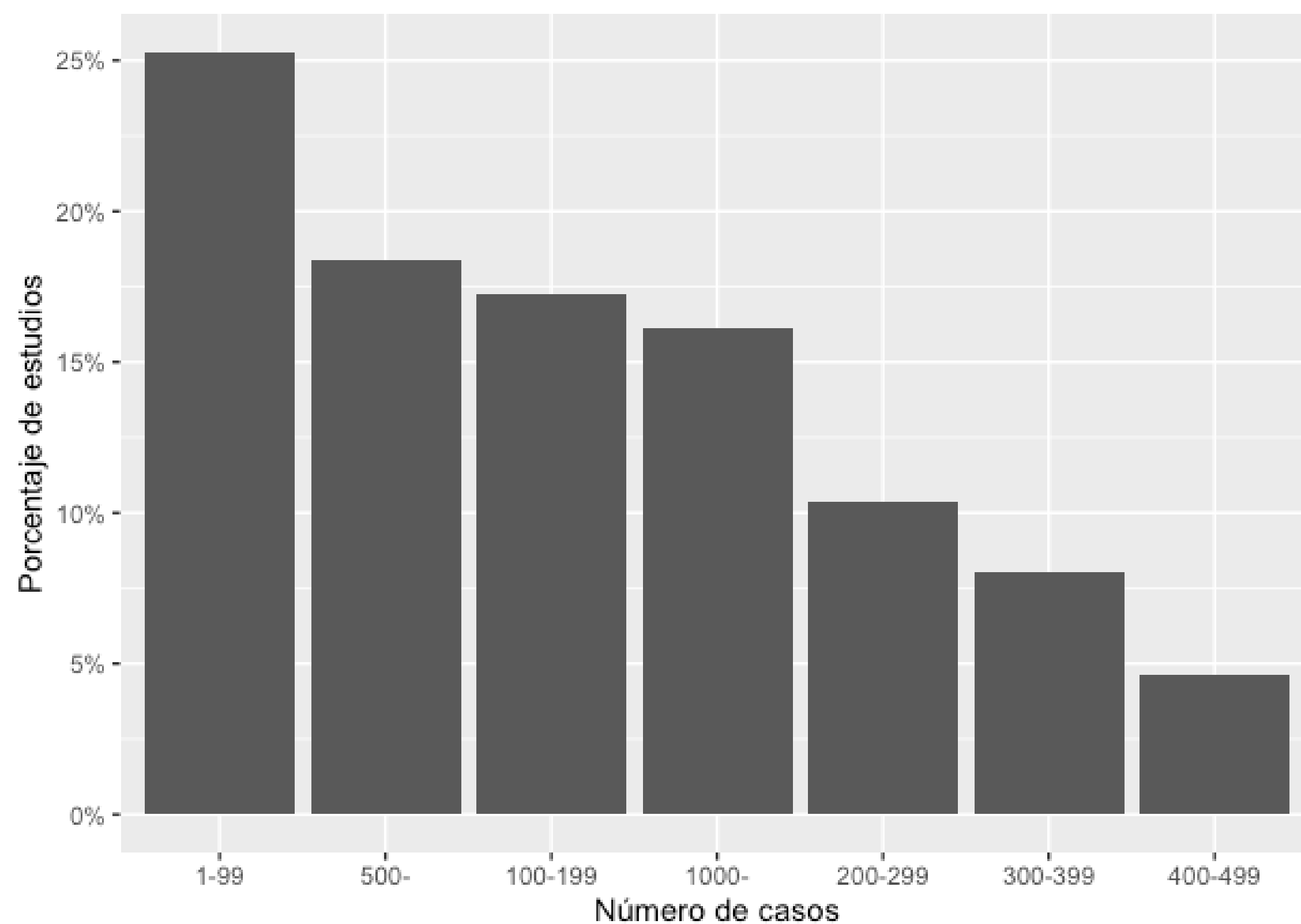
Figura 12: Número de repeticiones de cada ítem del corpus en los estudios sobre los correlatos acústicos del acento léxico analizados por Roettger y Gordon (2017)



Nota. Gráfico elaborado a partir de la información recogida en la base de datos del proyecto *Studies on acoustic correlates of word stress*, de M. K. Gordon y T. B. Roettger, 18 de abril de 2018, OSF (<https://doi.org/10.17605/OSF.IO/9R2CD>).

Multiplicando el número de palabras por el de sus repeticiones y por el número de hablantes, puede conocerse el total de casos analizados en los estudios sobre los correlatos acústicos del acento léxico considerados por Roettger y Gordon (2017). Así, en la figura 13 se constata que la situación más frecuente (en el 25 % de los trabajos) es que el número de casos analizados esté por debajo de 100.

Figura 13: Número total de casos analizados en los estudios sobre los correlatos acústicos del acento léxico recogidos en Roettger y Gordon (2017)



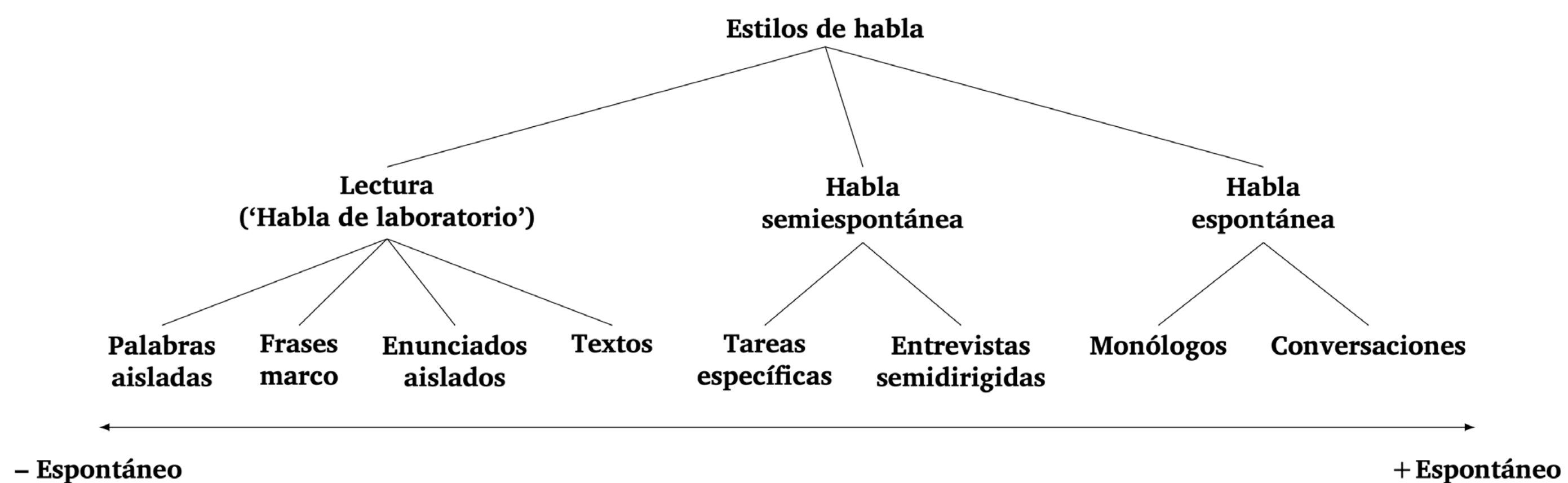
Nota. Gráfico elaborado a partir de la información recogida en la base de datos del proyecto *Studies on acoustic correlates of word stress*, de M. K. Gordon y T. B. Roettger, 18 de abril de 2018, OSF (<https://doi.org/10.17605/OSF.IO/9R2CD>).

En resumen, según las informaciones proporcionadas por Roettger y Gordon (2017) y recogidas en Gordon y Roettger (2018), en las publicaciones seleccionadas por estos autores el estudio de los correlatos acústicos del acento léxico se lleva a cabo, mayoritariamente, empleando entre una y 40 palabras diferentes, con un número de repeticiones de cada palabra que oscila entre una y seis y contando con un total de casos que en una cuarta parte de los trabajos publicados se sitúa por debajo de 100.

Estilos de habla

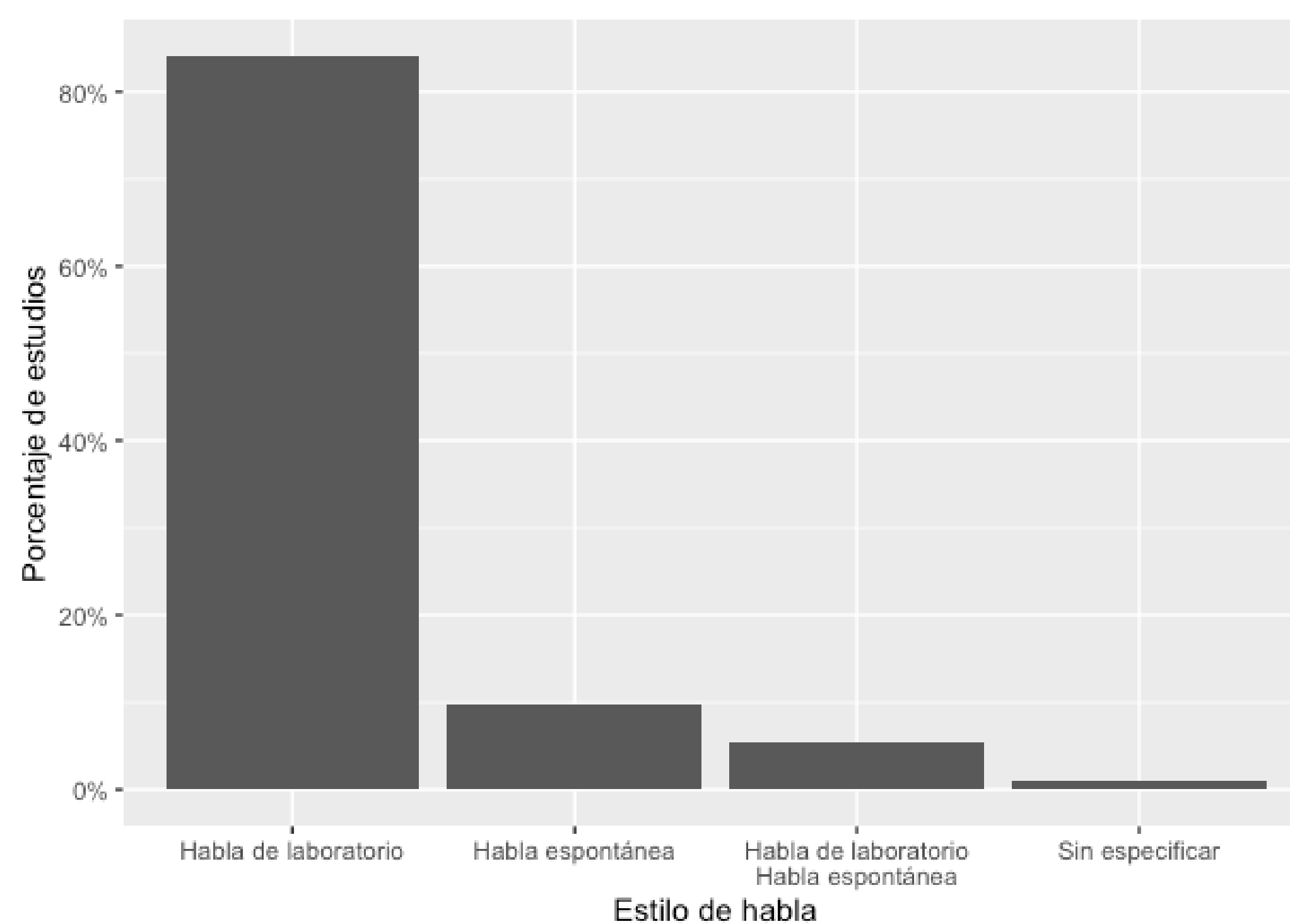
Un segundo aspecto sobre el que conviene reflexionar en lo que respecta al corpus es el tipo de materiales que en él se incluyen, lo que se relaciona con la elección de los estilos de habla que se tendrán en cuenta en una investigación. En la figura 14 se indican algunas de las posibilidades en cuanto a los materiales, organizadas en un continuo, desde el estilo menos espontáneo —la lectura de palabras aisladas— al más espontáneo —la conversación— que refleja, en líneas generales, el grado de planificación del discurso por parte del hablante o, en términos labovianos, el grado de atención que el locutor presta a su producción lingüística (Labov, 1972/1978).

Figura 14: Clasificación de los materiales que pueden formar parte de un corpus en función de los estilos de habla



En la información recabada por Roettger y Gordon (2017) se pone de manifiesto que, como se aprecia en la figura 15, en más del 80 % de los estudios sobre los correlatos acústicos del acento léxico se analiza lo que habitualmente se define como ‘habla de laboratorio’, es decir, la lectura de un corpus controlado y previamente planificado constituido por palabras, frases o textos.

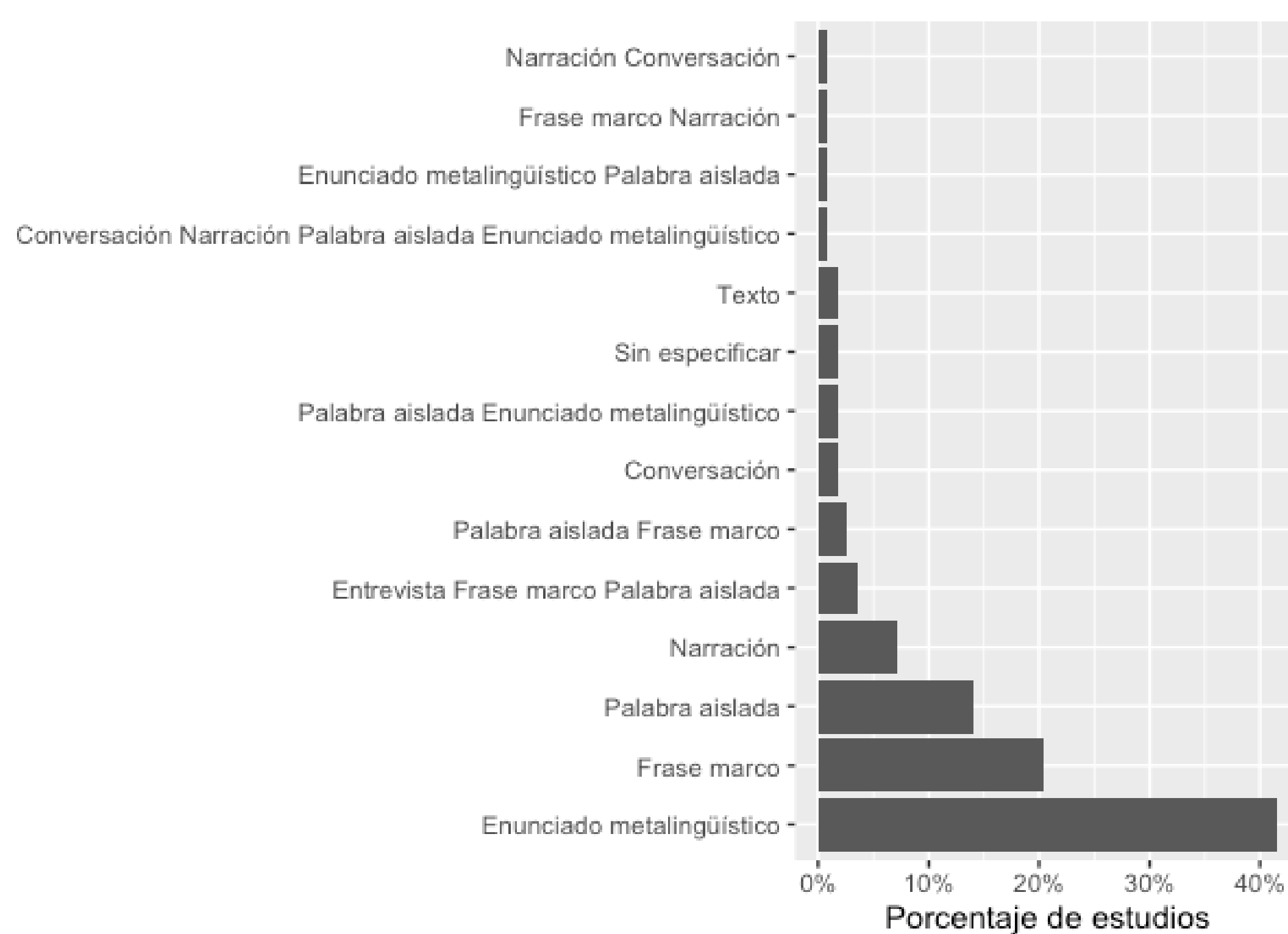
Figura 15: Estilos de habla en los corpus empleados en los estudios sobre los correlatos acústicos del acento léxico analizados por Roettger y Gordon (2017)



Nota. Gráfico elaborado a partir de la información recogida en la base de datos del proyecto *Studies on acoustic correlates of word stress*, de M. K. Gordon y T. B. Roettger, 18 de abril de 2018, OSF (<https://doi.org/10.17605/OSF.IO/9R2CD>).

Una visión algo más pormenorizada del contexto en el que aparecen los materiales analizados en los trabajos sobre los correlatos acústicos del acento léxico se muestra en la figura 16, elaborada a partir de la información recogida en la base de datos de Gordon y Roettger (2018). Como se puede observar, las tres opciones más habituales son, por este orden, los enunciados metalingüísticos, las frases marco y las palabras aisladas.

Figura 16: Contexto de aparición de los materiales del corpus en los estudios sobre los correlatos acústicos del acento léxico analizados por Roettger y Gordon (2017)



Nota. Gráfico elaborado a partir de la información recogida en la base de datos del proyecto *Studies on acoustic correlates of word stress*, de M. K. Gordon y T. B. Roettger, 18 de abril de 2018, OSF (<https://doi.org/10.17605/OSF.IO/9R2CD>).

El estilo de habla con un menor grado de espontaneidad que se encuentra en las investigaciones sobre los correlatos acústicos del acento léxico consiste, cuando se trata de lenguas de acento libre, en la lectura de palabras aisladas que constituyen pares mínimos acentuales, o incluso tripletas, como se ha ejemplificado en (1).

Si se necesita neutralizar el efecto del significado, los investigadores recurren a seudopalabras (también llamadas ‘logatomas’), esto es, palabras fonológicamente posibles porque reúnen todos los requisitos fonotácticos exigidos, pero que realmente no se dan en la lengua. En algunos experimentos se utilizan secuencias de sílabas idénticas en las que se introduce un contraste acentual (DE MORAES, 1987), mientras que en otros se crean conjuntos de seudopalabras diferentes, como podrían ser, en el caso del español, «ládebo», «ladebo» y «ladebó» (LLISTERRI et al., 2014).

A fin de evitar el denominado ‘efecto de lista’ propio de la lectura de palabras aisladas, es habitual que las palabras objeto de estudio se incluyan dentro de lo que se conoce como ‘frases marco’ o ‘frases portadoras’ (*carrier sentences*, en inglés). Frecuentemente, como se pone de manifiesto en los datos de Gordon y Roettger (2018) mostrados en la figura 16, las frases marco consisten en enunciados metalingüísticos del tipo «Digo la palabra _____ otra vez» (QUILIS, 1971) o «Digo _____ baixinho» (Fontes, 2013). Mediante esta estrategia se consigue contextualizar las palabras en las que el investigador está interesado en una estructura constante y, con ello, controlar determinadas

variables como la posición de la palabra en el enunciado o la distribución de los acentos. A pesar de que, evidentemente, están lejos de lo que se consideraría una producción espontánea, las frases marco favorecen hasta cierto punto una realización más natural que una lectura de palabras aisladas y por ello se han empleado en estudios sobre los correlatos acústicos del acento léxico en portugués y en español como los de Cruz et al. (2016), Face (1998–1999), Ferreira (2014) o Massini-Cagliari (1993), entre otros.

Otra posibilidad para crear un corpus controlado consiste en hacer uso de enunciados aislados, contruidos en función de las variables que se han considerado pertinentes en el diseño del experimento. Por ejemplo, en el corpus empleado por Llisterri et al. (2014) se pretende conseguir que la sílaba tónica objeto de estudio siempre aparezca precedida y seguida de dos sílabas átonas en enunciados con el mismo número de sílabas, como sucede en «Calculas el límite mal», «No le límites los gastos» y «No limité los tabiques». El análisis de aspectos específicos como, por ejemplo, el comportamiento de dos sílabas acentuadas adyacentes —es decir, el denominado ‘choque acentual’— requiere también diseñar un corpus con características bien controladas, como el propuesto por Barbosa y Arantes (2003) para el portugués de Brasil, que contiene oraciones como «Eu **comi** **bolor** sexta-feira à noite» y «Eu **comi** **bolo** sexta-feira à noite», o el creado por Atria (2009) para estudiar el choque acentual en español de Chile, en el que se encuentran enunciados como «Su **papá** **pasa**» y «Su **papa** **pasa**».

Los investigadores tienen también a su disposición una serie de tareas que permiten obtener realizaciones de las palabras elegidas como objeto de estudio con un mayor grado de naturalidad que en los casos que se han examinado hasta ahora, todos ellos basados en la lectura del corpus que se analiza. En algunos casos, los participantes en el experimento deben construir un enunciado a partir de una serie de palabras clave (TORREIRA et al., 2014) o responder a una pregunta con una determinada construcción sintáctica (ORTEGA-LLEBARIA; PRIETO, 2007); en otros, tienen que reemplazar una palabra por una seudopalabra (MASSINI-CAGLIARI, 1993; ORTEGA-LLEBARIA, 2006). También puede recurrirse a imágenes, solas o acompañadas de un texto, para conseguir que los hablantes emitan el enunciado deseado (RONQUEST, 2013; VOGEL et al., 2016).

Una aproximación diferente al modo de recabar los datos consiste en incorporar tareas que aparentemente no guardan ninguna relación con el propósito del experimento, para lograr que los hablantes centren su atención en el objetivo de la tarea más que en su propia producción lingüística, consiguiendo así un estilo que puede caracterizarse como semiespontáneo según la clasificación propuesta en la figura 14. Entre las tareas de este tipo más conocidas destaca la denominada ‘tarea del mapa’ (ANDERSON et al., 1991), en la que uno de los participantes tiene un mapa con un itinerario marcado, mientras que el otro cuenta con una versión del mismo mapa sin ningún itinerario; los dos hablantes, que no pueden ver sus mapas respectivos, deben cooperar para que el que tiene el mapa

sin itinerario dibuje el mismo camino que está trazado en el otro mapa. Basándose en el *Map-Task Corpus of the Open University of Israel*, recogido mediante este procedimiento, Silber-Varod et al. (2019) obtuvieron un total de 887 vocales para su estudio sobre la influencia del acento léxico en los valores de los formantes de las vocales en hebreo.

También formarían parte del estilo semiespontáneo las entrevistas que se realizan a los hablantes para conseguir intervenciones lo más largas posibles y disponer, con ello, de suficiente material para el análisis; se trata de un tipo de interacción que suele describirse como ‘entrevista semidirigida’, puesto que existe un cierto grado de planificación por parte del entrevistador. Esta técnica se ha usado, por ejemplo, en el estudio sobre los correlatos acústicos del acento léxico en portugués de Brasil llevado a cabo por Barbosa et al. (2013b).

Entre los estilos de habla que en la figura 14 se clasifican como espontáneos se encuentra la conversación. Por regla general, cuanto mayor es el grado de espontaneidad del corpus, menor es el control sobre las variables definidas en el diseño del experimento, pero Torreira y Ernestus (2011), por ejemplo, consiguieron extraer 727 vocales del español peninsular y 497 del francés europeo que cumplieran con los requisitos necesarios para estudiar el acento partiendo de las conversaciones del *Nijmegen Corpus of Casual Spanish* y del *Nijmegen Corpus of Casual French*. En este trabajo, al igual que en el de Silber-Varod et al. (2019), se pone de manifiesto que resulta factible investigar sobre el acento léxico empleando recursos ya disponibles creados inicialmente con otros objetivos.

Finalmente, en las publicaciones sobre los correlatos acústicos del acento léxico se ha considerado también la posibilidad de comparar diversos estilos de habla, como se hace, por ejemplo, en el ya mencionado trabajo de Barbosa et al. (2013b) sobre el portugués de Brasil, en el que se contrastan tres estilos de habla diferentes en los mismos locutores: entrevista semidirigida, lectura de textos y lectura de palabras aisladas.

Reflexiones finales

En las páginas anteriores se ha intentado mostrar que el análisis de los correlatos acústicos del acento léxico exige que el investigador tome, de entrada, una serie de decisiones metodológicas. Los elementos que se han mencionado no son, obviamente, los únicos que es necesario tener en cuenta a la hora de diseñar un experimento; como se ha indicado al principio, una lectura atenta y crítica de los trabajos ya realizados resulta sumamente útil para conocer los diversos factores que pueden incidir en los resultados.

Roettger y Gordon (2017) concluyen su revisión de la metodología empleada en las publicaciones sobre los correlatos acústicos del acento léxico, a la que ya se ha aludido en varias ocasiones, con recomendaciones centradas en tres aspectos: la transparencia del diseño experimental, el diseño del corpus y el tamaño de la muestra analizada.

La transparencia en el diseño del experimento hace referencia, para estos autores, a la necesidad de especificar con claridad todas las decisiones asumidas por el investigador, especialmente en lo que se refiere a la estructura prosódica del material analizado, con la finalidad de separar los efectos de una prominencia propia del nivel de la frase de los que se derivan del acento léxico en la palabra, diferenciando entre lo que en la terminología propia del modelo métrico-autosegmental (HUALDE, 2003) se conoce como *accent*, que se manifiesta mediante un movimiento tonal relacionado con la frase, y lo que se denomina *stress*, vinculado al acento léxico.

La importancia de esta distinción se retoma en las recomendaciones de Roettger y Gordon (2017) relacionadas con el diseño del corpus, pero no debe olvidarse, además, que las propiedades inherentes de los segmentos analizados, su contexto fonético y la estructura silábica en la que se enmarcan tienen potenciales repercusiones en los resultados de los estudios acústicos sobre el acento.

El tamaño de la muestra es el tercero de los factores mencionados por Roettger y Gordon (2017). Por una parte, la muestra debe reflejar adecuadamente las características de un conjunto de hablantes más amplio que el formado por los participantes en el experimento y, por otra, tiene que ser representativa de la lengua. Sin embargo, para estos autores la decisión sobre el número de casos se toma teniendo principalmente en cuenta criterios relacionados con la potencia de las pruebas estadísticas empleadas.

A los planteamientos de Roettger y Gordon (2017) pueden añadirse algunas reflexiones que se derivan de lo expuesto en el presente trabajo. Como se ha señalado, parece darse una cierta tendencia a dejar de lado estudios que no se publican en el contexto académico anglosajón, escritos en lenguas que no sean el inglés y, en ocasiones, que no aparezcan en lo que se ha dado en llamar 'revistas de impacto', como si la calidad de una investigación estuviera necesariamente ligada a la lengua en la que está redactada y al grado de difusión que llegue a alcanzar en determinadas plataformas. El desconocimiento de los trabajos que no cumplen estos supuestos requisitos lleva a una visión parcial del campo e incluso a presentar como nueva una información que, en mayor o menor medida, ya era conocida. Por ello, una revisión bibliográfica razonablemente exhaustiva y, especialmente, exenta de sesgos como los que se acaban de mencionar parece un punto de partida imprescindible en el momento de plantear un experimento.

En este trabajo también se ha insistido en la necesidad de contar con un perfil lingüístico lo más detallado posible de los participantes, tanto en los estudios sobre la primera lengua como en los que se centran en hablantes no nativos. Como se ha explicado, la variante geográfica de los hablantes monolingües puede repercutir en la realización fonética del acento léxico y bajo etiquetas como 'bilingüe' o 'estudiante de L2' se esconden realidades lingüísticas muy distintas.

Asimismo, se ha abordado la cuestión de los estilos de habla que se incluyen en el corpus analizado pues, como se ha visto, parece que gran parte de los datos de los que se dispone sobre los correlatos acústicos del acento léxico proceden de palabras aisladas y de enunciados relativamente breves, que tal vez no son completamente representativos de la lengua en su uso real.

Finalmente, al hilo de las propuestas de Roettger y Gordon (2017) sobre el diseño del corpus recogidas al principio de este apartado, en los estudios sobre el acento léxico tal vez sería pertinente precisar el modelo adoptado por el autor para evitar que, en ciertos casos, pueda llegarse a pensar que se asume como universal —por ejemplo, la noción de ‘acento tonal’ o la transcripción mediante el sistema ToBI (Aguilar et al., 2009; Frota et al., 2015)— algo que únicamente es propio de una determinada teoría.

En conjunto, como han señalado múltiples autores, el acento léxico es un fenómeno de naturaleza multiparamétrica sobre el que incide una gran diversidad de factores. Por tal motivo, en el presente trabajo se ha intentado poner de relieve que el número de decisiones metodológicas que debe tomar un investigador — sus ‘grados de libertad’, usando el término de Roettger (2019) — es considerable, lo que lleva a la necesidad de documentar tales decisiones de la forma más explícita posible para que, siguiendo lo que parece ser una tendencia creciente también en el campo de la fonética (Roettger et al., 2019), los experimentos sean reproducibles. Contar con una información detallada sobre la metodología permite también, como apuntan Roettger y Gordon (2017), valorar la pertinencia de comparar resultados obtenidos en experimentos sobre lenguas diferentes, lo que constituye una innegable ventaja tanto para los estudios contrastivos como para los que se orientan hacia la tipología fonética.

Referencias

AGUILAR, L.; DE LA MOTA, C.; PRIETO, P. *Sp_ToBI training materials*. Grup d’Estudis de Prosòdia, 2009. Disponible em: http://prosodia.upf.edu/sp_tobi/en/index.php. Acceso em: 07 jun. 2022.

ANDERSON, A. H.; BADER, M.; BARD, E. G.; BOYLE, E.; DOHERTY, G.; GARROD, S.; ISARD, S.; KOWTKO, J.; MCALLISTER, J.; MILLER, J.; SOTILLO, C.; THOMPSON, H. S. Y.; WEINERT, R. The HCRC Map Task corpus. *Language and Speech*, v. 34, n. 4, p. 351-366, 1991. DOI: <https://doi.org/10.1177/002383099103400404>

ATRIA, J. J. Estrategias de resolución de choques acentuales en el castellano hablado en Santiago de Chile. *Onomázein*, v. 19, n. 1, p. 11-31, 2009. Disponible em: http://onomazein.lettras.uc.cl/Articulos/19/1_Atria.pdf. Acceso em: 07 jun. 2022.

AVELLO, P. Assessing learners' changes in foreign accent during Study Abroad. In: PÉREZ VIDAL, C.; LÓPEZ-SERRANO, S.; AMENT, J.; THOMAS-WILHELM, D. J. (ed.). *Learning context effects: study abroad, formal instruction and international immersion classrooms*. Language Science Press, 2018. p. 131-154. DOI: <https://doi.org/10.5281/zenodo.1300630>

BADIA I MARGARIT, A. M. Notes sobre el castellà parlat per catalans. *Llengua i cultura als Països Catalans*. 4. ed. Edicions 62, 1975. p. 145-153.

BARBOSA, P. A.; ALBANO, E. C. Brazilian Portuguese. *Journal of the International Phonetic Association*, v. 34, n. 2, p. 227-232, 2004. DOI: <https://doi.org/10.1017/S0025100304001756>

BARBOSA, P. A.; ARANTES, P. Investigation of non-pitch-accented phrases in Brazilian Portuguese: No evidence favoring stress shift. In: SOLÉ, M. J.; RECASENS, D.; ROMERO, J. (ed.). *15th International Congress of Phonetic Sciences*. Barcelona, Spain, August 3-9, 2003. p. 475-478. ICPHS Archive. Disponible em: https://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS2003/p15_0475.html. Acesso em: 07 jun. 2022.

BARBOSA, P. A.; ERIKSSON, A.; ÅKESSON, J. Cross-linguistic similarities and differences of lexical stress realisation in Swedish and Brazilian Portuguese. In: ASU, E. L.; LIPPUS, P. (ed.). *Nordic prosody: Proceedings of the XIth Conference, Tartu 2012*. Peter Lang, 2013a. p. 97-106.

BARBOSA, P. A.; ERIKSSON, A.; ÅKESSON, J. On the robustness of some acoustic parameters for signalling word stress across styles in Brazilian Portuguese. In: BIMBOT, F.; CERISARA, C.; FOUGERON, C.; GRAVIER, G.; LAMEL, L.; PELLEGRINO, F.; PERRIER, P. (ed.). *INTERSPEECH 2013, 14th Annual Conference of the International Speech Communication Association*. Lyon, France, August 25-29, 2013b. p. 282-286. ISCA Archive. Disponible em: https://www.isca-speech.org/archive/interspeech_2013/i13_0282.html. Acesso em: 07 jun. 2022.

BARBOSA, P. A.; MADUREIRA, S. *Manual de fonética acústica experimental: aplicações a dados do português*. São Paulo: Cortez, 2015.

BILINGUALISM AND PSYCHOLINGUISTICS RESEARCH GROUP. (s.f.). *LEAP-Questionnaire*. Northwestern University. Disponible em: <https://bilingualism.northwestern.edu/leapq/>. Acesso em: 07 jun. 2022.

BIRDSONG, D.; GERTKEN, L. M.; AMENGUAL, M. *Bilingual Language Profile: An easy-to-use instrument to assess bilingualism*. University of Texas at Austin, Center for Open Educational Resources and Language Learning (COERLL). 2012, 20 de enero. Disponible em: <https://sites.la.utexas.edu/bilingual/>. Acesso em: 07 jun. 2022.

BOERSMA, P.; WEENINK, D. *Praat: Doing phonetics by computer* (Versión 6.1.40) [Programa informático]. University of Amsterdam. 2021. Disponível em: <http://www.praat.org>. Acesso em: 07 jun. 2022.

COLANTONI, L.; MARASCO, O.; STEELE, J.; SUNARA, S. Learning to realize prosodic prominence in French and Spanish. In: MILLER, R. T.; MARTIN, K. I.; EDDINGTON, C. M.; HENERY, A.; MARCOS, N.; TSENG, A. M.; TUNINETTI, A.; WALTER, D. (ed.). *Selected Proceedings of the 2012 Second Language Research Forum: Building bridges between disciplines*. 2014. p. 15-29. Cascadilla Proceedings Project. Disponível em: <http://www.lingref.com/cpp/slrf/2012/abstract3082.html>. Acesso em: 07 jun. 2022.

CONSEJO DE EUROPA. *Marco común europeo de referencia para las lenguas: aprendizaje, enseñanza, evaluación*. Ministerio de Educación, Cultura y Deportes, Instituto Cervantes; Anaya, 2002. Disponível em: https://cvc.cervantes.es/ensenanza/biblioteca_ele/marco/. Acesso em: 07 jun. 2022.

CRUZ, R. C. F.; BORGES, B. do S. P.; FERREIRA, J. É. Q.; RILLIARD, A.; FONTEL, E. da S. Análise acústica dos correlatos prosódicos do acento lexical do Português Brasileiro. *Journal of Speech Sciences*, v. 5, n. 2, p. 159-175, 2016. Disponível em: http://revistas.iel.unicamp.br/ojs_joss/index.php/journalofspeechsciences/article/view/121. Acesso em: 07 jun. 2022.

CRYSTAL, D. Stress. *A first dictionary of linguistics and phonetics*. André Deutsch, 1980. p. 332-333.

FACE, T. L. Efectos segmentales del acento en español. *Boletín de Lingüística*, v. 14, p. 18-32, 1998-1999.

FERREIRA, L. P. A duração como correlato acústico do acento de palavra no português brasileiro e no espanhol: desafios para o ensino de suprasegmentais e preparação de material didático. *Signum: Estudos da Linguagem*, v. 17, n. 1, p. 74-101, 2014. DOI: <https://doi.org/10.5433/2237-4876.2014v17n1p74>

FONTES, M. A. S. Pistas acústicas e a percepção do acento lexical em português brasileiro. *Intercâmbio*, v. 27, p. 100-109, 2013. Disponível em: <https://revistas.pucsp.br/intercambio/article/view/17692>. Acesso em: 07 jun. 2022.

FROTA, S.; OLIVEIRA, P.; CRUZ, M.; VIGÁRIO, M. *P-ToBI: Tools for the transcription of Portuguese prosody*. CLUL/FLUL, Laboratório de Fonética, 2015. Disponível em: <http://labfon.letra.sulisboa.pt/InAPoP/P-ToBI/index.html>. Acesso em: 07 jun. 2022.

GARDE, P. *L'accent*. Presses Universitaires de France, 1968.

- GARRIDO, J. M. Using large corpora and computational tools to describe prosody: An exciting challenge for the future with some (important) pending problems to solve. In: FELDHAUSEN, I.; FLIESSBACH, J.; DEL M. VANRELL, M. (ed.). *Methods in prosody: A Romance language perspective*. Language Science, 2018. p. 3-43. DOI: <https://doi.org/10.5281/zenodo.1441335>
- GERTKEN, L. M.; AMENGUAL, M.; BIRDSONG, D. Assessing language dominance with the Bilingual Language Profile. In: LECLERCQ, P.; EDMONDS, A.; HILTON, H. (ed.). *Measuring L2 proficiency: Perspectives from SLA*. Multilingual Matters, 2014. p. 208-225.
- GORDON, M. K.; ROETTGER, T. B. Acoustic correlates of word stress: A crosslinguistic survey. *Linguistics Vanguard*, v. 3, n. 1, Artículo 20170007, 2017. DOI: <https://doi.org/10.1515/lingvan-2017-0007>
- GORDON, M. K.; ROETTGER, T. B. *Studies on acoustic correlates of word stress: An online corpus*. 2018, 18 de abril. DOI: <https://doi.org/10.17605/OSF.IO/9R2CD>
- HOFFMANN, C. *An introduction to bilingualism*. Longman, 1991.
- HUALDE, J. I. El modelo métrico y autosegmental. In: PRIETO, P. (ed.). *Teorías de la entonación*. Ariel, 2003. p. 155-184.
- INTERNATIONAL PHONETIC ASSOCIATION. *Handbook of the International Phonetic Association: A guide to the use of the International Phonetic Alphabet*. Cambridge: Cambridge University Press, 1999.
- ISBELL, D. R. Assessing pronunciation for research purposes with listener-based numerical scales. In: KANG, O.; GINTHER, A. (ed.). *Assessment in second language pronunciation*. Routledge, 2018. p. 89-111. DOI: <https://doi.org/10.4324/9781315170756>
- JULIÀ, J. *Fonètica aplicada catalana: dels fonaments a les aplicacions de les ciències fonètiques*. Ariel, 2005.
- KAUSHANSKAYA, M.; BLUMENFELD, H. K.; MARIAN, V. The Language Experience and Proficiency Questionnaire (LEAP-Q): Ten years later. *Bilingualism: Language and Cognition*, v. 23, n. 5, p. 945-950, 2020. DOI: <https://doi.org/10.1017/s1366728919000038>
- KIM, M. *The phonetics of stress manifestation: Segmental variation, syllable constituency and rhythm*. Tesis doctoral, Stony Brook University. SUNY Digital Repository, 2011. <http://hdl.handle.net/1951/56031>

LABOV, W. The isolation of contextual styles. *Sociolinguistic patterns*. Basil Blackwell, 1978. p. 70-109. Obra original publicada en 1972.

LADEFOGED, P. *Phonetic data analysis: An introduction to fieldwork and instrumental techniques*. Blackwell, 2003.

LI, P.; ZHANG, F.; YU, A.; ZHAO, X. Language History Questionnaire (LHQ3): An enhanced tool for assessing multilingual experience. *Bilingualism: Language and Cognition*, v. 23, n. 5, p. 938-944, 2020. DOI: <https://doi.org/10.1017/S1366728918001153>

LLISTERRI, J.; MACHUCA, M. J.; RÍOS, A.; SCHWAB, S. El acento léxico en contexto: datos acústicos. In: CONGOSTO, Y.; MONTERO, M. L.; SALVADOR, A. (ed.). *Fonética experimental, educación superior e investigación*. Arco/Libros, 2014. v. 1, p. 357-376.

MARTÍNEZ CELDRÁN, E.; FERNÁNDEZ PLANAS, A. M.; CARRERA, J. Castilian Spanish. *Journal of the International Phonetic Association*, v. 33, n. 2, p. 255-259, 2003. DOI: <https://doi.org/10.1017/S0025100303001373>

MASSINI-CAGLIARI, G. Sobre a natureza fonética do acento em português. *DELTA: Documentação e Estudos em Linguística Teórica e Aplicada*, v. 9, n. 2, p. 195-216, 1993. Disponível em: <http://ken.pucsp.br/delta/article/view/45514>. Acesso em: 07 jun. 2022.

DE MORAES, J. A. Correlats acoustiques de l'accent de mot en portugais bresilien. *Proceedings XIth ICPHS: The Eleventh International Congress of Phonetic Sciences*. August 1-8, 1987, Tallinn, Estonia, U.S.S.R., Academy of Sciences of the Estonian S.S.R., v. 3, p. 313-316, 1987.

MUNRO, M. J.; DERWING, T. M. Processing time, accent, and comprehensibility in the perception of native and foreign-accented speech. *Language and Speech*, v. 38, n. 3, p. 289-306, 1995. DOI: <https://doi.org/10.1177/002383099503800305>

NORRIS, J.; ORTEGA, L. Defining and measuring SLA. In: DOUGHTY, C. J.; LONG, M. H. (ed.). *The handbook of second language acquisition*. Blackwell, 2003. p. 716-761. DOI: <https://doi.org/10.1002/9780470756492.ch21>

ORTEGA-LLEBARIA, M. Phonetic cues to stress and accent in Spanish. In: DÍAZCAMPOS, M. (ed.). *Selected Proceedings of the 2nd Conference on Laboratory Approaches to Spanish Phonetics and Phonology*. Cascadilla Proceedings Project, 2006. p. 104-118. Disponível em: <http://www.lingref.com/cpp/lasp/2/abstract1329.html>. Acesso em: 07 jun. 2022.

ORTEGA-LLEBARIA, M.; PRIETO, P. Disentangling stress from accent in Spanish: Production patterns of the stress contrast in deaccented syllables. In: PRIETO, P.; MASCARÓ, J.; SOLÉ, M. J. (ed.). *Segmental and prosodic issues in Romance phonology*. John Benjamins, 2007. p. 155-176. DOI: <https://doi.org/10.1075/cilt.282.11ort>

ORTEGA-LLEBARIA, M.; PRIETO, P. Acoustic correlates of stress in Central Catalan and Castilian Spanish. *Language and Speech*, v. 54, n. 1, p. 73-97, 2011. DOI: <https://doi.org/10.1177/0023830910388014>

QUILIS, A. Caracterización fonética del acento español. *Travaux de Linguistique et de Littérature*, v. 9, n. 1, p. 53-72, 1971.

R CORE TEAM. *R: A language and environment for statistical computing* (Versión 4.0.4) [Programa informático]. Vienna, R Foundation for Statistical Computing. 2021. Disponível em: <https://www.r-project.org/>. Acesso em: 07 jun. 2022.

ROETTGER, T. B. Researcher degrees of freedom in phonetic research. *Laboratory Phonology: Journal of the Association for Laboratory Phonology*, v. 10, n. 1, artículo 1, 2019. DOI: <https://doi.org/10.5334/labphon.147>

ROETTGER, T. B.; GORDON, M. K. Methodological issues in the study of word stress correlates. *Linguistics Vanguard*, v. 3, n. 1, artículo 20170007, 2017. DOI: <https://doi.org/10.1515/lingvan-2017-0006>

ROETTGER, T. B.; WINTER, B.; BAAYEN, H. Emergent data analysis in phonetic sciences: Towards pluralism and reproducibility. *Journal of Phonetics*, v. 73, p. 1-7, 2019. DOI: <https://doi.org/10.1016/j.wocn.2018.12.001>

RONQUEST, R. E. An acoustic examination of unstressed vowel reduction in Heritage Spanish. In: HOWE, C.; BLACKWELL, S. E.; LUBBERS QUESADA, M. (ed.). *Selected Proceedings of the 15th Hispanic Linguistics Symposium*. Cascadilla Proceedings Project, 2013. p. 157-171. Disponível em: <http://www.lingref.com/cpp/hls/15/abstract2882.html>. Acesso em: 07 jun. 2022.

SANTIAGO, F.; MAIRANO, P. The role of lexical stress on vowel duration and vowel space in two varieties of Spanish. In: KLESSA, K.; BACHAN, J.; WAGNER, A.; KARPIŃSKI, M.; ŚLEDZIŃSKI, D. (ed.). *9th International Conference on Speech Prosody 2018*. Poznań, Poland, 13-16 June 2018. p. 453-457. ISCA Archive. DOI: <https://doi.org/10.21437/SpeechProsody.2018-92>

SCHWAB, S. Acoustic correlates of French accentuation in advanced Spanish learners of French. *COPAL, Concordia Working Papers in Applied Linguistics*, v. 5, p. 612-623, 2014. Disponível em: http://doe.concordia.ca/copaldocuments/39_Schwab_Vol5.pdf. Acesso em: 07 jun. 2022.

SILBER-VAROD, V.; KHORSHIDI, N.; LEVI, L.; AMIR, N. The influence of lexical stress on formant values in spontaneous Hebrew speech. In: CALHOUN, S.; ESCUDERO, P.; TABAIN, M.; WARREN, P. (ed.). *Proceedings of the 19th International Congress of Phonetic Sciences*. Melbourne, Australia 2019. p. 3538-3542. Australasian Speech Science and Technology Association, 2019. Disponível em: https://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS2019/papers/ICPhS_3587.pdf. Acesso em: 07 jun. 2022.

THE BRAIN, LANGUAGE, AND COMPUTATION (BLC) Lab. (s. f.). *Language History Questionnaire 3.0*. The Hong Kong Polytechnic University. Disponível em: <https://blclab.org/lhq3/>. Acesso em: 07 jun. 2022.

TORREIRA, F.; ERNESTUS, M. Realization of voiceless stops and vowels in conversational French and Spanish. *Laboratory Phonology: Journal of the Association for Laboratory Phonology*, v. 2, n. 2, p. 331-353, 2011. DOI: <https://doi.org/10.1515/labphon.2011.012>

TORREIRA, F.; SIMONET, M.; HUALDE, J. I. Quasi-neutralization of stress contrasts in Spanish. In: CAMPBELL, N.; GIBBON, D.; HIRST, D. (ed.). *7th International Conference on Speech Prosody 2014*. 20-23 May 2014, Dublin, Ireland, 2014. p. 197-201. ISCA Archive. DOI: <https://doi.org/10.21437/SpeechProsody.2014-27>

TRASK, R. L. Stress. *A dictionary of phonetics and phonology*. Routledge, 1996. p. 336. DOI: <https://doi.org/10.4324/9780203695111>

TREFFERS-DALLER, J. What defines language dominance in bilinguals? *Annual Review of Applied Linguistics*, v. 5, p. 375-393, 2019. DOI: <https://doi.org/10.1146/annurev-linguistics-011817-045554>

VAN HEUVEN, V. J. Acoustic correlates and perceptual cues of word and sentence stress: Towards a cross linguistic perspective. In: GOEDEMANS, R.; HEINZ, J.; VAN DER HULST, H. (ed.). *The study of word stress and accent: Theories, methods and data*. Cambridge University Press, 2018. p. 15-59. DOI: <https://doi.org/10.1017/9781316683101.002>

VAN HEUVEN, V. J.; TURK, A. Phonetic correlates of word and sentence stress. In: GUSSENHOVEN, C.; CHEN, A. (ed.). *The Oxford handbook of language prosody*. Oxford: Oxford University Press, 2020. p. 150-165. DOI: <https://doi.org/10.1093/oxfordhb/9780198832232.013.8>

VOGEL, I.; ATHANASOPOULOU, A.; PINCUS, N. Prominence, contrast, and the Functional Load Hypothesis: An acoustic investigation. In: HEINZ, J.; GOEDEMANS, R.; VAN DER HULST, H. (ed.). *Dimensions of phonological stress*. Cambridge: Cambridge University Press, 2016. p. 123-167. DOI: <https://doi.org/10.1017/9781316212745.006>

Alguns aspectos metodológicos no estudo do acento lexical

Joaquim Llisterri

O estudo experimental do acento lexical

O acento lexical (na terminologia anglo-saxã *lexical stress*, *word stress*, ou, em alguns casos, simplesmente *stress*) é muitas vezes caracterizado como uma proeminência que cria um contraste entre uma unidade acentuada e unidades não acentuadas adjacentes (CRYSTAL, 1980; TRASK, 1996). Há uma tendência em se considerar que a unidade sobre a qual o acento recai é a sílaba, como fica claro, por exemplo, no fato de que no Alfabeto Fonético Internacional a marca que indica a presença de acento com uma marca que não está na vogal, mas no início da sílaba tônica (International Phonetic Association, 1999, p. 15, 22).

Sem entrar em detalhes sobre a definição das unidades prosódicas, pode-se dizer que a palavra é o domínio em que o acento lexical opera, portanto, nas línguas que são tradicionalmente conhecidas como de ‘acento livre’ (GARDE, 1968, p. 5-6), a posição do acento lexical na palavra em alguns casos dá origem a diferenças de significado, de modo que o acento adquire um valor contrastivo, como mostram os exemplos de (1a) para o espanhol e de (1b) para o português.

(1) a. *límite* [ˈlimite] (substantivo) / *limite* [liˈmite] (1.^a e 3.^a pessoa do singular do presente do subjuntivo do verbo *limitar*) / *limité* [limiˈte] (1.^a pessoa do singular do pretérito indefinido do verbo *limitar*) (MARTÍNEZ CELDRÁN et al., 2003, p. 257)

b. *sábia* [ˈsabiɐ] (adjetivo) / *sabia* [saˈbiɐ] (3.^a pessoa do singular do pretérito imperfeito do indicativo do verbo *saber*) / *sabiá* [sabiˈa] (substantivo) (BARBOSA; ALBANO, 2004, p. 230)

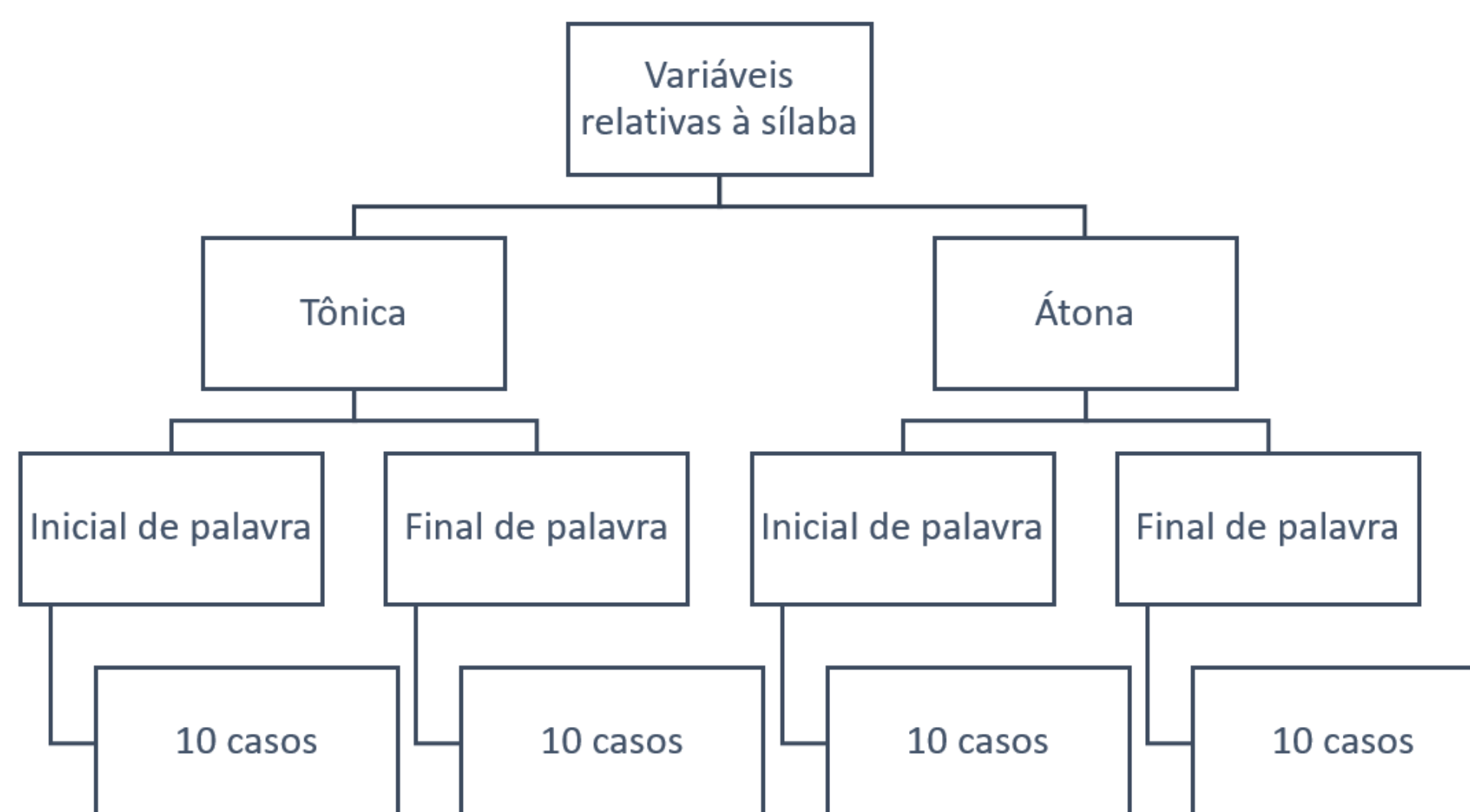
Boa parte dos estudos fonéticos experimentais sobre o acento lexical tem se preocupado em estabelecer seus correlatos acústicos ou perceptuais (VAN HEUVEN, 2018; VAN HEUVEN; TURK, 2020) e, em menor grau, articulatórios. Por ‘correlatos’ entendem-se aquelas propriedades de natureza fonética que estão relacionadas à produção pelo falante ou à percepção pelo ouvinte da proeminência que realça a sílaba tônica em relação às átonas. A realização de pesquisas sobre esses correlatos requer a tomada de uma série de decisões, algumas das quais serão discutidas a seguir. O leitor interessado em se aprofundar nos aspectos metodológicos encontrará informações relevantes no capítulo 5 (“Experimentação em Fonética Acústica”) de Barbosa e Madureira (2015), bem como nas seções dedicadas ao desenho de experimentos nas inúmeras publicações que abordaram o estudo do acento lexical.

Para analisar os correlatos acústicos do acento lexical, é necessário fazer perguntas relacionadas aos falantes, o corpus sobre o qual trabalhar, os procedimentos e as ferramentas utilizados para a aquisição e análise dos dados e o tratamento que será dado a eles. No que diz respeito aos falantes, o seu número e perfil linguístico são especialmente relevantes, bem como fatores de interesse

sociolinguístico como sexo, idade ou nível sociocultural. Estabelecer o corpus de trabalho implica tomar decisões sobre os estilos de fala que serão considerados e sobre o método mais adequado para obter amostras de diferentes estilos, bem como sobre as variáveis que serão levadas em consideração no desenho do corpus e o número de casos que estarão disponíveis para análise.

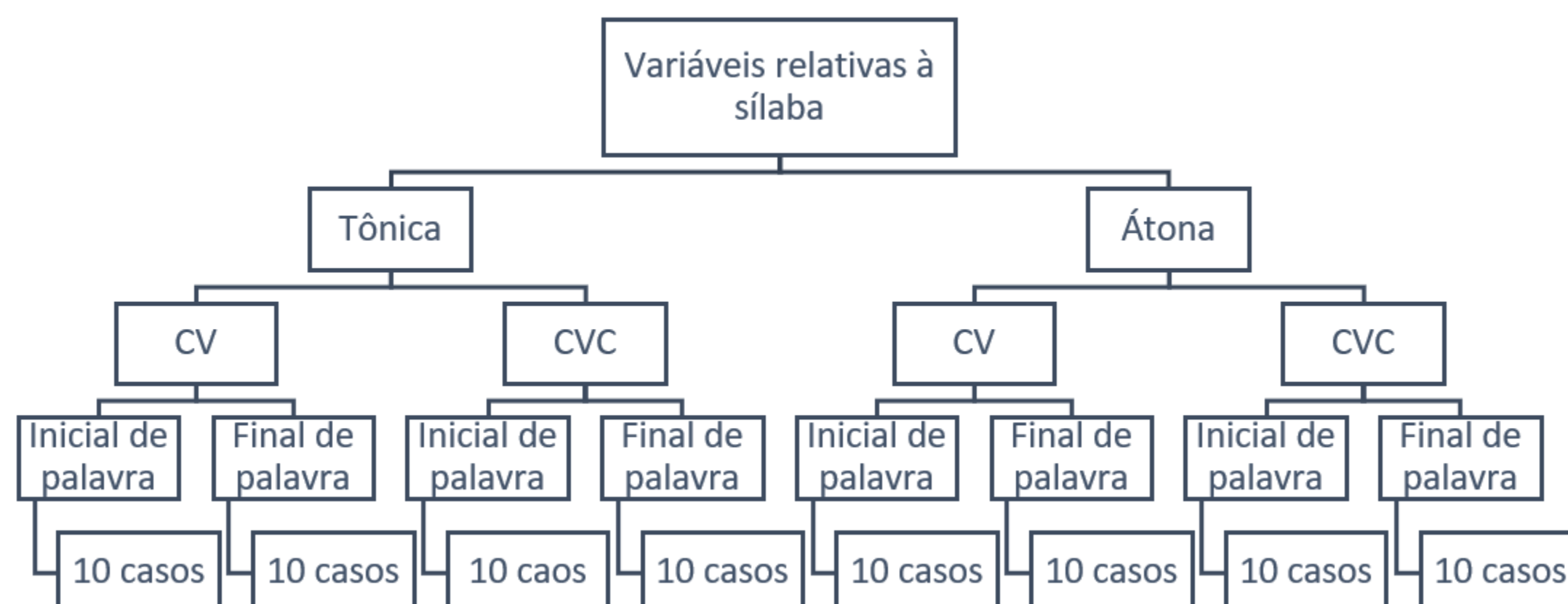
Deve-se notar que todos esses aspectos, que são explicados com mais detalhes na seção 2, estão, em maior ou menor grau, interrelacionados. Por exemplo, em um experimento no qual o efeito sobre os correlatos acústicos do acento da posição da sílaba dentro da palavra é investigado e se decide usar apenas palavras de duas sílabas, sílabas tônicas e átonas devem ser incluídas tanto em posição inicial quanto na posição final da palavra, conforme mostra a figura 1. Se você quiser ter 10 casos de cada combinação possível das duas variáveis (tonicidade e posição na palavra), o corpus será composto por 40 itens.

Figura 1: Desenho experimental de um estudo com duas variáveis: tonicidade da sílaba e posição da sílaba na palavra



Ao adicionar uma nova variável, por exemplo, a estrutura silábica, limitada a sílabas do tipo consoante-vogal (CV) e consoante-vogal-consoante (CVC), o número de casos que precisarão ser analisados sobe para 80, como pode ser verificado na figura 2.

Figura 2: Desenho experimental de um estudo com três variáveis: tonicidade silábica, estrutura silábica e posição da sílaba na palavra



Legenda de cima para baixo: variáveis relativas à sílaba; tônica, átona; estruturas de sílaba (CV, CVC); Inicial de palavra, final de palavra; 10 casos.

Se, além disso, fosse realizado o mesmo experimento comparando dois estilos de fala, o número de itens chegaria a 160 (80 para cada estilo). Assumindo que, em vez de 10 casos de cada combinação possível das variáveis, o pesquisador considera necessário ter 30 casos, um delineamento com duas variáveis implicaria analisar 120 itens, um com três variáveis suporia um corpus de 240 itens e, se introduzisse uma quarta variável, o número subiria para 480 itens para cada falante. Assim, em um estudo envolvendo 10 falantes, seriam analisadas 4.800 realizações.

Com este exemplo fictício, e talvez excessivamente simples, pretende-se destacar que à medida que aumenta o número de variáveis consideradas em um desenho experimental, aumenta também o tamanho do corpus e, com ele, o volume de dados que devem ser tratados. Pode-se argumentar que, atualmente, isso não é um problema, dada a possibilidade de realizar as análises por meio de procedimentos como segmentação e rotulagem automáticas e por meio de programas que permitem obter, também automaticamente, os valores dos parâmetros acústicos. No entanto, não se deve esquecer que, por um lado, as ferramentas de segmentação e marcação não estão disponíveis para todos os idiomas e que, por outro, os processos automáticos muitas vezes exigem revisão manual. No trabalho de Garrido (2018), são discutidas com mais detalhes as vantagens e desvantagens do tratamento automático de corpus para estudos prosódicos.

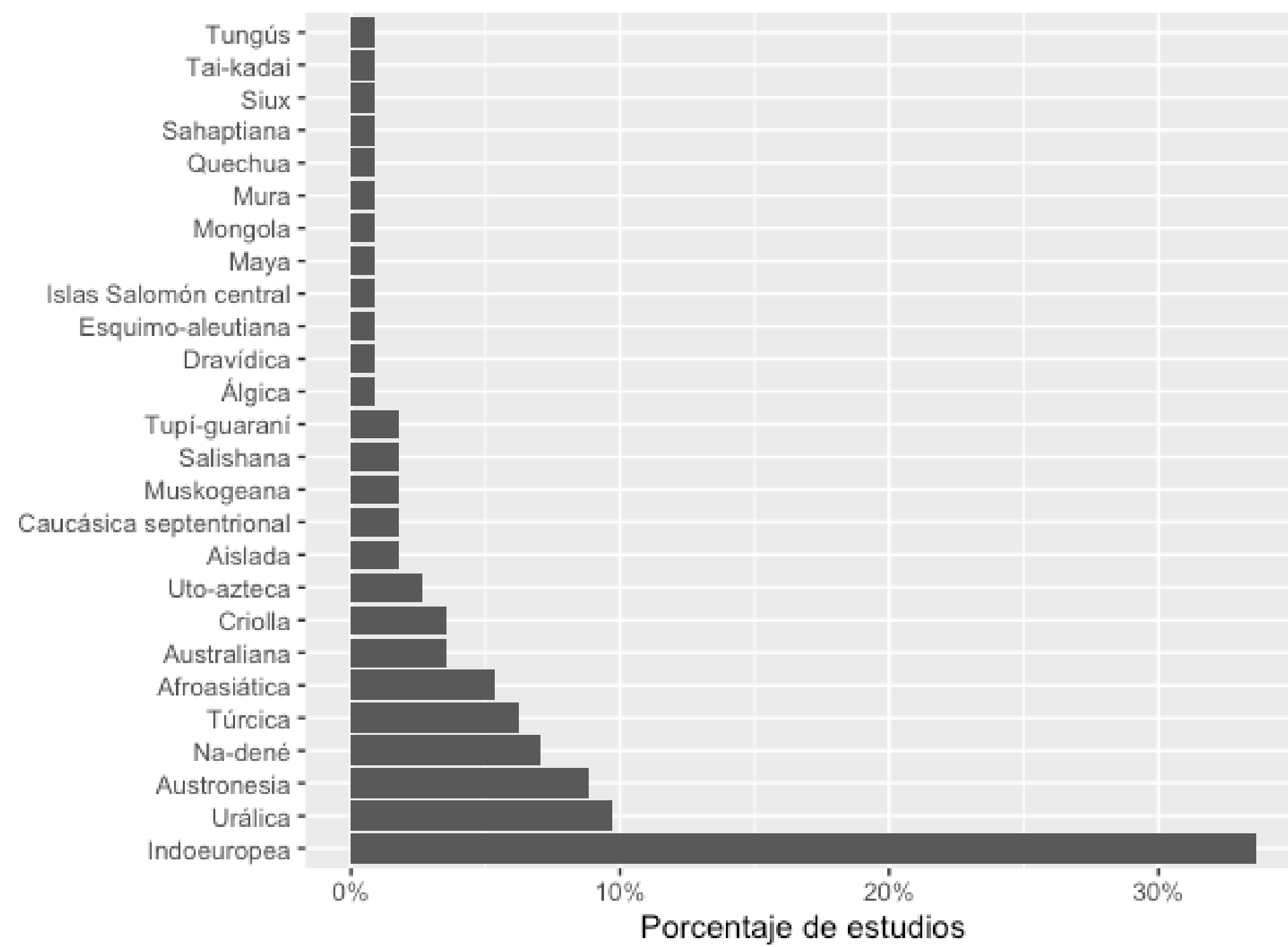
O estudo dos correlatos acústicos do acento lexical

Conforme explicado, projetar um experimento para estabelecer os correlatos acústicos do acento lexical envolve tomar uma série de decisões metodológicas. A seguir, são apresentados alguns dos relacionados aos falantes (seção 2.1) e o corpus utilizado (seção 2.2).

No entanto, antes de entrar em questões específicas, talvez seja conveniente começar com uma visão geral dos principais aspectos metodológicos que afetam o estudo experimental do acento lexical na perspectiva da análise acústica. Para isso, é extremamente útil o trabalho de Roettger e Gordon (2017), no qual são examinadas 86 publicações sobre os correlatos acústicos do acento lexical em 76 línguas pertencentes a 26 famílias linguísticas. O artigo é complementado com uma base de dados *on-line* (GORDON; ROETTGER, 2018) que contém as informações extraídas das publicações revisadas e com um conjunto de *scripts* em R (R Core Team, 2021) para reproduzir os cálculos feitos pelos autores.

Uma primeira pergunta que deve ser feita — pelo menos, do ponto de vista de um linguista — é em quais línguas se baseia o conhecimento atual sobre os parâmetros acústicos que indicam a presença de acento lexical. A Figura 3, com base em dados de Gordon e Roettger (2018), mostra que, como seria de esperar, aproximadamente um terço dos estudos incluídos na revisão de Roettger e Gordon (2017) foram realizados em idiomas que pertencem à família indo-européia.

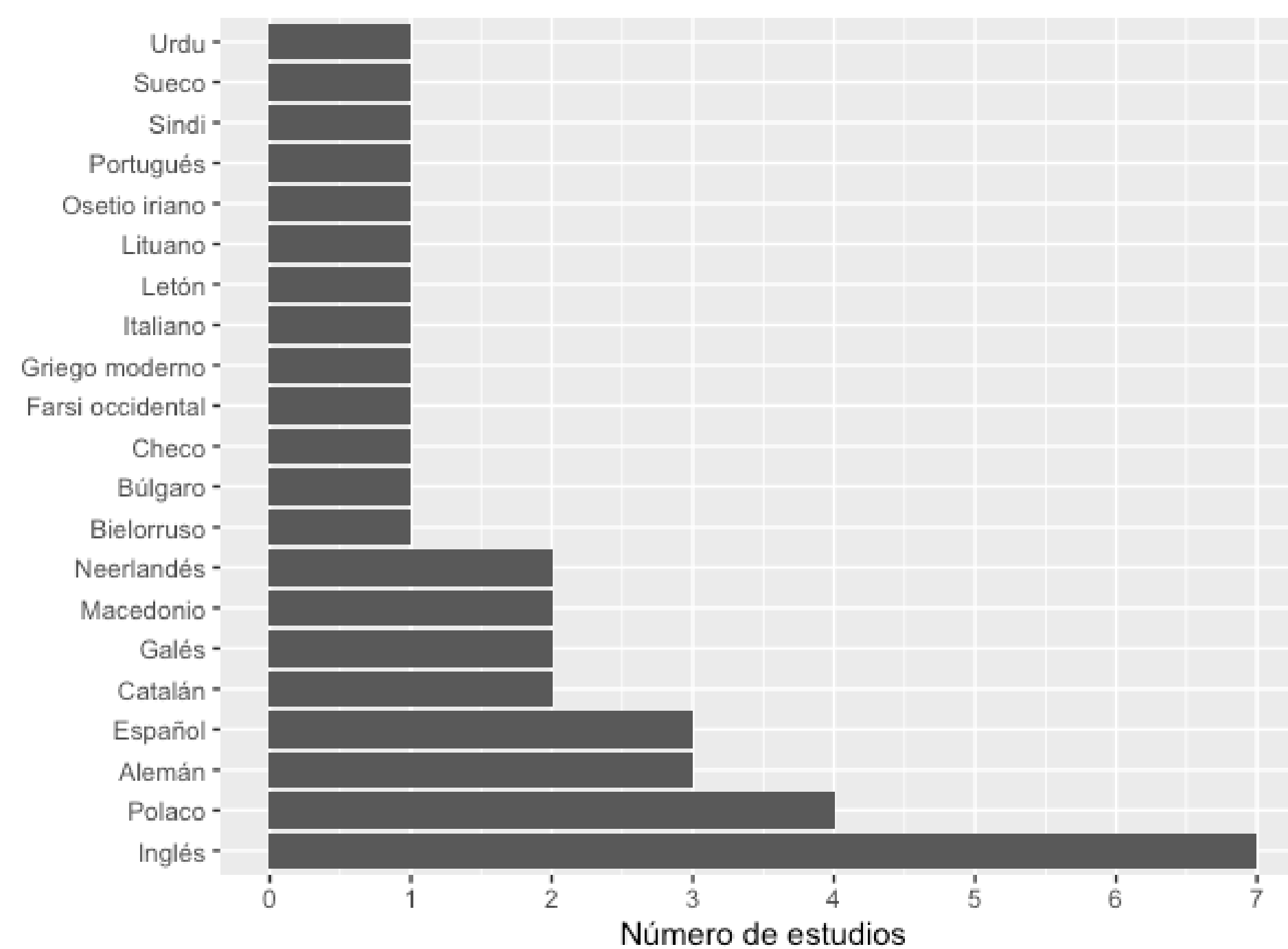
Figura 3: Famílias linguísticas representadas nos estudos sobre os correlatos acústicos do acento lexical analisados por Roettger e Gordon (2017)



Nota. Studies on acoustic correlates of word stress

Também não é de estranhar que nos trabalhos considerados por Roettger e Gordon (2017), o inglês seja a língua a que se dedica o maior número de estudos, como se pode verificar na Figura 4.

Figura 4: Línguas indo-europeias representadas nos estudos sobre os correlatos acústicos do acento lexical analisados por Roettger e Gordon (2017)



Nota. Studies on acoustic correlates of word stress

Para o espanhol, Roettger e Gordon (2017) selecionaram três publicações escritas em inglês, duas das quais também tratam de outros idiomas além do espanhol (ORTEGA-LLEBARIA, 2006; ORTEGA-LLEBARIA; PRIETO, 2011; VOGEL *et al.*, 2016), como acontece na única publicação escolhida em português, uma contribuição em inglês em que o português é comparado ao sueco (BARBOSA *et al.*, 2013a).

Diante dos dados, pode-se pensar que, ao analisar os métodos utilizados nas pesquisas sobre o acento lexical (ROETTGER; GORDON, 2017) e ao avaliar os resultados obtidos (GORDON; ROETTGER, 2017), há uma certa tendência a favor dos trabalhos em inglês e os publicados neste idioma quando se trata de contribuições relacionadas a outros idiomas.

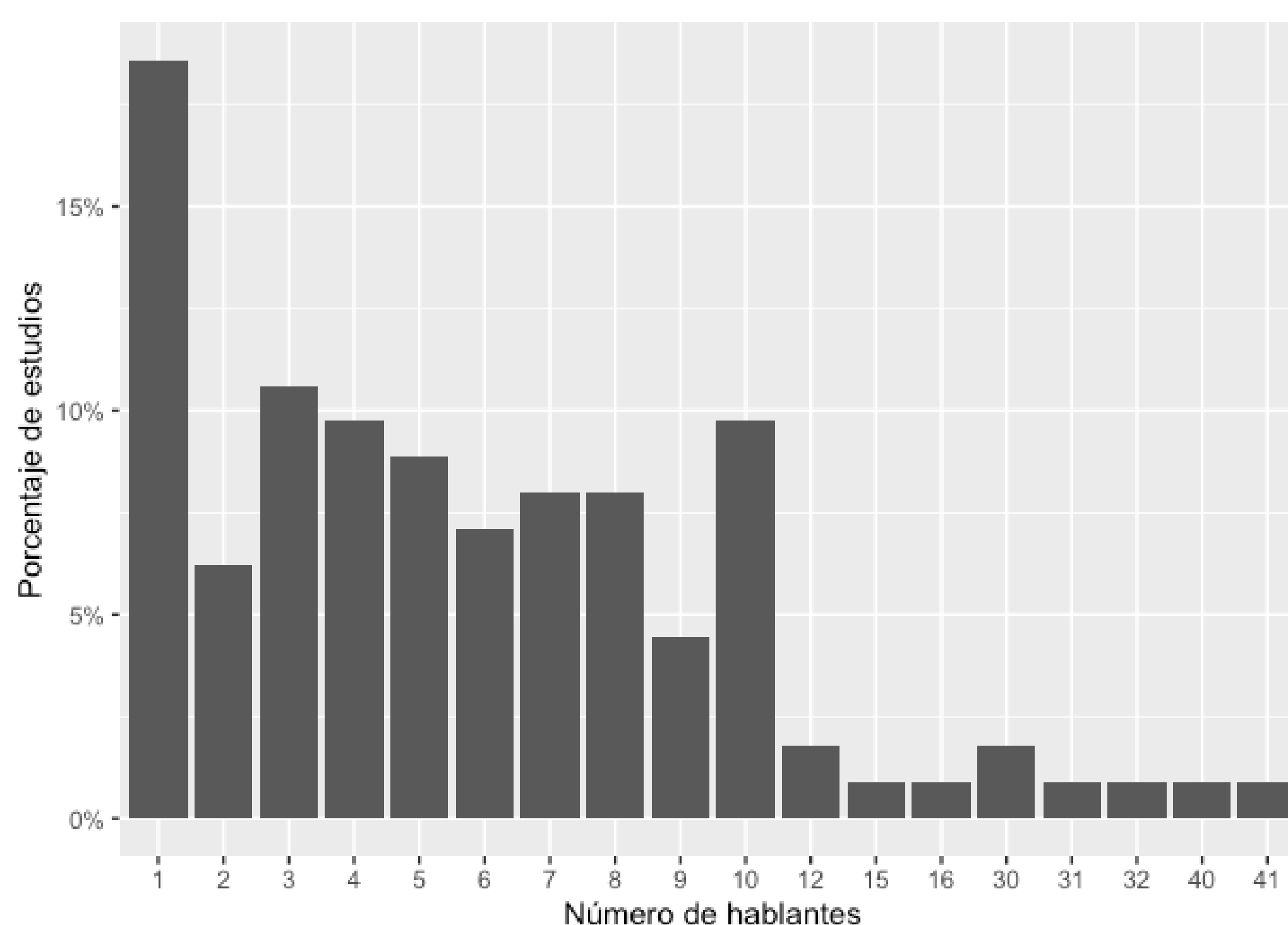
Falantes

A seleção de palestrantes que fornecerão os dados para um experimento não deve ser considerada uma tarefa trivial. Embora possa ser condicionada por fatores de natureza prática como a disponibilidade para participar no estudo, a distância geográfica ou o financiamento disponível se a participação for paga, também é importante ter em conta fatores sobre os quais o investigador tem um certo grau de controle, como o número de falantes, que é tratado na seção 2.1.1 e, muito especialmente, seu perfil linguístico, que é abordado na seção 2.1.2.

2.1.2 Número de falantes

Entre os aspectos metodológicos examinados por Roettger e Gordon (2017), está o número de falantes em que se baseiam os dados coletados nas publicações incluídas em sua revisão. Como salientam os autores, e como também pode ser visto na Figura 5, “A maioria dos estudos emprega dados de um a dez falantes, mas variam muito dentro dessa faixa. Por outro lado, o modelo (20 estudos) é um único falante” (ROETTGER; GORDON, 2017, § 3.2).

Figura 5: Número de falantes em estudos sobre os correlatos acústicos do acento lexical analisados por Roettger e Gordon (2017)



Nota. Gráfico feito a partir das informações coletadas no banco de dados do projeto Studies on acústico correlates of word stress, de MK Gordon e TB Roettger, 18 de abril de 2018, OSF (<https://doi.org/10.17605/OSF.IO/9R2CD>).

Legenda: Eixo Y: Percentual de estudios; Eixo X: Número de falantes.

Como exemplo, o trabalho sobre português selecionado por Roettger e Gordon (2017) é baseado em 10 falantes nativos de português brasileiro e o mesmo número de falantes nativos de sueco (BARBOSA *et al.*, 2013a), enquanto, no caso das três publicações sobre o espanhol escolhido por esses mesmos autores, em um deles há cinco falantes de espanhol peninsular (ORTEGA-LLEBARIA, 2006), um outro com 10 falantes de espanhol peninsular e 10 de catalão (ORTEGA-LLEBARIA; PRIETO, 2011) e, na terceira, são analisados dados de nove falantes de espanhol peninsular, 10 gregos, 8 húngaros e 9 turcos (VOGEL *et al.*, 2016).

Ao tomar decisões sobre o número de falantes, é pertinente lembrar a recomendação de Ladefoged (2003) para o trabalho de campo: “Idealmente você quer cerca de meia dúzia de falantes de cada sexo. Se você conseguir encontrar 12 ou até 20 membros de cada sexo, tanto melhor” (p. 14). Como pode ser observado na figura 5, poucas publicações ultrapassam esse mínimo de 12 falantes proposto pelo Ladefoged.

Perfil linguístico dos falantes

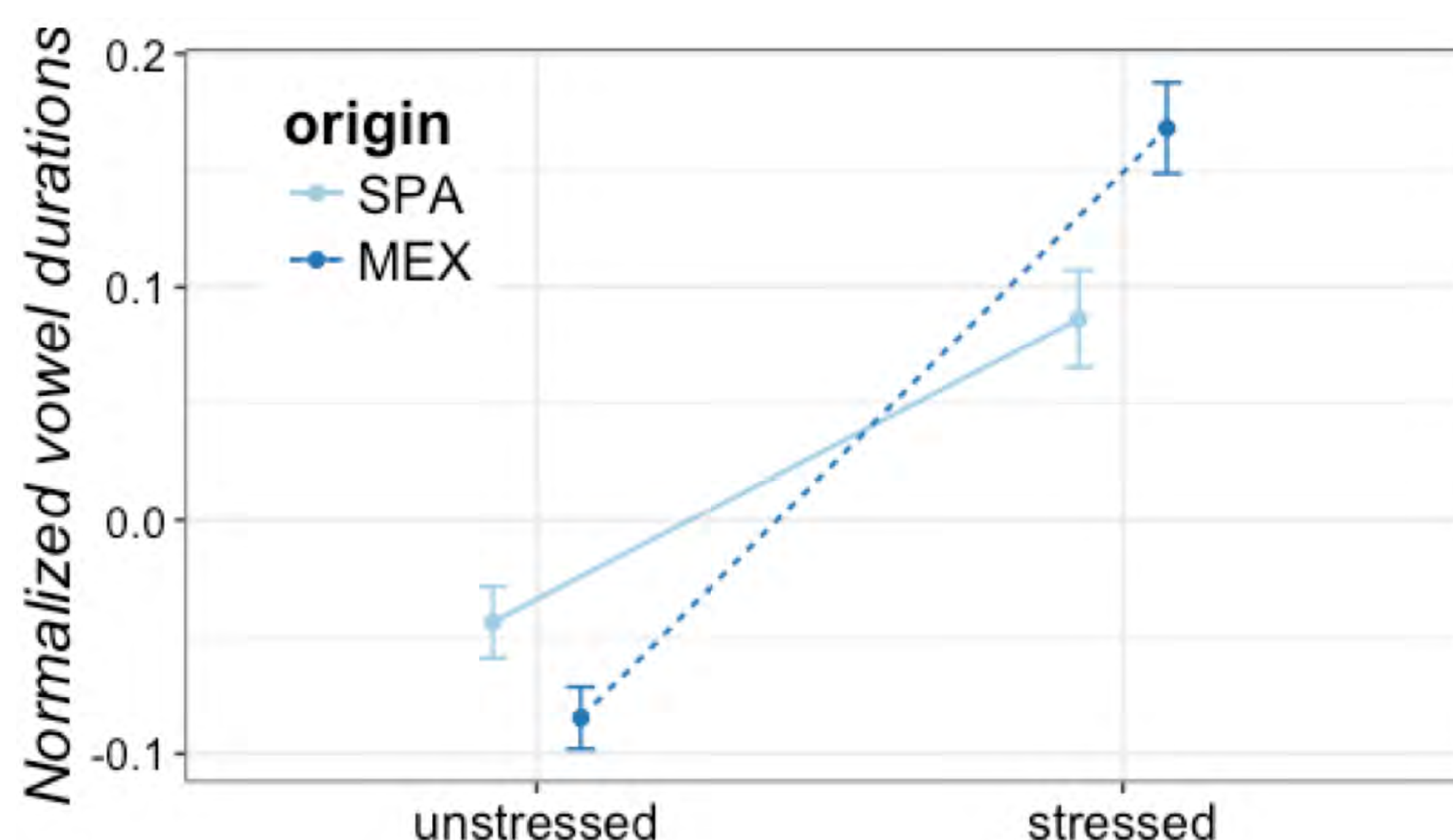
Em estudos fonéticos que visam descrever os correlatos acústicos do acento lexical, é lógico que sejam utilizados falantes monolíngues da língua em que o pesquisador está interessado. Assim, por exemplo, no trabalho de Vogel *et al.* (2016), incluído entre os selecionados por Roettger e Gordon (2017), é a seguinte descrição dos participantes do experimento:

Todos os falantes que participaram de nosso estudo eram estudantes universitários com idades entre 18 e 30 anos, sem relato de problemas de linguagem/audição. Além disso, eles eram monolíngues na medida do possível, embora se entenda que os estudantes universitários tenham estudado pelo menos uma outra língua como parte de sua educação. Para cada idioma, dez falantes de origens de dialeto padrão semelhantes foram registrados em suas universidades locais: húngaro – Debrecen, espanhol – Madri, turco – Istambul e grego – Atenas. (VOGEL et al., 2016, p. 132).¹

Essa citação deixa claro, em primeiro lugar, que dependendo da idade e do nível de escolaridade é muito possível que os falantes escolhidos para um determinado estudo tenham conhecimento de outras línguas. Decidir até que ponto uma pessoa capaz de usar mais de uma língua pode ser descrita como bilíngue ou multilíngue depende muito da definição de bilinguismo ou multilinguismo que o pesquisador adota, questão complexa que será retomada mais adiante. Um segundo aspecto que merece destaque é aquele relacionado à variedade geográfica utilizada pelos falantes, uma vez que alguns dos dados disponíveis sugerem que a manifestação acústica do acento lexical não precisa necessariamente ser homogênea ao longo de um domínio linguístico.

A Figura 6 reproduz os resultados obtidos por Santiago e Mairano (2018) sobre o efeito do acento lexical na duração das vogais em um grupo de 12 estudantes universitários falantes de espanhol na Cidade do México e em outro de 10 falantes de espanhol que estudavam em universidade em Madri. Como pode ser visto, a diferença de duração entre vogais átonas e tônicas é mais acentuada em falantes de espanhol mexicanos do que em falantes peninsulares; especificamente, “os falantes de espanhol reduzem as vogais átonas em 7% em relação às suas contrapartes tônicas, enquanto os falantes de mexicanos as reduzem em 13%” (SANTIAGO; MAIRANO, 2018, p. 455).

Figura 6: Diferenças de duração entre vogais tônicas e átonas no espanhol mexicano e no espanhol peninsular



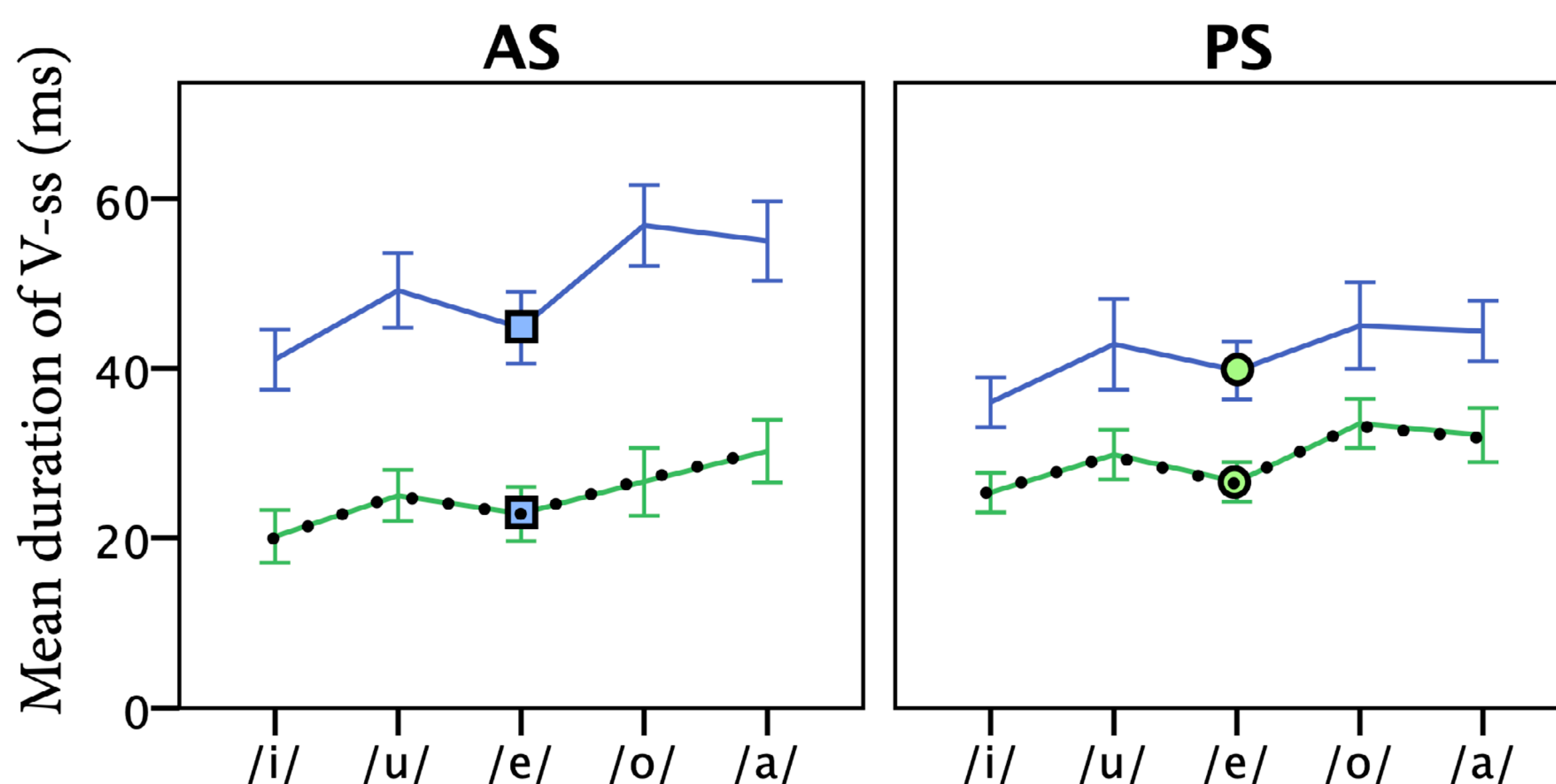
Nota. SPA = Espanhol peninsular; MEX = Espanhol de México. Figura extraída de “The role of lexical stress on vowel duration and vowel space in two varieties of Spanish”, de F. Santiago e P. Mairano, 2018, In: K. Klessa, J. Bachan, A. Wagner, M. Karpiński y D. Śledziński (Eds.), *9th International Conference on Speech Prosody 2018*, p. 455, ISCA Archive (<https://doi.org/10.21437/SpeechProsody.2018-92>).

Legenda: Eixo Y: Durações de vogais normalizadas, Origem; Eixo X: Átonas, Tônicas.

¹ Traduzido do original em inglês.

Um fenômeno semelhante é encontrado ao comparar a duração das vogais em função do acento lexical em cinco falantes de espanhol de Lima e cinco falantes de espanhol nascidos ou educados em Buenos Aires estudados por Kim (2011). A Figura 7 mostra que a diferença de duração entre vogais tônicas e átonas é menor na variedade peruana do que na argentina.

Figura 7: Diferenças de duração entre vogais tônicas e átonas no espanhol argentino e no espanhol peruano



Nota. V-ss = Parte estável da vogal; AS = Falantes de espanhol da Argentina; PS = Falantes de espanhol do Perú. A linha contínua corresponde às vogais tônicas e a linha pontilhada às átonas. Figura retirada de *The phonetics of stress manifestation: Segmental variation, syllable constituency and rhythm*, de M. Kim, 2011, Tese de doutorado, Stony Brook University, p. 121, SUNY Digital Repository (<http://hdl.handle.net/1951/56031>).

Esses dois exemplos sugerem que a variante geográfica constitui uma variável pertinente às características dos falantes monolíngues, portanto, sua não consideração pode afetar a homogeneidade do grupo de falantes analisados em um experimento.

Para estudar os correlatos acústicos do acento lexical, falantes bilíngues também têm sido utilizados. Como já foi sugerido, este termo pode abranger perfis linguísticos muito diversos; por exemplo, Hoffmann (1991, p. 16-17) coleta até 15 casos de uso de duas línguas em que se poderia falar de bilinguismo dependendo de como o termo é definido. Ainda como exemplo, na descrição dos participantes do trabalho de Ortega-Llebaria e Prieto (2011), um daqueles que Roettger e Gordon (2017) incluíram em sua resenha, lê-se o seguinte:

Dez falantes nativos de espanhol castelhano (7 mulheres e 3 homens) e 10 falantes nativos de catalão central (10 mulheres) participaram do experimento. As idades variavam de 20 a 32 anos. Todos os sujeitos tinham diplomas universitários e falavam uma variedade educada de sua língua nativa, que usavam regularmente com seus pais e irmãos. Os dez sujeitos catalães aprenderam espanhol como segunda língua na escola aos cinco anos ou mais tarde, enquanto apenas cinco dos sujeitos espanhóis também sabiam catalão como segunda língua, que aprenderam na escola secundária, ou seja, aos dez anos ou mais (ORTEGA-LLEBARIA; PRIETO, 2011, p. 80).²

² Traduzido do original em inglês.

Dos 15 falantes que podem ser definidos como bilíngues, 10 aprenderam a segunda língua a partir dos cinco anos e cinco começaram a aprendê-la a partir dos 10 anos, o que introduz certa assimetria entre os dois grupos. Nestas situações, é útil recorrer ao conceito de ‘dominância linguística’ (TREFFERS-DALLER, 2019), que, com a informação disponível, permitiria caracterizar 10 falantes como bilíngues catalão-espanhol, com catalão como língua dominante, e cinco falantes como espanhol bilíngue-Catalão, cuja língua dominante é o espanhol.

Ao trabalhar com bilíngues, é conveniente levar em consideração, além da dominância dos falantes, as diferenças entre os idiomas que podem afetar o fenômeno estudado. Assim, se for analisado o efeito do acento lexical sobre os valores de frequência dos formantes vocálicos, o fato de ocorrer um processo de redução vocálica no catalão central, ausente no espanhol, poderia afetar as realizações átonas em espanhol de um bilíngue com uma forte dominância catalã (BADIA I MARGARIT, 1975, p. 150), uma vez que no catalão central /e/, /ɛ/ e /a/ são realizados como [ə] na posição átona e /o/ e /ɔ/ são realizados como [u] nessa posição (JULIÀ, 2005, p. 160).

Como um todo, o termo ‘falante bilíngue’ abrange uma ampla gama de perfis e histórias linguísticas, de modo que a dominância dos falantes, juntamente com outros fatores, é uma variável que deve ser considerada se o objetivo é ter um grupo relativamente homogêneo de participantes em um experimento. Por esta razão, várias ferramentas foram desenvolvidas para caracterizar com algum detalhe o perfil de um falante bilíngue. Um deles é o Bilingual Language Profile (BLP) [Perfil de Bilinguismo Linguístico], um questionário de autoavaliação composto por 19 questões que podem ser respondidas em papel ou *on-line* – com pontuação automática neste último caso – e que possui versões para 15 pares de idiomas (BIRDSONG *et al.*, 2012; GERTKEN *et al.*, 2014).

Uma terceira categoria de falantes que foi contabilizada para a análise dos correlatos acústicos do acento lexical corresponde àquela das pessoas que aprendem ou adquirem uma segunda língua ou uma língua estrangeira (L2/LE). Por exemplo, no trabalho de Schwab (2014, p. 80³) sobre a produção de acento lexical em francês falado por alunos cuja primeira língua é o espanhol, os participantes são descritos da seguinte forma:

[...] oito alunos nativos de espanhol avançados de francês (B2–C1) (3 homens e 5 mulheres), morando em Genebra, com pelo menos 2 anos passados em um país de língua francesa. Os falantes de espanhol eram da Península Ibérica (Aragão, La Rioja, Galiza, La Rioja, Extremadura e Andaluzia) e o francês era a segunda língua estrangeira depois do inglês.

Como se vê, embora muitas vezes se faça referência a uma L2, em pesquisas com alunos que aprendem uma língua diferente da sua primeira língua e essa língua não é o inglês, é muito provável que a L2 seja na verdade uma terceira língua (L3).

³ Traduzido do original em inglês.

Também é interessante observar as características dos participantes em uma análise da produção de sotaque em francês e espanhol por estudantes canadenses anglófonos fornecida por Colantoni et al. (2014) e que está reproduzido na Figura 8.

Figura 8: Características dos participantes de um estudo sobre a aquisição de sotaques em francês e espanhol por estudantes anglófonos canadenses

	French		Spanish	
	Learners	Native Speakers	Learners	Native Speakers
n	9	4	8	3
Gender (M;F)	2;7	1;3	4;4	2;1
Mean (range) Age at Acquisition onset	6 (5-12)	-	17 (11-26)	-
Mean (range) Immersion (months)	23 (19-40)	-	25 (18-29)	-
Immersion (Months)	19 (0-150)	-	12 (0-66)	-

Nota. Tabela extraída de “Learning to realize prosodic prominence in French and Spanish”, de L. Colantoni, O. Marasco, J. Steele e S. Sunara, 2014, In: R. T. Miller, K. I. Martin, C. M. Eddington, A. Henery, N. Marcos, A. M. Tseng, A. Tuninetti y D. Walter (Eds.), *Selected Proceedings of the 2012 Second Language Research Forum: Building bridges between disciplines*, p. 18, Cascadia Proceedings Project (<https://bit.ly/3xnunHb>).

Legenda: Linha superior: Francês, Espanhol; Aprendizes, Falantes nativos. Primeira coluna: n (número), Gênero (M/F), Média (intervalo) Idade do início da aquisição, Média (intervalo) Imersão (meses), Imersão (meses)

A especificidade do sistema educacional canadense explica as diferenças na idade em que os alunos franceses e espanhóis começaram a aprender essas línguas, bem como os diferentes tempos de imersão (COLANTONI et al., 2014, p. 18), dois fatores que tornam os dois grupos de participantes do experimento não exatamente idênticos.

Tal como acontece com os falantes bilíngues, o perfil linguístico das pessoas que aprendem ou adquirem uma L2/LE pode ser muito diversificado e influenciado por diversos fatores; entre as mais relevantes vale destacar a idade em que a língua começa a ser aprendida, o tempo de estudo ou contato com a L2/LE, o contexto em que a língua foi aprendida e as variantes geográficas da L2/LE a que o aluno foi exposto. Por esta razão, muitas vezes é conveniente fazer uma avaliação do nível de competência linguística – e, idealmente, competência fonética – dos participantes de um experimento.

Uma primeira possibilidade é simplesmente considerar que o nível de conhecimento da L2/LE é dado pelo número de anos de estudo da língua ou pelo estágio educacional em que o aluno se encontra; isso implica supor que, por exemplo, todos os alunos do primeiro ano de uma licenciatura em espanhol em uma determinada universidade em um país que não é de língua espanhola têm o mesmo nível de competência, o que, obviamente, nem sempre corresponde totalmente à realidade.

A segunda opção, como no já citado trabalho de Schwab (2014) e em muitos outros, consiste em determinar o grau de competência em relação a um sistema de níveis globalmente aceito. No contexto europeu, é comum o uso do Quadro Europeu Comum de Referência (CEFR), no qual são estabelecidas três etapas, cada uma delas dividida em dois níveis: usuário básico - A1 (Inicial) e A2 (Elementar) -, usuário independente - B1 (Intermediário) e B2 (Pós-Intermediário) - e usuário proficiente - C1 (Domínio operacional efetivo) e C2 (Domínio Pleno). Para obter a certificação correspondente a cada nível, é necessário passar por um teste padronizado (CONSELHO DA EUROPA, 2002). No entanto, como bem apontam Norris e Ortega (2003), esse tipo de teste não foi inicialmente concebido para responder às necessidades dos pesquisadores da área de aquisição de segunda língua:

Por exemplo, Shohamy (1994) observou que testes destinados à tomada de decisão educacional, como o Teste de Inglês como Língua Estrangeira (TOEFL), são frequentemente usados por pesquisadores de SLA como medidas de aprendizagem ou aquisição, embora tais testes tenham sido concebidos como indicadores de habilidades globais de linguagem acadêmica. Da mesma forma, medidas holísticas de proficiência, como as Diretrizes do ACTFL (1986) e procedimentos relacionados, podem ser usadas como base para atribuir aos alunos condições de pesquisa instrucional, mesmo que as pontuações em tais medidas não tenham nada a ver com os formulários específicos da L2 ou habilidades investigadas (NORRIS; ORTEGA, 2003, p. 739).⁴

Conseqüentemente, pode acontecer que o nível refletido nessas avaliações não coincida com o conhecimento real do componente linguístico — o componente fônico, no caso do acento lexical — pelo qual o pesquisador está interessado.

Uma terceira via seria a explicada no estudo de Colantoni et al. (2014) mencionado acima: “A proficiência dos alunos é baseada na sua própria avaliação auto-referida confirmada pelos autores durante a análise dos dados” (COLANTONI et al., 2014, p. 18, n. 4)⁵. Nesse caso, o grau de conhecimento da L2/LE é determinado por meio de um questionário de autoavaliação, desenvolvido pelos autores de cada estudo, ou por meio de instrumentos existentes. Quando esta técnica é utilizada, é o participante do experimento que fornece as informações iniciais sobre como ele percebe seu nível de competência e uso da L2/LE.

Entre os questionários de autoavaliação de proficiência linguística voltados especificamente para falantes que conhecem mais de um idioma, merecem destaque o Language History Questionnaire (LHQ) e o Language Experience and Proficiency Questionnaire (LEAP-Q). A primeira delas, descrita em sua última versão no trabalho de Li et al. (2020), é composta por 24 questões e possui traduções para 14 idiomas; se o questionário online (The Brain, Language, and Computation Lab, s.d.) for usado,

⁴ Traduzido do original em inglês.

⁵ Traduzido do original em inglês.

a pontuação será calculada automaticamente. O LEAP-Q, para o qual estão disponíveis questionários adaptados em 26 idiomas, inclui nove perguntas gerais e sete perguntas específicas para cada um dos idiomas avaliados (KAUSHANSKAYA et al., 2020); pode ser respondido em documento Word com possibilidade de exportação dos dados para Excel e também está disponível on-line (Grupo de Pesquisa em Bilinguismo e Psicolinguística, s. f.).

Os três procedimentos que acabamos de descrever permitem caracterizar, cada um com as suas vantagens e limitações, o conhecimento geral da L2/LE, mas, muitas vezes, nem todos os membros de um grupo de participantes numa experiência com o mesmo pressuposto grau de competência — por estarem no mesmo estágio do sistema educacional, por possuírem o mesmo certificado de nível, ou por terem respondido de forma semelhante a um questionário de autoavaliação— necessariamente possuem uma competência fônica homogênea. Por isso, em estudos fonéticos como os relacionados ao acento lexical, pode ser aconselhável realizar também uma avaliação do grau de sotaque estrangeiro.

Nesse caso, a dimensão avaliada é o que se conhece como sotaque (*accent*, em inglês). Para Munro e Derwing (1995), dois dos autores que, sem dúvida, exerceram maior influência na caracterização do sotaque não nativo, “grau de sotaque refere-se a quão forte é percebido o sotaque estrangeiro do locutor”⁶ (MUNRO; DERWING, 1995, p. 291). As respostas dos avaliadores podem ser coletadas por meio de uma escala conhecida como VAS (Visual Analogue Scale [Escala Visual Analógica]) na qual, conforme mostra a figura 9, o grau de sotaque é estabelecido deslizando o cursor sobre uma linha contínua.

Figura 9: Avaliação do grau de sotaque estrangeiro usando uma escala VAS (Visual Analogue Scale [Escala Visual Analógica])

15. Ahora escucha con atención las grabaciones siguientes y evalúa cómo de fuerte es el acento, si es que lo tiene. *

NO TIENE ACENTO TIENE MUCHISIMO ACENTO

Lectura	<input type="range"/>
Narración	<input type="range"/>

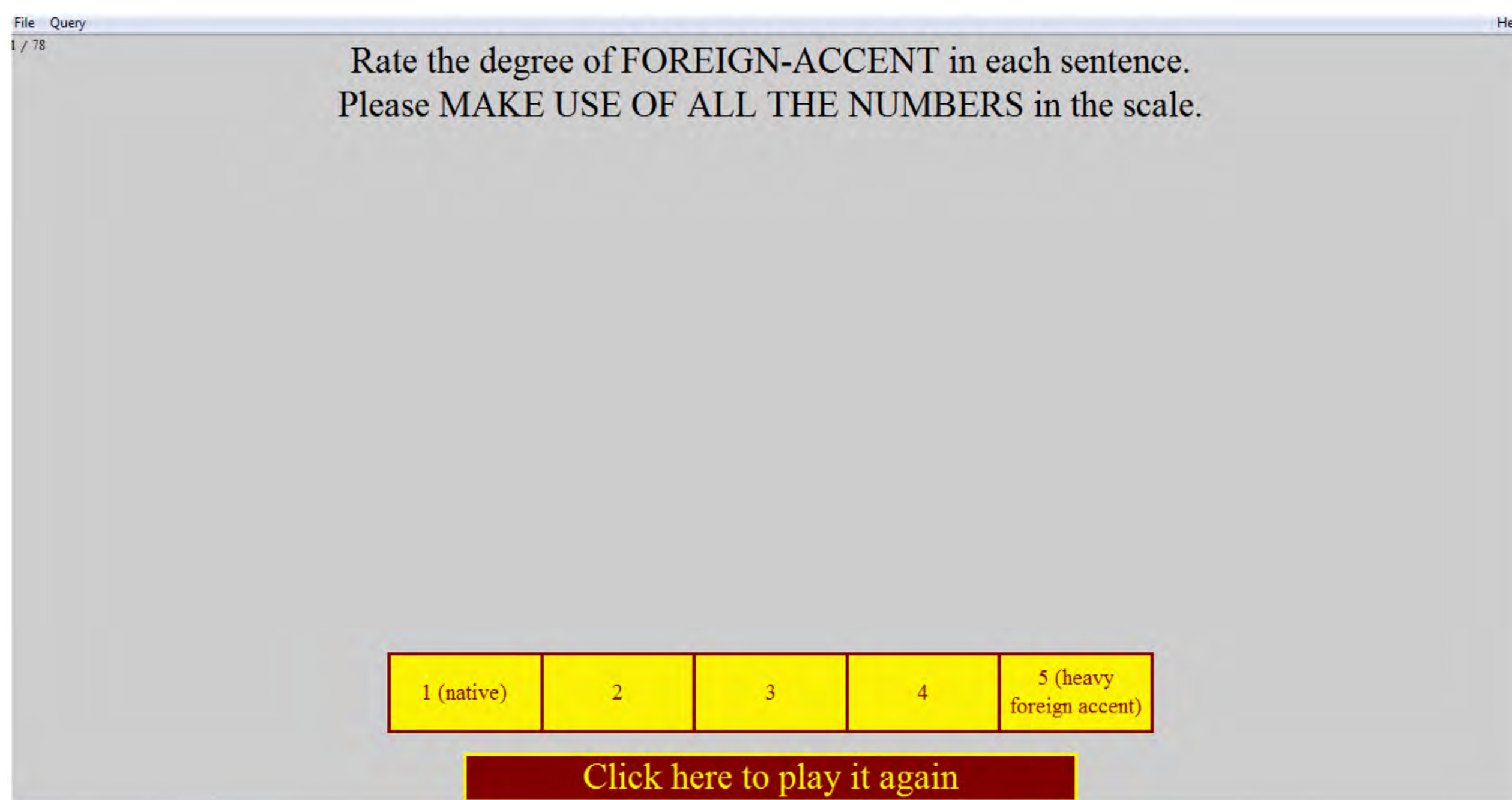
Nota. Figura extraída de “The contribution of prosody to foreign accent: A study of Spanish as a foreign language”, de C. C. Silva e P. Barbosa, 2017, *Loquens*, 4(2), artigo e041, p. 6 (<http://dx.doi.org/10.3989/loquens.2017.041>).

Legenda: Não tem sotaque/tem muito sotaque; Leitura; Narração.

⁶ Traduzido do original em inglês.

Também podem ser utilizadas escalas Likert, em que a pontuação é dividida em categorias discretas, como pode ser observado na figura 10, do estudo de Avello (2018), em que foi utilizada uma escala de 5 pontos, embora também seja de uso comum classificações de 1 a 7 ou de 1 a 9 (ISBELL, 2018, p. 89). É interessante mencionar que o Praat, um programa de domínio público com ampla base de usuários (BOERSMA; WEENINK, 2021), permite que experimentos sejam projetados usando esse tipo de escala.

Figura 10: Avaliação do grau de sotaque estrangeiro usando uma escala Likert



Nota. Figura extraída de “Assessing learners’ changes in foreign accent during Study Abroad”, de P. Avello, 2018, *In*: C. Pérez Vidal, S. López Serrano, J. Ament e D. J. Thomas-Wilhelm (Eds.), *Learning context effects: Study abroad, formal instruction and international immersion classrooms*, p. 142, Language Science Press (<https://doi.org/10.5281/zenodo.1300630>).

Legenda: Intervalo de grau de SOTAQUE-ESTRANGEIRO em cada sentença./ Por favor, FAÇA USO DOS NÚMEROS na escala./ 1(nativo), 2, 3, 4, 5 (sotaque estrangeiro forte)/ Clique aqui para tocar novamente.

Corpus

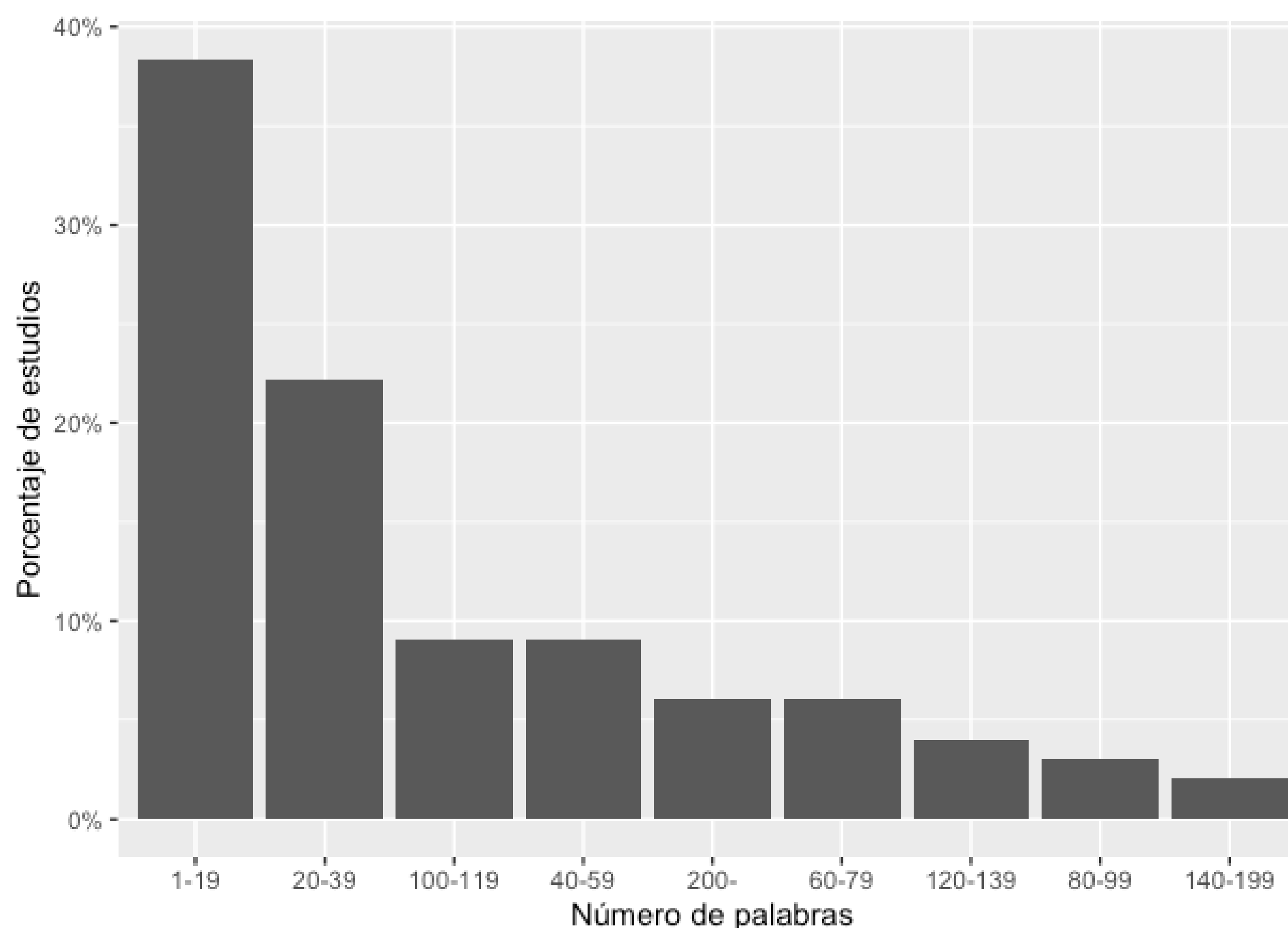
Uma segunda área em que surgem problemas de natureza metodológica quando se estudam experimentalmente os correlatos acústicos do acento lexical diz respeito ao desenho do corpus que será utilizado para obter as informações necessárias. Primeiramente, são apresentados alguns dados sobre o número de casos incluídos nas análises (seção 2.2.1) e, em seguida, são discutidos alguns aspectos relacionados ao estilo de fala considerado na pesquisa (seção 2.2.2).

Número de casos

Na revisão das publicações sobre os correlatos acústicos do acento lexical realizada por Roettger e Gordon (2017), aborda-se especificamente o número de casos analisados nos estudos selecionados pelos autores. Os dados sobre o número de palavras diferentes presentes no corpus utilizados em cada estudo que são recolhidos na figura 11 mostram que “a maioria dos estudos do nosso corpus

utiliza entre 1 e 40 itens lexicais diferentes. Dependendo de como um item lexical foi definido (tipo de palavra ou forma de palavra) e o corpus (fala de laboratório versus fala espontânea), alguns estudos analisam muitas centenas de palavras” (ROETTGER; GORDON, 2017, § 3.2)⁷.

Figura 11: Número de palavras do corpus utilizadas nos estudos sobre os correlatos acústicos do acento lexical analisados por Roettger e Gordon (2017)



Nota. Gráfico elaborado a partir das informações coletadas no banco de dados do projeto *Studies on acoustic correlates of word stress*, de M. K. Gordon e T. B. Roettger, 18 de abril de 2018, OSF (<https://doi.org/10.17605/OSF.IO/9R2CD>).

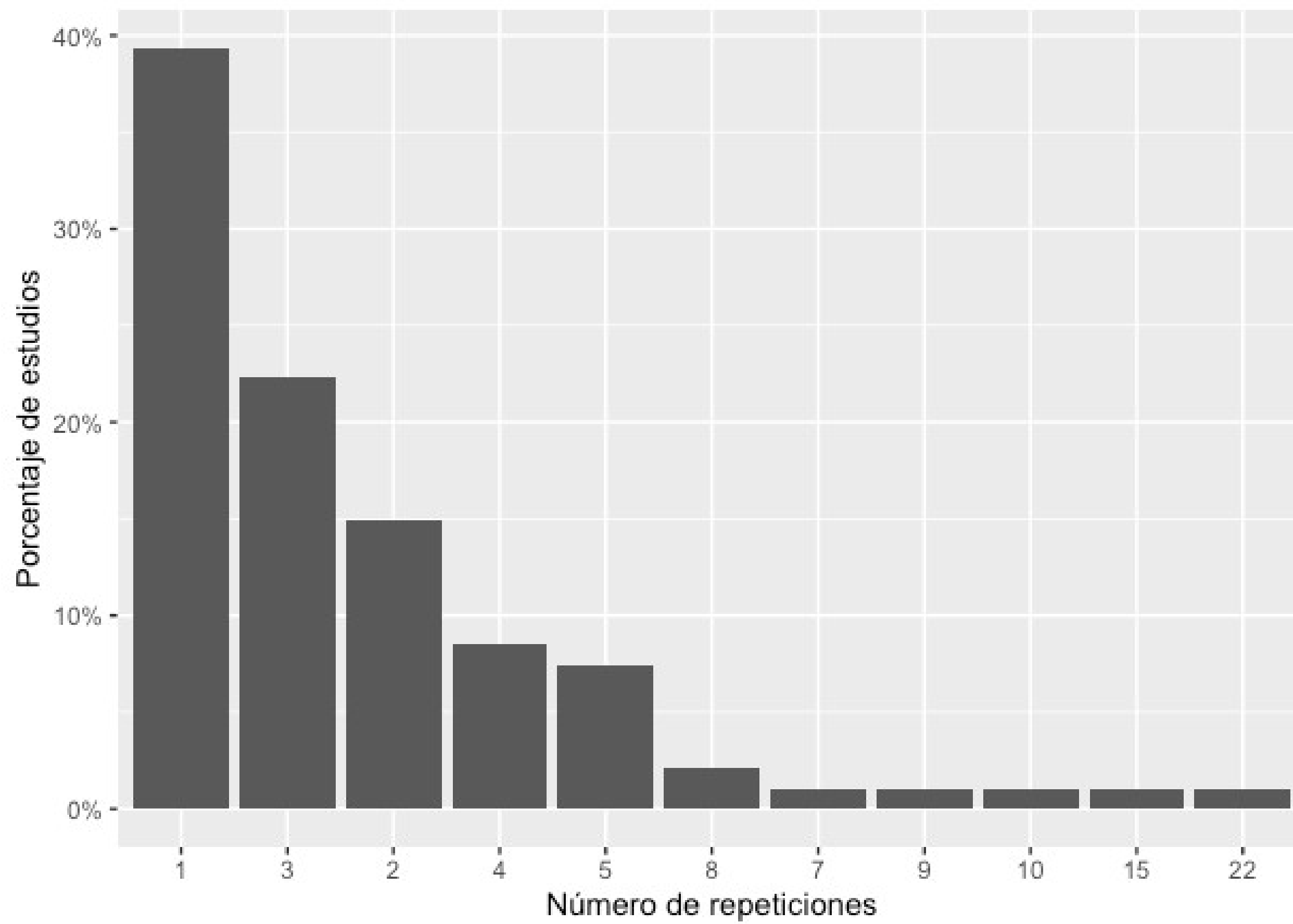
Legenda: Eixo Y: Percentuais de estudos. Eixo X: Número de palavras.

Embora uma forma comum de ter um maior volume de dados seja obter várias repetições da mesma palavra, na análise realizada por Roettger e Gordon (2017) observa-se que “a maioria dos estudos utiliza entre 1 e 6 repetições com uma única repetição sendo o mais comum”⁸ (ROETTGER; GORDON, 2017, § 3.2), conforme mostrado na Figura 12.

⁷ Traduzido do original em inglês.

⁸ Traduzido do original em inglês.

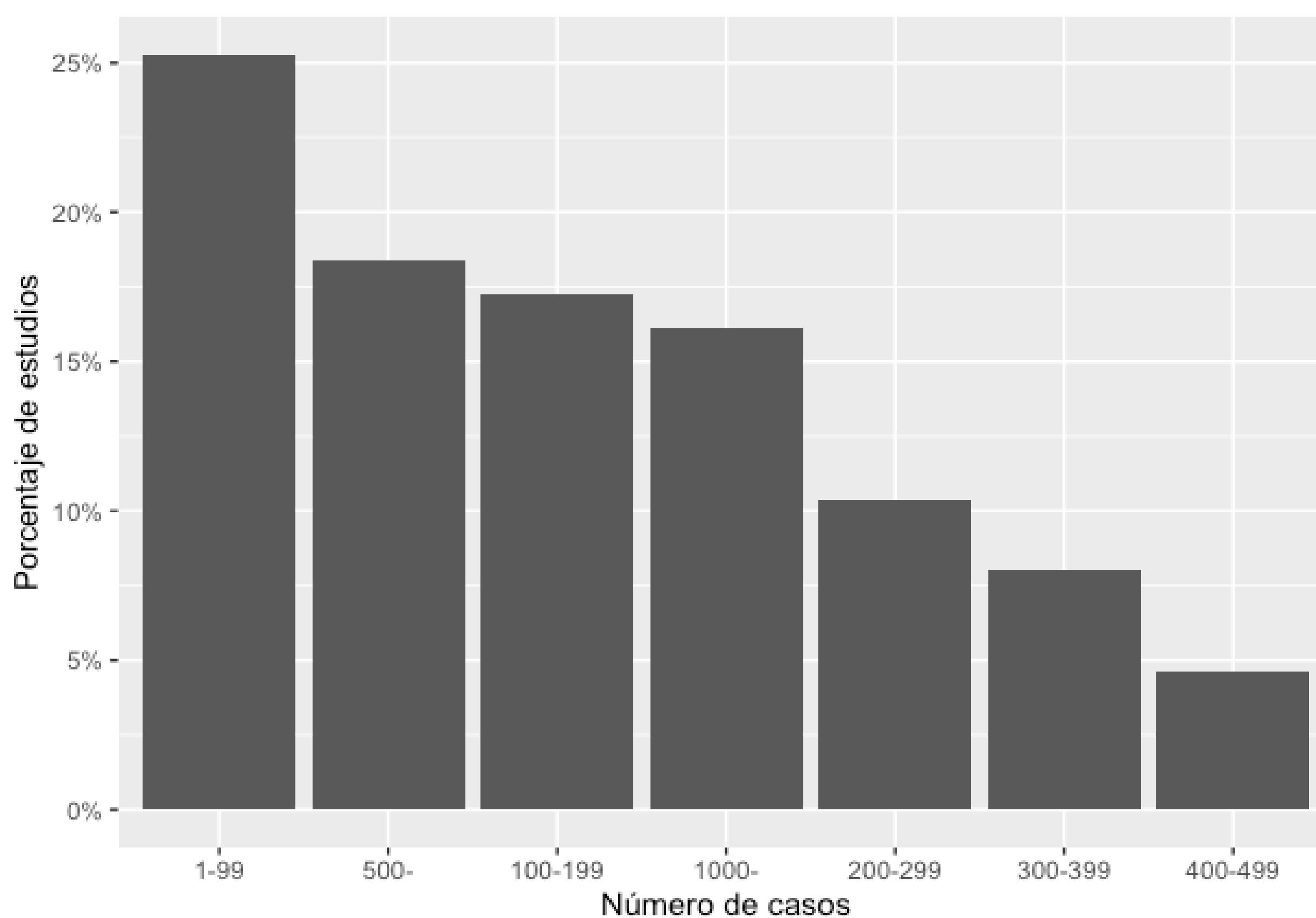
Figura 12: Número de repetições de cada item do corpus nos estudos sobre os correlatos acústicos do acento lexical analisados por Roettger e Gordon (2017)



Nota. Gráfico feito a partir das informações coletadas no banco de dados do projeto *Studies on acústico correlates of word stress*, de MK Gordon e TB Roettger, 18 de abril de 2018, OSF (<https://doi.org/10.17605/OSF.IO/9R2CD>).
Legenda: Eixo Y: Percentuais de estudos. Eixo X: Número de repetições.

Multiplicando o número de palavras por suas repetições e pelo número de falantes, pode-se conhecer o número total de casos analisados nos estudos sobre os correlatos acústicos do acento lexical considerados por Roettger e Gordon (2017). Assim, a Figura 13 mostra que a situação mais frequente (em 25% dos trabalhos) é que o número de casos analisados é inferior a 100.

Figura 13: Número total de casos analisados nos estudos sobre os correlatos acústicos do acento lexical coletados em Roettger e Gordon (2017)



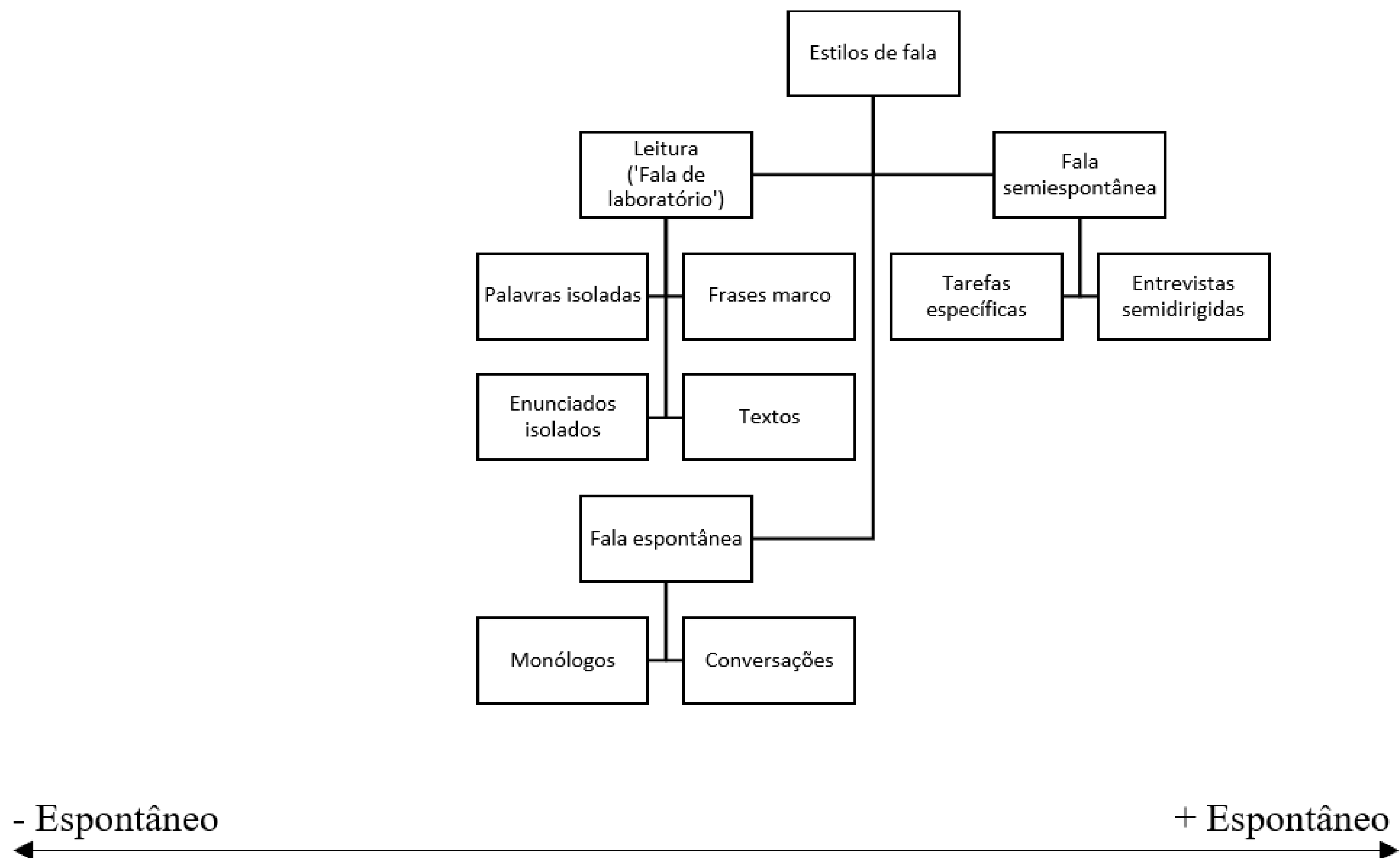
Nota. Gráfico feito a partir das informações coletadas no banco de dados do projeto Studies on acústico correlates of word stress, de MK Gordon e TB Roettger, 18 de abril de 2018, OSF (<https://doi.org/10.17605/OSF.IO/9R2CD>).

Em resumo, de acordo com as informações fornecidas por Roettger e Gordon (2017) e coletadas em Gordon e Roettger (2018), nas publicações selecionadas por esses autores, o estudo dos correlatos acústicos do acento lexical é realizado, em sua maioria, utilizando entre 0 e 40 palavras diferentes, com um número de repetições de cada palavra que varia entre uma e seis e com um total de casos que em um quarto dos trabalhos publicados é inferior a 100.

2.3.2 Estilos de fala

Um segundo aspecto que deve ser refletido em relação ao corpus é o tipo de material nele incluído, que está relacionado à escolha dos estilos de fala que serão levados em consideração em uma investigação. A Figura 14 mostra algumas das possibilidades em termos de materiais, organizados em um *continuum* desde o estilo menos espontâneo — ler palavras isoladas — até o mais espontâneo — falar — que reflete amplamente o grau de planejamento do discurso pelo falante ou, em termos labovianos, o grau de atenção que o falante presta à sua produção linguística (LABOV, 1972/1978).

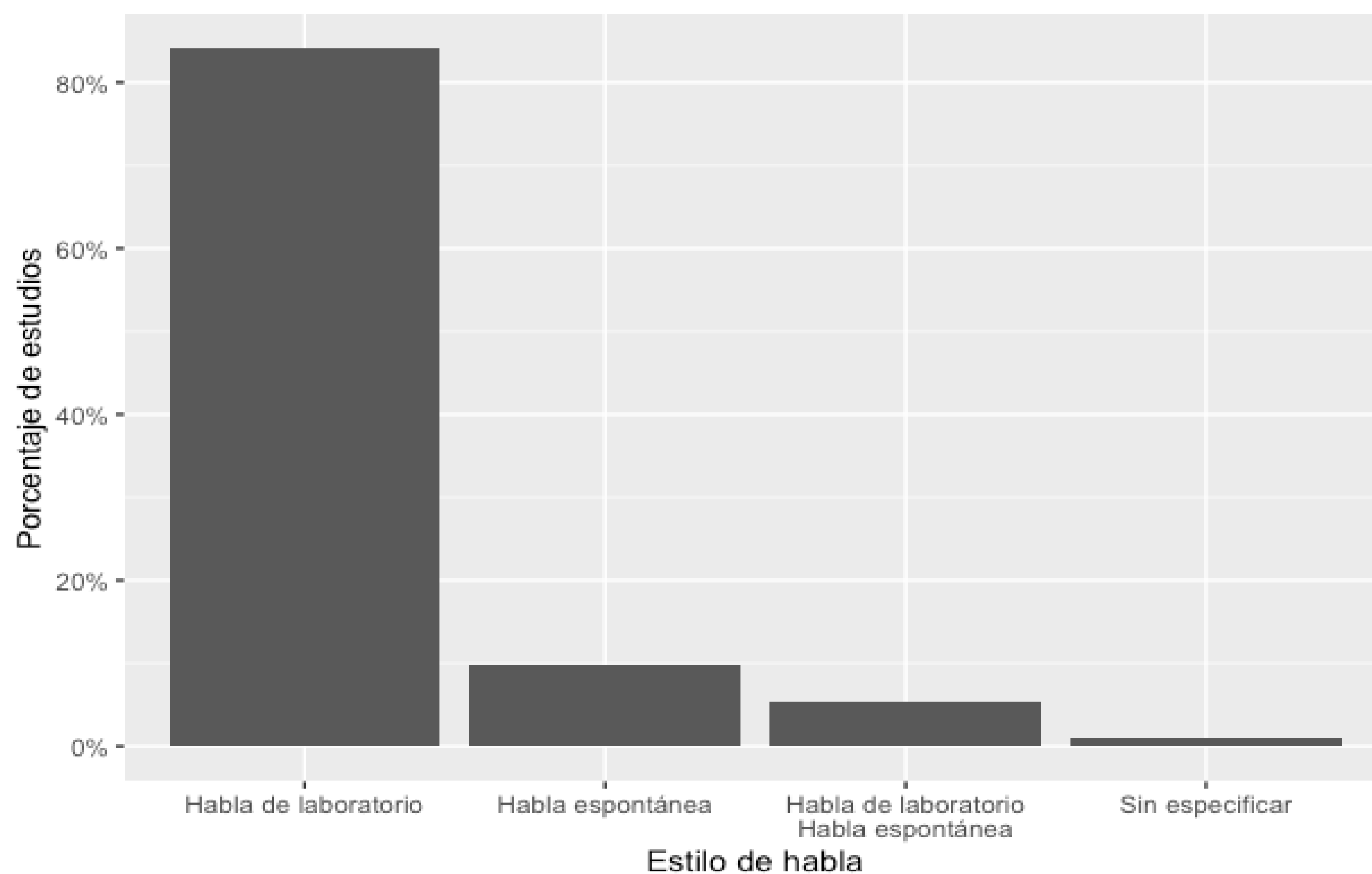
Figura 14: Classificação de materiais que podem fazer parte de um corpus com base em estilos de fala



Legenda: De cima para baixo: Estilos de Fala; Leitura ('fala de laboratório'), Fala semiespontânea, Fala espontânea; Palavras isoladas, Frases-veículo, Enunciados isolados, Textos, Tarefas específicas, entrevistas semidirigidas, monólogos, conversações.

As informações coletadas por Roettger e Gordon (2017) mostram que, como pode ser visto na figura 15, em mais de 80% dos estudos sobre os correlatos acústicos do acento lexical, analisa-se o que normalmente é definido como 'fala', ou seja, a leitura de um corpus controlado e previamente planejado composto por palavras, frases ou textos.

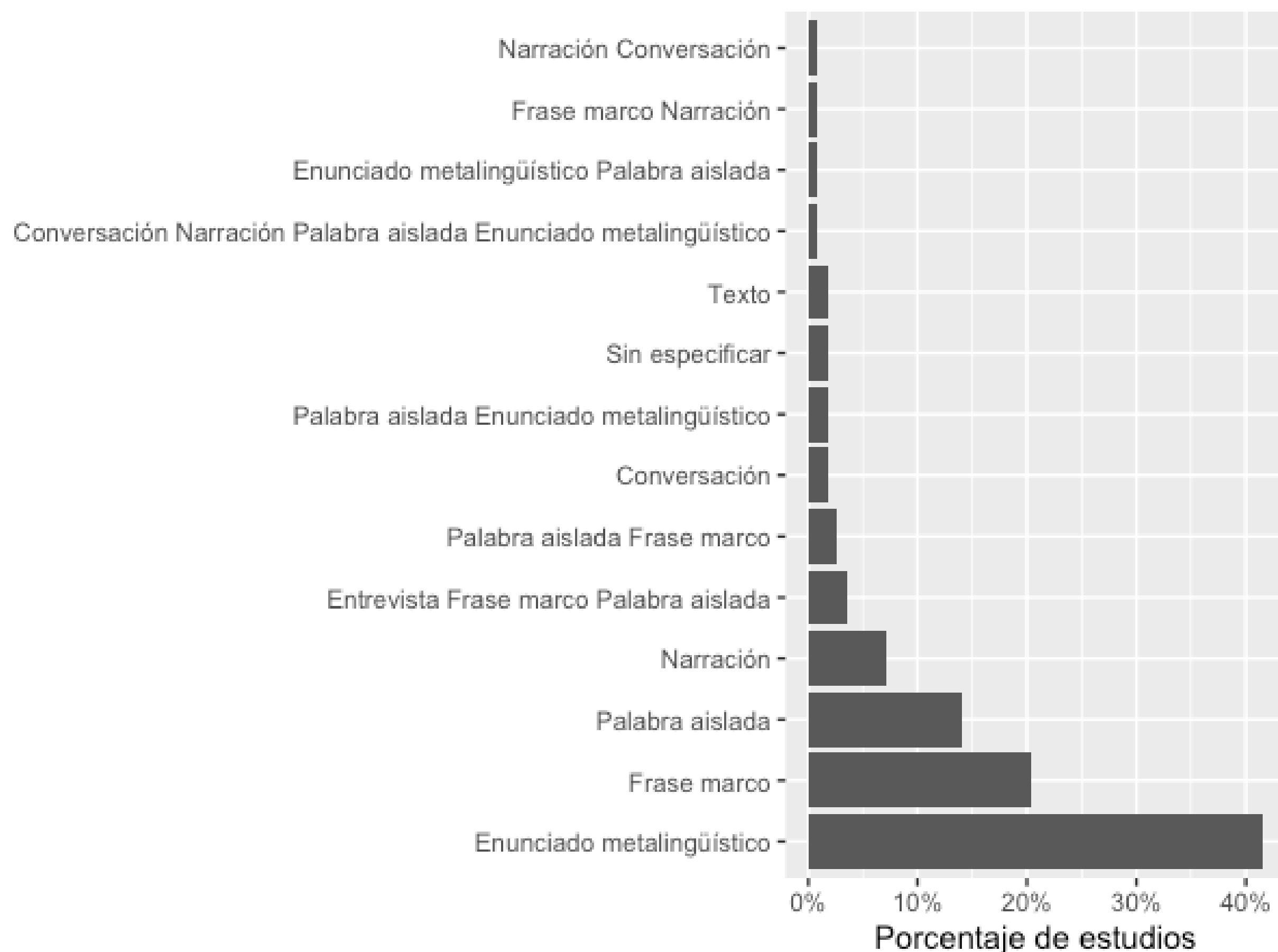
Figura 15: Estilos de fala nos corpora utilizados nos estudos sobre os correlatos acústicos do acento lexical analisados por Roettger e Gordon (2017)



Nota. Gráfico feito a partir das informações coletadas no banco de dados do projeto Studies on acústico correlates of word stress, de MK Gordon e TB Roettger, 18 de abril de 2018, OSF (<https://doi.org/10.17605/OSF.IO/9R2CD>).
Legenda: Eixo Y: Percentuais de estudos. Eixo X: Fala de laboratorio, Fala espontânea, Fala de laboratorio/Fala espontânea, Sem especificação.

Uma visão um pouco mais detalhada do contexto em que aparecem os materiais analisados nos trabalhos sobre os correlatos acústicos do acento lexical é apresentada na figura 16, elaborada a partir das informações coletadas na base de dados de Gordon e Roettger (2018). Como pode ser visto, as três opções mais comuns são, nesta ordem, enunciados metalinguísticos, frases-quadro e palavras isoladas.

Figura 16: Contextos em que aparecem os materiais do corpus nos estudos sobre os correlatos acústicos do acento lexical analisados por Roettger e Gordon (2017)



Nota. Gráfico feito a partir das informações coletadas no banco de dados do projeto Studies on acústico correlates of word stress, de MK Gordon e TB Roettger, 18 de abril de 2018, OSF (<https://doi.org/10.17605/OSF.IO/9R2CD>).

Legenda: Eixo Y: Narração/elicitación Conversación, Frase-Veículo elicitação, Enunciado metalingüístico Palabra aislada, Conversación elicitação palabra aislada enunciado metalingüístico, Conversación elicitação palabra aislada enunciado metalingüístico, Texto, Sem especificação, Palabra aislada enunciado metalingüístico, Conversación, Palabra aislada frase-veículo, Entrevista Frase-veículo Palabra aislada, Elicitación, Palabra aislada, Frase-veículo, Enunciado metalingüístico. Eixo X: Percentual de estudos.

O estilo de fala com menor grau de espontaneidade encontrado nas pesquisas sobre os correlatos acústicos do acento lexical consiste, nas línguas de acento livre, na leitura de palavras isoladas que constituem pares de acentos mínimos, ou mesmo tri-acentuais, como exemplificado em (1).

Se for necessário neutralizar o efeito de sentido, os pesquisadores recorrem a pseudopalavras (também chamadas de ‘logatoms’), ou seja, palavras que são fonologicamente possíveis porque atendem a todos os requisitos fonotáticos exigidos, mas que de fato não ocorrem na língua. Em alguns experimentos, são utilizadas sequências de sílabas idênticas nas quais é introduzido um contraste acentual (DE MORAES, 1987), enquanto em outros são criados conjuntos de pseudopalavras diferentes, como, no caso do espanhol, “ládebo”, “ladebo” e “ladebó” (LLISTERRI et al., 2014).

Para evitar o chamado ‘efeito lista’ típico da leitura de palavras isoladas, é comum que as palavras em estudo sejam incluídas no que é conhecido como ‘frases-veículo’ ou ‘frases transportadoras’ (em inglês). Frequentemente, como evidenciado nos dados de Gordon e Roettger (2018) apresentados

na figura 16, as frases-quadro consistem em enunciados metalinguísticos do tipo “eu digo a palavra _____ de novo” (QUILIS, 1971) ou “eu digo _____ baixinho” (FONTES, 2013). Por meio dessa estratégia, é possível contextualizar as palavras nas quais o pesquisador está interessado em uma estrutura constante e, com isso, controlar determinadas variáveis como a posição da palavra na frase ou a distribuição dos acentos. Apesar de estarem obviamente distantes do que seria considerado uma produção espontânea, as frases-veículo favorecem, em certa medida, uma realização mais natural do que uma leitura de palavras isoladas e por isso têm sido utilizadas em estudos sobre os correlatos acústicos de acento lexical em português e espanhol como os de Cruz et al. (2016), Face (1998-1999), Ferreira (2014) ou Massini-Cagliari (1993), entre outros.

Outra possibilidade de criar um corpus controlado é fazer uso de enunciados isolados, construídos com base nas variáveis que foram consideradas relevantes no desenho do experimento. Por exemplo, no corpus utilizado por Llisterri et al. (2014), o objetivo é garantir que a sílaba tônica em estudo apareça sempre precedida e seguida de duas sílabas átonas em frases com o mesmo número de sílabas, como em “Calculas el límite mal”, “No le límites los gastos” e “No limité los tabiques”. A análise de aspectos específicos, por exemplo, o comportamento de duas sílabas tônicas contíguas — isto é, o chamado ‘choque tônico’ — também requer a construção de um corpus com características bem controladas, como o proposto por Barbosa e Arantes (2003) para o português brasileiro, que contém frases como “Eu comi **bolor** sexta-feira à noite” e “Eu comi **bolo** sexta-feira à noite”, ou a criada por Atria (2009) para estudar o choque de acento no espanhol chileno, em que afirmações como “Su **papá** pasa” e “Su **papa** pasa”.

Os pesquisadores também têm à sua disposição uma série de tarefas que lhes permitem obter realizações das palavras escolhidas como objeto de estudo com um grau de naturalidade maior do que nos casos examinados até agora, todos eles baseados na leitura do corpus que foi analisado. Em alguns casos, os participantes do experimento devem construir um enunciado a partir de uma série de palavras-chave (TORREIRA et al., 2014) ou responder a uma pergunta com certa construção sintática (ORTEGA-LLEBARIA; PRIETO, 2007); em outros, eles têm que substituir uma palavra por uma pseudopalavra (MASSINI-CAGLIARI, 1993; ORTEGA-LLEBARIA, 2006). Imagens, sozinhas ou acompanhadas de texto, também podem ser usadas para fazer com que os falantes façam a declaração desejada (RONQUEST, 2013; VOGEL et al., 2016).

Uma abordagem diferente para a forma de coleta de dados consiste em incorporar tarefas que aparentemente não têm relação com o objetivo do experimento, a fim de fazer com que os falantes concentrem sua atenção no objetivo da tarefa e não em sua própria produção linguística, alcançando assim um estilo que pode ser caracterizado como semi-espontâneo de acordo com a classificação proposta na figura 14. Dentre as tarefas mais conhecidas desse tipo, destaca-se a chamada ‘tarefa do mapa’ (ANDERSON et al., 1991), em que um dos participantes possui um mapa com itinerário

marcado, enquanto o outro possui uma versão do mesmo mapa sem nenhum itinerário; os dois falantes, que não podem ver seus respectivos mapas, devem cooperar para que aquele com o mapa sem itinerário trace o mesmo caminho que é traçado no outro mapa. Com base no *Map-Task Corpus da Open University of Israel* [Corpus da Tarefa do Mapa da Universidade Aberrta de Israel], coletado por esse procedimento, Silber-Varod et al. (2019) obtiveram um total de 887 vogais para seu estudo sobre a influência do acento lexical nos valores dos formantes vocálicos em hebraico.

Também fazem parte do estilo semi-espontâneo as entrevistas que são realizadas com os palestrantes para conseguir intervenções as mais longas possíveis e assim ter material suficiente para análise. Este é um tipo de interação que muitas vezes é descrito como uma ‘entrevista semidirigida’, pois há um certo grau de planejamento por parte do entrevistador. Essa técnica foi utilizada, por exemplo, no estudo dos correlatos acústicos do acento lexical do português brasileiro realizado por Barbosa et al. (2013b).

A conversação está entre os estilos de fala classificados como espontâneos na Figura 14. Como regra geral, quanto maior o grau de espontaneidade do corpus, menor o controle sobre as variáveis definidas no desenho do experimento, mas Torreira e Ernestus (2011), por exemplo, conseguiram extrair 727 vogais do espanhol peninsular e 497 do francês, que preencheram os requisitos necessários para estudar o acento com base em conversas do Nijmegen Corpus de Espanhol Casual e do Nijmegen Corpus de Francês Casual. Neste trabalho, assim como no de Silber-Varod et al. (2019), fica claro que é viável investigar o acento lexical utilizando recursos já disponíveis inicialmente criados para outros fins.

Por fim, nas publicações sobre os correlatos acústicos do acento lexical, também foi considerada a possibilidade de comparar diferentes estilos de fala, como é feito, por exemplo, no já citado trabalho de Barbosa et al. (2013b) sobre o português brasileiro, em que três diferentes estilos de fala são contrastados nos mesmos falantes: entrevista semidirigida, leitura de textos e leitura de palavras isoladas.

Reflexões finais

Nas páginas anteriores, procurou-se mostrar que a análise dos correlatos acústicos do acento lexical exige que o pesquisador tome, desde o início, uma série de decisões metodológicas. Os elementos que foram mencionados obviamente não são os únicos que precisam ser levados em consideração ao projetar um experimento. Como indicado no início, uma leitura atenta e crítica do trabalho já realizado é extremamente útil para conhecer os diversos fatores que podem afetar os resultados.

Roettger e Gordon (2017) concluem sua revisão da metodologia utilizada nas publicações sobre os correlatos acústicos do acento lexical, já aludido em diversas ocasiões, com recomendações

focadas em três aspectos: a transparência do desenho experimental, o desenho do corpus e o tamanho da amostra analisada.

A transparência no desenho do experimento remete, para esses autores, à necessidade de especificar claramente todas as decisões tomadas pelo pesquisador, principalmente no que diz respeito à estrutura prosódica do material analisado, a fim de separar os efeitos de uma proeminência típica do nível da frase daqueles que derivam do acento lexical na palavra, diferenciando entre o que na terminologia do modelo métrico-autossegmental (HUALDE, 2003) é conhecido como acento, que se manifesta por um tom de movimento relacionado à frase, e o que se chama acento, ligado ao acento lexical.

A importância dessa distinção é retomada nas recomendações de Roettger e Gordon (2017) relacionadas ao desenho do corpus, mas também não se deve esquecer que as propriedades inerentes aos segmentos analisados, seu contexto fonético e a estrutura silábica em que são enquadrados tem repercussões potenciais nos resultados dos estudos acústicos sobre o acento.

O tamanho da amostra é o terceiro dos fatores citados por Roettger e Gordon (2017). Por um lado, a amostra deve refletir adequadamente as características de um grupo de falantes mais amplo do que aquele formado pelos participantes do experimento e, por outro, deve ser representativo da língua. No entanto, para esses autores, a decisão sobre o número de casos é feita principalmente levando-se em conta critérios relacionados ao poder dos testes estatísticos utilizados.

Às abordagens de Roettger e Gordon (2017) podem ser acrescentadas algumas reflexões que derivam do que é exposto neste trabalho. Como foi apontado, parece haver uma certa tendência a ignorar estudos que não são publicados no contexto acadêmico anglo-saxão, escritos em outros idiomas que não o inglês e, às vezes, que não aparecem no que tem sido chamado de 'periódicos de impacto', como se a qualidade de uma pesquisa estivesse necessariamente ligada à linguagem em que é escrita e ao grau de divulgação que alcança em determinadas plataformas. O desconhecimento das obras que não atendem a esses supostos requisitos leva a uma visão parcial do campo e até mesmo a apresentar como novas informações que, em maior ou menor grau, já eram conhecidas. Por isso, uma revisão bibliográfica razoavelmente exaustiva e, principalmente, livre de vieses como o que acaba de ser mencionado parece um ponto de partida essencial no planejamento de um experimento.

Este trabalho também insistiu na necessidade de se ter um perfil linguístico o mais detalhado possível dos participantes, tanto nos estudos sobre a primeira língua quanto naqueles que focam em falantes não nativos. Como foi explicado, a variante geográfica dos falantes monolíngues pode ter repercussões na realização fonética do acento lexical e realidades linguísticas muito diferentes são escondidas sob rótulos como 'bilíngue' ou 'aprendiz de L2'.

Da mesma forma, a questão dos estilos de fala que compõem o corpus analisado foi abordada, pois, como visto, parece que boa parte dos dados disponíveis sobre os correlatos acústicos do acento lexical são oriundos de palavras isoladas e de frases relativamente curtas, que podem não ser totalmente representativas da linguagem em uso real.

Por fim, em consonância com as propostas de Roettger e Gordon (2017) sobre o desenho do corpus coletadas no início desta seção, nos estudos sobre acento lexical pode ser pertinente especificar o modelo adotado pelo autor para evitar que, em certos casos, pode-se pensar que se assume como universal – por exemplo, a noção de ‘acento tonal’ ou transcrição através do sistema ToBI (AGUILAR et al., 2009; FROTA et al., 2015) – algo que é único de uma determinada teoria.

Como um todo, como muitos autores têm apontado, o acento lexical é um fenômeno de natureza multiparamétrica que é afetado por uma grande variedade de fatores. Por isso, neste artigo procuramos destacar que o número de decisões metodológicas que um pesquisador deve tomar – seus ‘graus de liberdade’, usando o termo de Roettger (2019) – é considerável, o que leva à necessidade de documentar tais decisões o mais explicitamente possível para que, seguindo o que parece ser uma tendência crescente também no campo da fonética (ROETTGER et al., 2019), os experimentos sejam reproduzíveis. Ter informações detalhadas sobre a metodologia também permite, como apontam Roettger e Gordon (2017), avaliar a relevância de comparar resultados obtidos em experimentos em diferentes línguas, o que é uma vantagem inegável tanto para estudos contrastivos quanto para aqueles voltados para a tipologia fonética.

Referências

AGUILAR, L.; DE LA MOTA, C.; PRIETO, P. *Sp_ToBI training materials*. Grup d’Estudis de Prosòdia, 2009. Disponível em: http://prosodia.upf.edu/sp_tobi/en/index.php. Acesso em: 07 jun. 2022.

ANDERSON, A. H.; BADER, M.; BARD, E. G.; BOYLE, E.; DOHERTY, G.; BOTMA, B.; NOSKE, R. (ed.). *Phonological Explorations: Empirical, Theoretical and Diachronic Issues*. Berlin: Mouton de Gruyter.

GARROD, S.; ISARD, S.; KOWTKO, J.; MCALLISTER, J.; MILLER, J.; SOTILLO, C.; THOMPSON, H. S. Y.; WEINERT, R. The HCRC Map Task corpus. *Language and Speech*, v. 34, n. 4, p. 351-366, 1991. DOI: <https://doi.org/10.1177/002383099103400404>

ATRIA, J. J. Estrategias de resolución de choques acentuales en el castellano hablado en Santiago de Chile. *Onomázein*, v. 19, n. 1, p. 11-31, 2009. Disponível em: http://onomazein.lettras.uc.cl/Articulos/19/1_Atria.pdf. Acesso em: 07 jun. 2022.

AVELLO, P. Assessing learners' changes in foreign accent during Study Abroad. In: PÉREZ VIDAL, C.; LÓPEZ-SERRANO, S.; AMENT, J.; THOMAS-WILHELM, D. J. (ed.). *Learning context effects: study abroad, formal instruction and international immersion classrooms*. Language Science Press, 2018. p. 131-154. DOI: <https://doi.org/10.5281/zenodo.1300630>

BADIA I MARGARIT, A. M. Notes sobre el castellà parlat per catalans. *Llengua i cultura als Països Catalans*. 4. ed. Edicions 62, 1975. p. 145-153.

BARBOSA, P. A.; ALBANO, E. C. Brazilian Portuguese. *Journal of the International Phonetic Association*, v. 34, n. 2, p. 227-232, 2004. DOI: <https://doi.org/10.1017/S0025100304001756>

BARBOSA, P. A.; ARANTES, P. Investigation of non-pitch-accented phrases in Brazilian Portuguese: No evidence favoring stress shift. In: SOLÉ, M. J.; RECASENS, D.; ROMERO, J. (ed.). *15th International Congress of Phonetic Sciences*. Barcelona, Spain, August 3-9, 2003. p. 475-478. ICPHS Archive. Disponible em: https://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS2003/p15_0475.html. Acesso em: 07 jun. 2022.

BARBOSA, P. A.; ERIKSSON, A.; ÅKESSON, J. Cross-linguistic similarities and differences of lexical stress realisation in Swedish and Brazilian Portuguese. In: ASU, E. L.; LIPPUS, P. (ed.). *Nordic prosody: Proceedings of the XIth Conference, Tartu 2012*. Peter Lang, 2013a. p. 97-106.

BARBOSA, P. A.; ERIKSSON, A.; ÅKESSON, J. On the robustness of some acoustic parameters for signalling word stress across styles in Brazilian Portuguese. In: BIMBOT, F.; CERISARA, C.; FOUGERON, C.; GRAVIER, G.; LAMEL, L.; PELLEGRINO, F.; PERRIER, P. (ed.). *INTERSPEECH 2013, 14th Annual Conference of the International Speech Communication Association*. Lyon, France, August 25-29, 2013b. p. 282-286. ISCA Archive. Disponible em: https://www.isca-speech.org/archive/interspeech_2013/i13_0282.html. Acesso em: 07 jun. 2022.

BARBOSA, P. A.; MADUREIRA, S. *Manual de fonética acústica experimental: aplicações a dados do português*. São Paulo: Cortez, 2015.

BILINGUALISM AND PSYCHOLINGUISTICS RESEARCH GROUP. (s.f.). *LEAP-Questionnaire*. Northwestern University. Disponible em: <https://bilingualism.northwestern.edu/leapq/>. Acesso em: 07 jun. 2022.

BIRDSONG, D.; GERTKEN, L. M.; AMENGUAL, M. *Bilingual Language Profile: An easy-to-use instrument to assess bilingualism*. University of Texas at Austin, Center for Open Educational Resources and Language Learning (COERLL). 2012, 20 de enero. Disponible em: <https://sites.la.utexas.edu/bilingual/>. Acesso em: 07 jun. 2022.

BOERSMA, P.; WEENINK, D. *Praat: Doing phonetics by computer* (Versión 6.1.40) [Programa informático]. University of Amsterdam. 2021. Disponível em: <http://www.praat.org>. Acesso em: 07 jun. 2022.

COLANTONI, L.; MARASCO, O.; STEELE, J.; SUNARA, S. Learning to realize prosodic prominence in French and Spanish. In: MILLER, R. T.; MARTIN, K. I.; EDDINGTON, C. M.; HENERY, A.; MARCOS, N.; TSENG, A. M.; TUNINETTI, A.; WALTER, D. (ed.). *Selected Proceedings of the 2012 Second Language Research Forum: Building bridges between disciplines*. 2014. p. 15-29. Cascadilla Proceedings Project. Disponível em: <http://www.lingref.com/cpp/slrf/2012/abstract3082.html>. Acesso em: 07 jun. 2022.

CONSEJO DE EUROPA. *Marco común europeo de referencia para las lenguas: aprendizaje, enseñanza, evaluación*. Ministerio de Educación, Cultura y Deportes, Instituto Cervantes; Anaya, 2002. Disponível em: https://cvc.cervantes.es/ensenanza/biblioteca_ele/marco/. Acesso em: 07 jun. 2022.

CRUZ, R. C. F.; BORGES, B. do S. P.; FERREIRA, J. É. Q.; RILLIARD, A.; FONTEL, E. da S. Análise acústica dos correlatos prosódicos do acento lexical do Português Brasileiro. *Journal of Speech Sciences*, v. 5, n. 2, p. 159-175, 2016. Disponível em: http://revistas.iel.unicamp.br/ojs_joss/index.php/journalofspeechsciences/article/view/121. Acesso em: 07 jun. 2022.

CRYSTAL, D. Stress. *A first dictionary of linguistics and phonetics*. André Deutsch, 1980. p. 332-333.

FACE, T. L. Efectos segmentales del acento en español. *Boletín de Lingüística*, v. 14, p. 18-32, 1998-1999.

FERREIRA, L. P. A duração como correlato acústico do acento de palavra no português brasileiro e no espanhol: desafios para o ensino de suprasegmentais e preparação de material didático. *Signum: Estudos da Linguagem*, v. 17, n. 1, p. 74-101, 2014. DOI: <https://doi.org/10.5433/2237-4876.2014v17n1p74>

FONTES, M. A. S. Pistas acústicas e a percepção do acento lexical em português brasileiro. *Intercâmbio*, v. 27, p. 100-109, 2013. Disponível em: <https://revistas.pucsp.br/intercambio/article/view/17692>. Acesso em: 07 jun. 2022.

FROTA, S.; OLIVEIRA, P.; CRUZ, M.; VIGÁRIO, M. *P-ToBI: Tools for the transcription of Portuguese prosody*. CLUL/FLUL, Laboratório de Fonética, 2015. Disponível em: <http://labfon.letra.sulisboa.pt/InAPoP/P-ToBI/index.html>. Acesso em: 07 jun. 2022.

GARDE, P. *L'accent*. Presses Universitaires de France, 1968.

- GARRIDO, J. M. Using large corpora and computational tools to describe prosody: An exciting challenge for the future with some (important) pending problems to solve. In: FELDHAUSEN, I.; FLIESSBACH, J.; DEL M. VANRELL, M. (ed.). *Methods in prosody: A Romance language perspective*. Language Science, 2018. p. 3-43. DOI: <https://doi.org/10.5281/zenodo.1441335>
- GERTKEN, L. M.; AMENGUAL, M.; BIRDSONG, D. Assessing language dominance with the Bilingual Language Profile. In: LECLERCQ, P.; EDMONDS, A.; HILTON, H. (ed.). *Measuring L2 proficiency: Perspectives from SLA*. Multilingual Matters, 2014. p. 208-225.
- GORDON, M. K.; ROETTGER, T. B. Acoustic correlates of word stress: A crosslinguistic survey. *Linguistics Vanguard*, v. 3, n. 1, Artículo 20170007, 2017. DOI: <https://doi.org/10.1515/lingvan-2017-0007>
- GORDON, M. K.; ROETTGER, T. B. *Studies on acoustic correlates of word stress: An online corpus*. 2018, 18 de abril. DOI: <https://doi.org/10.17605/OSF.IO/9R2CD>
- HOFFMANN, C. *An introduction to bilingualism*. Longman, 1991.
- HUALDE, J. I. El modelo métrico y autosegmental. In: PRIETO, P. (ed.). *Teorías de la entonación*. Ariel, 2003. p. 155-184.
- INTERNATIONAL PHONETIC ASSOCIATION. *Handbook of the International Phonetic Association: A guide to the use of the International Phonetic Alphabet*. Cambridge: Cambridge University Press, 1999.
- ISBELL, D. R. Assessing pronunciation for research purposes with listener-based numerical scales. In: KANG, O.; GINTHER, A. (ed.). *Assessment in second language pronunciation*. Routledge, 2018. p. 89-111. DOI: <https://doi.org/10.4324/9781315170756>
- JULIÀ, J. *Fonètica aplicada catalana: dels fonaments a les aplicacions de les ciències fonètiques*. Ariel, 2005.
- KAUSHANSKAYA, M.; BLUMENFELD, H. K.; MARIAN, V. The Language Experience and Proficiency Questionnaire (LEAP-Q): Ten years later. *Bilingualism: Language and Cognition*, v. 23, n. 5, p. 945-950, 2020. DOI: <https://doi.org/10.1017/s1366728919000038>
- KIM, M. *The phonetics of stress manifestation: Segmental variation, syllable constituency and rhythm*. Tesis doctoral, Stony Brook University. SUNY Digital Repository, 2011. <http://hdl.handle.net/1951/56031>
- LABOV, W. The isolation of contextual styles. *Sociolinguistic patterns*. Basil Blackwell, 1978. p. 70-109. Obra original publicada en 1972.

LADEFOGED, P. *Phonetic data analysis: An introduction to fieldwork and instrumental techniques*. Blackwell, 2003.

LI, P.; ZHANG, F.; YU, A.; ZHAO, X. Language History Questionnaire (LHQ3): An enhanced tool for assessing multilingual experience. *Bilingualism: Language and Cognition*, v. 23, n. 5, p. 938-944, 2020. DOI: <https://doi.org/10.1017/S1366728918001153>

LLISTERRI, J.; MACHUCA, M. J.; RÍOS, A.; SCHWAB, S. El acento léxico en contexto: datos acústicos. In: CONGOSTO, Y.; MONTERO, M. L.; SALVADOR, A. (ed.). *Fonética experimental, educación superior e investigación*. Arco/Libros, 2014. v. 1, p. 357-376.

MARTÍNEZ CELDRÁN, E.; FERNÁNDEZ PLANAS, A. M.; CARRERA, J. Castilian Spanish. *Journal of the International Phonetic Association*, v. 33, n. 2, p. 255-259, 2003. DOI: <https://doi.org/10.1017/S0025100303001373>

MASSINI-CAGLIARI, G. Sobre a natureza fonética do acento em português. *DELTA: Documentação e Estudos em Linguística Teórica e Aplicada*, v. 9, n. 2, p. 195-216, 1993. Disponível em: <http://ken.pucsp.br/delta/article/view/45514>. Acesso em: 07 jun. 2022.

DE MORAES, J. A. Correlats acoustiques de l'accent de mot en portugais bresilien. *Proceedings XIth ICPhS: The Eleventh International Congress of Phonetic Sciences*. August 1-8, 1987, Tallinn, Estonia, U.S.S.R., Academy of Sciences of the Estonian S.S.R., v. 3, p. 313-316, 1987.

MUNRO, M. J.; DERWING, T. M. Processing time, accent, and comprehensibility in the perception of native and foreign-accented speech. *Language and Speech*, v. 38, n. 3, p. 289-306, 1995. DOI: <https://doi.org/10.1177/002383099503800305>

NORRIS, J.; ORTEGA, L. Defining and measuring SLA. In: DOUGHTY, C. J.; LONG, M. H. (ed.). *The handbook of second language acquisition*. Blackwell, 2003. p. 716-761. DOI: <https://doi.org/10.1002/9780470756492.ch21>

ORTEGA-LLEBARIA, M. Phonetic cues to stress and accent in Spanish. In: DÍAZCAMPOS, M. (ed.). *Selected Proceedings of the 2nd Conference on Laboratory Approaches to Spanish Phonetics and Phonology*. Cascadilla Proceedings Project, 2006. p. 104-118. Disponível em: <http://www.lingref.com/cpp/lasp/2/abstract1329.html>. Acesso em: 07 jun. 2022.

ORTEGA-LLEBARIA, M.; PRIETO, P. Disentangling stress from accent in Spanish: Production patterns of the stress contrast in deaccented syllables. In: PRIETO, P.; MASCARÓ, J.; SOLÉ, M. J. (ed.). *Segmental and prosodic issues in Romance phonology*. John Benjamins, 2007. p. 155-176. DOI: <https://doi.org/10.1075/cilt.282.11ort>

- ORTEGA-LLEBARIA, M.; PRIETO, P. Acoustic correlates of stress in Central Catalan and Castilian Spanish. *Language and Speech*, v. 54, n. 1, p. 73-97, 2011. DOI: <https://doi.org/10.1177/0023830910388014>
- QUILIS, A. Caracterización fonética del acento español. *Travaux de Linguistique et de Littérature*, v. 9, n. 1, p. 53-72, 1971.
- R CORE TEAM. *R: A language and environment for statistical computing* (Versión 4.0.4) [Programa informático]. Vienna, R Foundation for Statistical Computing. 2021. Disponível em: <https://www.r-project.org/>. Acesso em: 07 jun. 2022.
- ROETTGER, T. B. Researcher degrees of freedom in phonetic research. *Laboratory Phonology: Journal of the Association for Laboratory Phonology*, v. 10, n. 1, artículo 1, 2019. DOI: <https://doi.org/10.5334/labphon.147>
- ROETTGER, T. B.; GORDON, M. K. Methodological issues in the study of word stress correlates. *Linguistics Vanguard*, v. 3, n. 1, artículo 20170007, 2017. DOI: <https://doi.org/10.1515/lingvan-2017-0006>
- ROETTGER, T. B.; WINTER, B.; BAAYEN, H. Emergent data analysis in phonetic sciences: Towards pluralism and reproducibility. *Journal of Phonetics*, v. 73, p. 1-7, 2019. DOI: <https://doi.org/10.1016/j.wocn.2018.12.001>
- RONQUEST, R. E. An acoustic examination of unstressed vowel reduction in Heritage Spanish. In: HOWE, C.; BLACKWELL, S. E.; LUBBERS QUESADA, M. (ed.). *Selected Proceedings of the 15th Hispanic Linguistics Symposium*. Cascadilla Proceedings Project, 2013. p. 157-171. Disponível em: <http://www.lingref.com/cpp/hls/15/abstract2882.html>
- SANTIAGO, F.; MAIRANO, P. The role of lexical stress on vowel duration and vowel space in two varieties of Spanish. In: KLESSA, K.; BACHAN, J.; WAGNER, A.; KARPIŃSKI, M.; ŚLEDZIŃSKI, D. (ed.). *9th International Conference on Speech Prosody 2018*. Poznań, Poland, 13-16 June 2018. p. 453-457. ISCA Archive. DOI: <https://doi.org/10.21437/SpeechProsody.2018-92>
- SCHWAB, S. Acoustic correlates of French accentuation in advanced Spanish learners of French. *COPAL, Concordia Working Papers in Applied Linguistics*, v. 5, p. 612-623, 2014. Disponível em: http://doe.concordia.ca/copaldocuments/39_Schwab_Vol5.pdf. Acesso em: 07 jun. 2022.

SILBER-VAROD, V.; KHORSHIDI, N.; LEVI, L.; AMIR, N. The influence of lexical stress on formant values in spontaneous Hebrew speech. In: CALHOUN, S.; ESCUDERO, P.; TABAIN, M.; WARREN, P. (ed.). *Proceedings of the 19th International Congress of Phonetic Sciences*. Melbourne, Australia 2019. p. 3538-3542. Australasian Speech Science and Technology Association, 2019. Disponível em: https://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS2019/papers/ICPhS_3587.pdf. Acesso em: 07 jun. 2022.

THE BRAIN, LANGUAGE, AND COMPUTATION (BLC) Lab. (s. f.). *Language History Questionnaire 3.0*. The Hong Kong Polytechnic University. Disponível em: <https://blclab.org/lhq3/>. Acesso em: 07 jun. 2022.

TORREIRA, F.; ERNESTUS, M. Realization of voiceless stops and vowels in conversational French and Spanish. *Laboratory Phonology: Journal of the Association for Laboratory Phonology*, v. 2, n. 2, p. 331-353, 2011. DOI: <https://doi.org/10.1515/labphon.2011.012>

TORREIRA, F.; SIMONET, M.; HUALDE, J. I. Quasi-neutralization of stress contrasts in Spanish. In: CAMPBELL, N.; GIBBON, D.; HIRST, D. (ed.). *7th International Conference on Speech Prosody 2014*. 20-23 May 2014, Dublin, Ireland, 2014. p. 197-201. ISCA Archive. DOI: <https://doi.org/10.21437/SpeechProsody.2014-27>

TRASK, R. L. Stress. *A dictionary of phonetics and phonology*. Routledge, 1996. p. 336. DOI: <https://doi.org/10.4324/9780203695111>

TREFFERS-DALLER, J. What defines language dominance in bilinguals? *Annual Review of Applied Linguistics*, v. 5, p. 375-393, 2019. DOI: <https://doi.org/10.1146/annurev-linguistics-011817-045554>

VAN HEUVEN, V. J. Acoustic correlates and perceptual cues of word and sentence stress: Towards a cross linguistic perspective. In: GOEDEMANS, R.; HEINZ, J.; VAN DER HULST, H. (ed.). *The study of word stress and accent: Theories, methods and data*. Cambridge University Press, 2018. p. 15-59. DOI: <https://doi.org/10.1017/9781316683101.002>

VAN HEUVEN, V. J.; TURK, A. Phonetic correlates of word and sentence stress. In: GUSSENHOVEN, C.; CHEN, A. (ed.). *The Oxford handbook of language prosody*. Oxford: Oxford University Press, 2020. p. 150-165. DOI: <https://doi.org/10.1093/oxfordhb/9780198832232.013.8>

VOGEL, I.; ATHANASOPOULOU, A.; PINCUS, N. Prominence, contrast, and the Functional Load Hypothesis: An acoustic investigation. In: HEINZ, J.; GOEDEMANS, R.; VAN DER HULST, H. (ed.). *Dimensions of phonological stress*. Cambridge: Cambridge University Press, 2016. p. 123-167. DOI: <https://doi.org/10.1017/9781316212745.006>

Chapter 2

Towards a comprehensive analysis of Franconian tonal accent

Björn Köhnlein

I. Introduction

Franconian tonal accent, spoken in parts of Belgium, Germany, and the Netherlands, is an opposition between two-word accents, which are commonly referred to as Accent 1 and Accent 2. I provide two minimal pairs in (1), one between monosyllabic items (1a), and one between disyllabic items (1b). The accents are indicated with superscripts at the end of the syllable carrying the accent; these syllables are always stressed since the tonal accent is restricted to stressed syllables.

Tone accent minimal pairs from Mayen (Schmidt 1986)

a. [man¹] ‘basket’ [man²] ‘man’

b. [tɔu¹və] ‘pigeons’ [tɔu²və] ‘baptisms’

The phenomenon has been studied for far more than a century, the earliest known source being Hardt (1843). During this long research history, it has been difficult to pin down the phonetic and phonological aspects of tonal accent. While our understanding of the phonetic characteristics of accent have improved considerably over the past few decades, there is still considerable disagreement about what the ‘correct’ synchronic phonological analysis of tonal accent should be. (This same is true for the diachronic analysis of tonal accent, which I will not discuss here; see Köhnlein, 2020, for an overview).

The main reason that the phonology of the accents is still somewhat elusive can arguably be attributed to their phonetic complexity, which makes them particularly interesting for our understanding of word-prosodic typology, a central part of which is the study of tone and stress across languages. Within prosodic typology, it is often maintained that there is a substantial difference between tone and stress, with clear differences between these two types of phonological properties (e.g., Hyman, 2006, 2009). Tone-accent systems like Franconian, however, are not necessarily a straightforward fit for these typological categories for the same reasons that make them difficult to analyze in the first place.

Intuitively, it seems appropriate to say that the Franconian tonal accent can be placed somewhere between tone and stress. On one hand, relevant dialects appear to be uncontroversially tonal, in the sense that the two accents are generally realized with different tonal melodies, such as a falling tone for Accent 1 and a high-level tone for Accent 2 in statements (there is more variation, some of which I will discuss in §2), yet the occurrence of the tonal contrast is restricted to stressed syllables and tends to disappear in non-prominent sentence positions. Furthermore, the tonal differences between Accent 1 and Accent 2 are typically accompanied by other phonetic correlates, the details of which can differ across dialects: for instance, Accent 1 and Accent 2 often differ durationally, they can show phonological differences in vowel quality (such as diphthongal versus monophthongal

realizations depending on the accent), whether a word receives Accent 1 or Accent 2 can interact with the strength of following consonants, etc. Such differences in duration, as well as vowel and consonant quality, are prototypical of stress systems rather than tone systems. Furthermore, it is well known that intonational tone can be relevant for marking stressed syllables phonetically in languages with purely intonational systems, such as English or Portuguese. Would this suggest, then, that Franconian only has an intricate stress system rather than a lexical tone contrast?

In this context, there is another important additional consideration that I have only discussed implicitly so far: When linguists talk about the location of stress, they typically refer to the syllable as the unit that carries stress (sometimes, people still refer to, e.g., stressed vowels, but this typically implicitly also includes the consonants in the same syllable). In fact, there is wide agreement that stress is a property of syllables, not of units below the level of the syllable (Hyman, 2006, 2009). If this is indeed the case cross-linguistically, it would be analytically non-trivial to assume a difference in stress between, say, the first or the second part of a long vowel – indeed, such contrasts should be impossible if stress is always a property of syllables. With this generalization in mind, another look at the examples in (1) reveals why it would be controversial to describe Franconian tonal accent as a purely stress-based phonological opposition: The contrast between Accent 1 and Accent 2 occurs within syllables, which is at least atypical (or maybe even impossible?) in stress systems, but perfectly expectable in lexical-tone systems. So, is this contrast based on phonological tone? Or two types of stress? Or something else?

Based on a brief overview of some of the most important correlates of tonal accent and the morphological role of accent in §2, I aim to provide an introduction to the tenets of analytical approaches to Franconian tonal accent, with a focus on what I refer to as the *metrical approach*. In a nutshell, there currently are two competing analytical approaches. On one hand, the tonal accent has been analyzed as a contrast between an accent that carries a lexical tone (typically Accent 2) versus a lexically toneless accent (typically Accent 1). This is the traditional analysis in theoretical phonology, which I refer to as the *tonal approach*. Over the past decade, however, an alternative metrical approach to accent has been proposed. Scholars following the metrical approach typically assume that there is a more abstract, structural difference between the two accents, such as two types of metrical feet. I will discuss the basics of the metrical approach in §3, taking my own work as a concrete example. In §4, I will briefly compare it to the tonal approach and conclude the paper.

2. Background on Franconian

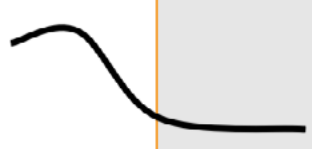
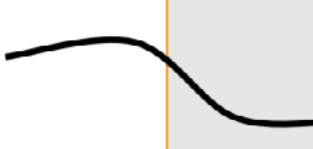
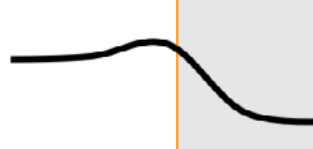
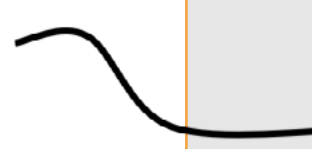
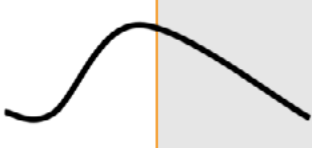
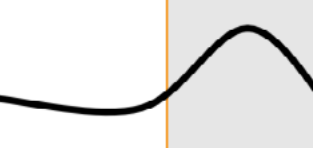

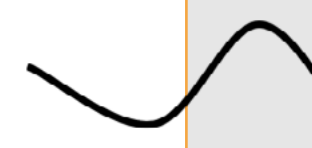
This section provides some background on Franconian tonal accent and discusses the phenomena that – at least in my view – should be captured in a principled way in a comprehensive analysis. I start by discussing distributional tenets and basic tonal melodies (§2.1), continue with a brief review

of additional phonetic and phonological correlates across dialects (§2.2), and then focus on some of the morphological functions of tonal accent (§2.3).

2.1 Basic tonal melodies

In most dialects, the realization of the tone-accent contrast requires either a long vowel/diphthong or a sequence of a short vowel plus a sonorant coda consonant. Sequences of short vowels and coda obstruents, on the other hand, typically do not display an accentual contrast. Figure 1 (left side) displays the tonal realizations of Accent 1 and Accent 2 in the Cologne dialect (e.g. Gussenhoven & Peters, 2004; Peters, 2006), a prototypical tone accent dialect, in non-final focus position of declarative and interrogative utterances. It should be noted that there can be considerable variation of the realization in the realization of the accents, depending on the position in the sentence (non-final, final), prominence (focal, pre-focal, post-focal), and pragmatic context (in addition to declaratives and interrogatives, there can be other contexts, such as continuation intonation); furthermore, the accent contrast is often neutralized out of focus. For the purposes of this overview, however, I focus on the most basic, default realizations only.

Figure 1: Idealized tonal contours in Cologne (Rule A) and Arzbach (Rule B), focus, non-final position; nuclear contour non-shaded, overall post-nuclear contour shaded.

Context	Cologne (Rule A)		Arzbach (Rule B)	
	Accent 1	Accent 2	Accent 1	Accent 2
Declarative, non-final				
Interrogative, non-final				

As shown in Figure 1, the focal tone contours for Accent 1 are an early fall in declarative sentences and an early rise in interrogative sentences; Accent-2 tonal movements occur relatively later, with a high-level pitch in the accented syllable for declaratives and a late fall into the post-tonic syllable (shown as grey-shaded in Figure 1), and low a level-pitch and a late rise for Accent 2. The main difference between the Accent-1 contours and the Accent-2 contours in Cologne thus lies in the timing of the relevant pitch movements – early for Accent 1, late for Accent 2.

Dialects with tonal realizations of the accents like Cologne have been referred to as *Rule A* in the literature, a classification based on Wiesinger (1970). There is a variation of the tonal realization across dialects, often minor but sometimes considerable. The arguably most extreme type of variation occurs in so-called *Rule-B* dialects, as first described by Bach (1921) for Arzbach Franconian. As shown in Figure 1 (right side), Arzbach has opposite tonal melodies in comparison to Cologne, at least in declaratives: Accent 1 is produced with a level tone and a late fall, as opposed to Cologne,

where Accent 1 has an early fall; conversely, Accent 2 is realized as an early fall in Arzbach while it realized as late fall in Cologne. While Köhnlein (2011) generally confirms Bach's observations, it also relativizes them: My fieldwork demonstrated that the reversal is restricted to declaratives but cannot be observed in interrogatives. In fact, the melodies in interrogatives are very similar between Cologne and Arzbach. In both dialects, Accent 1 rises earlier than Accent 2. In my work, I have referred to this as a *semi-reversal* of tonal melodies. The Arzbach facts are typologically peculiar from various perspectives: Generally, such tonal reversals between closely related dialects are exceedingly rare (yet see Kingston 2011 for Athabaskan). Moreover, the Arzbach dialect appears to be the only documented language to date that displays an alignment reversal of word accents in one dialect: That is, Accent 1 *falls later* than Accent 2 in declaratives but shows an *earlier rise* than Accent 2 in interrogatives.

2.2 Additional phonetic and phonological correlates

Typically, pitch appears to be the main phonetic correlate of the tone-accent opposition across dialects; yet other factors can play a role in the realization of the accents. Here, I will briefly discuss the effects of duration, vowel quality, and consonant voicing. Among these correlates, duration is the one correlate most consistently attested across dialects. Phonetically, what we typically find is that level tones are longer than falling tones, which in most dialects translates to Accent 2 being longer than Accent 1. Intuitively, this might appear to be somewhat surprising since it is well-established that producing an audible contour tone will require more time than producing a level tone; yet this is only obvious for very short durations. As argued in Köhnlein (2015), this relationship might be reversed (at least *ceteris paribus*) for longer durations, such as long vowels or sequences of short vowels and coda sonorants.

For Cologne, Gussenhoven & Peters (2004) report that duration can be the main correlate of the tone-accent opposition in post-focal positions where the tonal contrast is non-distinctive but where Accent 2 (level tone) is still considerably longer than Accent 1 (falling tone), particularly in the final position of a phrase. An example is provided in (2), where Accent 1 is marked as long and Accent 2 as overlong:

(2) Durational differences between Accent 1 and Accent 2 in the Cologne dialects

a. [da:x¹] 'day.dat' [da::x²] 'day.nom'

Some dialects also show differences in vowel quality. In the most widespread pattern, vowels with contour tones (typically Accent 1) correspond to more diphthongal vowels, and vowels with level tones (typically Accent 2) correspond to more monophthongal vowels (e.g., Dols, 1953; Wiesinger, 1970; Cajot, 2006; Gussenhoven, 2012; Köhnlein, 2017). Two examples from Maastricht (Gussenhoven, 2012), where high vowels diphthongize if they have Accent 1, are given in (3).

(3) Differences in vowel quality between Accent 1 and Accent 2 in Maastricht Franconian

- | | | |
|---|------------------------|-------------|
| a. [blɛif ¹] ‘stay.1.PS.SG’ | [bli: ² və] | ‘stay-INF’ |
| b. [dɔuf ¹] ‘pigeon.SG’ | [du: ² və] | ‘pigeon-PL’ |

Lastly, accentuation can interact with the voicing quality of word-medial and word-final obstruents. This is unsurprising from a historical perspective since consonant voicing is an important predictor of whether a word received Accent 1 or Accent 2 after the genesis of tonal accent (Köhnlein, 2020, for overview and references). In many dialects, later phonological processes (such as word-medial consonant weakening or final devoicing) have obscured these interactions over time, but some dialects still preserve relevant interactions. One of these dialects is (again) Cologne Franconian, where disyllabic words with word-medial voiced consonants always receive Accent 1, and words with word-medial voiceless consonants typically receive Accent 2 (even though they may receive Accent 1 in some cases). Some examples, taken from Münch (1904), are provided in (4). Examples (4a-c), which feature disyllabic words with intervocalic voiceless consonants, always surface with Accent 1, while the items (4d-f), which contain voiced intervocalic consonants, surface with Accent 2:

(4) Interaction of accent and word-medial consonant voicing in Cologne (Münch, 1904)

- | | | | |
|---------------------------|-------------|---------------------------|------------|
| a. [o: ¹ ʏə] | ‘eye’ | d. [lo: ² fə] | ‘to run’ |
| b. [fre: ¹ zə] | ‘to freeze’ | e. [ri: ² sə] | ‘to tear’ |
| c. [drø: ¹ mə] | ‘to dream’ | f. [ʃli: ² sə] | ‘to split’ |

To give another example of interactions between tonal accent and consonant voicing, the Moresnet dialect shows predictable synchronic alternations between Accent 1 and Accent 2 based on the voicing of post-vocalic consonants. Specifically, Moresnet is among those dialects that display an accent opposition, not only in long vowels/diphthongs and short vowels plus sonorants but also in short vowels plus obstruents (Jongen 1972). As discussed in van Oostendorp (2017), sequences of short vowels plus underlyingly voiced obstruents will receive Accent 1, while sequences of short vowels plus underlyingly voiceless obstruents will receive Accent 2. In the surface realization, however, this voicing contrast can be neutralized by final devoicing, a common phonological process in Dutch and German dialects. This is demonstrated in (5). (5a) shows that short vowels followed by underlyingly voiced obstruents have Accent 1, even if the obstruent is devoiced; (5b) demonstrates that short vowels plus underlyingly voiceless obstruents receive Accent 2.

(5) Interaction of accent and underlying voicing in Moresnet

- | | |
|------------------------------------|---|
| a. /bɛd/ [bɛt ¹] ‘bed’ | /bɛd/ + /ə/ [bɛ ¹ də] ‘beds’ |
| b. /zɛs/ [zɛs ²] ‘six’ | /tɛpəx/ [tɛ ² pəx] ‘carpet’ |

2.3 Morphological relevance of tonal accent

In every dialect, the accents of certain items will have to be memorized by speakers, as they are synchronically unpredictable. For instance, the difference between [man¹] ‘basket’ and [man²] ‘man’, which I have introduced in (1), will arguably have to be learned. In addition to the phonetic and phonological correlates discussed in §2.2, however, there are also certain synchronic morphological generalizations. A pattern that is found across all dialects concerns the alternation between Accent 1 and Accent 2 in morphologically conditioned accent minimal pairs: In such cases, the morphologically complex form will *always* receive Accent 1, while the morphologically simplex form will *always* receive Accent 2. I provide two examples in (6), taken from my own fieldwork on the Arzbach dialect:

(6) Shifts from Accent 2 to Accent 1 in morphological alternations

- a. [ʃta:n²] ‘stone’ [ʃta:n¹] ‘stones’
b. [ət is brait²] ‘it is broad’ [ən brait¹ haus²] ‘a.NOM broad.NOM house’

(6a) shows an alternation between a singular form with Accent 2 versus a plural with Accent 1; (6b), on the other hand, provides different forms of the adjective meaning ‘broad’: Here, the more basic, predicative form receives Accent 2, while the attributive neuter singular form has Accent 1. Generally, we can regard the attributive forms as more complex than their predicative counterparts since they sometimes also contain additional segmental material, depending on gender or plurality – as an example, consider the Arzbach form [brai¹də], which is the realization of ‘broad’ for masculine singular and all plural forms.

To end this section, I will mention one more example of the morphonological relevance of accent that does not concern tone-accent minimal pairs per se but still indicates a predictable interaction of morphological structure and accentuation. A variety of German dialects delete certain word-final plosives in plural forms, among them many Franconian-tone accent dialects. Some examples from Arzbach are provided in (7a-c). As can be observed, in all of these cases, the morphologically simplex singular forms contain a word-final [t], which is deleted in corresponding plural forms. In the morphological literature, alternations where simplex forms contain more phonological information than more complex forms are referred to as *subtraction*. That is, it looks like the respective morphemes – in this case, the plural morpheme – remove phonological content rather than adding it.

(7) Consonant subtraction and word-medial deletion in Arzbach Franconian

- a. [hʊnt²] ‘dog’ [hʊn¹] ‘dogs’
b. [vant²] ‘wall’ [vɛn¹] ‘walls’
c. [hant²] ‘hand’ [hɛn¹] ‘hands’
d. [kɪnt²] ‘child’ [kɪnɐ¹] ‘children’
e. [lant²] ‘country’ [lɛnɐ¹] ‘countries’
f. [rɪnt²] ‘cow’ [rɪnɐ¹] ‘cows’

With regard to tonal accent, subtraction is relevant since the respective plural forms not only miss a final consonant present in corresponding singulars but also switch the accent of the item from Accent 2 to Accent 1. Crucially, this pattern is not an idiosyncrasy of the Arzbach dialect but it holds true across dialects: While we obviously cannot know all forms in all dialects, there appear to be no known counterexamples to the generalization that subtraction coincides with an accent change from Accent 2 to Accent 1. A few more general notes on subtraction in Franconian dialects: First, what kinds of segmental sequences are subject to subtractions differs across dialects; see Köhnlein (2018) for a more detailed discussion. Furthermore, note that some alternations also show differences in the vowel quality, which is due to an independent fronting process commonly referred to as *umlaut*; since *umlaut* is not of immediate relevance for the purposes of this paper, I will ignore it from now on. Lastly, consider (7d-f), which shows that consonant subtraction coincides with word-medial deletion; that is, consonants that are subtracted word-finally will also be deleted in the same context when the sequence in question is followed by another vowel – the examples in (7d-f) are plural forms that contain a suffixal vowel, but relevant sequences are also absent in monomorphemic words, such as [vunə¹] ‘miracle’, whose Standard German cognate is *Wunder*, pronounced with [d].

3. A metrical analysis of tonal accent

In §2, I have shown a set of phenomena that are related to tonal accent, concerning the tonal realizations themselves, as well as additional phonetic correlates (durational differences), phonological correlates (interactions with consonant voicing), as well as morphophonological aspects of accentuation (Accent 1 as the morphological default; interaction of tonal accent with subtraction). In this section, I provide the basics of how my analytical approach to tonal accent aims to account for all of these different phenomena in a unified way. As briefly mentioned in the introduction, I follow what I have previously referred to as a *metrical approach* to tonal accent. The basic conceptual idea behind the metrical approach is to provide an analysis of tonal accent that does not only focus on the tonal contrasts but aims to incorporate all the different correlates of accent in a unified way. It should be noted that my specific analysis is not the only metrical framework that has been proposed for the analysis of tonal accent (other metrical work includes, e.g., Hermans, 2012, or Kehrein, 2017). Here I will focus on my own work without meaning to imply that my metrical analysis is necessarily superior to those of my colleagues.

My approach is based on the tenets of the autosegmental metrical theory, specifically on the general foot inventory that has been proposed in works like Hayes (1995). For instance, it is widely recognized that there is a difference between so-called syllabic trochees and moraic trochees, two types of feet where the left branch is strong (the head) and the right branch weak (the dependent). Feet thus typically have a head and a dependent, that is, they are preferably binary – the syllable

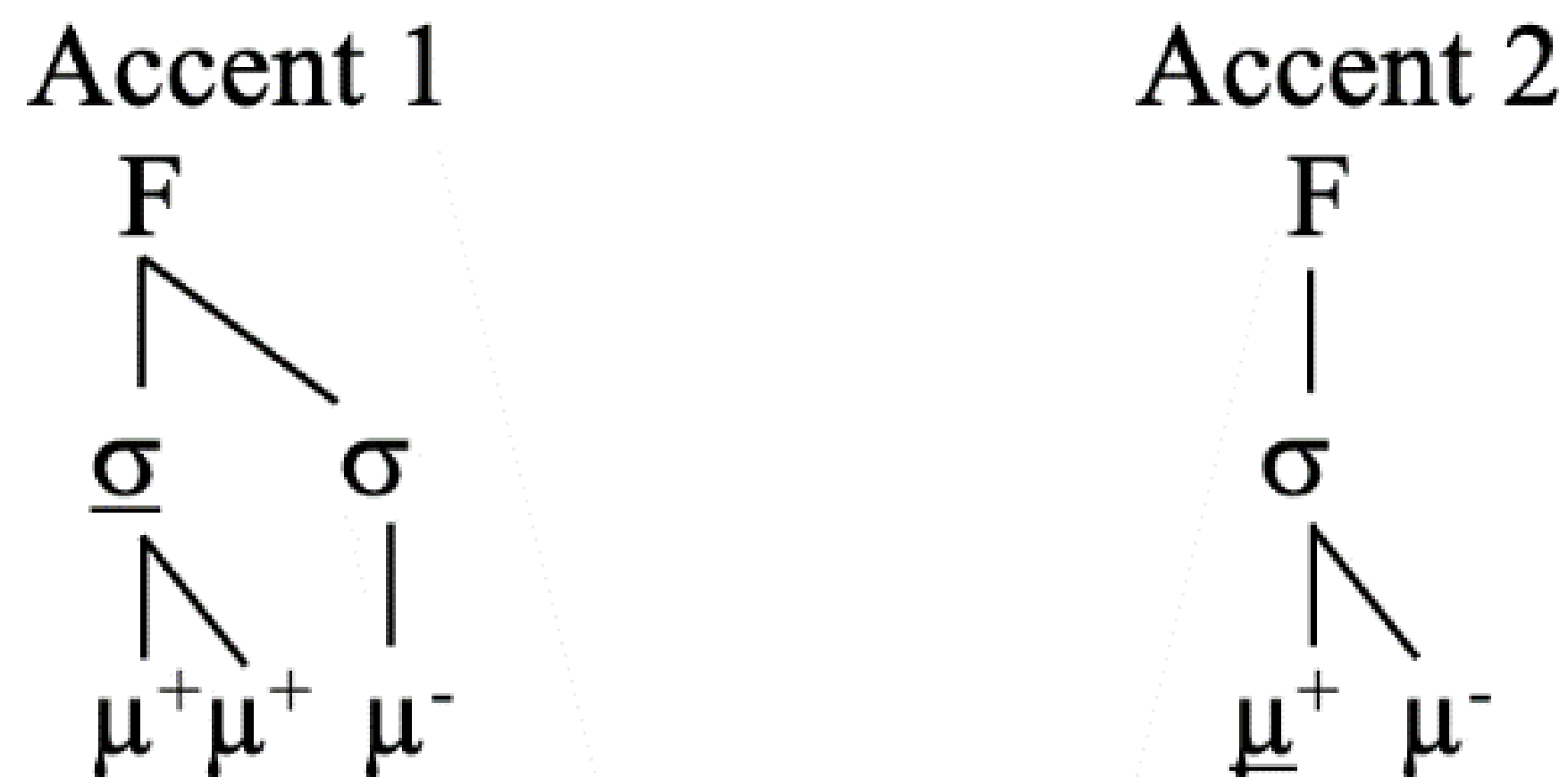
containing the foot head will be the stressed syllable. For syllabic trochees, which count syllables (hence the name *syllabic* trochee), binarity is established at the syllable level. For moraic trochees, which count moras (hence the name *moraic* trochee), binarity is established at the mora level. While the notion of the syllable surely needs no introduction, let me mention that moras are timing units located between the segment and the syllables: Short vowels have one mora, long vowels have two moras, and coda consonants may or may not have moras, depending on the language (Hayes, 1989). In Franconian, codas in stressed syllables can be moraic, and the maximal number of moras per syllable is two. In moraic theory, it is often stipulated that onset consonants cannot be moraic, and while there is some controversy surrounding this question, there is no evidence that onset consonants are moraic in Franconian.

In traditional, parameter-based autosegmental metrical theory, it is typically assumed that languages have exactly one type of foot in their inventory, either syllabic trochees, moraic trochees, or iambic feet (these feet are right-dominant). My work deviates from this assumption; specifically, I claim that languages can have more than one foot type, and Franconian is among these languages (in constraint-based theories like Optimality Theory, it is unsurprising that languages could have more than one foot). In what follows, I discuss the two-foot types I propose for Franconian and how they generally capture the contrastive tonal melodies (§3.1). In a second step, I demonstrate how my foot-based analysis makes it possible to account for the additional, non-tonal correlates of accent discussed in §2.

3.1 Foot inventory and tonal mapping

The main ingredient of my synchronic analysis of Franconian tonal accent is that Accent 1 and Accent 2 correspond to two types of feet (Köhnlein, 2011, 2016, 2017, 2018, 2019, 2020; Köhnlein & Cameron, 2020). Specifically, I claim that some items in Franconian are footed as moraic trochees, and this corresponds to Accent 2 on the surface. Other items, however, are footed as syllabic trochees, which corresponds to Accent 1. The relevant foot structures are given in (8). The trees contain a set of diacritics (underlining, superscript pluses and minuses), which I use for purposes of illustration below; yet they have no status in the theory.

(8) Syllabic trochees (Accent 1) and moraic trochees (Accent 2) in Franconian



First of all, what decides whether a foot is a syllabic trochee or an iambic trochee? I assume that the headedness of a foot is determined at the highest level where the foot branches, that is, the level where it becomes binary. Accent-1 feet are binary at the syllable level, and since the syllable level is higher than the mora level, the initial syllable will be the head of the foot (shown as underlined); the second syllable is the weak branch of the foot, i.e., the foot dependent. Accent-2 feet, on the other hand, contain only one syllable, so the foot cannot be binary at the syllable level, yet it does contain two moras, and it thus is binary at the moraic level. This, then, means that the foot will be a moraic trochee, with the first mora being the foot head (underlined) and the second mora being the foot dependent.

In a second step, I argue that the level of headedness (syllable for Accent 1 versus mora for Accent 2) has an influence on the prosodic 'strength' of the moras in the stressed syllable. Consider the syllabic Accent-1 foot first: The foot head, the first syllable, dominates two moras, and I claim that these moras inherit the metrical strength of the head – in (8, left side), I indicate this by giving them superscript pluses. The mora dominated by the second syllable is metrically weak, as it is linked to the dependent syllable of the foot; it thus receives a superscript minus. Now consider the moraic Accent-2 foot in (8, right side): Here, the first mora is the foot head, which makes it strong (superscript plus), and the second mora is the dependent of the foot, which makes it weak (superscript minus). When comparing the two feet, we can note that the syllabic trochee (= Accent 1) has a disyllabic domain, and both moras in the first syllable are metrically strong because they are dominated by the foot head, the first syllable. The moraic trochee (= Accent 2) is a monosyllabic, bimoraic domain, which means that one mora in the stressed syllable is strong and the other one is weak. Accordingly, Accent-1 syllables have two strong moras, and Accent-2 syllables have one strong and one weak mora – this difference is crucial for my analysis of tonal accent. To see how these foot structures map onto existing words, consider the examples in (9) from Cologne (Münch, 1904):

(9) Tone-accent word pairs from Cologne Franconian

- a. [da:x¹] ‘day-nom.’ [da:x²] ‘day-dat.’
b. [i:¹zə] ‘iron’ [ri:²sə] ‘to tear’

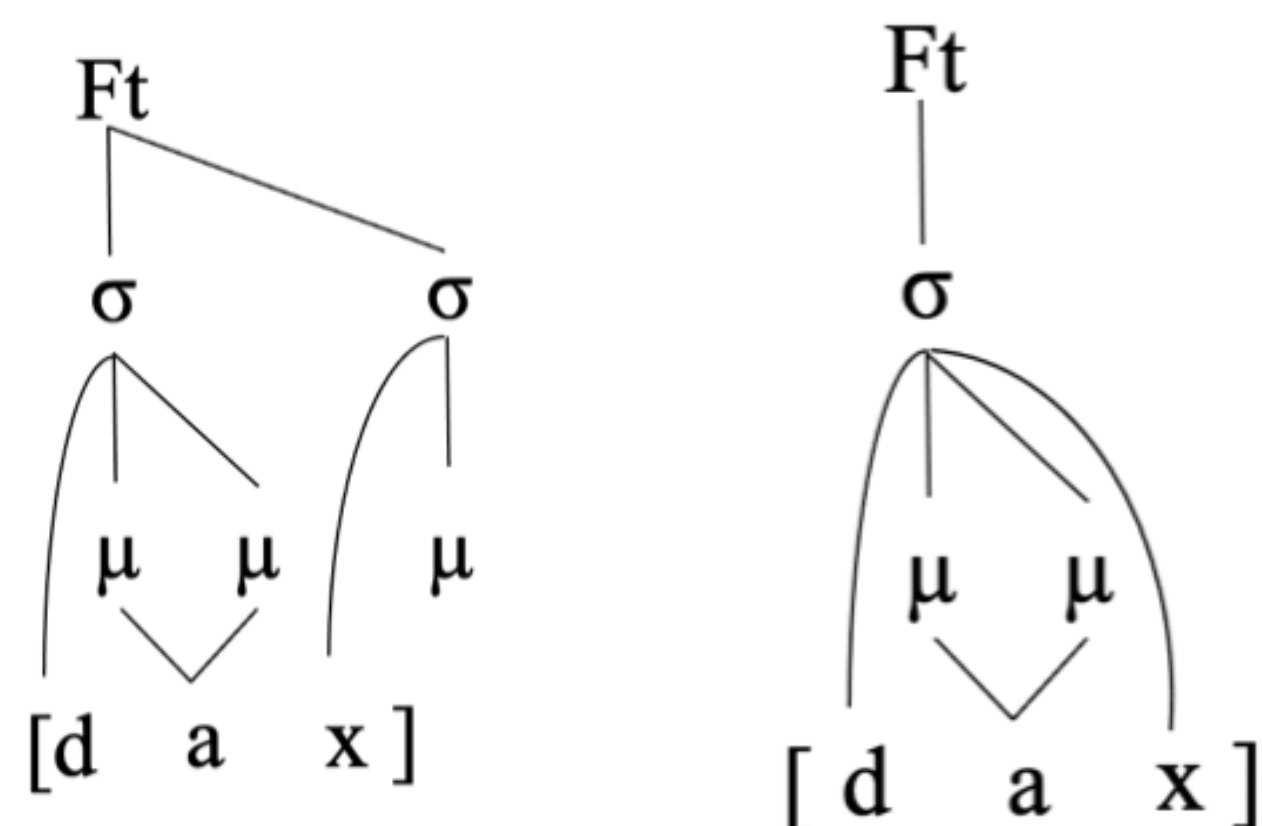
For disyllabic words, such as in (10b), modelling the contrast in footing is rather straightforward, as shown in (10), taken from Köhnlein & Cameron (2020). For items with Accent 1 (10, left side), both syllables are included in the foot, which is a syllabic trochee that spans two syllables. For items with Accent 2 (10, right side), only the first syllable is part of the foot, which leads to a moraic trochee that spans two moras but only one syllable; the second syllable remains unparsed by the foot. In the surface representation, this unparsed syllable will link directly to the prosodic-word node (not shown in the representations in (10)).

(10) Accent 1 (syllabic trochee, left) vs. Accent 2 (moraic trochee, right) for disyllabic words



The situation is somewhat more complex for words with only one vowel, such as the items in (11). Accent-2 items (11, right side) are footed as ‘default’ moraic trochees, i.e., one vowel corresponds to one syllable. While this is arguably straightforward, it is less obvious how disyllabic Accent-1 feet are mapped onto the relevant segmental structures. For such cases, I propose that monosyllabic words with Accent 1 are only *apparently* monosyllabic; yet in their metrical representation, they contain a so-called *empty-headed* second syllable, as shown in (11, left side). Essentially, an empty-headed syllable can be thought of to contain an unpronounced vowel, which I represent as a mora that is not linked to any segmental content. Undoubtedly, postulating such a representation assumes an additional level of abstraction, which should always be a reason for caution. I think it is justified based on various reasons, two of which I will mention here: First, while not being universally accepted in phonological theory, empty-headed syllables, as well as empty categories in general, have been successfully used for decades as analytical tools in phonological theory (e.g., Cavarani & van Oostendorp, 2017, for an overview). Second, Accent-1 items often behave in a similar fashion to overtly disyllabic words; I will discuss a few pieces of evidence in §3.2.

(11) Accent 1 (syllabic trochee, left) vs. Accent 2 (moraic trochee, right) for ‘apparently’ monosyllabic words



An obvious follow-up question is how to derive the tonal contrast from these two types of feet. I focus here on phrase-final declarative melodies in the Cologne dialects for purposes of illustration and ignore variation across dialects and across different prosodic and pragmatic contexts – the tonal aspects of the contrast in Rule A and Rule B are discussed in more detail in Köhnlein (2016), and Köhnlein (2011) contains a comprehensive analysis for four different types of Franconian dialects.

Recall that the basic idea in the metrical approach to tonal accent is that there is no lexical tone in Franconian; instead, the tonal contrasts between the accents are derived from differences in the association of post-lexical intonational tones. In the Cologne dialect, focus in declaratives is marked with a H*L intonational pitch accent (the asterisk indicates that H* is the tone that has to be associated with a stressed syllable), similar to what we know from prototypical intonational languages. The main rule for tone assignment in Cologne is that an intonational tone has to be licensed by a strong mora, that is, a mora that is connected to a foot head. Furthermore, a mora cannot host more than one tone.

Figure 2: Association of intonational tones to moras depending on foot structure in the Cologne dialect.

	Accent 1		Accent 2	
Declarative, focus, non-final				
	H* μ+	L μ+	H* μ+	μ-

The tonal mapping for Cologne is shown in Figure 2. For Accent 1, a syllabic trochee, the two moras in the stressed syllable are both linked to the foot head, which makes them metrically strong; as a consequence, each mora can host a tone, which means that Accent 1 will be realized as a falling

tone in the accented syllable. Accent 2, on the other hand, is a moraic trochee, so only the first mora (= the foot head) is strong, while the second mora (= the foot dependent) is weak. Since tones have to be licensed by strong moras, Accent-2 syllables can only host one tone, H*; the low tone following H* cannot be linked to the weak mora and will have to be realized later in the utterance, that is, on a following item. The second mora, however, receives the tone from the first mora via autosegmental spreading; the weak mora can thus share a tone with a strong preceding mora, but it cannot host a tone on its own. This leads to a high-level tone with a late fall towards the post-tonic low tone, which, in turn, derives the tonal opposition between Accent 1 and Accent 2 in Cologne.

3.2 Additional correlates of accent

3.2.1 Duration

As discussed in §2.2, various Franconian dialects have durational differences alongside the tone-accent contrast. In Cologne, where Accent-1 syllables are longer than Accent-2 syllables, the durational contrast can, in some contexts, even be the main correlate of the opposition rather than tone (Gussenhoven & Peters, 2004). Following Köhnlein (2016), I attribute the durational differences between the accents to the two types of feet. Inspired by previous work by Prince (1980) and Odden (1997) for Estonian, I assume that in some languages, a foot can have a certain duration associated with it. In a nutshell, if a foot has two syllables, the duration will be distributed across these two syllables. If a foot has one syllable, the duration of the foot will be expressed in only one syllable. Accordingly, we expect the stressed syllable of a monosyllabic foot to be longer than the stressed syllable of a disyllabic foot. When applied to Franconian, this means that stressed syllables of bimoraic, monosyllabic Accent-2 feet will be longer than stressed syllables of disyllabic Accent-1 feet, resulting in oppositions of the type [da:x¹] ‘day.dat’ (with an empty-headed second syllable whose duration is not expressed) vs. [da:x²] ‘day.nom’, or [i:zə] ‘iron’ (disyllabic foot, stressed syllable long) vs. [ri:²sə] ‘to tear’ (monosyllabic foot, stressed syllable overlong).

3.2.2 Vowel quality

Across dialects, there is a tendency for long vowels with Accent 1 to develop into diphthongs and for diphthongs with Accent 2 to become monophthongs. My account for this tendency is very similar to my analysis of the tonal mapping, under the assumption that diphthongs are the combination of two vowels in the same syllable. Along these lines, a diphthong like [ɛi] would be a combination of [ɛ] and [i]. Accent 1, which has two strong moras in the accented syllable, is preferably diphthongal so that each strong mora can dominate one vocalic node; this is similar to Accent 1 preferably having two tones (such as H*L) in the accented syllable. Accent 2, on the other hand, only has one strong mora, so it preferably dominates long vowels where the second, weak mora does not license its

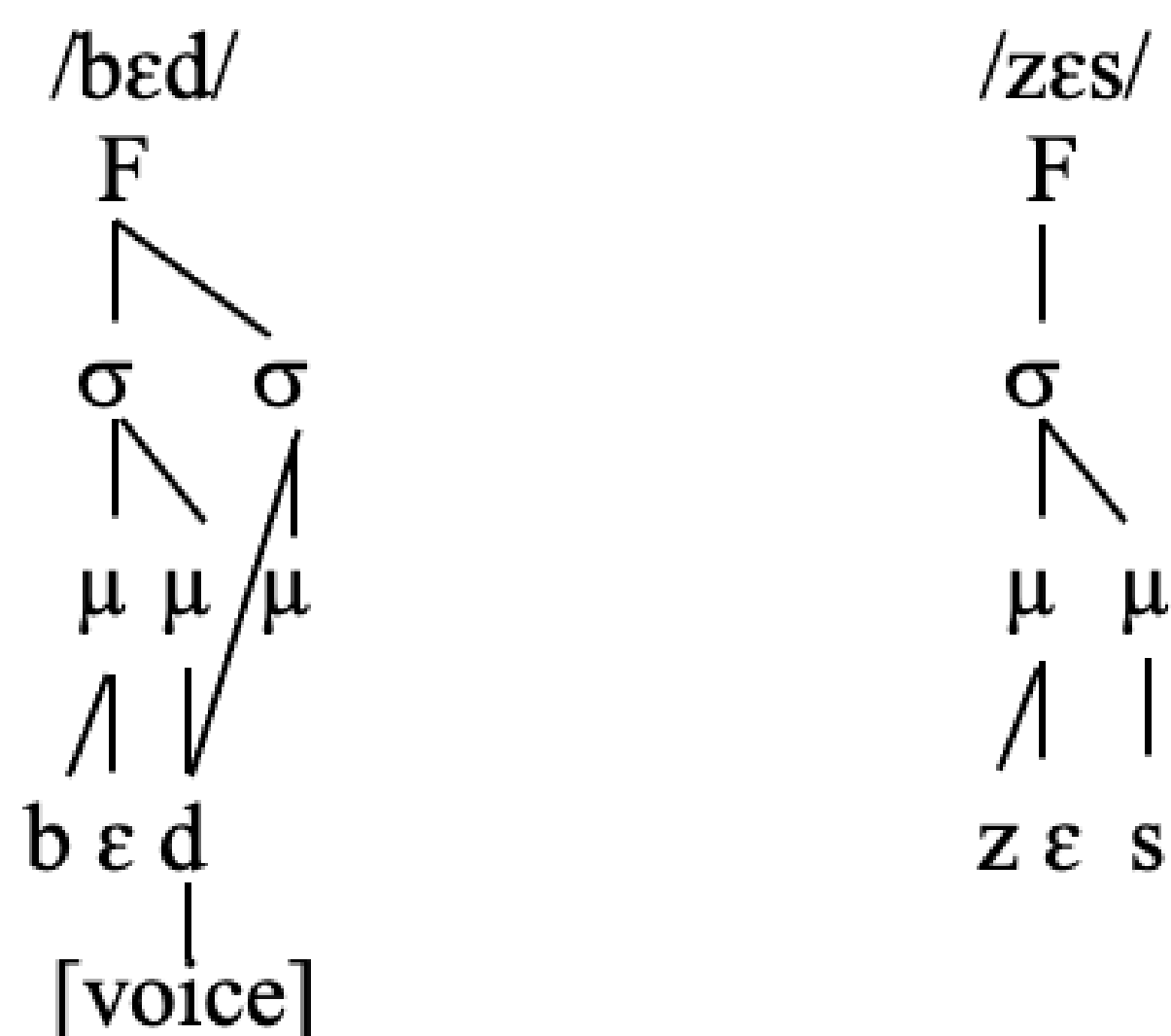
own vowel but receives the vowel from the first, strong mora via spreading; this is comparable to preferably having one tone (such as H*) in the accented syllable. This, then, results in alternations such as [blɛif¹] ‘stay.1.ps.sg’ vs. [bli:²və] ‘stay-inf.’ in the Maastricht dialect (see Köhnlein, 2017, for a detailed analysis).

3.2.3 Consonant voicing

It is well-established that there is a cross-linguistic tendency for foot-medial onsets to contain weaker, so-called *lenis* consonants than onsets that are not foot-medial; some examples from English concern foot-medial flapping, non-aspiration of plosives, the distribution of /h/ (Davis & Cho, 2003) or fricative voicing (Honeybone, 2012). These cross-linguistic observations can be straightforwardly implemented into my foot-based analysis of Franconian: Recall that in Cologne, disyllabic words with foot-medial voiced consonants always receive Accent 1, while words with word-medial voiceless consonants preferably have Accent 1. In my analysis, this follows from disyllabic feet (= Accent 1) preferably being built across a voiced consonant (which is weaker, i.e., more *lenis* than a voiceless consonant). Accordingly, an item like [i:¹zə] ‘iron’ will surface with Accent 1 (10, left side). Conversely, words with strong word-medial, voiceless consonants, such as [ri:²sə] ‘to tear’, preferably receive a monosyllabic foot that only covers the first syllable (10, right side). That way, the foot does not span across a strong, so-called *fortis* consonant – accordingly, the item will receive Accent 2 (see Köhnlein & Cameron 2020, in prep. for a more detailed analysis).

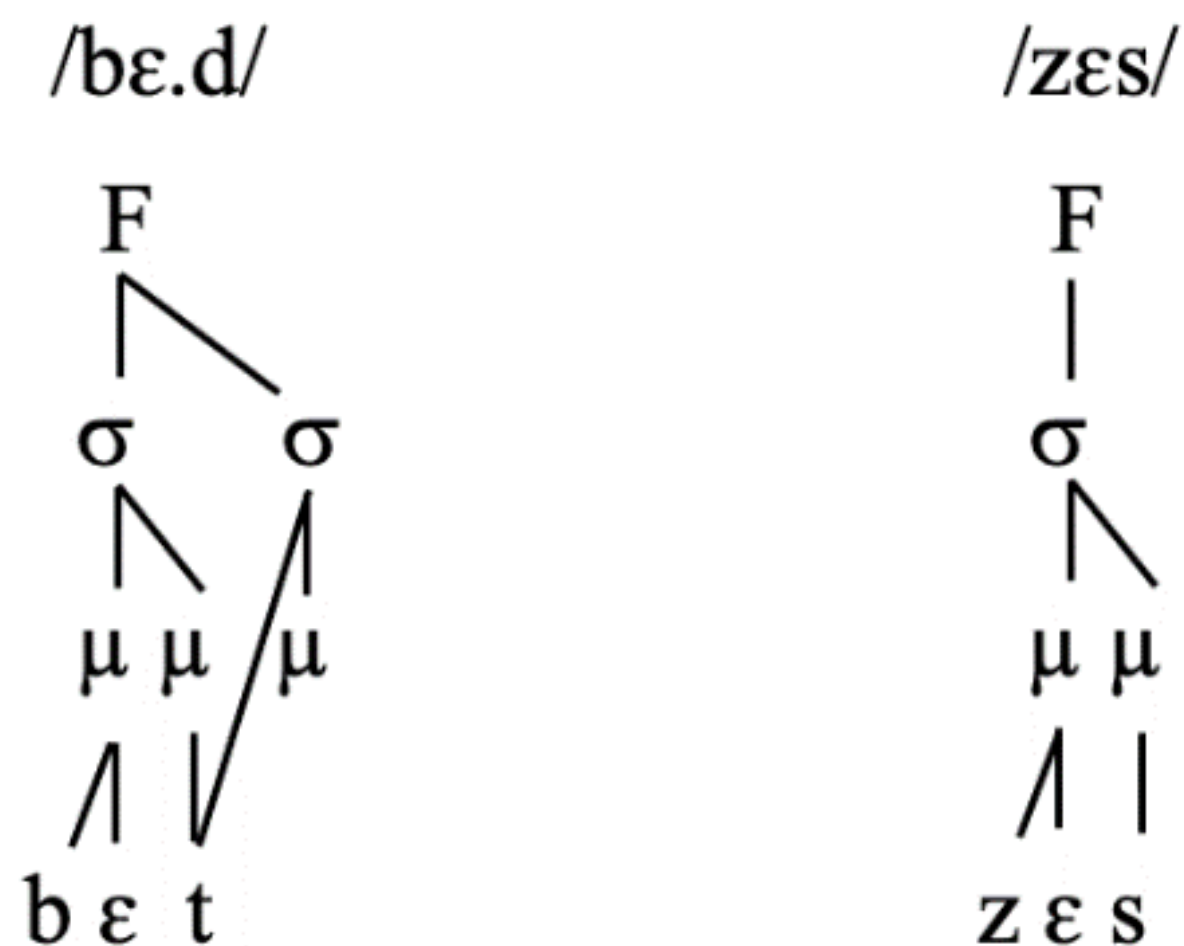
A related, if not identical analysis is possible for the Moresnet facts discussed in §2.2, where items with word-final voiced obstruents after short vowels receive Accent 1, and items with word-final voiceless obstruents receive Accent 2. Again, I follow the approach in Köhnlein (2016), which in turn is based on an analysis in van Oostendorp (2017). As with most dialects of Dutch, Moresnet has final devoicing, but it applies only at a later stage of the multi-level grammar. In the first step in (12, left), which corresponds to the root level, the final devoicing of voiced obstruents, as in /bɛd/’bed’, can be avoided by placing a word-final obstruent in the onset of an empty-headed syllable – at the root level of the grammar, final devoicing only applies to coda consonants. In items ending in voiceless consonants, such as /zɛs/ ‘six’ (12, right), the consonant can be syllabified as a coda – it is already voiceless anyway, so there is no ‘need’ to create an empty-headed syllable.

(12) Root-level representations in Moresnet



In a second step, consider the word-level representations in (13). The foot structure, inherited from the root level, is preserved; yet at this level, final devoicing affects all word-final obstruents. As discussed in van Oostendorp (2017), this difference between root and word-level corresponds to cross-linguistic observations, where final devoicing sometimes affects all coda consonants and sometimes only word-final consonants. As a result, /bɛd/ surfaces as [bɛt¹], but it still receives Accent 1 because of the disyllabic foot with an empty-headed syllable that was built at the root level (13, left). [zɛs²], on the other hand, stays monosyllabic, and this item thus surfaces with Accent 2 (13, right).

(13) Word-level representations in Moresnet

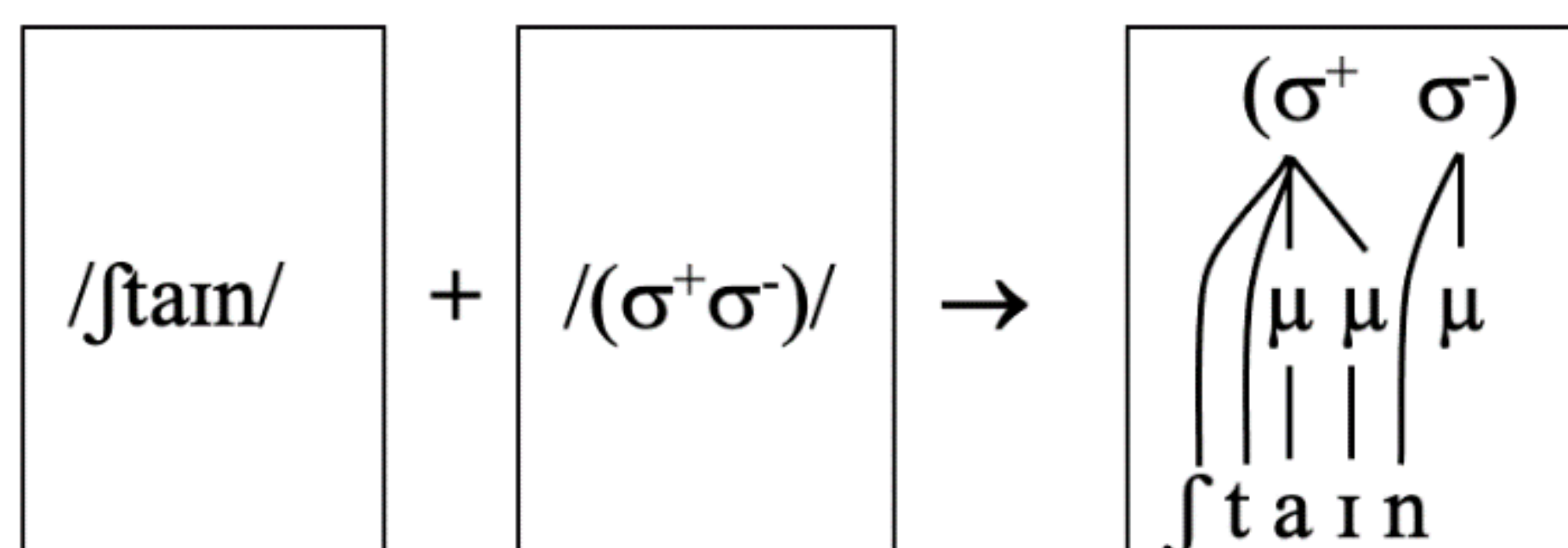


3.2.4 General morphological relevance and subtraction

As discussed in §2, Accent 1 is the morphologically marked accent. In my approach, this follows from the assumption that Accent 1, being a disyllabic trochee, has a ‘bigger’ foot than Accent 2, which comes with a monosyllabic, bimoraic trochee. In that sense, the bigger Accent-1 foot can incorporate and override the smaller Accent-2 foot. Representationally, we can think of morphologically conditioned Accent 1 as a foot is stored as a morpheme (such as a plural morpheme), which then

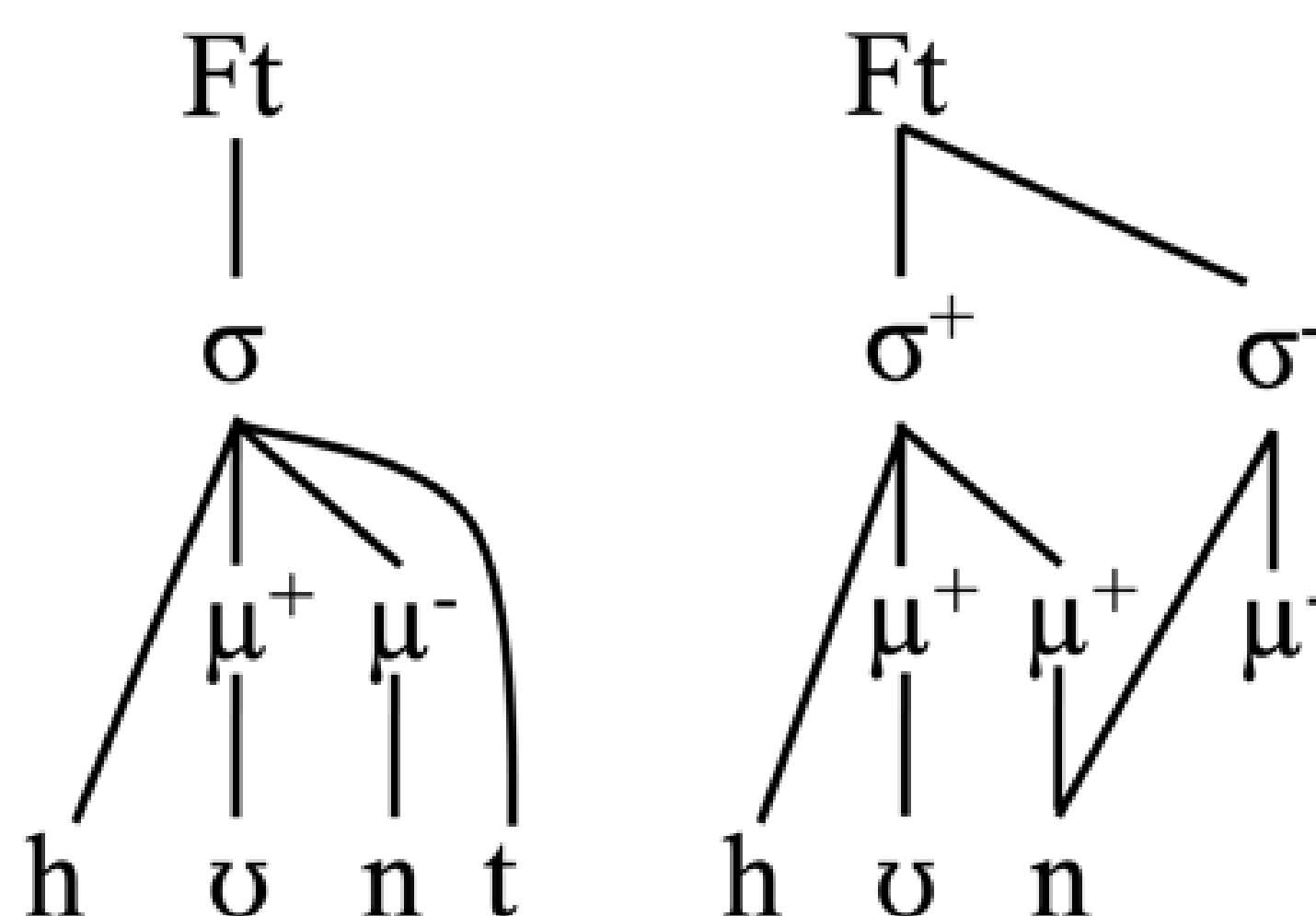
associates with segmental structure from a given stem. This is shown in (14) for the derivation of the Arzbach alternation between [ʃtaɪn²] ‘stone-sg’ and [ʃtaɪn¹] ‘stone-pl’. The lexical entry for ‘stone’ is /ʃtaɪn/ (left box in (14)), and as a default, this will be parsed as a monosyllabic foot, leading to Accent 2. The plural morpheme associated with /ʃtaɪn/ is a disyllabic, trochaic foot, indicated as / ($\acute{\sigma}^+$ $\acute{\sigma}^-$) / in the middle box. When the plural morpheme is combined with the segmental string, this leads to a disyllabic foot with an empty-headed syllable, as shown in (14, right box); this, of course, will correspond to Accent 1.

(14) Plural derivation of [ʃtaɪn¹] ‘stone-pl’ in Arzbach



Note that my foot-based analysis of morphological alternations can easily be extended to subtractive plurals. Recall another example from Arzbach, the alternation between [hʊnt²] ‘dog’ and [hʊn¹] ‘dogs’ (see §2.3 for more data). Similar to the plural for ‘stone’, the plural morpheme for ‘dogs’ is a disyllabic foot. The singular by itself will be parsed as a default bimoraic, monosyllabic foot and will thus receive Accent 2 (15, left). The plural, however, will have to incorporate the disyllabic foot of the plural morpheme, which leads to the structure in (15, right). Not only does the resulting item receive Accent 1, but this footing also leads to the deletion of the onset plosive in the second, empty-headed syllable because the dialect does not tolerate the nasal-plosive sequence if the plosive occurs in the onset of an empty-headed syllable. This then makes the nasal ambisyllabic (that is, it belongs to both the stressed syllable as a coda and the unstressed syllable as an onset).

(15) Subtractive morphology in Arzbach Franconian



The deletion of the word-final plosives in plurals finds a perfect parallel in overtly disyllabic words, as evidenced by the alternation between [lant²] ‘country’ and [lɛnɐ¹] ‘countries’. In the case at hand, the plural form is overtly disyllabic, and the word-medial plosive in the singular form is still deleted because it would occur at the onset of an unstressed syllable (Köhnlein, 2018, for a detailed analysis of subtraction across dialects).

4. Brief comparison and conclusion

In the preceding section, I have demonstrated how my foot-based approach to tonal accent makes it possible to account for multiple phenomena related to accentuation in a principled way. That is, my analysis captures not only the tonal mapping but also shows how the difference between monosyllabic feet (Accent 1) and disyllabic feet (Accent 2) interacts in predictable ways with durational differences between the accents and vowel alternations. It also straightforwardly accounts for interactions of accent with consonant voicing and captures morphological alternations, including subtractive morphology. Furthermore, the analysis of all the interactions in question are not arbitrarily associated with foot structure but follow from established cross-linguistic tendencies.

The available general analytical alternative, an approach with lexical tone (e.g., Gussenhoven & Peters, 2004), can successfully capture the contrastive tonal melodies. At least in its current version, I believe that it has nothing substantial to say about the synchronic relevance of any of the additional correlates, at least not from a phonological perspective (yet, see Gussenhoven & Peters, 2019, for a discussion of apparent arguments in favor of a tonal approach, and Köhnlein, 2019, for a brief reply). Interactions with duration have been referred to as phonetic enhancement of the lexical-tone contrast, but the enhancement approach itself is arbitrary and not tied to the structure of the accents in a meaningful way. Interactions with consonant voicing are difficult to capture in the tonal approach since surface voicing does not predict the quality of preceding tones: For instance, Cologne has falls in declaratives and rises in interrogatives for Accent 1, and there is no way to directly connect these opposite contours to the voicing of word-medial consonants. Furthermore, existing tonal approaches always mark Accent 2 lexically, which is at odds with the fact that morphological complexity correlates with Accent 1.

While I believe that my metrical analysis is considerably more adequate than the tonal approach, a possible counterargument to my analysis might be found in the fact that it relies on more abstract representations than the tonal alternative – just how much abstraction is justified or needed in phonological theory, however, is an open question. In addition to its empirical coverage, I would also like to point out that my analysis of Accent 1 as a disyllabic foot can be helpful in modeling that apparently monosyllabic Accent-1 words (which have an empty-headed second syllable) often behave similarly to overtly disyllabic words, as can, for instance, be observed in their behavior in subtracting

dialects. In this context, it should also be mentioned that my foot-based approach to Franconian tonal accent has strong parallels to metrical analyses of other languages with comparable accentual systems, such as, for instance, certain varieties of North Germanic (such as Morén-Duolljá, 2013; Iosad, 2016a, 2016b; Kaldhol & Köhnlein, 2021), Scottish Gaelic (such as Iosad, 2015; Morrison, 2019), or Uspanteko (Köhnlein, 2019; Köhnlein & Zhu, 2019), which I believe, strengthens my proposal from a typological perspective – even if different metrical analysis do not necessarily rely on the same set of representational tones. Assuming that the metrical approach to tonal accent is on the right track, future work will thus have to focus on narrowing down what the necessary and sufficient theoretical tools are to capture accentual contrasts across languages.

Acknowledgments

I would like to thank all organizers of the *Simpósio Internacional de Prosódia e Bilinguismo*, and specifically Amanda Post, for inviting me to present my work and including the present chapter in this volume. This research is supported by the National Science Foundation (BCS-1845107)

References

- BACH, A. Die Schärfung in der moselfränkischen Mundart von Arzbach (Unterwesterwaldkreis). *Beiträge zur Geschichte der Deutschen Sprache und Literatur*, n. 45, p. 266-290, 1921.
- BOTMA, B.; NOSKE, R. (ed.). *Phonological Explorations: Empirical, Theoretical and Diachronic Issues*. Berlin: Mouton de Gruyter, 2012.
- CAJOT, J. Phonologisch bedingter Polytonverlust: Eine tonlose Enklave südlich von Maastricht. In: VAAN, M. (ed.). *Germanic tone accents. Proceedings of the First International Workshop on Franconian Tone Accents. Zeitschrift für Dialektologie und Linguistik. Supplement 131*. Wiesbaden: Steiner, 2006. p. 11-23.
- CAVIRANI, E.; OOSTENDORP, M. VAN. On silent markedness. In: SAMUELS, B. (ed.). *Beyond Markedness in formal phonology*. Amsterdam: John Benjamins, 2017. p. 101-120.
- DAVIS, S.; CHO, M.-H. The distribution of aspirated stops and /h/ in American English and Korean: An alignment approach with typological implications. *Linguistics*, v. 41, p. 607-652, 2003.
- DOLS, W. *Sittardse diftongering*. Sittard, 1953.
- GUSSENHOVEN, C. The lexical tone contrast of Roermond Dutch in Optimality Theory. In: HORNE, M. (ed.). *Prosody: Theory and Experiment*. Amsterdam: Kluwer, 2000. p. 129-167.

GUSSENHOVEN, C. *Quantity or durational enhancement of tone: the case of Maastricht Limburgian high vowels*. 2012. p. 241-254.

GUSSENHOVEN, C.; PETERS, J. A tonal analysis of Cologne Schärfung. *Phonology*, v. 21, n. 2, p. 251-285, 2004.

GUSSENHOVEN, C.; PETERS, J. Franconian Tones Fare Better as Tones Than as Feet: A Reply to Köhnlein (2016). *Phonology*, v. 36, n. 3, p. 497-530, 2019.

HERMANS, B. The phonological representation of the Limburgian tonal accents. In: BOTMA, B.; NOSKE, R. (ed.). *Phonological Explorations: Empirical, Theoretical and Diachronic Issues*. Berlin: Mouton de Gruyter, 2012. p. 227-244.

HARDT, M. Vocalismus der Sauer-Mundart. In: *Programm des Königlich-Großherzoglichen Progymnasiums zu Echternach, herausgegeben beim Schlusze des Schuljahres 1842-43*. Trier: Fr. Lintz. 1-29, 1843.

HAYES, B. Compensatory lengthening in moraic phonology. *Linguistic Inquiry*, v. 20, p. 253-306, 1989.

HAYES, B. *Metrical stress theory: Principles and case studies*. Chicago: University of Chicago Press, 1995.

HONEYBONE, P. Lenition in English. In: NEVALAINEN, T.; TRAUGOTT, E. C. (ed.). *Handbook on the history of English: Rethinking approaches to the history of English*. Oxford: Oxford University Press, 2012. p. 773-787.

HYMAN, L. M. Word-prosodic typology. *Phonology*, v. 23, n. 2, p. 225-257, 2006.

HYMAN, L. M. How (not) to do phonological typology: the case of pitch-accent. *Language Sciences*, v. 31, n. 2, p. 213-238, 2009.

IOSAD, P. Pitch accent and prosodic structure in Scottish Gaelic: reassessing the role of contact. In: HILPERT, M.; DUKE, J.; MERTZLUFFT, C.; ÖSTMAN, J.-O.; RIEßLER, M. (ed.). *New Trends in Nordic and General Linguistics*. Berlin: Mouton de Gruyter, 2015. p. 28-54.

IOSAD, P. Prosodic structure and suprasegmental features: Short-vowel stød in Danish. *Journal of Comparative Germanic Linguistics*, v. 19, n. 3, p. 221-268, 2016a.

IOSAD, P. Tonal stability and tonogenesis in North Germanic. In: GILES, I.; CHAPOT, L.; COOIJMANS, C.; FOSTER, R.; TESIO, B. (ed.). *Beyond Borealism: New perspectives on the North*. London: Norvik Press, 2016b. p. 80-98.

JONGEN, R. *Phonologie der Moresneter Mundart: eine Beschreibung der segmentalen und prosodischen Wortformdiakrise*. Assen: Van Gorcum, 1972.

KALDHOL, N. H.; KÖHNLEIN, B. North Germanic Tonal Accent is Equipollent and Metrical: Evidence from Compounding. *Proceedings of the 2020 Annual Meeting on Phonology*. 2021.

KEHREIN, W. There's no tone in Cologne: against tone segment interactions in Franconian. In: KEHREIN, W.; KÖHNLEIN, B.; BOERSMA, P.; OOSTENDORP, M. van (ed.). *Segmental structure and tone*. Berlin: Walter de Gruyter, 2017. p. 147-194.

KÖHNLEIN, B. *Rule reversal revisited: Synchrony and diachrony of tone and prosodic structure in the Franconian dialect of Arzbach*. 2011. Utrecht: LOT. Dissertation Series 274.

KÖHNLEIN, B. Contrastive foot structure in Franconian tone accent dialects. *Phonology*, v. 33, n. 1, p. 87-123, 2016.

KÖHNLEIN, B. Synchronic alternations between monophthongs and diphthongs in Franconian tone accent dialects: a metrical approach. In: KEHREIN, W.; KÖHNLEIN, B.; BOERSMA, P.; OOSTENDORP, M. (ed.). *Segmental structure and tone*. Berlin: Walter de Gruyter, 2017. p. 211-236.

KÖHNLEIN, B. Why metrical approaches to tonal accent are worth pursuing. *Phonology*, v. 36, n. 3, p. 531-535, 2019.

KÖHNLEIN, B. Metrically conditioned pitch accent in Uspanteko. In: GOEDEMANS, R.; HEINZ, J.; VAN DER HULST, H. (ed.). *The study of word stress and accent: theories, methods and data*. Cambridge: Cambridge University Press, 2019. p. 293-322.

KÖHNLEIN, B. Tone accent in North and West Germanic. In: PAGE, R.; PUTNAM, M. (ed.). *The Cambridge Handbook of Germanic Linguistics*. Cambridge: Cambridge University Press, 2020. p. 143-166.

KÖHNLEIN, B.; CAMERON, I. A foot-based approach to interactions of tonal accent and consonantal strength. *Proceedings of NELS 50*, v. 2, p. 119-132, 2020.

KÖHNLEIN, B.; ZHU, Y. Restricting the Power of Cophonologies: A Representational Solution to Stem Allomorphy in Uspanteko. *Proceedings of the 2018 Annual Meeting on Phonology*. San Diego, 2019.

MORRISON, D. A. Metrical structure in Scottish Gaelic: tonal accent, glottalisation and overlength. *Phonology*, v. 36, n. 3, p. 391-432, 2019.

MÜNCH, F. *Grammatik der ripuarisch-fränkischen Mundart*. Bonn: Cohen, 1904.

ODDEN, D. Some theoretical issues in Estonian prosody. In: LEHISTE, I.; ROSS, J. (ed.). *Estonian prosody: Papers from a symposium*. Tallinn: Institute of Estonian language, 1997. p. 165-195.

OOSTENDORP, M. V. Tone, Final Devoicing and Assimilation in Moresnet. In: KEHREIN, W.; KÖHNLEIN, B.; BOERSMA, P.; OOSTENDORP, M. (ed.). *Segmental Structure and Tone*. Berlin: Walter de Gruyter, 2017. p. 237-252.

PETERS, J. The Cologne word accent revisited. In: DE VAAN, M. (ed.). *Germanic Tone Accents. Proceedings of the First International Workshop on Franconian Tone Accents, Leiden, 13-14 June 2003*. Zeitschrift für Dialektologie und Linguistik – Beiheft 131. Stuttgart: Franz Steiner Verlag, 2006. p. 107-133.

PRINCE, A. Quantitative Consequences of Rhythmic Organization. In: DEATON, K.; NOSKE, M.; ZIOLKOWSKI, M. (ed.). *CLS 26-II: Papers from the Parasession on the Syllable in Phonetics and Phonology*. Chicago: Chicago Linguistic Society, 1980. p. 355-398.

SCHMIDT, J. E. *Die Mittelfränkischen Tonakzente (Rheinische Akzentuierung)*. Mainzer Studien zur Sprach- und Volksforschung 8. Stuttgart: Franz Steiner Verlag, 1986.

WIESINGER, P. *Phonetisch-phonologische Untersuchungen zur Vokalentwicklung in den deutschen Dialekten*. Bd. 1: Die Langvokale im Hochdeutschen. Bd. 2: Die Diphthonge im Hochdeutschen. *Studia Linguistica Germanica*. 2. Berlin: Walter de Gruyter, 1970.

Rumo a uma análise abrangente do acento tonal do francônio

Björn Köhnlein

I. Introdução

O acento tonal do Francônio, falado em partes da Bélgica, Alemanha e Holanda, é uma oposição entre dois acentos de palavras, que são comumente referidos como Acento 1 e Acento 2. Forneço dois pares mínimos em (1), um entre itens monossilábicos (1a), um entre itens dissilábicos (1b). Os acentos são indicados com sobrescritos no final da sílaba que carrega o acento; essas sílabas são sempre tônicas, pois o acento tonal é restrito às sílabas tônicas.

Pares mínimos de acento tonal de Mayen (SCHMIDT, 1986):

- a. [man¹] ‘cesto’ [man²] ‘homem’
- b. [tɔu¹və] ‘pombas’ [tɔu²və] ‘batismo’

O fenômeno tem sido estudado há muito mais de um século, sendo a fonte mais antiga conhecida Hardt (1843). Durante esta longa história de pesquisa, tem sido difícil definir os aspectos fonéticos e fonológicos do acento tonal. Embora nossa compreensão das características fonéticas do acento tenha melhorado consideravelmente nas últimas décadas, ainda há um considerável desacordo sobre qual deve ser a análise fonológica sincrônica ‘correta’ do acento tonal (isso vale para a análise diacrônica do acento tonal, que não discutirei aqui; veja Köhnlein, 2020 para uma visão geral).

A principal razão pela qual a fonologia dos acentos ainda é um tanto elusiva pode ser atribuída à sua complexidade fonética, o que os torna particularmente interessantes para nossa compreensão da tipologia palavra-prosódica, cuja parte central é o estudo do tom e do acento entre as línguas. Dentro da tipologia prosódica, muitas vezes se sustenta que há uma diferença substancial entre tom e acento, com diferenças claras entre esses dois tipos de propriedades fonológicas (por exemplo, HYMAN, 2006, 2009). Sistemas tom-acentuados como o do Francônio, no entanto, não são necessariamente um ajuste direto para essas categorias tipológicas, pelas mesmas razões que as tornam difíceis de analisar em primeiro lugar.

Intuitivamente, parece apropriado dizer que o acento tonal do francônio pode ser colocado em algum lugar entre o tom e o acento. Por um lado, dialetos relevantes parecem ser sem dúvida tonais, no sentido de que os dois acentos são geralmente realizados com diferentes melodias tonais, como um tom descendente para o Acento 1 e um tom de nível alto para o Acento 2 em declarações (há mais variação, algumas das quais discutirei no §2), mas a ocorrência do contraste tonal é restrita às sílabas tônicas e tende a desaparecer em posições de frase não proeminentes. Além disso, as diferenças tonais entre o Acento 1 e o Acento 2 são tipicamente acompanhadas por outros correlatos fonéticos, cujos detalhes podem diferir entre os dialetos: como realizações ditongais *versus* monotongais dependendo do acento), se uma palavra recebe Acento 1 ou Acento 2 pode interagir com a força das consoantes seguintes, etc. Tais diferenças na duração, bem como na qualidade das vogais e consoantes, são prototípicas de sistemas acentuais, e não de sistemas tonais. Além disso, sabe-se

que o tom entoacional pode ser relevante para a marcação fonética de sílabas tônicas em línguas com sistemas puramente entonacionais, como o inglês ou o português. Isso sugeriria, então, que o francônio tem apenas um intrincado sistema de acentos, em vez de um contraste de tom lexical?

Nesse contexto, há outra consideração adicional importante que discuti apenas implicitamente até agora: quando os linguistas falam sobre a localização do acento, eles normalmente se referem à sílaba como a unidade que carrega o acento (às vezes, as pessoas ainda se referem, por exemplo, a vogais tônicas, mas isso normalmente inclui implicitamente também as consoantes na mesma sílaba). De fato, há um amplo consenso de que o acento é uma propriedade das sílabas, não de unidades abaixo do nível da sílaba (HYMAN, 2006, 2009). Se este for de fato o caso, de modo interlinguístico, seria pouco costumeiro em termos de análise assumir uma diferença de acento entre, digamos, a primeira ou a segunda parte de uma vogal longa – de fato, tais contrastes deveriam ser impossíveis se o acento fosse sempre uma propriedade das sílabas. Com essa generalização em mente, outra olhada nos exemplos em (1) revela porque seria controverso descrever o acento tonal do francônio como uma oposição fonológica puramente baseada em acento: o contraste entre o Acento 1 e o Acento 2 ocorre dentro de sílabas, que é em menos atípico (ou talvez até impossível?) em sistemas de acento, mas perfeitamente esperado em sistemas de tom lexical. Então, esse contraste é baseado no tom fonológico? Ou dois tipos de acento? Ou outra coisa?

Com base em uma breve visão geral de alguns dos correlatos mais importantes do acento tonal e do papel morfológico do acento no §2, pretendo fornecer uma introdução aos princípios das abordagens analíticas do acento tonal do francônio, com foco no que me refiro como a abordagem métrica. Em poucas palavras, existem atualmente duas abordagens analíticas concorrentes. Por um lado, o acento tonal foi analisado como um contraste entre um acento que carrega um tom lexical (tipicamente Acento 2) *versus* um acento lexicalmente sem tom (tipicamente Acento 1). Esta é a análise tradicional em fonologia teórica, que chamo de abordagem tonal. Ao longo da última década, no entanto, uma abordagem métrica alternativa para acento foi proposta. Os estudiosos que seguem a abordagem métrica normalmente assumem que há uma diferença estrutural mais abstrata entre os dois acentos, como dois tipos de pés métricos. Discutirei os fundamentos da abordagem métrica no §3, tomando meu próprio trabalho como exemplo concreto. No §4, vou compará-lo brevemente com a abordagem tonal e concluir o texto.



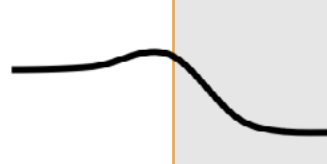
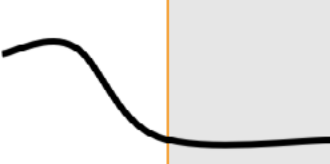
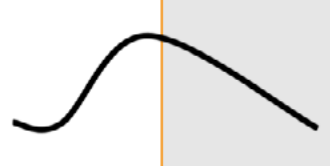


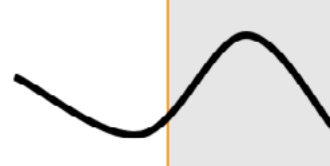
2. Revisão da literatura sobre Francônio

Esta seção fornece algumas informações sobre o acento tonal do francônio e discute os fenômenos que – pelo menos na minha opinião – devem ser capturados de uma forma baseada em princípios em uma análise abrangente. Começo discutindo princípios distributivos e melodias tonais básicas (§2.1), continuo com uma breve revisão de correlatos fonéticos e fonológicos adicionais entre dialetos (§2.2) e, em seguida, foco em algumas das funções morfológicas do acento tonal (§2.3).

2.1 Melodias tonais básicas

Na maioria dos dialetos, a realização do contraste tom-acentos requer uma vogal longa/ ditongo ou uma sequência de uma vogal curta mais uma consoante de coda soante. Sequências de vogais curtas e obstruintes de coda, por outro lado, normalmente não apresentam contraste acentuado. A Figura 1 (lado esquerdo) exibe as realizações tonais de Acento 1 e Acento 2 no dialeto de Colônia (por exemplo, GUSSENHOVEN; PETERS, 2004, PETERS, 2006), um acento prototípico do dialeto de tom, em posição de foco não final de enunciados declarativos e interrogativos. Deve-se notar que pode haver variação considerável na realização dos acentos, dependendo da posição na frase (não final, final), proeminência (focal, pré-focal, pós-focal) e contexto pragmático (além das declarativas e interrogativas, pode haver outro contexto, como a entonação de continuação); além disso, o contraste do acento é muitas vezes neutralizado fora de foco. Para os propósitos desta visão geral, no entanto, concentro-me apenas nas realizações padrão mais básicas.

Figura 1: Contornos tonais idealizados em Colônia (Regra A) e Arzbach (Regra B), foco, posição não final; contorno nuclear não sombreado, contorno pós-nuclear geral sombreado.

Contexto	Colônia (Regra A)		Arzbach (Regra B)	
	Acento 1	Acento 2	Acento1	Acento 2
Declarativa, não-final				
Interrogativa, não-final				

Conforme mostrado na Figura 1, os contornos de tom focal para o Acento 1 são uma queda inicial nas frases declarativas e um tom ascendente inicial nas frases interrogativas; movimentos tonais de Acento-2 ocorrem relativamente mais tarde, com tom de nível alto na sílaba acentuada para declarativas e uma queda tardia na sílaba pós-tônica (mostrada como sombreado em cinza na Figura 1), e tom de nível baixo e um tom ascendente tardio para Acento 2. A principal diferença entre os contornos de Acento-1 e os contornos de Acento-2 em Colônia está no tempo dos movimentos de afinação relevantes – cedo para Acento 1, tarde para Acento 2.

Dialetos com realizações tonais dos acentos como os de Colônia têm sido referidos como Regra A na literatura, uma classificação baseada em Wiesinger (1970). Há variação da realização tonal entre os dialetos, muitas vezes menores, mas às vezes consideráveis. O tipo de variação indiscutivelmente mais extremo ocorre nos chamados dialetos de Regra-B, conforme descrito pela primeira vez por Bach (1921) para o francônio de Arzbach. Como mostra a Figura 1 (lado direito), Arzbach tem melodias tonais opostas em relação à Colônia, pelo menos nas declarativas: o Acento 1 é produzido com um tom nivelado e uma queda tardia, ao contrário de Colônia, onde o Acento 1 tem uma queda precoce;

inversamente, o Acento 2 é realizado como um declínio inicial em Arzbach, enquanto realizado como declínio tardio em Colônia. Embora Köhnlein (2011) geralmente confirme as observações de Bach, também as relativiza: meu trabalho de campo demonstrou que a inversão é restrita a declarativas, mas não pode ser observada em interrogativas. De fato, as melodias nas interrogativas são muito semelhantes entre Colônia e Arzbach. Em ambos os dialetos, o Acento 1 surge antes do Acento 2. Em meu trabalho, me refiro a isso como uma semi-reversão de melodias tonais. Os fatos de Arzbach são tipologicamente peculiares de várias perspectivas: Geralmente, tais inversões tonais entre dialetos intimamente relacionados são muito raras e extremamente raras (ainda ver Kingston (2011) para Athabaskan). Além disso, o dialeto de Arzbach parece ser o único idioma documentado até o momento que exibe uma inversão de alinhamento de acentos de palavras em um dialeto: isto é, o Acento 1 cai depois do Acento 2 em declarativas, mas mostra um tom ascendente mais cedo do que o Acento 2 em interrogativas.

2.2 Correlatos fonéticos e fonológicos adicionais

Normalmente, o tom parece ser o principal correlato fonético da oposição tom-acento nos dialetos; ainda outros fatores podem desempenhar um papel na realização dos acentos. Aqui, discutirei brevemente os efeitos da duração, qualidade das vogais e vocalização das consoantes. Entre esses correlatos, a duração é o correlato mais consistentemente atestado em todos os dialetos. Foneticamente, o que normalmente encontramos é que os tons de nível são mais longos do que os tons decrescentes, o que na maioria dos dialetos significa que o Acento 2 é mais longo que o Acento 1. Intuitivamente, isso pode parecer um tanto surpreendente, já que está bem estabelecido que produzir um tom de contorno exigirá mais tempo do que produzir um tom de nível; no entanto, isso só é óbvio para durações muito curtas. Conforme argumentado em Köhnlein (2015), essa relação pode ser invertida (pelo menos *ceteris paribus*) para durações mais longas, como vogais longas ou sequências de vogais curtas e codas sonorantes.

Para os dialetos de Colônia, Gussenhoven e Peters (2004) relatam que a duração pode ser o principal correlato da oposição tom-acento em posições pós-focais onde o contraste tonal não é distintivo, mas onde o Acento 2 (tom nivelado) ainda é consideravelmente mais longo do que o Acento 1 (tom descendente), particularmente na posição final de uma frase. Um exemplo é fornecido em (2), onde o Acento 1 é marcado como longo e o Acento 2 como super longo:

- (2) Diferenças de duração entre Acento 1 e Acento 2 nos dialetos de Colônia
a. [da:x¹] 'day.dat' [da::x²] 'day.nom'

Alguns dialetos também mostram diferenças na qualidade das vogais. No padrão mais difundido, vogais com tons de contorno (tipicamente Acento 1) correspondem a vogais mais ditongais, e vogais com tons de nível (tipicamente Acento 2) correspondem a vogais mais monotongais (por exemplo,

DOLS, 1953; WIESINGER, 1970; CAJOT, 2006; GUSSENHOVEN, 2012; KOHNLEIN, 2017). Dois exemplos de Maastricht (Gussenhoven 2012), onde vogais altas formam ditongos se tiverem Acento 1, são dados em (3).

(3) Diferenças na qualidade da vogal entre o Acento 1 e o Acento 2 no francônio de Maastricht

- | | |
|--|------------------------------------|
| a. [bleif ¹] ‘ficar.1.PS.SG’ | [bli: ² və] ‘ficar-INF’ |
| b. [dɔuf ¹] ‘pomba.SG’ | [du: ² və] ‘pomba-PL’ |

Por fim, a acentuação pode interagir com a qualidade de vozeamento das obstruintes do meio e do final da palavra. Isso não é surpreendente de uma perspectiva histórica, uma vez que a sonoridade consonantal é um importante preditor de uma palavra receber o Acento 1 ou o Acento 2 após a gênese do acento tonal (Köhnlein, 2020 para visão geral e referências). Em muitos dialetos, processos fonológicos posteriores (como o enfraquecimento da consoante medial da palavra ou o enfraquecimento final) obscureceram essas interações ao longo do tempo, mas alguns dialetos ainda preservam interações relevantes. Um desses dialetos é (novamente) o francônio de Colônia, onde as palavras dissílabas com consoantes sonoras no meio da palavra sempre recebem o Acento 1, e as palavras com consoantes surdas no meio da palavra normalmente recebem o Acento 2 (mesmo que possam receber o Acento 1 em alguns casos). Alguns exemplos, retirados de Münch (1904), são fornecidos em (4). Os exemplos (4a-c), que apresentam palavras dissílabas com consoantes intervocálicas surdas, sempre aparecem com Acento 1, enquanto os itens (4d-f), que contêm consoantes intervocálicas surdas, aparecem com Acento 2:

(4) Interação do acento e vocalização da consoante medial de palavra em Colônia (MÜNCH, 1904)

- | | |
|--------------------------------------|---------------------------------------|
| a. [o: ¹ ʏə] ‘olho’ | d. [lo: ² fə] ‘correr’ |
| b. [fre: ¹ zə] ‘congelar’ | e. [ri: ² sə] ‘lacrimejar’ |
| c. [drø: ¹ mə] ‘sonhar’ | f. [ʃli: ² sə] ‘dividir’ |

Para dar outro exemplo de interações entre o acento tonal e a sonoridade consonantal, o dialeto Moresnet mostra alternâncias sincrônicas previsíveis entre o Acento 1 e o Acento 2 com base na sonoridade das consoantes pós-vocálicas. Especificamente, Moresnet está entre aqueles dialetos que exibem uma oposição de acento não apenas em vogais longas/ditongos e vogais curtas mais sonorantes, mas também em vogais curtas mais obstruintes (JONGEN, 1972). Conforme discutido em van Oostendorp (2017), sequências de vogais curtas mais obstruintes vozeadas subjacentes receberão o Acento 1, enquanto sequências de vogais curtas mais obstruintes desvozeadas subjacentes receberão o Acento 2. Na realização de superfície, no entanto, esse contraste de sonoridade pode ser neutralizado pela sonoridade final, um processo fonológico comum nos dialetos holandeses e alemães. Isso é demonstrado em (5). (5a) mostra que vogais curtas seguidas de obstruintes sutilmente sonoras têm Acento 1, mesmo que a obstruinte seja surda; (5b) demonstra que vogais curtas mais obstruintes desvozeadas recebem Acento 2.

(5) Interação de acento e vozeamento subjacente em Moresnet

a. /bɛd/ [bɛt¹] ‘cama’ /bɛd/ + /ə/ [bɛ¹də] ‘camas’

b. /zɛs/ [zɛs²] ‘seis’ /tɛpəx/ [tɛ²pəx] ‘carpete’

2.3 Relevância morfológica do acento tonal

Em todos os dialetos, os acentos de certos itens terão que ser memorizados pelos falantes, pois são sincronicamente imprevisíveis. Por exemplo, a diferença entre [man¹] ‘cesto’ e [man²] ‘homem’, que apresentei em (1), provavelmente terá que ser aprendida. Além dos correlatos fonéticos e fonológicos discutidos no §2.2, no entanto, existem também certas generalizações morfológicas sincrônicas. Um padrão encontrado em todos os dialetos diz respeito à alternância entre o Acento 1 e o Acento 2 em pares mínimos de acentos morfológicamente condicionados: nesses casos, a forma morfológicamente complexa sempre receberá o Acento 1, enquanto a forma morfológicamente simples sempre receberá o Acento 2. Forneço dois exemplos em (6), retirados de meu próprio trabalho de campo sobre o dialeto de Arzbach:

(6) Mudanças de Acento 2 para Acento 1 em alternâncias morfológicas

a. [ʃta:n²] ‘pedra’ [ʃta:n¹] ‘pedras’

b. [ət is brait²] ‘isto é amplo’ [ən brait¹ haus²] ‘um.NOM amplo.NOM
casa’

(6a) mostra uma alternância entre uma forma singular com Acento 2 *versus* um plural com Acento 1; (6b), por outro lado, fornece diferentes formas do adjetivo que significa ‘amplo’. Aqui, a forma predicativa mais básica recebe o Acento 2, enquanto a forma atributiva singular neutra tem o Acento 1. Geralmente, podemos considerar as formas atributivas como mais complexas do que suas contrapartes predicativas, uma vez que às vezes também contêm material segmentar adicional, dependendo do gênero ou pluralidade – como exemplo, considere a forma de Arzbach [brai¹də], que é a realização de ‘amplo’ para o singular masculino e todas as formas plurais. Para encerrar esta seção, mencionarei mais um exemplo da relevância morfológica do acento que não diz respeito aos pares mínimos tom-acento *per se*, mas ainda indica uma interação previsível de estrutura morfológica e acentuação. Uma variedade de dialetos alemães exclui certas plosivas de finais de palavras em formas plurais, entre eles muitos dialetos de acento tonal do Francônio. Alguns exemplos de Arzbach são fornecidos em (7a-c); como pode ser observado, em todos esses casos, as formas singulares morfológicamente simples contêm um final de palavra [t], que é suprimido nas formas plurais correspondentes. Na literatura morfológica, as alternâncias em que as formas simples contêm mais informações fonológicas do que as formas mais complexas são chamadas de subtração. Ou seja, parece que os respectivos morfemas – neste caso o morfema plural – removem o conteúdo fonológico, ao invés de adicioná-lo.

(7) Subtração consonantal e deleção medial de palavras no francônio de Arzbach

a. [hʊnt ²]	‘cão’	[hʊn ¹]	‘cães’
b. [vant ²]	‘parede’	[vɛn ¹]	‘paredes’
c. [hant ²]	‘mão’	[hɛn ¹]	‘mãos’
d. [kɪnt ²]	‘criança’	[kɪnɛ ¹]	‘crianças’
e. [lant ²]	‘país’	[lɛnɛ ¹]	‘países’
f. [rɪnt ²]	‘vaca’	[rɪnɛ ¹]	‘vacas’

Com relação ao acento tonal, a subtração é relevante, pois as respectivas formas plurais não apenas perdem uma consoante final presente nos singulares correspondentes, mas também mudam o acento do item de Acento 2 para Acento 1. Fundamentalmente, esse padrão não é uma idiosincrasia do Arzbach mas é verdade em todos os dialetos. Embora obviamente não possamos conhecer todas as formas em todos os dialetos, parece não haver contraexemplos conhecidos para a generalização de que a subtração coincide com uma mudança de Acento 2 para Acento 1. Algumas notas mais gerais sobre subtração nos dialetos do francônio. Primeiro, quais tipos de sequências segmentares estão sujeitas a subtrações difere entre os dialetos; ver Köhnlein (2018) para uma discussão mais detalhada. Além disso, observe que algumas alternâncias também mostram diferenças na qualidade das vogais, o que se deve a um processo de frontalização independente comumente chamado de *trema*; já que o *trema* não é de relevância imediata para os propósitos deste artigo, vou ignorá-lo de agora em diante. Por último, considere (7d-f), que mostra que a subtração consonantal coincide com a exclusão do meio da palavra; isto é, consoantes que são subtraídas no final da palavra também serão deletadas no mesmo contexto quando a sequência em questão for seguida por outra vogal – os exemplos em (7d-f) são formas plurais que contêm uma vogal sufixal, mas sequências relevantes são também ausentes em palavras monomorfêmicas, como [vunɛ¹] ‘milagre’, cujo cognato alemão padrão é *Wunder*, pronunciado com [d].

3. Uma análise métrica do acento tonal

Na §2, mostrei um conjunto de fenômenos relacionados ao acento tonal, no que diz respeito às próprias realizações tonais, bem como correlatos fonéticos adicionais (diferenças duracionais), correlatos fonológicos (interações com vozeamento consonantal), bem como aspectos morfofonológicos de acentuação (Acento 1 como padrão morfológico; interação do acento tonal com a subtração). Nesta seção, forneço os fundamentos de como minha própria abordagem analítica do acento tonal visa explicar todos esses diferentes fenômenos de maneira unificada. Conforme mencionado brevemente na introdução, sigo o que anteriormente chamei de abordagem métrica para acento tonal. A ideia conceitual básica por trás da abordagem métrica é fornecer uma análise do acento tonal que não se concentre apenas nos contrastes tonais, mas visa incorporar todos os diferentes correlatos do acento de forma unificada. Deve-se notar que minha análise específica

não é a única estrutura métrica que foi proposta para a análise do acento tonal (outros trabalhos métricos incluem, por exemplo, Hermans (2012) ou Kehrein (2017)). Aqui vou me concentrar em meu próprio trabalho, sem querer sugerir que minha análise métrica seja necessariamente superior às de meus colegas.

Minha abordagem é baseada nos princípios da teoria métrica autossegmental, especificamente no inventário geral de pés que foi proposto em trabalhos como Hayes (1995). Por exemplo, é amplamente reconhecido que existe uma diferença entre os chamados troqueus silábicos e os troqueus moraicos, dois tipos de pés em que o ramo esquerdo é forte (a cabeça) e o ramo direito é fraco (o dependente). Os pés, portanto, têm tipicamente um núcleo e um dependente, ou seja, são preferencialmente binários – a sílaba que contém o núcleo do pé será a sílaba tônica. Para troqueus silábicos, que contam sílabas (daí o nome troqueu silábico), a binaridade é estabelecida no nível da sílaba. Para os troqueus moraicos, que contam moras, daí o nome troqueu moraico), a binaridade é estabelecida no nível mora. Embora a noção de sílaba certamente não precise de introdução, deixe-me mencionar que moras são unidades de tempo localizadas entre o segmento e as sílabas: vogais curtas têm uma mora, vogais longas têm duas moras e consoantes coda podem ou não ter moras, dependendo da língua (HAYES, 1989). Em francônio, codas em sílabas tônicas podem ser moraicas, e o número máximo de moras por sílaba é dois. Na teoria moraica, muitas vezes é estipulado que as consoantes iniciais não podem ser moraicas e, embora haja alguma controvérsia em torno dessa questão, não há evidências de que as consoantes iniciais sejam moraicas em francônio.

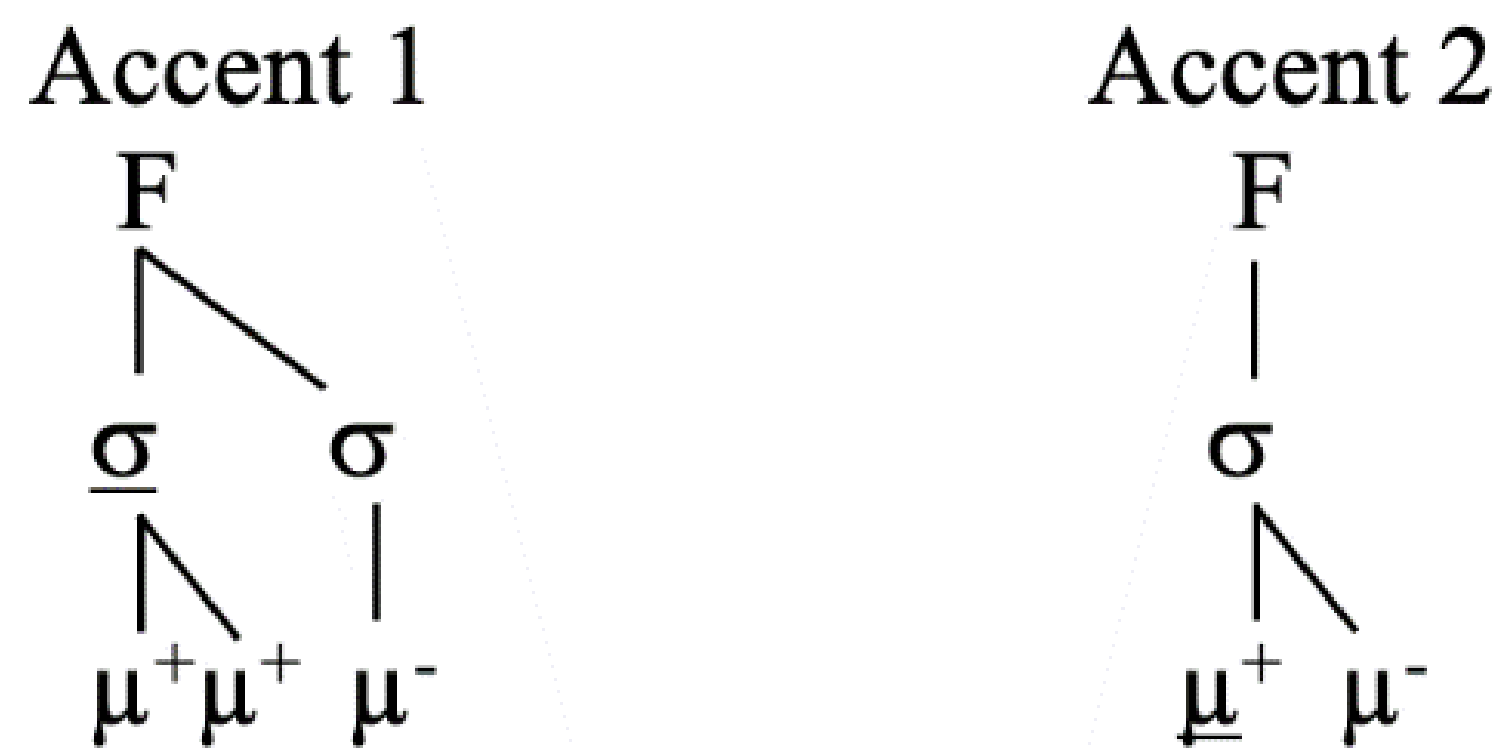
Na teoria métrica autossegmental tradicional, baseada em parâmetros, normalmente é assumido que as línguas têm exatamente um tipo de pé em seu inventário, troqueus silábicos, troqueus moraicos ou pés iâmbicos (esses pés são dominantes à direita). Meu trabalho se desvia dessa suposição; especificamente, afirmo que as línguas podem ter mais de um tipo de pé, e o francônio está entre essas línguas (em teorias baseadas em restrições como a Teoria da Otimidade, não é surpreendente que as línguas possam ter mais de um pé). A seguir, discuto os dois tipos de pés que proponho para o francônio e como eles geralmente capturam as melodias tonais contrastantes (§3.1). Em uma segunda etapa, demonstro como minha análise baseada no pé torna possível explicar os correlatos adicionais e não-tonais de acento discutidos na §2.

3.1 Inventário do pé e mapeamento tonal

O principal ingrediente da minha análise sincrônica do acento tonal do francônio é que o Acento 1 e o Acento 2 correspondem a dois tipos de pés (KÖHNLEIN, 2011, 2016, 2017, 2018, 2019, 2020; KÖHNLEIN; CAMERON, 2020). Especificamente, afirmo que alguns itens em francônio formam o pé como troqueus moraicos, e isso corresponde ao Acento 2 na superfície. Outros itens, no entanto, formam pés como troqueus silábicos, o que corresponde ao Acento 1. As estruturas de pé relevantes

são cinco, mostradas em (8). As árvores contêm um conjunto de sinais diacríticos (sublinhados, sobrescritos positivos e negativos), que utilizo para fins de ilustração abaixo; no entanto, eles não têm status na teoria.

(8) Troqueus silábicos (Acento 1) e troqueus moraicos (Acento 2) em francônio



Legenda: Árvores: Primeira linha: Acento 1, Acento 2

Em primeiro lugar, o que decide se um pé é um troqueu silábico ou um troqueu iâmbico? Suponho que a cabeça de um pé é determinada no nível mais alto onde o pé se ramifica, ou seja, o nível em que se torna binário. Os pés de Acento-1 são binários no nível da sílaba e, como o nível da sílaba é mais alto que o nível da mora, a sílaba inicial será a cabeça do pé (mostrada como sublinhada); a segunda sílaba é o ramo fraco do pé, ou seja, o pé dependente. Os pés de Acento-2, por outro lado, contêm apenas uma sílaba, de modo que o pé não pode ser binário no nível da sílaba; no entanto, contém duas moras e, portanto, é binário no nível moraico. Isso, então, significa que o pé será um troqueu moraico, com a primeira mora sendo a cabeça do pé (sublinhado) e a segunda mora sendo o pé dependente.

Em uma segunda etapa, argumento que o nível de cabeça (sílabas para o Acento 1 versus mora para o Acento 2) tem influência na “força” prosódica das moras na sílaba tônica. Considere o pé do acento silábico-1 primeiro: o núcleo do pé, a primeira sílaba, domina duas moras, e afirmo que essas moras herdam a força métrica do núcleo – em (8, lado esquerdo), indico isso dando-lhes sobrescrito bônus. A mora dominada pela segunda sílaba é metricamente fraca, pois está ligada à sílaba dependente do pé; recebe assim um sinal de menos sobrescrito. Agora considere o pé moraico do Acento-2 (8, lado direito): Aqui, a primeira mora é a cabeça do pé, o que a torna forte (sobrescrito mais), e a segunda mora é a dependente do pé, o que a torna fraca (sobrescrito menos). Ao comparar os dois pés, podemos notar que o troqueu silábico (= Acento 1) tem domínio dissilábico, e ambas as moras na primeira sílaba são metricamente fortes porque são dominadas pela cabeça do pé, a primeira sílaba. O troqueu moraico (= Acento 2) é um domínio monossilábico, bimoraico, o que significa que uma mora na sílaba tônica é forte e a outra é fraca. Assim, as sílabas com Acento 1 têm duas moras fortes e as sílabas com Acento 2 têm uma mora forte e uma mora fraca – essa diferença é crucial para minha análise do acento tonal. Para ver como essas estruturas dos pés mapeiam as palavras existentes, considere os exemplos em (9) de Colônia (MÜNCH, 1904):

(9) Pares de palavras de tom-acento do francônio de Colônia

- a. [da:x¹] ‘dia-nom.’ [da:x²] ‘dia-dat.’
 b. [i:¹zə] ‘ferro’ [ri:²sə] ‘lacrimejar’

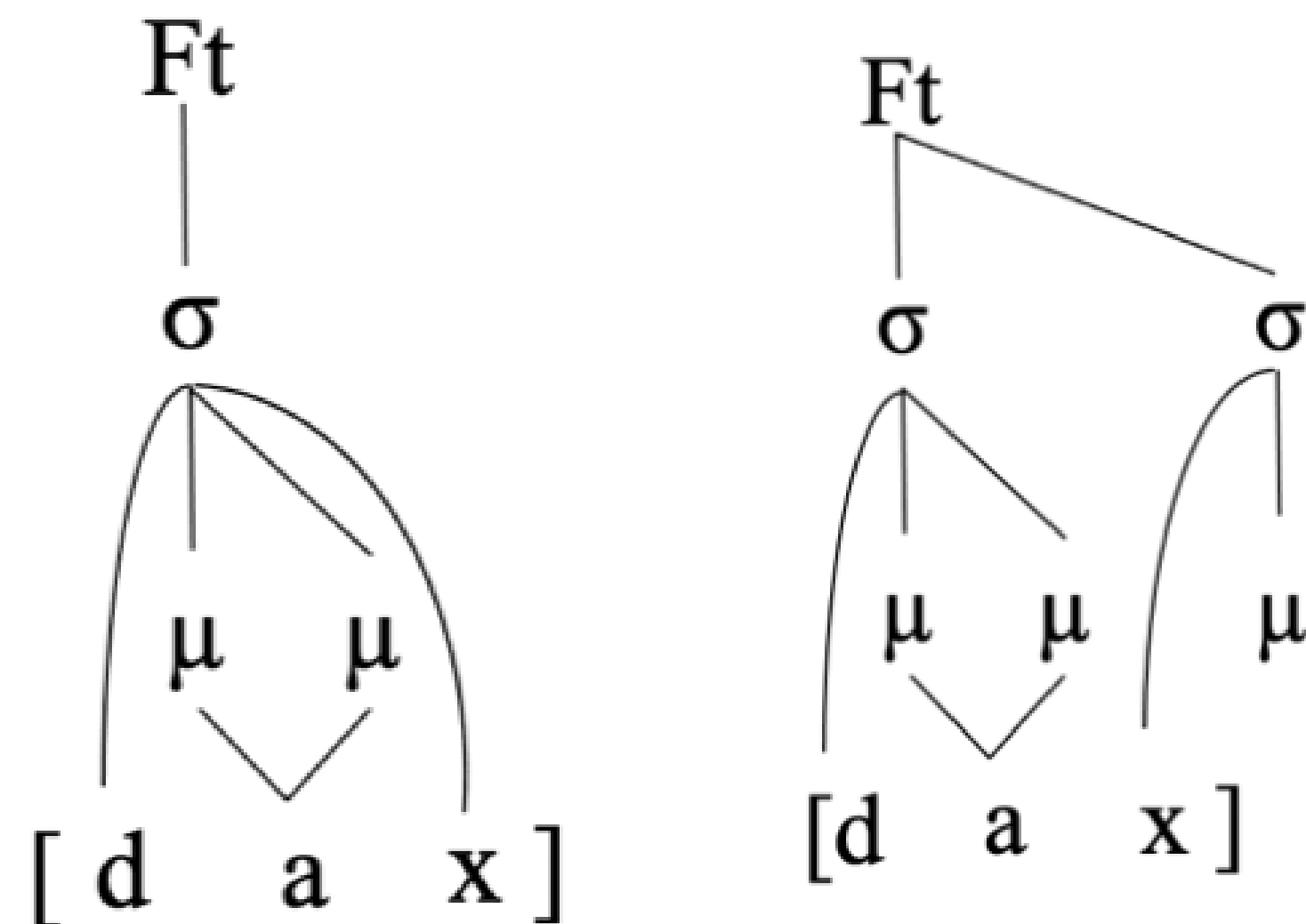
Para palavras dissílabas, como em (10b), modelar o contraste na formação de pés é bastante simples, como mostrado em (10), retirado de Köhnlein e Cameron (2020). Para itens com Acento 1 (10, lado esquerdo), ambas as sílabas são incluídas no pé, que é um troqueu silábico que se estende por duas sílabas. Para itens com Acento 2 (10, lado direito), apenas a primeira sílaba faz parte do pé, o que leva a um troqueu moraico que se estende por duas moras, mas apenas uma sílaba; a segunda sílaba permanece não analisada pelo pé. Na representação de superfície, esta sílaba não analisada se ligará diretamente ao nó da palavra prosódica (não mostrado nas representações em (10)).

(10) Acento 1 (troqueu silábico, à esquerda) vs. Acento 2 (troqueu moraico, à direita) para palavras dissílabas



A situação é um pouco mais complexa para palavras com apenas uma vogal, como os itens em (11). Os itens de Acento-2 (11, lado direito) são colocados como troqueus moraicos ‘padrão’, ou seja, uma vogal corresponde a uma sílaba. Embora isso seja indiscutivelmente direto, é menos óbvio como os pés dissilábicos do Acento-1 são mapeados nas estruturas segmentares relevantes. Para tais casos, proponho que palavras monossilábicas com Acento 1 sejam apenas aparentemente monossilábicas; ainda em sua representação métrica; eles contêm a chamada segunda sílaba de cabeça vazia, como mostrado (11, lado esquerdo). Essencialmente, uma sílaba de cabeça vazia pode ser pensada como contendo uma vogal não pronunciada, que represento como uma mora que não está vinculada a nenhum conteúdo segmentar. Sem dúvida, postular tal representação pressupõe um nível adicional de abstração, que deve sempre ser motivo de cautela. Acho que se justifica por vários motivos, dois dos quais mencionarei aqui: primeiro, embora não sejam universalmente aceitos na teoria fonológica, as sílabas de cabeça vazia, assim como as categorias vazias em geral, têm sido usadas com sucesso há décadas como ferramentas em teoria fonológica (por exemplo, CAVIRANI; VAN OOSTENDORP, 2017 para visão geral). Em segundo lugar, os itens de Acento-1 geralmente se comportam de maneira semelhante a palavras abertamente dissílabas. Discutirei algumas evidências em §3.2.

(11) Acento 1 (troqueia silábica, esquerda) vs. Acento 2 (trochee moraico, à direita) para palavras 'aparentemente' monossilábicas



Uma questão óbvia de acompanhamento é como derivar o contraste tonal desses dois tipos de pés. Concentro-me aqui nas melodias declarativas finais de frase nos dialetos de Colônia para fins de ilustração e ignoro a variação entre dialetos e em diferentes contextos prosódicos e pragmáticos – os aspectos tonais do contraste na Regra A e na Regra B são discutidos em mais detalhes em Köhnlein (2016), e Köhnlein (2011) contém uma análise abrangente para quatro tipos diferentes de dialetos do francônio.

Lembre-se de que a ideia básica na abordagem métrica do acento tonal é que não há tom lexical em francônio; em vez disso, os contrastes tonais entre os acentos são derivados de diferenças na associação de tons entoacionais pós-lexicais. No dialeto de Colônia, o foco nos declarativos é marcado com um acento entoacional H*L (o asterisco indica que H* é o tom que deve ser associado a uma sílaba tônica), semelhante ao que conhecemos das línguas entoacionais prototípicas. A regra principal para atribuição de tom em Colônia é que um tom entoacional deve ser licenciado por uma mora forte, ou seja, uma mora que está ligada a uma cabeça de pé. Além disso, uma mora não pode hospedar mais de um tom.

Figura 2: Associação de tons entoacionais a moras dependendo da estrutura do pé no dialeto de Colônia.

	Acento 1		Acento 2	
Declarativa, foco, não-final				
	H* μ ⁺	L μ ⁺	H* μ ⁺	L μ ⁻

O mapeamento tonal para Colônia é mostrado na Figura 2. Para o Acento 1, um troqueu silábico, as duas moras na sílaba tônica estão ambas ligadas à cabeça do pé, o que as torna metricamente fortes; como consequência, cada mora pode hospedar um tom, o que significa que o Acento 1 será realizado como um tom descendente na sílaba acentuada. O Acento 2, por outro lado, é um troqueu moraico, então apenas a primeira mora (= a cabeça do pé) é forte, enquanto a segunda mora (= o pé dependente) é fraca. Uma vez que os tons têm que ser licenciados por moras fortes, as sílabas de Acento-2 só podem hospedar um tom, H*; o tom baixo após H* não pode ser vinculado à mora fraca e deverá ser realizado posteriormente no enunciado, ou seja, em um item seguinte. A segunda mora, no entanto, recebe o tom da primeira mora via espalhamento autossegmental; a mora fraca pode, assim, compartilhar um tom com uma mora precedente forte, mas não pode hospedar um tom por conta própria. Isso leva a um tom de alto nível com uma queda tardia em direção ao tom pós-tônico baixo, que por sua vez deriva a oposição tonal entre Acento 1 e Acento 2 em Colônia.

3.2 Correlatos adicionais de acento

3.2.1 Duração

Conforme discutido em §2.2, vários dialetos do francônio têm diferenças de duração concomitantes ao contraste tom-acento. Em Colônia, onde as sílabas com Acento-1 são mais longas que as sílabas com Acento-2, o contraste de duração pode, em alguns contextos, até ser o principal correlato da oposição, ao invés do tom (GUSSENHOVEN; PETERS, 2004). Seguindo Köhnlein (2016), atribuo as diferenças de duração entre os acentos aos dois tipos de pés. Inspirado em trabalhos anteriores de, por exemplo, Prince (1980) e Odden (1997) para estoniano, suponho que, em algumas línguas, um pé pode ter uma certa duração associada a ele. Em poucas palavras, se um pé tiver duas sílabas, a duração será distribuída entre essas duas sílabas. Se um pé tiver uma sílaba, a duração do pé será expressa em apenas uma sílaba. Assim, esperamos que a sílaba tônica de um pé monossilábico seja mais longa do que a sílaba tônica de um pé dissilábico. Quando aplicado ao francônio, isso significa que as sílabas tônicas de pés bimoraicos e monossilábicos de Acento-2 serão mais longas do que as sílabas tônicas de pés dissilábicos de Acento-1, resultando em oposições do tipo [da:x1] <dia.dat> (com uma cabeça vazia – segunda sílaba cuja duração não é expressa) vs. [da:x2] ‘dia.nom’, ou [i:1zə] ‘ferro’ (pé dissilábico, sílaba tônica longa) vs. [ri:2sə] ‘lacrimar’ (pé monossilábico, sílaba tônica muito longa).

3.2.2 Qualidade da vogal

Em todos os dialetos, há uma tendência de que vogais longas com Acento 1 se transformem em ditongos e ditongos com Acento 2 se tornem monotongos. Minha explicação para essa tendência é muito semelhante à minha análise do mapeamento tonal, sob a suposição de que ditongos são a

combinação de duas vogais na mesma sílaba. Nessa linha, um ditongo como [ei] seria uma combinação de [ɛ] e [i]. O Acento 1, que tem duas moras fortes na sílaba acentuada, é preferencialmente ditongal, de modo que cada mora forte pode dominar um nó vocálico; isso é semelhante ao Acento 1, de preferência com dois tons (como H*L) na sílaba acentuada. O Acento 2, por outro lado, tem apenas uma mora forte, portanto domina preferencialmente as vogais longas onde a segunda, a mora fraca, não licencia sua própria vogal, mas recebe a vogal da primeira, mora forte via espalhamento; isso é comparável a ter preferencialmente um tom (como H*) na sílaba acentuada. Isso, então, resulta em alternâncias como [bleif1] ‘ficar.1.ps.sg’ vs. [bli:2və] ‘ficar-inf.’ no dialeto de Maastricht (veja Köhnlein, 2017 para uma análise detalhada).

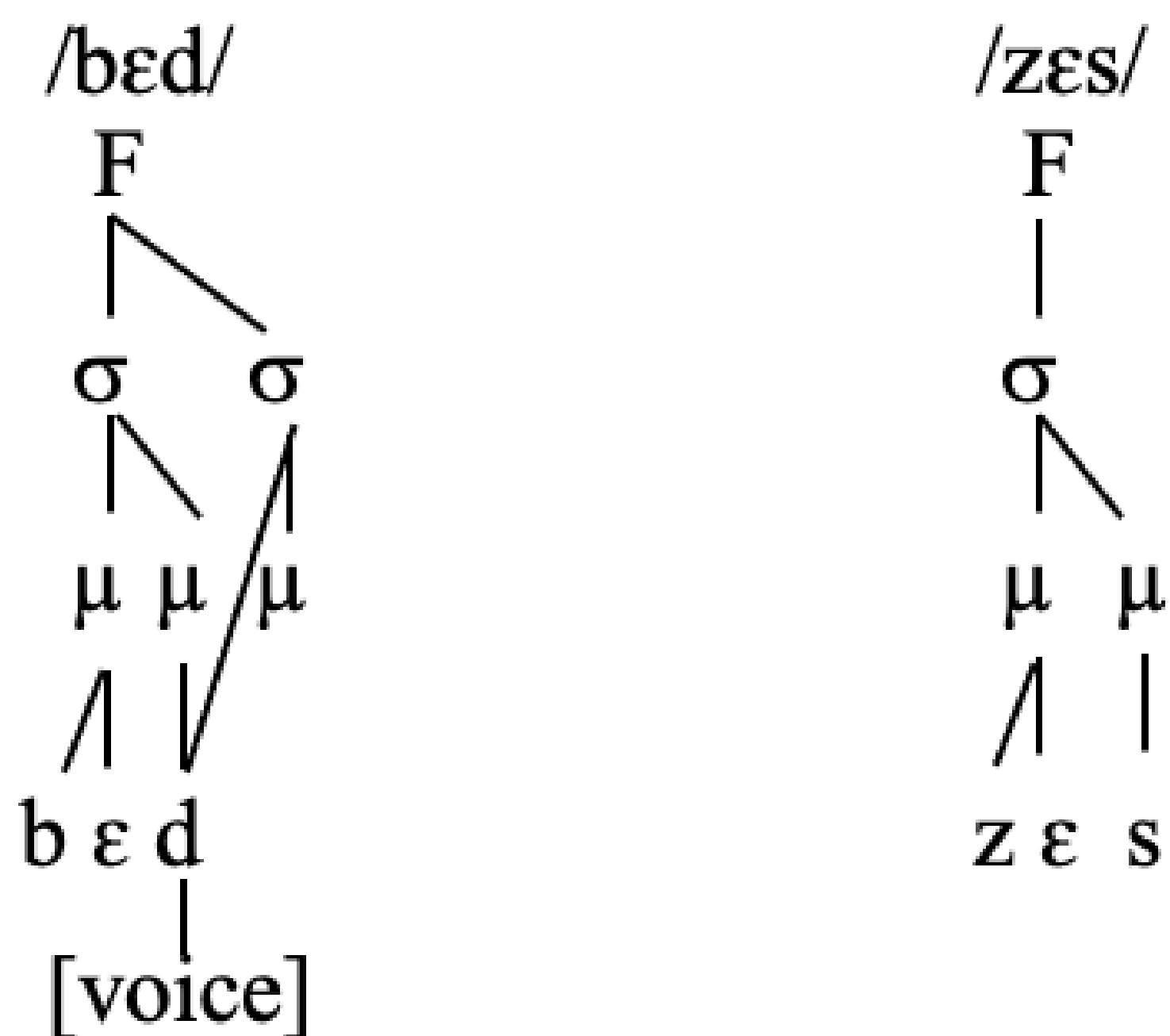
3.2.3 Vozeamento consonantal

Está bem estabelecido que existe uma tendência interlinguística de onsets mediais do pé conterem consoantes lenis mais fracas do que os onsets que não são mediais do pé; alguns exemplos do inglês referem-se ao *flapping* medial do pé, não aspiração de plosivas, a distribuição de /h/ (DAVIS; CHO, 2003) ou vozeamento fricativo (HONEYBONE, 2012). Essas observações interlinguísticas podem ser implementadas diretamente em minha análise do francônio com base no pé. Lembre-se de que, em Colônia, palavras dissílabas com consoantes sonoras mediais do pé sempre recebem o Acento 1, enquanto palavras com consoantes surdas mediais da palavra preferencialmente têm o Acento 1. Em minha análise, isto decorre de pés dissilábicos (= Acento 1) sendo construídos preferencialmente através de uma consoante sonora (que é mais fraca, ou seja, mais lenis do que uma consoante surda). Assim, um item como [i:1zə] ‘ferro’ aparecerá com Acento 1 (exemplo 10, lado esquerdo). Por outro lado, palavras com consoantes fortes no meio da palavra, surdas, como [ri:2sə] ‘lacrimejar’, recebem preferencialmente um pé monossilábico que cobre apenas a primeira sílaba (exemplo 10, lado direito). Dessa forma, o pé não atravessa uma forte, chamada consoante fortis – portanto, o item receberá Acento 2 (veja KÖHNLEIN; CAMERON, 2020, em preparação para uma análise mais detalhada).

Uma análise relacionada, se não idêntica, é possível para os fatos de Moresnet discutidos em §2.2, onde itens com obstruintes sonoras no final da palavra após vogais curtas recebem o Acento 1, e os itens com obstruintes surdos no final da palavra recebem o Acento 2, abordado em Köhnlein (2016), que, por sua vez, se baseia em uma análise em van Oostendorp (2017). Como a maioria dos dialetos do holandês, o Moresnet tem desvozeamento final, mas se aplica apenas em um estágio posterior da gramática multinível. No primeiro passo (12, à esquerda), que corresponde ao nível da raiz, o ensurdecimento final de obstruintes sonoras, como em /bɛd/ ‘cama’, pode ser evitado colocando-se uma obstruinte final de palavra no onset de uma sílaba com cabeça vazia – no nível de raiz gramatical, o desvozeamento final só se aplica a consoantes em coda. Em itens que terminam em consoantes surdas, como /zɛs/ ‘seis’ (12, à direita), a consoante pode ser silabada como uma

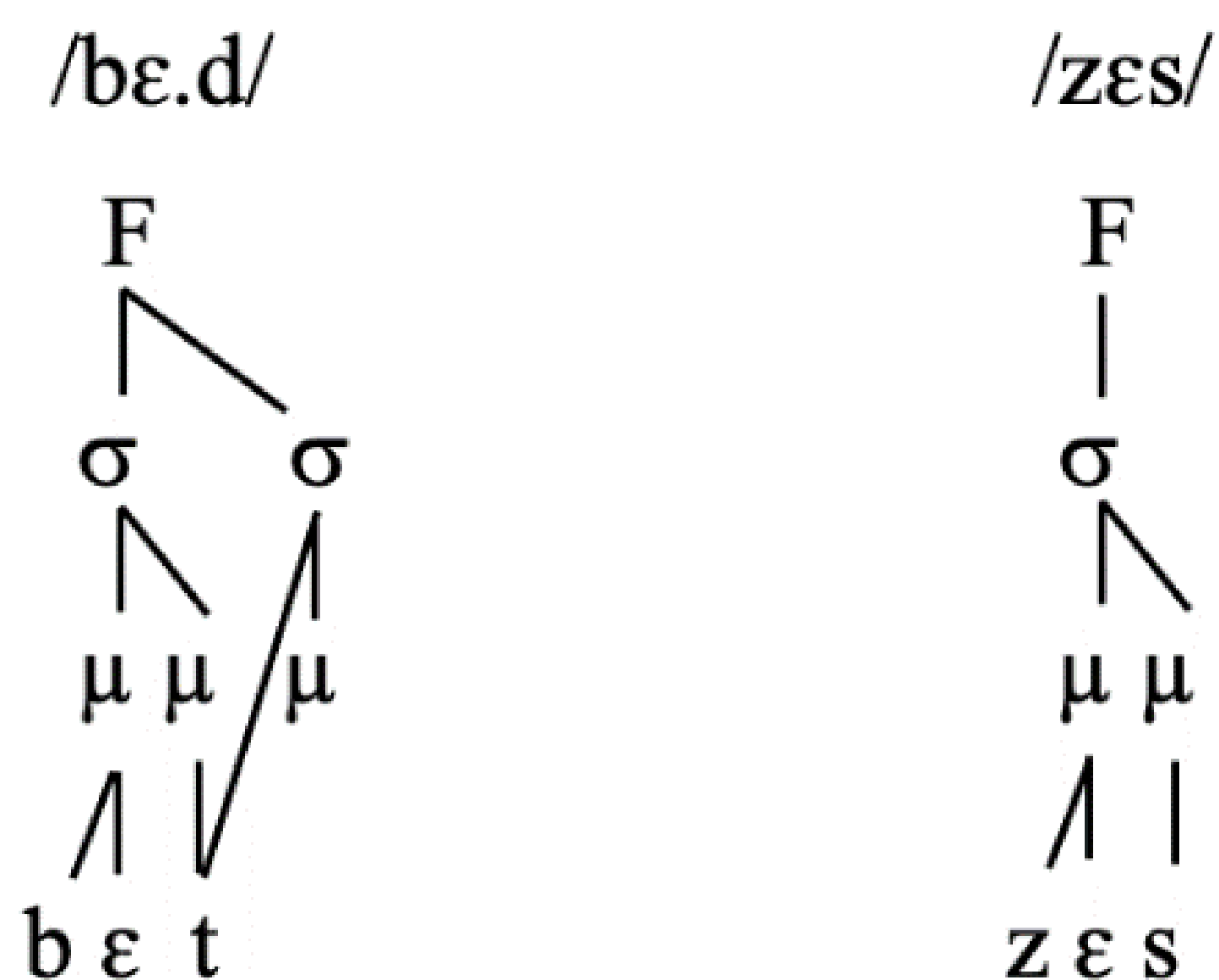
coda – ela já é surda de qualquer maneira, então não há ‘necessidade’ de criar uma sílaba de cabeça vazia.

(12) Representações no nível da raiz em Moresnet



Em uma segunda etapa, considere as representações em nível de palavra em (13). A estrutura do pé, herdada do nível da raiz, é preservada; no entanto, nesse nível, o desvozeamento final afeta todas as obstruintes finais de palavra. Conforme discutido em van Oostendorp (2017), essa diferença entre a raiz e o nível da palavra corresponde a observações interlinguísticas, onde o desvozeamento final às vezes afeta todas as consoantes de coda e às vezes apenas as consoantes finais de palavra. Como resultado, /bɛd/ aparece como [bɛt1], mas ainda recebe Acento 1 por causa do pé dissilábico com uma sílaba de cabeça vazia que foi construída no nível da raiz (13, esquerda). [zɛs2], por outro lado, permanece monossilábico, e este item aparece com Acento 2 (13, à direita).

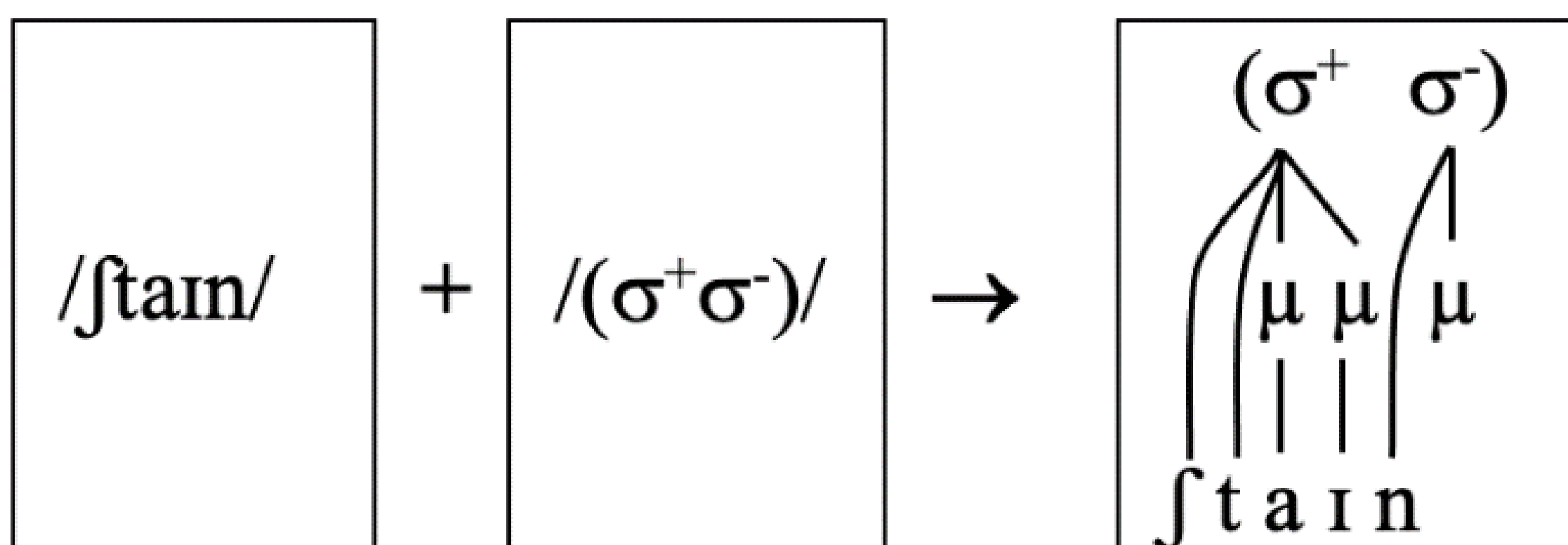
(13) Representação no nível de palavra em Moresnet



3.2.4 Relevância morfológica geral e subtração

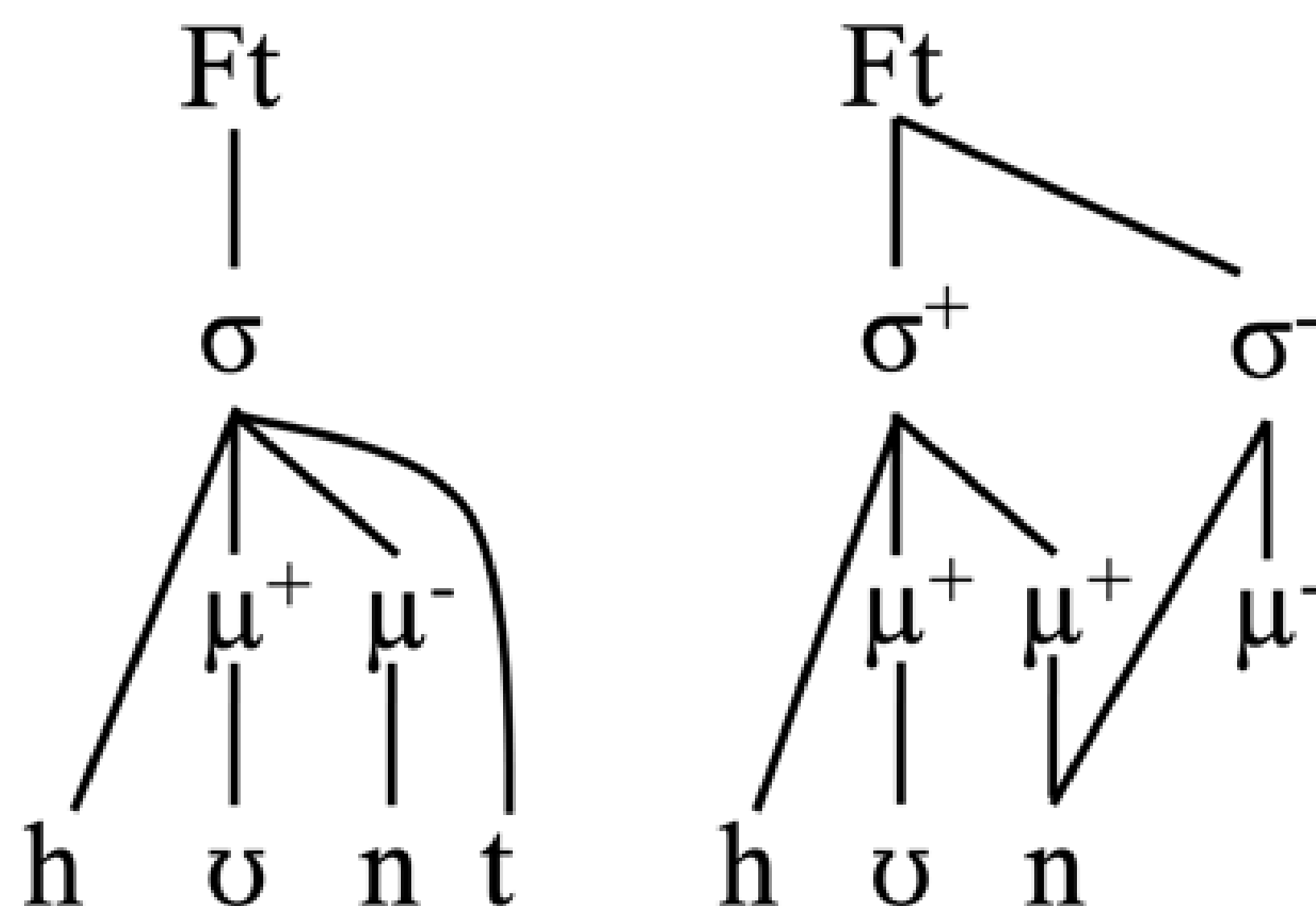
Conforme discutido no §2, o Acento 1 é o acento morfológicamente marcado. Na minha abordagem, isso decorre da suposição de que o Acento 1, sendo um troqueu dissilábico, tem um pé ‘maior’ que o Acento 2, que vem com um troqueu monossilábico e bimoraico. Nesse sentido, o pé maior do Acento-1 pode incorporar e substituir o pé menor do Acento-2. Do ponto de vista representacional, podemos pensar no Acento 1 morfológicamente condicionado como um pé é armazenado como um morfema (como um morfema plural), que então se associa à estrutura segmental de um determinado tronco. Isso é mostrado em (14) para a derivação da alternância de Arzbach entre [ʃtaɪn²] ‘pedra-sg’ e [ʃtaɪn¹] ‘pedra-pl’. A entrada lexical para ‘pedra’ é /ʃtaɪn/ (caixa esquerda em (14)), e como padrão, isso será analisado como um pé monossilábico, levando ao Acento 2. O morfema plural associado a /ʃtaɪn/ é um dissilábico, pé trocaico, indicado como /(σ^+ σ^-)/ na caixa do meio. Quando o morfema plural é combinado com a sequência segmental, isso leva a um pé dissilábico com uma sílaba de cabeça vazia, como mostrado (14, caixa à direita); isso, é claro, corresponderá ao Acento 1.

(14) Derivação plural de [ʃtaɪn¹] ‘pedra-pl’ em Arzbach



Observe que minha análise baseada em pés de alternâncias morfológicas pode ser facilmente estendida para plurais subtrativos. Lembre-se de outro exemplo de Arzbach, a alternância entre [hʊnt²] ‘cachorro’ e [hʊn¹] ‘cachorros’ (veja §2.3 para mais dados). Semelhante ao plural para ‘pedra’, o morfema plural para ‘cachorros’ é um pé dissilábico. O singular por si só será analisado como um pé monossilábico bimoraico padrão e, portanto, receberá o Acento 2 (15, à esquerda). O plural, no entanto, terá que incorporar o pé dissilábico do morfema plural, o que leva à estrutura (15, à direita). Não só o item resultante recebe o Acento 1, esta formação de pé também leva à exclusão da plosiva de onset na segunda sílaba de cabeça vazia, porque o dialeto não tolera a sequência nasal-plosiva se a plosiva ocorrer no início de uma sílaba com cabeça vazia. Isso então torna a nasal ambissilábica (isto é, pertence tanto à sílaba tônica como coda, quanto à sílaba átona como onset).

(15) Morfologia subtrativa em francônio de Arzbach



A exclusão das plosivas finais de palavra em plurais encontra um paralelo perfeito em palavras abertamente dissílabas, como evidência pela alternância entre [lant2] ‘país’ e [lɛnɐ1] ‘países’. No caso em questão, a forma plural é abertamente dissilábica, e a oclusiva medial da palavra na forma singular ainda é excluída, porque ocorreria no início de uma sílaba átona (Köhnlein, 2018 para uma análise detalhada da subtração entre dialetos).

4. Breve comparação e conclusão

Na seção anterior, demonstrei como minha abordagem baseada no pé para acento tonal torna possível explicar vários fenômenos relacionados à acentuação de uma maneira baseada em princípios. Ou seja, minha análise captura não apenas o mapeamento tonal, mas também mostra como a diferença entre pés monossilábicos (Acento 1) e pés dissílabos (Acento 2) interage de forma previsível com diferenças de duração entre os acentos e alternâncias vocálicas. Ela também explica diretamente as interações do acento com a sonoridade consonantal e captura as alternâncias morfológicas, incluindo a morfologia subtrativa. Além disso, a análise de todas as interações em questão não é associada arbitrariamente com a estrutura do pé, mas segue tendências translinguísticas estabelecidas.

A alternativa analítica geral disponível, uma abordagem com tom lexical (por exemplo, GUSSENHOVEN; PETERS, 2004), pode dar conta com sucesso das melodias tonais contrastantes. Pelo menos em sua versão atual, acredito que não há nada substancial a dizer sobre a relevância sincrônica de qualquer um dos correlatos adicionais, pelo menos não de uma perspectiva fonológica (ainda ver GUSSENHOVEN; PETERS, 2019 para uma discussão de argumentos aparentes a favor de uma abordagem tonal e KÖHNLEIN, 2019 para uma breve resposta). As interações com a duração têm sido referidas como aprimoramento fonético do contraste lexical-tonal, mas a abordagem de aprimoramento em si é arbitrária e não está vinculada à estrutura dos acentos de maneira mediana. As interações com o vozeamento consonantal são difíceis de capturar na abordagem tonal, uma vez que o vozeamento superficial não prevê a qualidade dos tons precedentes: contornos para a

vocalização de consoantes mediais de palavras. Além disso, as abordagens tonais existentes sempre marcam o Acento 2 lexicalmente, o que está em desacordo com o fato de que a complexidade morfológica se correlaciona com o Acento 1.

Embora eu acredite que minha análise métrica seja consideravelmente mais adequada do que a abordagem tonal, um possível contra-argumento para minha análise pode ser encontrado no fato de que ela se baseia em representações mais abstratas do que a alternativa tonal – quanta abstração é justificada ou necessária na teoria fonológica, no entanto, é uma questão em aberto. Além de sua cobertura empírica, também gostaria de salientar que minha análise do Acento 1 como um pé dissilábico pode ser útil na modelagem de que palavras aparentemente monossilábicas Acento-1 (que têm uma segunda sílaba de cabeça vazia) geralmente se comportam de forma semelhante às palavras abertamente dissílabas, como pode ser observado, por exemplo, em seu comportamento de dialetos subtrativos. Neste contexto, também deve ser mencionado que minha abordagem baseada no pé do acento tonal do francônio tem fortes paralelos com análises métricas de outras línguas com sistemas acentuais comparáveis, por exemplo, certas variedades do germânico do norte (como MORÉN-DUOLLJÁ, 2013; IOSAD, 2016a, 2016b; KALDHOL; KÖHNLEIN, 2021), gaélico escocês (como IOSAD, 2015; MORRISON, 2019), ou Uspanteko (KÖHNLEIN, 2019; KÖHNLEIN; ZHU, 2019), que acredito fortalecer minha proposta do ponto de vista tipológico – mesmo que análises métricas diferentes não dependam necessariamente do mesmo conjunto de tons representacionais. Assumindo que a abordagem métrica do acento tonal está no caminho certo, trabalhos futuros terão, portanto, que se concentrar em restringir quais são as ferramentas teóricas necessárias e suficientes para analisar contrastes acentuais entre as línguas.

Agradecimentos

Gostaria de agradecer a todos os organizadores do Simpósio Internacional de Prosódia e Bilinguismo, e especificamente à Amanda Post, pelo convite para apresentar meu trabalho e por incluir o presente capítulo neste volume. Esta pesquisa é apoiada pela National Science Foundation (BCS-1845107).

Referências

BACH, A. Die Schärfung in der moselfränkischen Mundart von Arzbach (Unterwesterwaldkreis). *Beiträge zur Geschichte der Deutschen Sprache und Literatur*, n. 45, p. 266-290, 1921.

BOTMA, B.; NOSKE, R. (ed.). *Phonological Explorations: Empirical, Theoretical and Diachronic Issues*. Berlin: Mouton de Gruyter, 2012.

CAJOT, J. Phonologisch bedingter Polytonieverlust – eine tonlose Enklave südlich von Maastricht. In: DE VAAN, M. (ed.). *Germanic Tone Accents*. 2006. p. 11-24.

CAVIRANI, E.; OOSTENDORP, M. VAN. On silent markedness. In: SAMUELS, B. (ed.): *Beyond Markedness in formal phonology*. Amsterdam: John Benjamins, 2017. p. 101-120.

DAVIS, S.; CHO, M.-H. The distribution of aspirated stops and /h/ in American English and Korean: An alignment approach with typological implications. *Linguistics*, v. 41, p. 607-652, 2003.

DOLS, W. *Sittardse diftongering*. Sittard, 1953.

GUSSENHOVEN, C. The lexical tone contrast of Roermond Dutch in Optimality Theory. In: HORNE, M. (ed.): *Prosody: Theory and Experiment*. Amsterdam: Kluwer, 2000. p. 129-167.

GUSSENHOVEN, C. *Quantity or durational enhancement of tone: the case of Maastricht Limburgian high vowels*. 2012. p. 241-254.

GUSSENHOVEN, C.; PETERS, J. A tonal analysis of Cologne Schärfung. *Phonology*, v. 21, n. 2, p. 251-285, 2004.

GUSSENHOVEN, C.; PETERS, J. Franconian Tones Fare Better as Tones Than as Feet: A Reply to Köhnlein (2016). *Phonology*, v. 36, n. 3, p. 497-530, 2019.

HERMANS, B. The phonological representation of the Limburgian tonal accents. In: BOTMA, B.; NOSKE, R. (ed.). *Phonological Explorations: Empirical, Theoretical and Diachronic Issues*. Berlin: Mouton de Gruyter, 2012, p. 227-244.

HARDT, M. Vocalismus der Sauer-Mundart. In: *Programm des Königlich-Großherzoglichen Progymnasiums zu Echternach, herausgegeben beim Schlusze des Schuljahres 1842-43*. Trier: Fr. Lintz. 1-29, 1843.

HAYES, B. Compensatory lengthening in moraic phonology. *Linguistic Inquiry*, v. 20, p. 253-306, 1989.

HAYES, B. *Metrical stress theory: Principles and case studies*. Chicago: University of Chicago Press, 1995.

HONEYBONE, P. Lenition in English. In: NEVALAINEN, T.; TRAUGOTT, E. C. (ed.). *Handbook on the history of English: Rethinking approaches to the history of English*. Oxford: Oxford University Press, 2012. p. 773-787.

HYMAN, L. M. Word-prosodic typology. *Phonology*, v. 23, n. 2, p. 225-257, 2006.

HYMAN, L. M. How (not) to do phonological typology: the case of pitch-accent. *Language Sciences*, v. 31, n. 2, p. 213-238, 2009.

- IOSAD, P. Pitch accent and prosodic structure in Scottish Gaelic: reassessing the role of contact. In: HILPERT, M.; DUKE, J.; MERTZLUFFT, C.; ÖSTMAN, J.-O.; RIEßLER, M. (ed.). *New Trends in Nordic and General Linguistics*. Berlin: Mouton de Gruyter. 2015. p. 28-54.
- IOSAD, P. Prosodic structure and suprasegmental features: Short-vowel stød in Danish. *Journal of Comparative Germanic Linguistics*, v. 19, n. 3, p. 221-268, 2016a.
- IOSAD, P. Tonal stability and tonogenesis in North Germanic. In: GILES, I.; CHAPOT, L.; COOIJMANS, C.; FOSTER, R.; TESIO, B. (ed.). *Beyond Borealism: New perspectives on the North*. London: Norvik Press, 2016b. p. 80-98.
- JONGEN, R. *Phonologie der Moersener Mundart: eine Beschreibung der segmentalen und prosodischen Wortformdiakrise*. Assen: Van Gorcum, 1972.
- KALDHOL, N. H.; KÖHNLEIN, B. North Germanic Tonal Accent is Equipollent and Metrical: Evidence from Compounding. *Proceedings of the 2020 Annual Meeting on Phonology*. 2021.
- KEHREIN, W. There's no tone in Cologne: against tone segment interactions in Franconian. In: KEHREIN, W.; KÖHNLEIN, B.; BOERSMA, P.; VAN OOSTENDORP, M. (ed.). *Segmental structure and tone*. Berlin: Walter de Gruyter, 2017. p. 147-194.
- KÖHNLEIN, B. *Rule reversal revisited: Synchrony and diachrony of tone and prosodic structure in the Franconian dialect of Arzbach*. 2011. Utrecht: LOT. Dissertation Series 274.
- KÖHNLEIN, B. Contrastive foot structure in Franconian tone accent dialects. *Phonology*, v. 33, n. 1, p. 87-123, 2016.
- KÖHNLEIN, B. Synchronic alternations between monophthongs and diphthongs in Franconian tone accent dialects: a metrical approach. In: KEHREIN, W.; KÖHNLEIN, B.; BOERSMA, P.; OOSTENDORP, M. (ed.). *Segmental structure and tone*. Berlin: Walter de Gruyter, 2017. p. 211-236.
- KÖHNLEIN, B. Why metrical approaches to tonal accent are worth pursuing. *Phonology*, v. 36, n. 3, p. 531-535, 2019.
- KÖHNLEIN, B. Metrically conditioned pitch accent in Uspanteko. In: GOEDEMANS, R.; HEINZ, J.; VAN DER HULST, H. (ed.). *The study of word stress and accent: theories, methods and data*. Cambridge: Cambridge University Press, 2019. p. 293-322.
- KÖHNLEIN, B. Tone accent in North and West Germanic. In: PAGE, R.; PUTNAM, M. (ed.). *The Cambridge Handbook of Germanic Linguistics*. Cambridge: Cambridge University Press, 2020. p. 143-166.

KÖHNLEIN, B.; CAMERON, I. A foot-based approach to interactions of tonal accent and consonantal strength. *Proceedings of NELS 50*, v. 2, p. 119-132, 2020.

KÖHNLEIN, B.; ZHU, Y. Restricting the Power of Cophonologies: A Representational Solution to Stem Allomorphy in Uspanteko. *Proceedings of the 2018 Annual Meeting on Phonology*. San Diego, 2019.

MORRISON, D. A. Metrical structure in Scottish Gaelic: tonal accent, glottalisation and overlength. *Phonology*, v. 36, n. 3, p. 391-432, 2019.

MÜNCH, F. *Grammatik der ripuarisch-fränkischen Mundart*. Bonn: Cohen, 1904.

ODDEN, D. Some theoretical issues in Estonian prosody. In: LEHISTE, I.; ROSS, J. (ed.). *Estonian prosody: Papers from a symposium*. Tallinn: Institute of Estonian language, 1997. p. 165-195.

OOSTENDORP, M. V. Tone, Final Devoicing and Assimilation in Moresnet. In: KEHREIN, W.; KÖHNLEIN, B.; BOERSMA, P.; OOSTENDORP, M. (ed.). *Segmental Structure and Tone*. Berlin: Walter de Gruyter, 2017. p. 237-252.

PETERS, J. The Cologne word accent revisited. In: DE VAAN, M. (ed.). *Germanic Tone Accents. Proceedings of the First International Workshop on Franconian Tone Accents, Leiden, 13-14 June 2003*. Zeitschrift für Dialektologie und Linguistik – Beiheft 131. Stuttgart: Franz Steiner Verlag, 2006. p. 107-133.

PRINCE, A. Quantitative Consequences of Rhythmic Organization. In: DEATON, K.; NOSKE, M.; ZIOLKOWSKI, M. (ed.). *CLS 26-II: Papers from the Parasession on the Syllable in Phonetics and Phonology*. Chicago: Chicago Linguistic Society, 1980. p. 355-398.

SCHMIDT, J. E. *Die Mittelfränkischen Tonakzente (Rheinische Akzentuierung)*. Mainzer Studien zur Sprach- und Volksforschung 8. Stuttgart: Franz Steiner Verlag, 1986.

WIESINGER, P. *Phonetisch-phonologische Untersuchungen zur Vokalentwicklung in den deutschen Dialekten*. Bd. 1: Die Langvokale im Hochdeutschen. Bd. 2: Die Diphthonge im Hochdeutschen. *Studia Linguistica Germanica*. 2. Berlin: Walter de Gruyter, 1970.



Bilingualism

Chapter 3

O usuário de L2 sob perspectiva bilíngue

Ricardo Augusto de Souza

Neste capítulo, proponho aos leitores algumas considerações sobre questões que se impõem para o estudo dos usuários de línguas adicionais, tendo por ponto de vista privilegiado a perspectiva engendrada pelos estudos sobre o processamento da linguagem por bilíngues acerca dos falantes plurilíngues. Em outras palavras, tenho a expectativa de que este escrito faça o compartilhamento do que entendo serem os aspectos principais da chave de compreensão acerca do falante bilíngue que nos é ofertada pela psicolinguística do bilinguismo.

Compõem também o campo da psicolinguística os estudos que abordam a ontogênese e os marcos desenvolvimentais que tipificam a aquisição de linguagem. Todavia, tal como deixarei explícito mais adiante neste capítulo, a Aquisição de Segunda Língua se constituiu como uma subárea dos estudos da linguagem que preservou certa autonomia em relação à psicolinguística do bilinguismo, a despeito da comunhão esparsa de aportes metodológicos e quadros conceituais. De modo decisivo, como buscarei demonstrar em minha argumentação, a chave de interpretação trazida pela psicolinguística do bilinguismo discrepa de modo fundamental daquela que pode ser tomada como a que funda os estudos sobre os falantes de L2 que se alinham até recentemente ao campo de Aquisição de Segunda Língua.

Por outro, como também buscarei apontar na argumentação do presente capítulo, em tempos recentes, estudos em Aquisição de Segunda Língua passam a encampar clara e ostensivamente a perspectiva bilíngue que se consolidou nos estudos psicolinguísticos do bilinguismo. Este é, creio eu, uma constatação que por si só reitera a relevância da consciência das fronteiras que outrora separavam um e outro campo e que presentemente se dissolvem, especialmente para o estudioso ou pesquisador jovem ou recém-chegado ao tema das línguas adicionais. Assim, julgo por entender que, no tempo presente da redação deste capítulo, esses jovens ou recém-chegados estudiosos e pesquisadores certamente dialogarão com fontes que podem orientar-se ainda na distinção que durante muito tempo conferiu às duas subáreas direções ortogonais. Sobretudo, entendo que esta chave de interpretação encampada no processamento da linguagem por bilíngues é hoje imprescindível aos pesquisadores que se debruçam sobre os fenômenos do uso de línguas não maternas, seja seu foco de estudos a prosódia, a fonologia e a oralidade, ou outros níveis e âmbitos da organização linguística. E creio eu ainda que ela é imprescindível até mesmo para os docentes dedicados à educação linguística em línguas adicionais.

O capítulo está organizado em três movimentos expositivos. Inicialmente, delineio a especificidade dos objetivos e dos objetos da Aquisição de Segunda Língua a partir do que entendo ser o conceito-chave que inaugura tal campo como uma subárea dos estudos da linguagem. Em seguida, exponho construtos e hipóteses que permeiam os estudos sobre os falantes de línguas adicionais, construtos esses que balizaram os debates constitutivos da própria existência da Aquisição de Segunda Língua como área de investigação científica com relativa autonomia teórica e metodológica. Passo, então, a

uma discussão acerca do campo da Psicolinguística do Bilinguismo, buscando delinear a especificidade do tratamento que esse campo oferece, através de estudos em processamento da linguagem por bilíngues, acerca dos construtos acima mencionados. Concluo o capítulo reiterando a justificativa que julgo pertinente à escrita deste capítulo para uma obra que compila estudos sobre um objeto que não é abordado especificamente neste capítulo: a prosódia e a fala em L2.

I. A aquisição de segunda língua

O alçamento do aprendiz de segunda língua e do falante de línguas adicionais a objetos da investigação linguística moderna ocorre a partir do conceito de interlíngua (SELINKER, 1972). O conceito de interlíngua operou uma reformulação da concepção acerca das diferenças entre as formas produzidas pelo aprendiz e falante de uma dada L2 e os falantes daquelas línguas como falantes nativos. Sob tal reformulação, tais diferenças passam a ser apreendidas como a manifestação observável de um sistema de conhecimento linguístico representado mentalmente tal qual uma língua materna, ainda que possivelmente tendo sua ontogênese modulada por fatores que não entram em cena na ontogênese da(s) primeira(s) língua(s), tais como efeitos do tipo treinamento/instrução, de estratégias de aprendizagem e de estratégias de comunicação (*op. cit.*).

O conceito de interlíngua expande a inteligibilidade de fenômenos que foram, até seu surgimento como operador teórico, majoritariamente categorizados como erros e falhas de aprendizagem do aprendiz de L2. Assim vistos, esses fenômenos eram debatidos primariamente em termos das escolhas didáticas que se mostrariam mais eficientes para evitá-los. A hipótese da interlíngua, por outro lado, situa o “erro” no uso de uma L2 como parte de um mecanismo psicológico subjacente à aquisição da linguagem, não como falha didática ou fracasso no engajamento do aprendiz com o processo de aprendizagem. E a elucidação dos componentes, fatores e variações desse mecanismo psicológico torna-se o objetivo do campo de estudos em Aquisição de Segunda Língua, estudos esses que, após a interlíngua selinkeriana, surgem tanto epistemologicamente (THOMAS, 2013a) quanto institucionalmente (THOMAS, 2013b).

Os estudos em Aquisição de Segunda Língua instaurados a partir dos anos de 1970 passam a ter por escopo o estudo do desenvolvimento e das alterações qualitativas dos estados de interlíngua. Para atingir seus objetivos, esses estudos enfocaram fatores propriamente linguísticos, fatores cognitivos do aprendiz, fatores relativos aos ambientes e contextos de aprendizagem e também as interações entre esses grupos de fatores (para panoramas da área, recomendo PAIVA, 2014 e SOUZA, 2021). O paradigma orientador desse programa investigativo, severamente criticado em mais de um momento de sua história (BLEY-VROMAN, 1983; GROSJEAN, 1989; ORTEGA, 2013), passou a ser majoritariamente a comparação entre o aprendiz de segunda língua com o aprendiz de primeira língua.

Em tal comparação, sempre fica evidente a diferença entre a altamente notória variabilidade das consequências da aquisição de L2, no tocante ao seu resultado último, e a aparente homogeneidade das consequências da aquisição da língua materna. Da aquisição de primeira língua, o resultado último, salvaguardadas situações de desenvolvimento neurológico, psicossocial e cognitivos altamente atípicos, o resultado é certa e inevitavelmente um falante nativo da variante linguística da comunidade de fala dos cuidadores do bebê-aprendiz, inevitabilidade que alcança ao menos até o marco a partir do qual desigualdades no acesso a bens socioculturais tão cruéis e vergonhosas como as existentes no Brasil produzem diferenças detectáveis nos usos de linguagem. Da aquisição de segunda língua, ao contrário, é inexistente qualquer expectativa inevitável e certa de resultados, assim como é rara a indiferenciação em relação a falantes nativos de uma dada comunidade de fala da L2 (condição usualmente denominada *nativelikehood* na literatura especializada em inglês), raridade diminuída apenas por circunstâncias de início muito precoce de aprendizagem da L2 associada à imersão em meio social onde ela é a língua dominante.

Passo, em seguida, à discussão sobre três noções basilares nos estudos sobre falantes de línguas adicionais, noções essas que nortearam o empreendimento científico da Aquisição de Segunda Língua. São elas os construtos “aprendizibilidade” e transferência linguística e a hipótese de que as interlínguas são sistemas deficitários nas representações gramaticais que os configuram.

2. Aprendizibilidade, transferência linguística e a hipótese do *déficit* representacional nas línguas adicionais

A aprendizibilidade é provavelmente o problema central dos estudos em aquisição de L2 fomentados por teoria linguística do fim da década de 1980 até a primeira década do presente século. Genericamente, o problema da aprendizibilidade pode ser formulado como a pergunta sobre ser a incorporação de uma determinada estrutura ou construção, ao repertório expressivo do aprendiz de uma L2, condição suficiente para que no sistema da L2 sendo internalizado pelo aprendiz ocorram generalizações para um conjunto de fenômenos gramaticais tidos como relacionados à estrutura ou construção em questão.

É através da teoria da gramática que se estabelecem hipóteses de existência de redes ou conjuntos de fenômenos gramaticais correlacionados e subjacentes à ocorrência de uma dada estrutura ou construção, sendo esta a razão da indispensabilidade da teoria linguística para os estudos sobre o problema da aprendizibilidade em L2. As transposições da teoria gramatical para a aquisição de L2 são encontradas em vários estudos, acerca de diferentes fenômenos. No início deste nosso século, problematizações sobre a aprendizibilidade de padrões de realização de estrutura argumental no contexto de línguas adicionais tiveram algum destaque nesse debate. Tal destaque pode ser verificado em Juffs (2000), Montrul (2001) e White (2003).

O trabalho de Montrul (2001) chama a atenção para um contraste estrutural entre o espanhol e o inglês, que ocorre também entre o português e o inglês. Trata-se de uma divergência no licenciamento de construções de leitura causativa com verbos que expressam diferentes modos de movimento (ex.: correr, andar, pular, saltar etc.). Esses predicadores ocorrem tanto no inglês quanto no português como verbos intransitivos, configurando orações nas quais ao sujeito sintático atribui-se uma leitura agentiva (ou seja, a interpretação semântica de agente do evento expresso pelo verbo. São exemplos em cada uma das duas línguas, respectivamente, os itens (1) e (2) abaixo.

(1)- The horse jumped.

(2)- O cavalo pulou.

Por outro lado, somente o inglês licencia uma construção na qual ao sujeito sintático se atribui a leitura de iniciador de um evento causativo, em relação ao qual a ação de movimento expressa pelo verbo passa a ser interpretada como um sub-evento causado por tal sujeito. Nesse sub-evento causado, o papel de agente da ação de movimento é atribuído ao referente de um sintagma nominal realizado sintaticamente como objeto direto do verbo. A denominação “alternância de movimento induzido” é usada para tal construção (LEVIN, 1993). Tal denominação encampa o pressuposto teórico derivacional, aludindo a uma interpretação da construção como um caso de alternância de transitividade. Sem nos comprometer com a explicação derivacional, a denominação alternância de movimento induzido parece-me interessante por parecer-me que relembrar-nos que um subgrupo dos usualmente intransitivos verbos de modo de movimento da língua inglesa pode surgir como verbos transitivos diretos, produzindo a interpretação aqui discutida. Essa configuração sintática lícita no inglês, exemplificada no item (3), é anômala no português, tal como indicado na notação usada no item (4).

(3)- The girl jumped her poney over the fence.

(4)- */?? A menina pulou seu pônei por sobre a cerca.

Os trabalhos de Juffs (2000) e Montrul (2001) lançaram hipóteses de que distinções gramaticais do tipo acima descrito, relacionadas à expressão linguística de estruturas argumentais, são apoiadas por um conjunto de fatores semântico-lexicais e morfossintáticos sutis o suficiente para constituir um possível problema de aprendizibilidade em L2, problema este associado visceralmente aos dois outros temas que nos propomos a aqui discutir. São esses dois temas igualmente centrais aos estudos sobre falantes de línguas adicionais que se orientam pela Aquisição de Segunda Língua e que têm por objetivo descrever as interlínguas e seus condicionantes. Trata-se da transferência linguística e a hipótese de que há déficit nas representações gramaticais que dão suporte aos sistemas de interlíngua.

A transferência linguística é um fenômeno que manteve interesse teórico ao longo de décadas de pesquisas sobre as línguas adicionais (para revisões, ver ODLIN, 1989; JARVIS; PAVLENKO, 2007), interesse ainda vivo. A motivação da longevidade desse interesse científico é que não se trata de um fenômeno que pode ser previsto *a priori*. Ou seja, quais diferenças na organização de duas línguas acarretarão o surgimento de uma estrutura ou construção da língua X no uso da língua Y é sempre uma questão a ser respondida por observação empírica, não se tratando de uma ocorrência autoevidente ou derivável logicamente da pura descrição da existência de uma diferença nas respectivas gramáticas.

Assim, estudos sobre o fenômeno da transferência linguística usualmente têm objetivos tais como identificar construções passíveis de transferência interlinguística, delimitar propriedades linguísticas que possam condicionar a transferibilidade interlinguística de certas construções, avaliar as possíveis relações entre transferência e aprendizibilidade, e descrever circunstâncias relacionadas às diferenças individuais e/ou às contingências nas condições de processamento linguístico que modulam a ocorrência ou não de episódios de transferência.

Apesar da longevidade da pesquisa sobre transferência nos estudos sobre a L2 e dos avanços coletivamente realizados, não há ainda uma teoria geral consensual que explique a variabilidade da transferência linguística. Portanto, pesquisas sobre a transferência interlinguística são sobretudo estudos teóricos sobre a natureza e os mecanismos deste fenômeno da linguagem humana. É o verdadeiro escopo desses estudos que justifica que pesquisadores observem um mesmo fato linguístico (estrutura ou construção), ou família de fatos linguísticos com algum âmbito de interconexão, ao longo de vários anos.

A hipótese do déficit representacional, por sua vez, é uma hipótese sobre os limites da aprendizibilidade de L2. Tal hipótese propõe que haja uma incapacidade, estruturalmente determinada e de natureza linguística, de que o aprendiz de L2 possa atingir uma representação gramatical na L2 comparável aos falantes nativos. O debate que motivou tal hipótese nos estudos sobre a L2 é o problema da aquisição de morfologia flexional em L2. A morfologia flexional, especialmente a concordância entre sujeito sentencial e verbo (auxiliar ou lexical) e a marcação de tempo e aspecto, em situações de produção de enunciados espontânea e não ensaiada, costuma revelar erros em relação às normas usuais dos falares nativos cultos mesmo entre falantes de L2 com elevada proficiência e fluência em suas línguas adicionais.

Tal aparente dificuldade com a morfologia flexional de L2 guarda grande interesse empírico e teórico. Sobretudo, ela implica problematizações importantes sobre a questão da transferência linguística. O desafio posto à hipótese da transferência linguística como mecanismo formador das representações de interlíngua advém do fato de que a dificuldade na aquisição de morfologia flexional aparentemente não se limita a diferenças tipológicas entre línguas. Não se trata de um problema

de aprendizibilidade restrito à situação de aquisição na L2 de morfemas que não são realizados na gramática da L1. Por exemplo, o sistema de marcação morfológica de tempo e concordância do inglês tem um análogo no português, sendo o sistema do português consideravelmente mais complexo, do ponto de vista descritivo, do que o sistema do inglês. Não obstante, verifica-se com frequência a omissão ou a supergeneralização anômala de marcação morfológica de tempo e concordância na produção espontânea de enunciados em inglês como L2 de seus usuários cujo L1 é o português, sugerindo que não há um efeito facilitatório da gramática da L1 impulsionado pelo abundante e relativamente consistente sistema de morfologia flexional na marcação de tempo e concordância do português.

A permanência de dificuldades na morfologia flexional de L2 ocorre mesmo em circunstâncias onde a língua adicional pode ser descrita como tendo uma morfologia manifesta mais enxuta do que uma língua previamente falada pelo aprendiz. Isto pode ser interpretado como uma deficiência na representação gramatical alcançável pelo aprendiz de L2. Essa é uma questão que se alinha a um problema teórico mais amplo: o quanto um sistema de interlíngua efetivamente pode ser tratado como uma língua natural. Trata-se, assim, de uma questão que se encontra no cerne da própria adequação de situar-se a L2 como um objeto de estudo de interesse para a Linguística teórica e descritiva.

Passemos, então, ao exame dos estudos do processamento da linguagem por bilíngues, para verificarmos as contribuições que esse nicho de investigação traz para o problema da transferência linguística e para a hipótese do *déficit* representacional nas línguas adicionais.

3. A Psicolinguística do bilinguismo

O estudo do processamento da linguagem é um dos campos de estudo para o qual convergem, como componentes centrais, conceitos e construtos erigidos na Psicologia Cognitiva. Trata-se de conceitos e construtos que subsidiam genericamente os modelos de processamento da informação, ou seja, modelos que buscam explicar os mecanismos mentais humanos responsáveis pela percepção de unidades significativas do ambiente e sua transformação em representações mentais (em outras palavras, conhecimento armazenado em memória). Através de modelos de processamento da informação, busca-se igualmente compreender os mecanismos que dão suporte ao acesso às representações mentais necessárias para o desempenho das atividades cognitivas humanas, assim como as interações entre as representações adquiridas e a apreensão de novas informações no ambiente.

Por dirigir-se especificamente para informações linguísticas, o estudo do processamento da linguagem propõe-se a desvendar os mecanismos subjacentes à compreensão e à produção de linguagem. Para realizar tal objetivo, além dos conceitos e construtos relevantes nos modelos de

processamento da informação em geral, tal estudo precisa apoiar-se em modelos sobre a constituição das unidades informacionais decisivas para a percepção e para produção da comunicação verbal. Portanto, o estudo do processamento da linguagem reúne com equilíbrio razoável teorias e quadros de análise da Psicologia Cognitiva e da Linguística. Ao lado da aquisição de linguagem, o processamento da linguagem compõe os objetos da Psicolinguística.

Os estudos em processamento da linguagem exploram hipóteses sobre processos que ocorrem durante a decodificação e a codificação de mensagens. Tais processos ocorrem muito rapidamente e envolvem operações cognitivas automatizadas, ou seja, que são feitas sem necessariamente a concorrência de foco de atenção intencional e consciente. A abordagem metodológica para a captura dos dados empíricos relevantes envolve comumente algum tipo de cronometria mental: a mensuração de tempos de resposta a estímulos. Essa abordagem metodológica é promissora para a investigação de hipóteses sobre a mecânica do acesso em tempo real a representações e processos mentais. Especificamente no tocante ao estudo da linguagem humana, ela contorna limites das informações obtidas por dados oriundos de julgamentos e percepções introspectivos. Tais limites são impostos pelo fato de que a autorreflexão introspectiva usualmente acessa representações e estados mentais que são o produto dos processos que transcorrem em tempo real de codificação e decodificação linguística. Por mais pistas que a análise de um produto revele sobre os processos de produção, há muitos detalhes importantíssimos para quem deseja conhecer os mecanismos da produção que não são revelados pela contemplação, ainda que muito engenhosa, do produto, seja ela introspectiva ou empírica.

Ao invés da problematização dos limites das dificuldades na aprendizagem, sempre mensurados em relação a uma norma atribuída abstratamente à competência idealizada do falante nativo, os estudos em psicolinguística do processamento da linguagem por bilíngues ocupam-se primariamente do impacto da vivência como usuário de mais de uma língua sobre a experiência linguística total de usuários de L2. Deriva-se disso a principal diferença entre a compreensão proposta pelos estudos em Aquisição de Segunda Língua que prevaleceram naquele campo entre os anos de 1970 e a primeira década do século atual e a Psicolinguística do Bilinguismo.

A Aquisição de Segunda Língua teve por objeto um falante a quem falta a semelhança com um falante nativo monolíngue tido como típico (e frequentemente idealizado), propondo-se o objetivo de explicar os condicionantes dessa falta. A Psicolinguística do Bilinguismo, por sua vez, assume como pressuposto de partido tal falta. Em uma paráfrase das famosas palavras de Grosjean (1989), um bilíngue não é dois monolíngues em uma única mente, pois o bilinguismo não permite, nem sob um prisma empírico nem tampouco lógico, um retorno ao estado de monolingüismo ou a chegada a algo análogo a tal estado como resultado de um processo de aprendizagem. Portanto, os estudos em Psicolinguística do Bilinguismo centram-se em investigações sobre a singularidade do uso da

linguagem por bilíngues, ocupando-se do que tal condição de falante acrescenta e/ou altera a seus usos e processos cognitivos mediados por linguagem, não no desvendamento do que neles está ausente: o monolinguismo.

Além disso, a Psicolinguística do Bilinguismo rompe barreiras impostas por uma concepção de bilíngue tradicional: o bilíngue como o indivíduo que adquiriu duas (ou mais) línguas simultaneamente, durante a primeira infância, no período de aquisição primária de linguagem. Essa concepção, da qual a Psicolinguística do Bilinguismo se afasta (e também a literatura sobre educação bilíngue, como eu viria a notar), delimita como bilíngue a idealização de um falante igualmente proficiente em duas línguas, e pode ser rastreada a Bloomfield (1933). O principal problema da definição tradicional é que ela reduz catastróficamente o espaço de interesse, perdendo de vista uma enormidade de fenômenos relevantes.

Na Psicolinguística do Bilinguismo encontra-se uma expansão crucial do problema da transferência linguística, ilustrada por trabalhos sobre os efeitos da L2 sobre a L1. São exemplos desses trabalhos Cook (1996), Cook (2003), Pavlenko e Jarvis (2002) e Jarvis e Pavlenko (2007). São trabalhos que constroem as noções como “transferência bidirecional” e de “multicompetência” para descrever o estado de conhecimento linguístico total do falante bilíngue. Ao recusar a abordagem tradicional, que toma a L1 como a única fonte de influências interlinguísticas, tais noções libertam-se da visão parcial dos usuários de L2 como indivíduos linguisticamente incompletos por não poderem ter sua competência descrita como idêntica ao que se supõe usualmente sobre o falante nativo monolíngue.

Na visão tradicional, a L1 é concebida como um sistema estável, ao qual se contrapõe a L2 instável, maleável e influenciável. Porém, a observação de que o falante de línguas adicionais pode manifestar um uso da L1 distinto do falante monolíngue daquela língua salienta que a totalidade, e não apenas parte do conhecimento linguístico dos falantes é maleável, influenciado por sua história de uso de línguas e, por isso, não estável, mas sim dinâmica.

Nesta perspectiva, diminui-se sensivelmente a relevância de explicações sobre supostos déficits representacionais dos falantes de L2 em relação a uma norma monolíngue idealizada. Não há menos no repertório linguístico do usuário de mais de uma língua. Há mais. Essa perspectiva compatibiliza-se prontamente com outros fenômenos do estudo do bilinguismo, como a mudança de código (*code-switch*) e a erosão linguística, ampliando o campo de fenômenos que guardam interesse. Esta ampliação abre espaços para a busca de modelos teóricos unificadores da aquisição de L2, dos processos subjacentes ao uso de mais de uma língua e também da perda de fluência e de capacidades linguísticas. As arquiteturas cognitivas propostas como mecanismos de processamento eram capazes de descrever e prever os “erros” do usuário de L2 com explicações mais econômicas e, principalmente, psicologicamente mais plausíveis, que o maquinário conceitual que tentava localizar essas explicações em caracterizações internas do sistema linguístico internalizado pelo aprendiz.

Voltemos, então, à hipótese do déficit representacional como característica do conhecimento adquirido pelo falante de L2 e acessado por tal falante no uso de sua(s) língua(s) adicional(ais). Essa hipótese pode ser associada a explicações que evocam o efeito da idade de aquisição.

O início da aprendizagem de uma segunda língua em tenra idade é o fator que mais prevê de altíssima proficiência em L2 (ou *nativelikeness*). O poder preditivo da idade de início da aquisição de L2 sobre a situação de *nativelikeness* é um consenso estabelecido ao longo de mais de quatro décadas de pesquisas (LONG, 2013).

O que não é absolutamente um consenso é a explicação desse efeito. Muitos pesquisadores em L2 aderem à hipótese de que o efeito da idade é reflexo de processos maturacionais de natureza neurológica. Esta hipótese é usualmente atraente para pesquisadores que compreendem haver determinações biológicas exclusivamente dedicadas à ontogênese da linguagem humana na cognição humana, uma vez que ela pode ser evocada como evidência independente da operação de tais processos desenvolvimentais exclusivamente linguísticos e biologicamente determinados. A hipótese prevê o encerramento de períodos ótimos, após certos marcos maturacionais, para o alcance de altíssimos níveis de proficiência em L2. Uma vez encerrado esse período crítico, o alcance desses níveis elevados de habilidade em segunda língua seria inviável.

Porém, há grandes dificuldades tanto na caracterização exaustiva de períodos críticos e também na explicitação detalhada dos mecanismos supostamente envolvidos (BIRDSONG, 2006), o que lança dúvidas sobre o real alcance explanatório dessa hipótese. Há uma hipótese alternativa à hipótese de períodos críticos determinados por fatores maturacionais (BYLUND; HYLSTENSTAM; ABRAHAMSSON, 2013; MACWHINNEY, 2017). Essa hipótese é pautada pela noção de entrincheiramento, que pode ser superficialmente definida como a forte estabilização da disponibilidade de uma representação mental, o que automatiza o acesso a ela. A perspectiva dessa hipótese alternativa é que, com a consolidação progressivamente mais forte do entrincheiramento da L1, em função da frequência e da constância de seu uso ao longo dos anos, mais recursos cognitivos teriam que ser alocados para a ruptura da automaticidade da implementação das rotinas de processamento condicionadas pela organização linguística da L1. Essa hipótese prevê, portanto, que quanto mais maduro um aprendiz de L2 for como usuário de sua L1, maior será o custo cognitivo para adequar-se aos padrões de organização e aos procedimentos de implementação da L2 e, conseqüentemente, menos semelhante ao desempenho de um nativo será o desempenho de tal aprendiz da L2.

Essa hipótese alternativa é nuançada, sendo na verdade proposta como uma constelação de fatores em interação (MACWHINNEY, 2017). Nessas interações dinâmicas, prevê-se a existência de fatores “protetores”, que seriam capazes de explicar a elevada proficiência alcançada por alguns aprendizes de L2, ainda que eles possam não necessariamente ser indistintos de falantes nativos da L2 na totalidade dos níveis de organização da linguagem, da articulação fonética à idiomatidade

das escolhas léxico-gramaticais no discurso (MACWHINNEY, 2013, 2017). De interesse particular para o pensamento que venho desenvolvendo nos anos mais recentes é que a constelação de fatores em interação dinâmica que a hipótese do entrenchamento aponta como alternativa para a explicação do fenômeno da correlação entre idade de início de aquisição de L2 e *nativelikeness* é prevalentemente relacionada às experiências de uso da L2. E o uso é exatamente o mecanismo geral previsto nos modelos probabilísticos e que entrelaçam representação mental na gramática das línguas naturais e modelos de arquiteturas cognitivas de implementação do conhecimento linguístico. Ou seja, perspectivas propriamente psicolinguísticas.

4. Breve conclusão

A principal motivação para esta exposição de caráter geral dentro de uma obra que reúne trabalhos especificamente voltados à prosódia é meu entendimento de que tal chave de compreensão discrepa relevantemente do entendimento acerca do falante de L2 que norteou os estudos em Aquisição de Segunda Língua entre os anos de 1970 e parte da primeira década do século atual. É somente mais recentemente que é possível verificar-se uma convergência nos debates e nos estudos das comunidades de pesquisa em aquisição de L2 e psicolinguística do bilinguismo, convergência esta que sugere que a discrepância entre as áreas avança célere para a obsolescência. Todavia, parece-me que a perspectiva mais tradicional da aquisição de L2 ainda prevalece em muitos estudos sobre a aprendizagem e o uso de línguas adicionais no nosso país. Presumindo meu acerto nesta avaliação, julgo que o tema aqui trabalhado se apresenta como uma contribuição ainda útil para o leitor brasileiro.

Referências

BIRDSONG, D. Age and second language acquisition and processing: A selective overview. *Language Learning*, v. 56, supl. 1, p. 9-49, 2006.

BLEY-VROMAN, R. The Comparative Fallacy in Interlanguage Studies: The Case of Systematicity. *Language Learning*, v. 33, n. 1, p. 1-17, 1983.

BLOOMFIELD, L. *Language*. New York: Holt Rinehart and Winston, 1933.

BYLUND, E.; HYLSTENSTAM, K.; ABRAHAMSSON, N. Age of acquisition effects or effects of bilingualism in second language ultimate attainment? In: GRANENA, G.; LONG, M. (org.). *Sensitive Periods, Language Aptitude, and Ultimate L2 Attainment*. Amsterdam/Philadelphia: John Benjamins, 2013.

COOK, V. Competence and multicompetence. In: BROWN, G.; MALMKJAER, K.; WILLIAMS, J. (org.). *Competence and Performance in Second Language Acquisition*. Cambridge/New York: Cambridge University Press, 1996.

COOK, V. The changing L1 in the L2 user's mind. In: COOK, V. (org.). *Effects of the Second Language on the First*. Bristol: Multilingual Matters, 2003.

GROSJEAN, F. Neurolinguists, Beware! The Bilingual is Not Two Monolinguals in One Person. *Brain and Language*, v. 36, n. 1, p. 3-15, 1989.

JARVIS, S.; PAVLENKO, A. *Cross-Linguistic Influences in Language and Cognition*. New York: Routledge, 2007.

JUFFS, A. An overview of the second language acquisition of links between verb semantics and morpho-syntax. In: ARCHIBALD, J. (org.). *Second Language Acquisition and Linguistic Theory*. Oxford: Blackwell, 2000.

LEVIN, B. *English Verb Classes and Alternations*. Chicago: The University of Chicago Press, 1993.

LONG, M. Maturational constraints on child and adult SLA. In: GRANENA, G.; LONG, M. (org.). *Sensitive Periods, Language Aptitude, and Ultimate L2 Attainment*. Amsterdam/Philadelphia: John Benjamins, 2013.

MACWHINNEY, B. Entrenchment in second language acquisition. In: SCHMID, H-J. (org.). *Entrenchment and the Psychology of Language Learning – How We Reorganize and Adapt Linguistic Knowledge*. Washington/Berlin: American Psychological Association/Mouton DeGruyter, 2017.

MONTRUL, S. Agentive verbs of manner-of-motion in Spanish and English as second Languages. *Studies in Second Language Acquisition*, v. 23, n. 2, p. 171-206, 2001.

ODLIN, T. *Language Transfer – Cross-linguistic Influence in Language Learning*. Cambridge/New York: Cambridge University Press, 1989.

ORTEGA, L. Ways forward for a bi/multilingual turn in SLA. In: MAY, S. (ed.). *The multilingual turn: Implications for SLA, TESOL and bilingual education*. New York, NY: Routledge, 2013.

PAVLENKO, A.; JARVIS, S. Bidirectional transfer. *Applied Linguistics*, v. 23, n. 2, p. 190-214, 2002.

PAIVA, V. L. M. O. *Aquisição de Segunda Língua*. São Paulo: Parábola, 2014.

SELINKER, L. Interlanguage. *IRAL – International Review of Applied Linguistics in Language Teaching*, v. 10, n. 1-4, p. 209-231, 1972.

SOUZA, R. A. *Segunda Língua: aquisição e conhecimento*. São Paulo: Parábola, 2021.

THOMAS, M. History of the study of second language acquisition. In: HERSCHENSOHN, J.; YOUNG-SCHOLTEN, M. (org.). *The Cambridge Handbook of Second Language Acquisition*. Cambridge/New York: Cambridge University Press, 2013a.

THOMAS, M. The doctorate in Second Language Acquisition: An institutional history. *Linguistic Approaches to Bilingualism*, v. 3, n. 4, p. 509-531, 2013b.

WHITE, L. *Second Language Acquisition and Universal Grammar*. Cambridge: Cambridge University Press, 2003.

Chapter 4

Implications of a usage-based approach to contact linguistics

Ad Backus

I. Introduction

Multilingualism is pervasive in the world, and comes in many manifestations. As a function of this diversity, there are also many ways in which multilingualism can affect people's lives. Considering this pervasiveness, it is somewhat surprising that linguistics, the discipline that studies language, has not actually reserved a central spot for the phenomenon among the things that need explaining about language. Within the sub-disciplines of sociolinguistics and psycholinguistics, multilingualism has become a primary topic of investigation, but this has only had a limited impact on linguistics in general, partly because those sub-disciplines are themselves often cornered off from 'theoretical linguistics'. While for many decades there has been a workable, though perhaps regrettable, division into three relatively self-contained sub-disciplines, a usage-based approach to linguistic theory calls for a much more serious integration of them. This paper aims to argue for that point.

Empirically, I will draw on work on language contact, mostly between Turkish and Dutch in the Netherlands, that I have been involved in as a primary investigator, team member, or supervisor in the past decades. Most of this work adopts a usage-based perspective on contact effects. By and large, work on Turkish-Dutch contact has shown that fifty years after the onset of migration, Turkish is maintained well in the community, perhaps more than could have been expected considering the fate of many immigrant languages rooted in labor migration (BACKUS, 2013). On the other hand, there are also signs that some degree of language shift is underway, and certainly, Turkish itself is undergoing lexical and grammatical changes in its immigration context. However, most of the information is based on fairly small-scale studies, and it is impossible to know whether what has been attested in these studies can be generalized to the community in general. Be that as it may, studies of heritage Turkish invariably demonstrate the abundant existence of code-switching with Dutch and quite a bit of Dutch influence on its grammar.

I will focus specifically on two kinds of studies here. First is work that investigated the degree to which bilingual speakers keep their languages separated or not (studies on code-switching, family language policy or socialization, and bilingual acquisition). Second, I will refer to findings from studies on the degree of grammatical change that can be found in Heritage Turkish (ranging from little evidence of change in some grammatical domains to lots of change in others). I will also draw on other work in the emergent paradigm of usage-based contact linguistics (AALBERSE *et al.*, 2020; ZENNER *et al.*, 2019; BACKUS; HAKIMOV to appear).

Section 1 below will review some relevant findings on language mixing and language separation, while Section 2 will summarize some of our findings regarding the contact-induced lexical and grammatical change. Section 3 will discuss these findings in the light of linguistic, psycholinguistic, and sociolinguistic insights into bilingualism, while Section 4 concludes the paper with a plea to try harder to integrate the dimensions of cognition and sociality into more inclusive linguistics.

2. Language separation

While separating our languages seems second nature to most of us, there are examples of bilingual speech that question this naturalness. Many studies of bilingual acquisition have confirmed that, for example, bilingual children will mix their languages, even if they grow up in families or communities that frown on code-switching. Implicitly or explicitly, these families may instill the norm that while learning to speak two or more languages is a good or necessary skill, these languages should be kept strictly separate in discourse. Parents often struggle to find out what is best for their children and fear the negative consequences of not speaking either language well. The study of language contact or Heritage Languages, with its focus on ‘contact-induced change’, ‘imperfect acquisition’, and ‘attrition’, often shows that at least one of the languages, often the one associated with the ethnic group, undergoes change, partly because parents couldn’t possibly provide their children with the same linguistic input monolingual children get in the country of origin. The available evidence tends to show that input matters: children produce what they have learned, and what they have learned depends on the input they get. On the other hand, this should not be interpreted to mean they only produce rote-learned chunks of language that they imitate. In fact, studies have often observed that children code-switch even though the input doesn’t provide them with examples of code-switching other than the alternation naturally produced when two parents each speak a different language. This code-switching, one could argue, shows that children don’t necessarily know that there are two languages, and if they do, they do not always know to which language a word belongs. It also shows that code-switching provides evidence for productivity: it shows which constructions the child can use to productively form novel utterances.

The ideology of language separation is almost universal, and so is an ideology of full proficiency. While efforts at bilingual socialization often subscribe to an ideology in which ultimate attainment, or perfect language mastery of preferably both languages, is the goal, the outcome may frustrate parents and other community members: manufacturing proficiency inevitably turns out not to be easy. Children simply learn to speak by letting input guide them and by unconsciously exploiting their mind’s response to it, and the outcome is likely to be different from what would count as full proficiency in a monolingual context.

To see how ingrained the notion of language separation is in all of us, note how virtually all concepts in linguistics take language boundaries for granted: ‘language’, ‘codeswitching’, ‘language alternation’, ‘multilingualism’, ‘proficiency’, ‘standardization’, ‘L1’ and ‘L2’. Even the term ‘translanguaging’, which was coined to combat the hidden assumption that we always know what a language is, cannot escape it. If even linguists cannot escape it, it’s a small wonder that ordinary language users can’t either. Our entire lives we keep reinforcing this ideology, usually unwittingly. That is why we can ask

each other how many languages we speak and why we can engage in language learning courses. However, the literature on bilingualism also contains some data that reflect deviations from this ideology, especially intense code-switching and mixed languages. Also, the fact noted above, and to be elaborated on below, that young children in ‘One Parent One Language’ families practice code-switching, provides an interesting challenge. Incidentally, perhaps the same holds for ordinary stylistic variation within a language.

In several recent studies (e.g. QUICK *et al.*, 2018; GASKINS *et al.*, 2019), we studied the language mixing of children who were growing up in such ‘OPOL’ families, combining German and English in three cases, and Polish and English in another. Although these children mixed their languages within utterances while they weren’t exposed to this kind of mixing in their input, they did show other familiar effects of conservative input-driven learning: their dominant language was the language they heard most often in the input (including a switch from German to English for one child after a prolonged stay at his grandparents’ place in Ireland), and it was also the language that functioned as the matrix language most often in bilingual utterances. The example below was, therefore, typical of a German-dominant child.

(1) ich will jetzt *the boat* machen

I want now *the boat* to.make

“I want to build *the boat* now”

One thing that is presumably essential here is that these children were too young to have much sociolinguistic awareness about language. If we consider the structure of their codeswitching, it is interesting to see that the asymmetry that we are familiar with from much adult code-switching, is only represented partially. Sometimes we see indeed the insertion of content words from the weaker language into constructions from the dominant language. At other times, however, we see function words from the weaker language functioning as the basis for constructions that host content words from the dominant language. One German-dominant child in Quick *et al.* (2018), for example, produced a lot of code-switched utterances featuring inserted German nouns into the English phrase ‘I want X’, in the middle of otherwise German discourse.

The most obvious type of language use that goes against the norm that languages have to be kept apart is code-switching. Demirçay (2017) discusses many examples of very intense switching patterns, in which it is often unclear whether a speaker is using Dutch or Turkish. The data were from conversations among second and third-generation young adults, who always knew each other well, speaking informally to each other. Broadly speaking, the recordings show constant back and forth switching between the two languages and the codeswitching, while sometimes the familiar insertion or alternation types (MUYSKEN, 2000) were often difficult to classify. The author labeled them, somewhat informally, ‘chunk-based complex code-switching’.

Generally speaking, the data give the appearance of a speaking style in which the languages are constantly being alternated. Monolingual stretches are rarely longer than two or three utterances. The following fragment exemplifies this; it's part of a conversation between speakers E and Ş, who were recalling one of E's recent experiences with a mutual friend. There is an intense alternation between parts that are in Turkish and parts that are in Dutch (here indicated through italics).

(2) Turkish-Dutch bilingual conversation

E: Öyle işte. *Ja sowieso. Zij zo ja canım acayip sarma çekti falan hatta eve bırakıyodum. Ik zo söyleseydin, aşağıda vardı. Zij zo valla de.*

"So that's it. Yes anyway. *She was like yes I really feel like stuffed vine leaves and so, I was actually dropping her off home. I was like if you had told me, we had some downstairs. She was like really.*"

Ş: *Hadden jullie het gemaakt?*

"Did you make it?"

E: *Wij hadden sowieso aşağıda vardı tencerede vardı ya.*

"We had anyway there was some in the pan downstairs"

Ş: *Ha iyi yapmıştın.*

"oh good, you had made it"

E: *Zij zo ja söyleseydin şapardım enzo.*

"she was like yes if you had told me i would have and so forth"

Ş: *O da mı biliyo sarmayı neyi?*

"she also knows about stuffed vine leaves and so on?"

E: *Ja man. Zij zo kesin bana öğret dit dat. Ik zo öğretiriz.*

"yeah man. *She was like you definitely have to teach me this and that. I was like I'll teach you.*"

When people mix languages like this, they are clearly not keeping their languages separate. At the same time, individual chunks can, of course, still be said to be in one particular language. The data contain monolingual utterances (classic cases of insertional code-switching in which Dutch words are inserted into Turkish utterances and vice versa) and alternation between Dutch and Turkish utterances (see the examples below), but overall the style of talk is marked by constant alternation between chunks in Turkish and chunks in Dutch, often within what could be seen as one utterance.

(3) Natalya'da bıraktık herşeyi. *Ja zij zo gelip gidemem. Ik moet alleen nog die 'handhavingsmiddelen' afmaken, sonra da 'biorechten' de yapamadık o da kaldı.*

"We had left everything at Natalya's. *Well, she's like 'I can't come.' I only have to finish that 'means of enforcement', and after that we couldn't do biolaws anymore, so it's still there*"

(4) *Wij doen salça in die iç.*

We do.PRES.1pl tomato.paste in that filling

"we put tomato paste in that filling"

The permeability of language borders can be looked at from a sociolinguistic and psycholinguistic perspective. Sociolinguistically, one could say that code-switching stands out for us, especially when it is as intense as in the Dutch-Turkish data just presented, because the ideology of separation is so completely ingrained in all of us that we find this kind of language use noteworthy. The ideology and the resulting practice of monolingual speech harden the borders so that we see monolingual language use as the default, as the thing that is normal. This in itself entrenches the fact that a particular word or expression belongs to a particular language as part of its meaning (or at least as part of the knowledge we have about that word), and lending importance or salience to this social meaning has the cognitive effect that we keep reinforcing the boundaries between languages in our minds. We constantly remind ourselves of the existence of different languages and which components belong to which language. What code-switching shows, though, is that the pressure of speedy communication (which CHRISTIANSEN; CHATER, 2016, refer to as the ‘now-or-never bottleneck’) may make us say things sometimes that go against separation, and this is where the psycholinguistic perspective comes in.

Language use is a constant back-and-forth between automatic production and conscious selection. Without relying on automaticity, we wouldn’t be able to communicate as fast and efficiently as we do; without relying on conscious selection, we wouldn’t be able to choose our words carefully when we need to. What happens in bilingual speech is that often an element may get activated that happens to be from the other language than the last uttered element was from, and the result is code-switching. The model of speech underlying all this is that all language use is the result of the division of labor between ease of activation and optimal selection as the two factors that govern everything we say. In some bilingual communication, this leads to what we as linguists perceive to be intense code-switching, but which for speakers is simply the use of language in a normal way: activation of easily activated units. In most other cases, this leads to monolingual language use, thanks to the well-practiced convention of keeping languages separate.

The usage-based insight here is that if the input is largely monolingual, or ‘bilingual but separate’ as in OPOL acquisition settings or bilingual communities with clear ‘rules’ for when to use which language, then also all multiword units and constructions that are stored in memory are ‘monolingual’; code-switching will then be rare and marked. In such settings, people often seem to flag their code-switches with metalinguistic ‘apologies’ (shown, for example, by Treffers-Daller’s 1994 analysis of switching between French and Dutch in Brussels; Belgium is notorious for its ‘language struggle’). If there is no normative pressure to separate, however, you are free to just use all resources you have acquired (and that you know your interlocutor has acquired too), and code-switching can develop into the ‘normal’ way of speaking, categorized by Myers-Scotton (1993) as ‘CS as the unmarked choice’. Under particular circumstances, this may develop into a conventionalized mixed lect, or even a mixed language.

For Heritage Turkish, we don't really know enough to generalize, as many different kinds of language use have been documented. Perhaps the combined studies show an abundance of code-switching rather than monolingual language use, but this may be a sampling issue: spontaneous conversation has mostly been investigated by linguists interested in code-switching, who, therefore, investigate data that show it. The same participants whose codeswitching we observed also spend much of their time in monolingual mode, mostly Dutch (at work or school, for example) but also Turkish, for example with older family members, with family and friends in Turkey, or with people of whom they know that they disapprove of language mixing. All we can say is that bilingual speech has been abundantly documented and that speakers indeed claim in language choice questionnaires that they often use both languages at the same time in informal in-group conversation (e.g. in AZAR *et al.*, 2019).

3. Lexical and grammatical change

As contact linguists, we are generally happy to suspend any doubts regarding the suitability of separating the languages and analyze what happens to the languages separately. As a result, we know a lot about lexical and grammatical changes undergone by minority languages, often supported by data that were collected in ways that forced speakers to produce monolingual speech, or by analyses of bilingual speech in which all parts of the data that are in the other language are, for now, ignored (in the extract represented above, for example, one could ignore all Dutch chunks and just analyze the structure of the parts that are in Turkish).

While this practice may ignore an essential aspect of bilingual speakers' language use, leaving its analysis to the separate research domain of 'code-switching', it is beyond doubt that it has also given us crucial insight into language change. No matter how well a minority or heritage language is maintained, it is virtually impossible for it to not undergo change, and much of that change will be contact-induced (though how much depends a bit on how the term 'contact-induced' is defined). Many speakers dislike such changes because they often get interpreted as indicating that the speakers do not take care of their language well. This kind of purism stems from the same ideological wellspring discussed above about keeping languages separate: the language should be kept intact and that also means it should be kept free of influence from the other language.

Every time a speaker practices simple insertional code-switching, e.g. the use of a Dutch noun or verb in a Turkish utterance, there is potential for lexical change in Turkish. Even if we accept that one use of a foreign word doesn't make that word an established loanword in the host language, a topic we will get back to, it is clear that, in general, the Turkish as spoken in the Netherlands is not the same as the Turkish spoken in Turkey, as the speakers in the Netherlands use a lot of Dutch words while speaking Turkish. They may avoid these words, for example, when in Turkey, just like

we can avoid informal slang when speaking our language in a formal setting, but if they routinely use these words when conversing informally in the vernacular, there is no reason to not conclude that Turkish in the Netherlands has undergone lexical change during its several decades of being a minority language: its vocabulary contains words that Turkish as spoken in Turkey does not have.

The big question is not so much *whether* languages change under contact conditions, but *what* changes and what does not, and how we can explain these patterns. The various empirical studies on Immigrant Turkish have shown, so far, that there seems to be widespread use of code-switching in informal in-group speech, of increasing complexity. The example below illustrates this complexity, with alternating chunks in Turkish and Dutch.

(5) *Ja zo 'n klein stuk-je bos var-mış ora-ya da ehm AVM mi ne kur-mak.*

Yes so a little piece-DIM forest there.is-EVID there-DAT too mall or what set-INF

isti-yo-muş ehm dinges o-ndan sonra hoe heet hij.. President. Ona da

want-PROG-EVID.3SG ehm thingie that-ABL after how name-3SG he president he-DAT too

şey yap-ma-mış-lar, izin ver-me-miş-ler. Maar da was zo 'n klein ehm

thing do-NEG-EVID-3PL permission give-NEG-EVID-3PL but that was so e little ehm

rotonde gibi bir şey-miş.

roundabout like something-EVID.3SG

'Yes there was such a little piece of forest ehm he wanted to build a mall hm or something and then what's his name President. They didn't do this thing to him, didn't allow him. But it was such a small ehm like a roundabout'

Turning to grammar, various domains have now been the subject of case studies, and together they provide conflicting evidence. There is quite some evidence for contact-induced grammatical change in some domains, notably subordinate clause construction, but little evidence for change in other domains, notably those that can be considered 'core syntax', such as subject expression, word order, or case marking.

We begin our survey with subordinate clauses. Turkish has two ways of forming a subordinate clause. The most unmarked form is the non-finite clause positioned before the matrix verb. This is the type of subordination often described for Turkish in typology textbooks, and not untypical for SOV languages, though somewhat exotic from the perspective of Indo-European languages. The other type resembles the finite subordinate clauses of, for example, English, and such clauses are usually positioned after the matrix verb. They typically occur in particular contexts, especially if the matrix verb is a verb of saying or cognition, and the subordinate clause is, therefore, a complement clause. Dutch normally forms its subordinate clauses in the same way: finite and positioned after the matrix verb.

(6) Non-finite subordination

[Ev-e gid-ince] sen-i ara-yacağ-ım.
house-DAT go-when you-ACC call-FUT-1sg
“I will call you when I go home.”

(7) Finite subordination:

Selin san-ıyor [sen dün sinema-ya git-ti-n].
Selin she.believes you yesterday cinema-DAT go-Past-2sg
“Selin believes that you went to the cinema yesterday.”

In Onar Valk (2015) it was established, through three different research methods, that the Dutch Turkish speakers made much more use of finite post-posed subordination than speakers in Turkey did. The same finding was obtained through the analysis of connected speech in interviews, through acceptability judgments, and through participants’ output in an elicited imitation task. In the latter, they had to retell a short fragment of speech, comprising three or four sentences and usually including a subordinate clause of any of the two types. Quite often, speakers from the Netherlands would use a finite subordinate clause in the retelling of a fragment that contained a non-finite clause. Turkish speakers in Turkey, who acted as a control group, rarely did this and sometimes showed the opposite pattern, replacing finite clauses with non-finite ones. The conclusion was that Heritage Turkish in the Netherlands has undergone a contact-induced syntactic change in the way it forms its subordinate clauses.

In stark contrast, various studies that looked for possible Dutch influence on Turkish in the domain of subject expression came up short. In Turkish, null subjects are the norm, as the person inflection on the verb identifies the person and number of the subject. Overt pronouns are used in special pragmatic conditions, primarily to indicate emphasis or a shift in topic. No matter whether the research was done through the analysis of a corpus of interviews (Doğruöz & Backus, 2009), narrative retellings (Azar et al., 2019), or a priming experiment (Sodacı et al. forthc), in all cases, the Turkish participants expressed subjects in more or less the same way as people in Turkey. There was no evidence of increased use of subject pronouns. The expression of the grammatical subject has proven robust in the studies we have done so far. Dutch uses subject pronouns categorically, but this marked difference between the languages clearly doesn’t automatically lead to grammatical influence on Turkish. This is not to say that this kind of influence doesn’t happen at all: we did find occasional examples of the unconventional use of subject pronouns, and the participants in these studies were all fairly highly educated. It is not impossible that these speakers are relatively close to the norms of Standard Turkish, also because they have a strong desire to speak ‘good Turkish’.

These Heritage Turkish studies are just a few examples of a large field of study that has documented contact-induced lexical and grammatical change, or the absence of it, around the world. Along with

studies documenting code-switching, it has made clear that there are many possible manifestations of bilingual speech, in the most general sense of that word. In my work, I have been involved in similar work more recently on the German of Germans living in the Netherlands, who for instance show increased use of Dutch-like word orders (placing a Prepositional Phrase adjunct more often at the end of a sentence, after the verb, rather than in the middle, thereby favoring an option already available in German but more similar to the dominant option in Dutch), Dutch-like Light Verb combinations, and Dutch-like ways of forming purposive subordinate clauses (BARKING; BACKUS; MOS FORTHC). General overviews (e.g. HEINE; KUTEVA, 2005; MATRAS, 2009, TORRES-CACOULLOS; TRAVIS, 2018) convey the general picture that many things change and many other things don't change in contact settings, and that how to define change, how to measure it, and how to account for it are far from settled issues.

4. Theoretical accounts of language change

The definition of change can be approached in various ways. Traditionally, linguistics implicitly sees languages as going through stages. A later stage will have constructions that an earlier stage didn't have, or conversely, will have lost a construction, and in this case, we can clearly talk about language change. Sociolinguistics has filled in some of the gaps in this historical linguistic picture, essentially focusing on the mechanisms that bring about and drive change. This is a natural consequence of sociolinguistics' focus on language variation, which can be seen as a snapshot of a system that might be undergoing continuous change. Sociolinguists have generally been cautious in interpreting variation as a sign of change, however, since variation may be more or less constant, e.g. because it indexes regional, social or generational variation. Some of this caution has also been incorporated into the tradition of contact linguistics, in which there is often resistance against assuming foreign influence too easily: if a construction was already possible in a language before contact, then continued or even increased usage of that construction during or after contact is not necessarily a case of language change, since it was already there. It would first have to be demonstrated, at the very least, that the way in which the construction is used has changed and that these changes can plausibly be interpreted as a direct influence from the other language. This would be the case, for example, if the lexical, syntactic, or pragmatic characteristics of a construction have changed, promoting uses that were previously unattested. The theoretical status of a mere change in frequency is debated: in most formal approaches to language frequency is an extralinguistic phenomenon that is not relevant for a theory of language change, while in a usage-based perspective, a change in the frequency with which a construction is used is the classic example of language change because the frequency is a causal factor in establishing a degree of entrenchment.

From a usage-based perspective, change is defined in terms that are strikingly different from the classic approach (BACKUS, 2014). Language is defined as an inventory of units. At the level

of 'language', this is a mere abstraction, accumulating over many speakers who share the same language. However, it is crucial to see that all individuals have their own inventory of units, and this inventory is dynamic. In a usage-based characterization of mental representation, learning never stops. The units that make up your inventory are not simply there or not there: if you 'have' a unit it is entrenched to a certain degree, and the degree to which it is entrenched differs from speaker to speaker, depending on that person's previous language experience. Degrees of entrenchment are important because, in turn, they affect the ease of activation, and the ease of activation affects how we talk, as it co-determines the units we select. Everyone's inventory is therefore affected by usage, i.e. what you actively say and what you passively hear, and because everybody's inventory is dynamic, change must be seen as a design feature of every individual speaker's competence and therefore of language.

However, while everybody's inventory must be different, there must also be considerable constraints on all this variation since communication normally proceeds pretty smoothly. We clearly do not differ so much from each other to the point that we cannot understand each other. Presumably because of the human need for group formation and for marking group membership, individual speakers who share membership of the same community will have fairly similar inventories, which, furthermore, change in similar ways. The mechanisms for this may be unconscious imitation (i.e. psycholinguistic alignment; PICKERING; GARROD, 2004) or conscious accommodation to other speakers in your group.

Specifying this general framework for the scenario of lexical borrowing, we can construct the following time course (BACKUS, 2014). It starts when a Turkish speaker in the Netherlands uses a particular Dutch word while speaking Turkish. This is the prototypical insertional code-switch: the foreign word is inserted into a host language utterance. Other people may also start using the same word independently; in their speech, too, it represents code-switching in this prototypical sense. People may also start adopting it from one another: you hear other Turkish speakers use the word, and you adopt this behavior of others. One could say that, in that case, it's not so much a code-switch anymore, as you're not adopting it from the other language but from another speaker of the same language: a classic case of sociolinguistic diffusion. At the same time, the speaker in question presumably knows very well that it's a Dutch word of Dutch origin and may well intend to use it because of connotations that come with being a Dutch word used while speaking Turkish: it may convey the social meaning of being in touch with life in the Netherlands. In this case, it would also be a typical code-switch. If the word continues to be used in Turkish utterances, it becomes a regular part of the Turkish lexicon of at least those people who use it, albeit one about which all speakers know that is of Dutch origin. By that time, it is an established loanword. At the same time, precisely because everybody knows it's a Dutch word, it might still be used with the pragmatic intentions that codeswitching often fulfills, such as an appeal to modernity or power. Presumably,

the more established the word becomes, and the more it is simply seen as the only or the best word for the concept it names, the less potential it has for this pragmatic use. It simply gets activated whenever the concept needs to be verbalized. Finally, if many speakers in the community routinely use the word, it would be logical to refer to it as an established Turkish word, and, because it is of Dutch origin, a loanword. It has been borrowed.

In this scenario, it would be impossible to say when the loanword status was reached unless we impose an arbitrary definition (for instance, a number of uses per person per x number of utterances, or usage by a certain percentage of the population, measures which would not be nonsensical, but virtually impossible to implement methodologically). What would make better sense, I think, is to accept some indeterminacy and reserve the term 'code-switching' for the synchronic selection of an item of foreign origin, and 'borrowing' as the name for the diachronic effect of those acts of synchronic selection.

For grammatical transfer or interference, the same scenario holds. A speaker may use a Dutch-like structure while speaking Turkish for the first time, such as a Dutch-like subordinate clause, and this would be a case of interference, and the same may happen to many other speakers, as all of them also know Dutch. Speakers will also be exposed to other speakers using the structural feature, and adopt it from them. Contrary to the use of a Dutch word, such structural interference is unlikely to be a matter of intention, which is why the connotation with intentionality and social indexicality that is always a possible part of the act of code-switching is largely absent from discussions of interference (as is also implied in the term 'interference' itself: it happens to you). As the individual speaker uses the structure over and over again, it becomes a conventional aspect of that speaker's Turkish more and more (a structural borrowing at the individual level), and as the same happens to other speakers, it can be said with more and more justification that Turkish has borrowed the structure. Once again, we cannot say exactly when the borrowing was established unless an arbitrary operationalization is imposed. 'Interference' (or 'transfer') is a name for the synchronic selection of a structure reinforced by a foreign equivalent; 'borrowing' is a name for the diachronic effect of those acts of synchronic selection.

5. Implications for theory

The similarities summarized above suggest that a more unified account of contact phenomena should be possible, rather than have the relatively separated subfields of code-switching studies on the one hand and accounts of contact-induced change on the other. Given that the empirical phenomena tend to occur together, in the same speakers and the same instances of language use, perhaps that would not be a bad idea to begin with. However, it would have implications regarding at least three topics.

First, the boundary between words and grammar may not be as hard as is generally assumed in linguistic theories. If lexical and structural borrowing co-occur and, more importantly, follow the same mechanisms of increasing entrenchment in speakers and conventionalization in their communities, it might be useful to investigate what they share rather than only how they differ. Usage-based linguistics presents all linguistic units as form-meaning units, and these can be of any size and complexity, and they vary in the degree to which they are lexically fixed or schematic. Typical lexical items are simple and specific; typical grammatical patterns are complex and schematic. A second issue is how we should see the relationship between synchrony and diachrony. If contact-induced change is the result of bilingual language use, that should not, in principle, be any different from what happens with all language use. Recall that in the description of the borrowing process above, it is not always clear whether an individual speaker takes a new unit or a new usage from speakers of the foreign language or from fellow speakers of the borrowing language. Finally, we should consider the impact of communication on linguistic competence. The locus of change is the individual mind, but the trigger for it is located in the communicative events the individual engages in.

Following usage-based logic, it would be useful to apply these thoughts to the general consensus about contact-induced change as it has emerged from decades of research on bilingualism from linguistic, sociolinguistic, and psycholinguistic perspectives. Generally, linguistic research has uncovered that some things are more borrowable than others and that certain principles exist that govern the process. Sociolinguistic studies have established that bilinguals often take things from the more powerful language because borrowing instantiates an act of identity: using something from the other language indexes an association with the culture that produces that language. Psycholinguistic studies have established that, cognitively, some things are easier to activate than others, or harder to inhibit, and that some things are easier to learn than others, or harder to forget. As a result, once frequent enough, foreign structures and words get activated almost automatically. The larger claim behind usage-based linguistics is that everything we call 'linguistic' might reduce to the combination of sociolinguistic and psycholinguistic perspectives. As far as borrowing is concerned, that should mean that something is borrowable if it is socially attractive or cognitively easy to activate.

The combination of cognitive and social considerations can perhaps help understand the fundamental question of why, in the same language pair and within the output of the same speakers, we often get substantial changes in one grammatical domain and not in another. We have seen that in Turkish-Dutch contact, for example, we see much more evidence of contact-induced change in subordination than in pro-drop or case marking. A usage-based logic would suggest that core syntactic domains host structures that are used all the time: thousands of utterances every day will instantiate them. As a result, they will be deeply entrenched for all speakers. This would hold, for example, for null pronouns in pragmatically neutral utterances in Turkish. Less frequently used

syntactic domains, on the other hand, host structures that are less well-entrenched. For Turkish speakers in the Netherlands, this may well hold for subordinate clauses, as these may be more typical of formal or academic discourse than for the everyday vernacular. Formal Turkish does not play a large role in the speech of most Turkish speakers in the Netherlands: most of the domains in which they use more formal registers will be experienced in Dutch. This would suggest that null pronouns are easily activated by these speakers, but non-finite subordinate clause structure is not. High frequency protects against variation and thus against change; low frequency makes a structure vulnerable. All of this happens more or less automatically, as grammatical choices are much less accessible for conscious attention than, for example, lexical items. Grammatical change is mostly a cognitive issue, though, under particular circumstances, attention can be focused on individual grammatical constructions. In that case, speakers may be able to consciously arrest change, or promote it, for that matter.

More in general, a usage-based perspective often requires answers to basic questions that refer to both cognition and sociality. If change is a design feature of the language, and language indeed changes all the time, then obviously we need to know why that is, and plenty of suggestions have been made throughout the modern era of linguistic theorizing, starting with 19th-century historical linguistics. As for contact-induced changes, again we see several suggested explanations, which are sometimes purely linguistic (e.g. an imbalance in the system), purely cognitive (e.g. reduction of cognitive load), and sometimes purely sociolinguistic (e.g. the power of a foreign language). A usage-based view commits us to try and attempt a unification of these different explanations since all of them make sense, and all of them have been empirically demonstrated. If we look at these explanations a bit closer, we start seeing similarities and points of contact between them. An imbalance in the system refers to a situation in which there are competing ways of saying the same thing. If reduction of cognitive load is driving a change, it is because, as a speaker, you want to keep things simple, and therefore you use what is familiar: you make use of the easy processing that high degrees of entrenchment make possible (you aim for low ‘prediction error’; CLARK, 2013). Finally, when foreign influence drives change, speakers may want to use what is semantically or socially useful and what others use. Obviously, it is valuable to study these factors in and by themselves, but they are not completely independent. A task that is especially for usage-based analysis is to see whether the linguistic explanations reduce to psycholinguistic ones, sociolinguistic ones, or a combination of the two.

6. Conclusion

What we can conclude from the above is that our studies on bilingual acquisition and socialization suggest that language borders are largely socially constructed. Young bilingual children mix their languages and stop doing so once they learn that it is appreciated by those around them if they keep

their languages separate. They continue doing so if code-switching is unmarked in their sociolinguistic environment. This community practice of mixing languages may itself be the inevitable result of the social and cognitive factors that co-determine language use. Lexical and grammatical changes are a hallmark of languages caught up in contact settings. Lexical changes are mostly explained by social factors and grammatical changes by cognitive factors, but this opposition is not a cut and dried one.

Together, this provides the outlines of a usage-based framework for contact effects. Cognition and sociality form the bedrock, as they do for all language use. Sociality ensures communication, which provides input for cognition to work with, and so on, in an endless feedback loop. This way of model building has at least two implications. One is that the need for linguistics, which has limited its object of study largely to verbal units in communication, to attend more fundamentally to the social and cognitive aspects of human behavior requires further discussion about what the discipline should be about, what its object of study is. The second implication is that the degree to which linguistics uncritically adopts the assumption, or ideology, that particular communities have particular languages blinds us a bit to what the variation we see in bilingual language use is telling us.

Language separation, I have argued, is the logical result of separated input and language mixing of mixed input. Co-occurrence of two or more units leads to the entrenchment of the combination into a larger unit, and this breeds further use of the same unit. These are cognitive mechanisms. However, separated input and separated use of two languages is also the result of the use of language as a marker of group membership, a social factor. The question then is why does language work like this. Perhaps it has to do with the sheer frequency of language use: we use it all day long, giving ample chance for cognition to do its work, and build many units that can be reused with high degrees of automaticity and low degrees of conscious attention. And perhaps it also has to do with its versatility: we use language as a cultural tool to convey unlimited shades of meaning, from semantics to social indexicality. Sociality is what drives why we use language in the first place.

In short, I have argued that a usage-based lens imposes on the analysis of language that both cognitive principles and social goals are taken into account together. Cognitive principles are important because communication is routine-driven behavior: degree of entrenchment determines the ease of activation, and ease of activation co-determines which units get selected in speech. Social goals are important because communication is also goal-oriented behavior: conscious selection determines what the best option is for conveying the intended meaning. Psycholinguistic and sociolinguistic explanations are therefore both needed, though perhaps not always at the same time. It is possible, and from a usage-based perspective even expected, that further articulation of such analyses will make a separate linguistic level of explanation unnecessary, as linguistic explanations will always reduce to a combination of cognitive and social factors.

References

- AALBERSE, S.; BACKUS, A.; MUYSKEN, P. *Heritage Languages. A Language Contact Approach*. Studies in Bilingualism 58. Amsterdam/Philadelphia: John Benjamins Publishing Company, 2020.
- AZAR, Z.; ÖZYÜREK, A.; BACKUS, A. Turkish-Dutch bilinguals maintain language-specific reference tracking strategies in elicited narratives. *International Journal of Bilingualism*. Advance online Publication. Published online 5 April 2019. DOI: <https://doi.org/10.1177/1367006919838375>
- BACKUS, A. Turkish as an immigrant language in Europe. In: BHATIA, T. K.; WILLIAM, C. R. (ed.). *The Handbook of Bilingualism and Multilingualism*. 2nd edition. Blackwell Handbooks in Linguistics, 2013. p. 770-790.
- BACKUS, A. Towards a usage-based account of language change: implications of contact linguistics for linguistic theory. In: NICOLAÏ, R. (ed.). *Questioning language contact. Limits of contact, contact at its limits*, 91-118. Leiden/Boston: Brill, 2014.
- BARKING, M.; BACKUS, A.; MOS, M. (forthc.). Comparing Forward and Reverse Transfer from Dutch to German. *International Journal of Bilingualism*.
- CHRISTIANSEN, M.; CHATER, N. The Now-or-Never Bottleneck: A Fundamental Constraint on Language. *Behavioral and Brain Studies*, v. 39, e62, 2016. DOI: <https://doi.org/10.1017/S0140525X1500031X,e62>
- CLARK, A. Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*, v. 36, p. 181-253, 2013. DOI: <https://doi.org/10.1017/S0140525X12000477>
- DEMIRÇAY, D. *Connected Languages: Effects of intensifying contact between Turkish and Dutch*. Utrecht: LOT, 2017.
- DOĞRUÖZ, S.; BACKUS, A. Innovative constructions in Dutch Turkish: An assessment of on-going contact-induced change. *Bilingualism: Language & Cognition*, v. 12, n. 1, p. 41-63, 2009.
- GASKINS, D.; BACKUS, A.; ENDESFELDER QUICK, A. Slot and frame schemas in the language of a Polish and English speaking child: the impact of usage patterns on the switch placement. *Languages*, 2019, 4, 8. DOI: <https://doi.org/10.3390/languages4010008>

- HAKIMOV, N.; BACKUS, A. (forthc.). Usage-Based Contact Linguistics: Effects of Frequency and Similarity in Language Contact. *Journal of Language Contact*, v. 13, p. 459-481, 2021.
- HEINE, B.; KUTEVA, T. *Language Contact and Grammatical Change* (Cambridge Approaches to Language Contact). Cambridge: Cambridge University Press, 2005.
- MATRAS, Y. *Language contact*. Cambridge: Cambridge University Press, 2009.
- MUYSKEN, P. *Bilingual Speech. A Typology of Code-Mixing*. Cambridge: Cambridge University Press, 2000.
- MYERS-SCOTTON, C. *Social Motivations for Codeswitching: Evidence from Africa*. Oxford: Oxford University Press, 1993.
- ONAR VALK, P. *Transformation in Dutch Turkish Subordination? Converging Evidence of Change Regarding Finiteness and Word Order in Complex Clauses*. PhD. Tilburg University, 2015.
- PICKERING, M. J.; GARROD, S. Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences*, 27, 169-190; discussion 190–226, 2004.
- QUICK, A. E.; LIEVEN, E.; BACKUS, A.; TOMASELLO, M. Constructively combining languages: The use of code-mixing in German-English bilingual child language acquisition. *Linguistic Approaches to Bilingualism*, v. 8, n. 3, p. 393-409, 2018. DOI: <https://doi.org/10.1075/lab.17008.qui>
- SODACI, H.; BACKUS, A.; KOOTSTRA, G. J. (forthc). Role of structural priming in contact-induced change: Subject pronoun expression in L1 Turkish by Turkish-Dutch bilinguals. *Journal of Cultural Cognitive Science*. Pre-print. DOI: <https://doi.org/10.31234/osf.io/2uxej>
- TORRES CACOULLOS, R.; TRAVIS, C. *Bilingualism in the Community. Code-Switching and Grammars in Contact*. Cambridge: Cambridge University Press, 2018.
- TREFFERS-DALLER, J. *Mixing two languages: French-Dutch contact in a comparative perspective*. Berlin: De Gruyter, 1994.
- ZENNER, E.; BACKUS, A.; WINTER-FROEMEL, E. (ed.). *Cognitive Contact Linguistics. Placing usage, meaning and mind at the core of contact-induced variation and change*. *Cognitive Linguistics Research*, Volume 62. Berlin: De Gruyter Mouton. I-VI + 1-342, 2019.

Implicações de um enfoque baseado no uso para a linguística de contato

Ad Backus

I. Introdução

O multilinguismo é difundido no mundo e manifesta-se de muitos modos. Em função dessa diversidade, também há muitas maneiras pelas quais o multilinguismo pode afetar a vida das pessoas. Considerando essa abrangência, é um tanto surpreendente que a linguística, a disciplina que estuda a linguagem, não tenha reservado um lugar central para o fenômeno entre as coisas que precisam ser explicadas sobre a linguagem. Dentro das subdisciplinas da sociolinguística e da psicolinguística, o multilinguismo tornou-se um tópico primário de investigação, mas isso só teve um impacto limitado na linguística em geral, em parte porque essas subdisciplinas estão muitas vezes afastadas da ‘linguística teórica’. Embora por muitas décadas tenha havido uma divisão viável, mesmo que lamentável, em três subdisciplinas relativamente independentes, uma abordagem baseada no uso da teoria linguística exige uma integração muito mais séria delas. Este artigo pretende argumentar sobre esse ponto.

Empiricamente, vou me basear no trabalho de contato linguístico, principalmente entre turco e holandês na Holanda, no qual estive envolvido como investigador principal, membro da equipe ou supervisor nas últimas décadas. A maior parte deste trabalho adota uma perspectiva baseada no uso dos efeitos de contato. Em geral, o trabalho sobre o contato turco-holandês mostrou que cinquenta anos após o início da migração, o turco é bem mantido na comunidade, talvez mais do que se poderia esperar, considerando o destino de muitas línguas imigrantes enraizadas na migração laboral (BACKUS, 2013). Por outro lado, também há sinais de que algum grau de mudança linguística está em andamento e, com certeza, o próprio turco está passando por mudanças lexicais e gramaticais em seu contexto de imigração. No entanto, a maioria das informações é baseada em estudos de pequena escala e é impossível saber se o que foi atestado nesses estudos pode ser generalizado para a comunidade em geral. Seja como for, os estudos da herança turca invariavelmente demonstram a existência abundante de *code-switching* com o holandês, e um pouco de influência holandesa em sua gramática.

Vou me concentrar especificamente em dois tipos de estudos aqui. O primeiro é o trabalho que investigou o grau em que os falantes bilíngues mantêm suas línguas separadas ou não (estudos sobre *codeswitching*, sobre política linguística familiar ou socialização e sobre aquisição bilíngue). Em segundo lugar, vou me referir às descobertas de estudos sobre o grau de mudança gramatical que podem ser encontrados no Heritage Turkish [Turco de Herança] (variando desde pouca evidência de mudança em alguns domínios gramaticais até muitas mudanças em outros). Também vou me basear em outros trabalhos no paradigma emergente da linguística de contato baseada no uso (AALBERSE *et al.*, 2020; ZENNER *et al.*, 2019; BACKUS; HAKIMOV a aparecer).

A Seção 1 abaixo revisará algumas descobertas relevantes sobre mistura e separação de línguas, enquanto a Seção 2 resumirá algumas de nossas descobertas sobre mudanças lexicais e gramaticais

induzidas pelo contato. A Seção 3 discutirá essas descobertas à luz dos *insights* linguísticos, psicolinguísticos e sociolinguísticos sobre o bilinguismo, enquanto a Seção 4 conclui o artigo, com um apelo para que se esforce mais para integrar as dimensões da cognição e da sociabilidade em uma linguística mais inclusiva.

2. Separação de línguas

Embora separar nossas línguas pareça natural para a maioria de nós, existem exemplos de fala bilíngue que questionam essa naturalidade. Muitos estudos de aquisição bilíngue confirmaram, por exemplo, que crianças bilíngues misturam suas línguas, mesmo que cresçam em famílias ou comunidades que desaprovam a troca de código (*code-switching*). Implícita ou explicitamente, essas famílias podem inculcar a norma de que, embora aprender a falar duas ou mais línguas seja uma habilidade boa ou necessária, essas línguas devem ser mantidas estritamente separadas no discurso. Os pais muitas vezes lutam para descobrir o que é melhor para seus filhos e temem as consequências negativas de não falar bem nenhum dos idiomas. O estudo do contato linguístico ou Línguas de Herança, com seu foco na ‘mudança induzida pelo contato’, ‘aquisição imperfeita’ e ‘atrito’, muitas vezes mostra que pelo menos uma das línguas, muitas vezes aquela associada ao grupo étnico, sofre mudanças, em parte porque os pais não poderiam fornecer a seus filhos o mesmo *input* linguístico que as crianças monolíngues recebem no país de origem. As evidências disponíveis tendem a mostrar que o *input* é importante: as crianças produzem o que aprenderam, e o que aprenderam depende do *input* que recebem. Por outro lado, isso não deve ser interpretado como se eles apenas produzissem fragmentos de linguagem aprendidos por imitação. De fato, estudos têm observado com frequência que as crianças trocam de código mesmo que o *input* não lhes forneça exemplos de troca de código além da alternância produzida naturalmente quando dois pais falam uma língua diferente. Essa troca de código, pode-se argumentar, mostra que as crianças não sabem necessariamente que existem duas línguas e, se sabem, nem sempre sabem a qual língua uma palavra pertence. Também mostra que a troca de códigos fornece evidências para a produtividade: mostra quais construções a criança pode usar para formar novos enunciados de forma produtiva.

A ideologia da separação linguística é quase universal, assim como uma ideologia de proficiência total. Quando os esforços de socialização bilíngue muitas vezes subscrevem uma ideologia em que o domínio perfeito da linguagem de preferência de ambas as línguas é o objetivo, o resultado pode frustrar os pais e outros membros da comunidade: produzir proficiência inevitavelmente acaba não sendo fácil. As crianças simplesmente aprendem a falar deixando o *input* guiá-las e explorando inconscientemente a resposta de sua mente a ele, e o resultado provavelmente será contrário ao de uma proficiência total de um contexto monolíngue.

Para ver como a noção de separação linguística está enraizada em todos nós, observe como virtualmente todos os conceitos em linguística assumem os limites linguísticos: ‘linguagem’, ‘troca de código’, ‘alternância linguística’, ‘multilinguismo’, ‘proficiência’, ‘padronização’, ‘L1’ e ‘L2’. Mesmo o termo ‘translinguagem’, que foi cunhado para combater a suposição oculta de que sempre sabemos o que é uma língua, não pode escapar dela. Se mesmo os linguistas não podem escapar disso, não é de se admirar que os usuários comuns da linguagem também não possam. Durante toda a nossa vida continuamos reforçando essa ideologia, geralmente sem querer. É por isso que podemos nos perguntar quantos idiomas falamos e por que podemos participar de cursos de aprendizado de idiomas. No entanto, a literatura sobre bilinguismo também contém alguns dados que refletem desvios dessa ideologia, especialmente a intensa troca de códigos e as línguas mistas. Além disso, o fato observado acima e a ser elaborado a seguir, de que crianças pequenas em famílias de “uma língua para o pai e outra língua para a mãe” praticam a troca de código, oferece um desafio interessante. Aliás, talvez o mesmo se aplique à variação estilística comum dentro de uma língua.

Em vários estudos recentes (por exemplo, QUICK *et al.*, 2018; GASKINS *et al.*, 2019), estudamos a mistura linguística de crianças que cresceram nessas famílias ‘OPOL’, combinando alemão e inglês em três casos, e polonês e inglês noutro. Embora essas crianças misturassem suas línguas dentro de enunciados enquanto não eram expostas a esse tipo de mistura em seus *inputs*, elas mostraram outros efeitos familiares da aprendizagem conservadora orientada por *inputs*: sua língua dominante era a língua que ouviam com mais frequência no *input* (incluindo uma mudança de alemão para inglês para uma criança após uma estadia prolongada na casa de seus avós na Irlanda), e também foi a língua que funcionou como a língua fonte mais frequentemente em enunciados bilíngues. O exemplo abaixo foi, portanto, típico de uma criança dominante alemã.

(1) *ich will jetzt the boat machen*
Eu quero agora o *barco* fazer
“Eu quero construir o *barco* agora”

Uma coisa que é presumivelmente essencial aqui é que essas crianças eram muito jovens para ter muita consciência sociolinguística sobre a linguagem. Se considerarmos a estrutura de seu *codeswitching*, é interessante ver que a assimetria com a qual estamos familiarizados de muitos *codeswitching* adultos é apenas parcialmente representada. Às vezes vemos de fato a inserção de palavras de conteúdo da língua mais fraca em construções da língua dominante. Outras vezes, porém, vemos palavras funcionais da língua mais fraca funcionando como base para construções que hospedam palavras de conteúdo da língua dominante. Uma criança dominante em alemão em Quick *et al.* (2018), por exemplo, produziu muitos enunciados com troca de código com substantivos alemães inseridos na frase em inglês ‘I want X’, no meio do discurso em alemão.

O tipo mais óbvio de uso de linguagem que vai contra a norma de que as linguagens devem ser mantidas separadas é a troca de código (*codeswitching*). Demirçay (2017) discute muitos exemplos de padrões de troca muito intensos, nos quais muitas vezes não está claro se um falante está usando holandês ou turco. Os dados foram de conversas entre jovens adultos de segunda e terceira geração, que sempre se conheceram bem, conversando informalmente entre si. De um modo geral, as gravações mostram constante alternância entre as duas línguas, e a comutação de códigos, embora às vezes dos tipos familiares de inserção ou alternância (MUYSKEN, 2000), muitas vezes eram difíceis de classificar. O autor os rotulou, um tanto informalmente, de ‘comutação de código complexa baseada em fragmentos’.

De um modo geral, os dados dão a aparência de um estilo de fala em que os idiomas são constantemente alternados. Os trechos monolíngues raramente são mais longos do que dois ou três enunciados. O fragmento a seguir exemplifica isso; é parte de uma conversa entre os falantes E e Ş, que estavam lembrando uma das experiências recentes de E com um amigo em comum. Há intensa alternância entre partes que estão em turco e partes que estão em holandês (aqui indicadas em itálico).

(2) Turkish-Dutch bilingual conversation

E: Öyle işte. *Ja sowieso. Zij zo ja* canım acayip sarma çekti falan hatta eve bırakıyodum. *Ik zo* söyleseydin, aşağıda vardı. *Zij zo* valla de.

“Então é isto. *Sim de qualquer modo. Ela disse como sim* eu estou a fim de folhas de parreira recheadas e tal, eu estava deixando ela em casa. *Eu disse assim* se você tivesse me dito, nós tínhamos umas lá embaixo. *Ela disse sério.*”

Ş: *Hadden jullie het gemaakt?*

“Vocês fizeram?”

E: *Wij hadden sowieso* aşağıda vardı tencerede vardı ya.

“Nós tínhamos de qualquer modo tinham algumas na panela lá embaixo”

Ş: *Ha iyi* yapmıştın.

“ah bom, vocês fizeram”

E: *Zij zo ja* söyleseydin şapardım enzo.

“*sela disse como* se você tivesse me dito eu teria comido e *assim por diante*”

Ş: *O da mı* biliyo sarmayı neyi?

“ela também sabe sobre folhas de parreira recheadas e tal?”

E: *Ja man. Zij zo* kesin bana öğret dit dat. *Ik zo* öğretiriz.

“*sim cara. Ela disse assim* você tem que me ensinar isso e *aquilo. Eu disse assim* eu vou te ensinar.”

Quando as pessoas misturam línguas assim, claramente não estão mantendo suas línguas separadas. Ao mesmo tempo, pode-se dizer que partes individuais ainda estão em uma língua

específica. Os dados contêm enunciados monolíngues, casos clássicos de *codeswitching* insercional em que palavras holandesas são inseridas em enunciados turcos (e vice-versa) e alternância entre enunciados holandês e turco (veja os exemplos abaixo), mas no geral o estilo de fala é marcado pela alternância constante entre fragmentos em turco e fragmentos em holandês, muitas vezes dentro do que poderia ser visto como um único enunciado.

(3) Natalya'da bıraktık herşeyi. *Ja zij zo gelip gidemem. Ik moet alleen nog die 'handhavingsmiddelen' afmaken, sonra da 'biorechten' de yapamadık o da kaldı.*

“Nós deixamos tudo na Natalya. Bem, ela estava como ‘Eu não posso ir.’ Eu tenho só que terminar aquele ‘meio de reforço’, e depois disso nós não poderemos mais fazer as ‘leis bio’, então continuam lá”

(4) *Wij doen salça in die iç.*

Nós fazemos.PRES.1pl extrato de tomate naquele recheio

“nós colocamos extrato de tomate no recheio”

A permeabilidade das fronteiras linguísticas pode ser vista a partir de uma perspectiva sociolinguística e psicolinguística. Sociolinguisticamente, pode-se dizer que a troca de códigos se destaca para nós, especialmente quando é tão intensa quanto nos dados holandês-turcos que acabamos de apresentar, porque a ideologia da separação está tão completamente enraizada em todos nós que achamos esse tipo de uso da linguagem digno de nota. A ideologia e a prática resultante da fala monolíngue endurecem as fronteiras, de modo que vemos o uso da língua monolíngue como o padrão, como o normal. Isso por si só consolida o fato de que uma determinada palavra ou expressão pertence a uma determinada língua como parte de seu significado (ou pelo menos como parte do conhecimento que temos sobre essa palavra), e dar importância ou saliência a esse significado social tem o efeito cognitivo de que continuamos reforçando os limites entre as línguas em nossas mentes. Constantemente nos lembramos da existência de diferentes línguas e quais componentes pertencem a qual língua. O que a troca de códigos mostra, no entanto, é que a pressão da comunicação rápida (que Christiansen e Chater (2016) se referem como o ‘gargalo agora ou nunca’) pode nos fazer dizer coisas às vezes que vão contra a separação, e é aí que a perspectiva psicolinguística entra.

O uso da linguagem é um constante vai e vem entre a produção automática e a seleção consciente. Sem confiar na automaticidade, não seríamos capazes de nos comunicar tão rápido e eficientemente como fazemos; sem confiar na seleção consciente, não seríamos capazes de escolher nossas palavras com cuidado quando precisamos. O que acontece na fala bilíngue é que muitas vezes pode ser ativado um elemento que vem de outra língua do que o último elemento proferido, e o resultado é a troca de código. O modelo de fala subjacente a tudo isso é que todo uso da linguagem é o resultado da divisão do trabalho entre facilidade de ativação e seleção ótima como os dois fatores que governam tudo o que dizemos. Em algumas comunicações bilíngues, isso leva ao que nós, como linguistas,

percebemos como uma intensa troca de códigos, mas que para os falantes é simplesmente o uso da linguagem de maneira normal: ativação de unidades facilmente ativadas. Na maioria dos outros casos, isso leva ao uso do idioma monolíngue, graças à convenção bem praticada de manter os idiomas separados.

A percepção baseada no uso aqui é que, se o *input* for, em grande parte, monolíngue, ou ‘bilíngue, mas separada’ como nas configurações de aquisição da OPOL ou comunidades bilíngues com ‘regras’ claras para quando usar qual idioma, então também todas as unidades multpalavras e construções que são armazenadas na memória são ‘monolíngues’; a troca de códigos será então rara e marcada. Em tais cenários, as pessoas muitas vezes parecem sinalizar suas trocas de código como ‘desculpas’ metalinguísticas (mostrada, por exemplo, pela análise de Treffers-Daller de 1994 sobre a troca entre francês e holandês em Bruxelas; a Bélgica é notória por sua ‘luta linguística’). Se não houver pressão normativa para separar, no entanto, você está livre para usar todos os recursos que adquiriu (e que você sabe que seu interlocutor também adquiriu), e a troca de código pode evoluir para a maneira ‘normal’ de falar, categorizada por Myers-Scotton (1993) como ‘*codeswitching* como a escolha não marcada’. Sob circunstâncias particulares, isso pode evoluir para um *lect* misto convencionalizado, ou mesmo uma língua mista.

Para o Turco de Herança, não sabemos o suficiente para generalizar, pois muitos tipos diferentes de uso da língua foram documentados. Talvez os estudos combinados mostrem uma abundância de *code-switching* em vez de uso de linguagem monolíngue, mas isso pode ser um problema de amostragem: a conversa espontânea tem sido investigada principalmente por linguistas interessados em *code-switching*, que, portanto, investigam dados que o mostram. Os mesmos participantes cuja troca de códigos também observamos passam grande parte do tempo em modo monolíngue, principalmente holandês (no trabalho ou escola, por exemplo), mas também turco, por exemplo, com familiares mais velhos, com familiares e amigos na Turquia ou com pessoas de quem eles sabem que desaprovam a mistura de línguas. Tudo o que podemos dizer é que a fala bilíngue foi amplamente documentada e que os falantes de fato afirmam em questionários de escolha de idiomas que costumam usar as duas línguas ao mesmo tempo em conversas informais em grupo (por exemplo, em Azar et al., 2019).

3. Mudança lexical e gramatical

Como linguistas de contato, geralmente ficamos felizes em suspender quaisquer dúvidas sobre a adequação de separar os idiomas e analisar o que acontece com os idiomas separadamente. Como resultado, sabemos muito sobre as mudanças lexicais e gramaticais sofridas pelas línguas minoritárias, muitas vezes apoiadas por dados que foram coletados de maneira que forçaram os falantes a produzir fala monolíngue, ou por análises de fala bilíngue em que todas as partes dos

dados que são no outro idioma são, por enquanto, ignorados (no extrato representado acima, por exemplo, pode-se ignorar todos os pedaços em holandês e apenas analisar a estrutura das partes que estão em turco).

Embora essa prática possa ignorar um aspecto essencial do uso da língua por falantes bilíngues, deixando sua análise para o domínio de pesquisa separado da ‘troca de código’, não há dúvida de que ela também nos deu uma visão crucial da mudança linguística. Não importa quão bem uma língua minoritária ou de herança seja mantida, é virtualmente impossível que ela não sofra mudanças, e grande parte dessa mudança será induzida pelo contato (embora o quanto dependa um pouco de como o termo ‘induzido pelo contato’ é definido). Muitos falantes não gostam dessas mudanças, porque muitas vezes são interpretadas como uma indicação de que os falantes não cuidam bem de sua língua. Esse tipo de purismo decorre da mesma fonte ideológica discutida acima sobre manter as línguas separadas: a língua deve ser mantida intacta e isso também significa que deve ser mantida livre da influência da outra língua.

Toda vez que um falante pratica a troca de código de inserção simples, por exemplo, o uso de um substantivo ou verbo holandês em um enunciado turco, há potencial para mudança lexical em turco. Mesmo que aceitemos que um uso de uma palavra estrangeira não faça dessa palavra um empréstimo estabelecido na língua de acolhimento, um tópico ao qual voltaremos, é claro que, em geral, o turco falado na Holanda não é o mesmo que o turco falado na Turquia, pois os falantes na Holanda usam muitas palavras holandesas enquanto falam turco. Eles podem evitar essas palavras, por exemplo, quando na Turquia, assim como podemos evitar gírias informais ao falar nossa língua em um ambiente formal, mas se eles usarem rotineiramente essas palavras ao conversar informalmente no vernáculo, não há razão para não concluir que o turco na Holanda sofreu mudanças lexicais durante suas várias décadas de língua minoritária: seu vocabulário contém palavras que o turco falado na Turquia não possui.

A grande questão não é tanto se as línguas mudam em condições de contato, mas o que muda e o que não muda e como podemos explicar esses padrões. Os vários estudos empíricos sobre os imigrantes turcos mostraram, até agora, que parece haver um uso generalizado de *codeswitching* na fala informal em grupo, de complexidade crescente. O exemplo abaixo ilustra essa complexidade, com partes alternadas em turco e holandês.

(5) *Ja zo ‘n klein stuk-je bos var-mış ora-ya da ehm AVM mi ne kur-mak.*

Sim então um pedacinho-DIM floresta existe-EVID lá-DAT muito pequeno ou o que fez-INF

isti-yo-muş ehm dinges o-ndan sonra hoe heet hij.. President. Ona da

quer-PROG-EVID.3SG ehm coisinha que-ABL depois como nome-3SG ele president ele-DAT também

şey yap-ma-miş-lar, izin ver-me-miş-ler. Maar da was zo 'n klein ehm
coisa faz-NEG-EVID-3PL permissão dar-NEG-EVID-3PL mas era isso e lpouco ehm
rotonde gibi bir şey-miş.
rotatória como algo-EVID.3SG

‘Sim havia esse pedaço de floresta ehm ele queria construir um *shopping* hm ou algo assim e então qual o nome dele *Presidente* e então qual é seu nome *Presidente*. Eles não fizeram essa coisa para ele, não o deixaram. Mas era tão pequeno ehm como uma rótula. *But it was such a small like a roundabout*’

Voltando à gramática, vários domínios já foram objeto de estudos de caso e, juntos, fornecem evidências conflitantes. Há bastante evidência de mudança gramatical induzida por contato em alguns domínios, notadamente construção de orações subordinadas, mas pouca evidência de mudança em outros domínios, notadamente aqueles que podem ser considerados ‘sintaxe central’, como expressão de sujeito, ordem de palavras ou marcação de caso.

Começamos nossa pesquisa com orações subordinadas. O turco tem duas maneiras de formar uma oração subordinada. A forma mais desmarcada é a cláusula não finita posicionada antes do verbo matriz. Este é o tipo de subordinação frequentemente descrito para o turco nos livros de tipologia, e não atípico para as línguas SOV, embora um tanto exótico da perspectiva das línguas indo-europeias. O outro tipo se assemelha às orações subordinadas finitas, por exemplo, do inglês, e essas orações geralmente são posicionadas após o verbo da matriz. Normalmente ocorrem em contextos particulares, especialmente se o verbo matriz for um verbo de dizer ou cognição, e a oração subordinada, portanto, for uma oração complementar. O holandês normalmente forma suas orações subordinadas da mesma maneira: finita e posicionada após o verbo matriz.

(6) Subordinação não finita

[Ev-e gid-ince] sen-i ara-yacağ-ım.
casa-DAT vai-quando você-ACC ligar-FUT-1sg
“Eu ligo para você quando eu for para casa.”

(7) Subordinação finita:

Selin san-ıyor [sen dün sinema-ya git-ti-n].
Selin ela.acredita você ontem cinema-DAT ir-Past-2sg
“Selin acredita que você foi no cinema ontem.”

Em Onar Valk (2015), foi estabelecido, por meio de três métodos de pesquisa diferentes, que os falantes de turco holandês fizeram muito mais uso da subordinação pós-posta finita do que os falantes da Turquia. O mesmo achado foi obtido por meio da análise do discurso conectado em

entrevistas, por meio de julgamentos de aceitabilidade e pela saída dos participantes em uma tarefa de imitação eliciada. Neste último, eles tiveram que recontar um pequeno fragmento de fala, composto de três ou quatro frases e geralmente incluindo uma oração subordinada de qualquer um dos dois tipos. Muitas vezes, os falantes da Holanda usavam uma oração subordinada finita na recontagem de um fragmento que continha uma oração não finita. Os falantes de turco na Turquia, que atuaram como grupo de controle, raramente faziam isso e, às vezes, mostravam o padrão oposto, substituindo orações finitas por não finitas. A conclusão foi que o Turco de Herança na Holanda sofreu uma mudança sintática induzida pelo contato no modo como forma suas orações subordinadas.

Contudo, vários estudos que buscaram uma possível influência holandesa sobre o turco no domínio da expressão do sujeito ficaram aquém. Em turco, os sujeitos nulos são a norma, pois a flexão da pessoa no verbo identifica a pessoa e o número do sujeito. Os pronomes abertos são usados em condições pragmáticas especiais, principalmente para indicar ênfase ou mudança de assunto. Não importa se a pesquisa foi feita através da análise de um corpus de entrevistas (DOĞRUÖZ; BACKUS, 2009), recontagens narrativas (AZAR et al., 2019) ou um experimento de *priming* (SODACI; BACKUS; KOOTSTRA, forc.), em todos os casos os participantes turcos expressaram assuntos em mais ou menos da mesma forma que as pessoas na Turquia. Não houve evidência de um aumento do uso de pronomes subjetivos. A expressão do sujeito gramatical mostrou-se robusta nos estudos que fizemos até agora. O holandês usa pronomes de sujeito categoricamente, mas essa diferença marcante entre os idiomas claramente não leva automaticamente à influência gramatical no turco. Isso não quer dizer que esse tipo de influência não aconteça: encontramos exemplos ocasionais do uso não convencional de pronomes subjetivos, e os participantes desses estudos eram todos bastante instruídos. Não é impossível que esses falantes estejam relativamente próximos das normas do turco padrão, também porque eles têm um forte desejo de falar ‘bom turco’.

Esses estudos do Turco de Herança são apenas alguns exemplos de um grande campo de estudo que documentou a mudança lexical e gramatical induzida pelo contato, ou a ausência dela, em todo o mundo. Juntamente com os estudos que documentam a troca de códigos, ficou claro que existem muitas manifestações possíveis da fala bilíngue, no sentido mais geral dessa palavra. Em meu próprio trabalho, estive envolvido em trabalhos semelhantes mais recentemente sobre o alemão de alemães que vivem na Holanda, que, por exemplo, mostram um uso crescente de ordens de palavras semelhantes ao holandês (colocando um adjunto de frase preposicional com mais frequência no final de uma frase, depois do verbo, em vez de no meio, favorecendo assim uma opção já disponível em alemão, mas mais semelhante à opção dominante em holandês), combinações de verbos leves semelhantes ao holandês e formas holandesas de formar orações subordinadas propositais (BARKING; BACKUS; MOS, forthc.). As visões gerais (por exemplo, HEINE; KUTEVA, 2005; MATRAS, 2009, TORRES-CACOULLOS; TRAVIS, 2018) transmitem a imagem geral de que

muitas coisas mudam e muitas outras não mudam nas configurações de contato, e como definir a mudança, como medi-la e como explicá-lo estão longe de serem questões resolvidas.

4. Relatos teóricos da mudança de linguagem

A definição de mudança pode ser abordada de várias maneiras. Tradicionalmente, a linguística vê implicitamente as línguas como passando por estágios. Um estágio posterior terá construções que um estágio anterior não tinha, ou inversamente terá perdido uma construção, e neste caso podemos falar claramente sobre mudança de linguagem. A sociolinguística preencheu algumas das lacunas desse quadro linguístico histórico, concentrando-se essencialmente nos mecanismos que provocam e impulsionam a mudança. Esta é uma consequência natural do foco da sociolinguística na variação linguística, que pode ser vista como uma observação instantânea de um sistema que pode estar passando por mudanças contínuas. Os sociolinguistas geralmente têm sido cautelosos ao interpretar a variação como um sinal de mudança, no entanto, uma vez que a variação pode ser mais ou menos constante, por exemplo, porque indexa a variação regional, social ou geracional. Parte dessa cautela também foi incorporada à tradição da linguística de contato, na qual muitas vezes há resistência em assumir influência estrangeira com muita facilidade: se uma construção já era possível em uma língua antes do contato, então o uso continuado ou mesmo aumentado dessa construção durante ou após o contato não é necessariamente um caso de mudança da língua, pois já estava lá. Deveria primeiro ser demonstrado, no mínimo, que a forma como a construção é usada mudou, e que essas mudanças podem ser interpretadas plausivelmente como influência direta da outra língua. Este seria o caso, por exemplo, se as características lexicais, sintáticas ou pragmáticas de uma construção se alterassem, promovendo usos antes não atestados. O *status* teórico de uma mera mudança na frequência é debatido: na maioria das abordagens formais, a frequência linguística é um fenômeno extralinguístico que não é relevante para uma teoria da mudança da linguagem, enquanto em uma perspectiva baseada no uso, uma mudança na frequência com que uma construção é usada é o exemplo clássico de mudança linguística, porque a frequência é um fator causal no estabelecimento do grau de enraizamento.

De uma perspectiva baseada no uso, a mudança é definida em termos que são notavelmente diferentes da abordagem clássica (BACKUS, 2014). A linguagem é definida como um inventário de unidades. No nível da 'linguagem', isso é uma mera abstração, acumulando-se sobre muitos falantes que compartilham a mesma língua. No entanto, é fundamental observar que todos os indivíduos possuem seu próprio estoque de unidades, e esse estoque é dinâmico. Em uma caracterização da representação mental baseada no uso, o aprendizado nunca para. As unidades que compõem seu inventário não estão simplesmente lá ou não estão lá: se você 'tem' uma unidade, ela está enraizada até certo ponto, e o grau em que ela está enraizada difere de falante para falante, dependendo do idioma anterior dessa pessoa, sua experiência. Graus de enraizamento são importantes porque,

por sua vez, afetam a facilidade de ativação, e a facilidade de ativação afeta a forma como falamos, pois co-determina as unidades que selecionamos. O inventário de todos é, portanto, afetado pelo uso, ou seja, o que você diz ativamente e o que ouve passivamente, e porque o inventário de todos é dinâmico, a mudança deve ser vista como uma característica de *design* da competência de cada falante individual e, portanto, da língua.

No entanto, embora o inventário de todos deva ser diferente, também deve haver restrições consideráveis em toda essa variação, uma vez que a comunicação normalmente ocorre sem problemas. Nós claramente não diferimos tanto um do outro que não possamos nos entender. Presumivelmente, devido à necessidade humana de formação de grupo e de marcação de membros do grupo, falantes individuais que compartilham a participação na mesma comunidade terão inventários bastante semelhantes, que além disso mudam de maneira semelhante. Os mecanismos para isso podem ser a imitação inconsciente (ou seja, alinhamento psicolinguístico; Pickering & Garrod, 2004) ou acomodação consciente a outros falantes do seu grupo.

Especificando esse quadro geral para o cenário de empréstimo lexical, podemos construir o seguinte curso de tempo (BACKUS, 2014). Começa quando um falante de turco na Holanda usa uma palavra holandesa específica enquanto fala turco. Esta é a chave de código de inserção prototípica: a palavra estrangeira é inserida em um enunciado da língua hospedeira. Outras pessoas também podem começar a usar a mesma palavra de forma independente; em seu discurso, também, representa uma mudança de código nesse sentido prototípico. As pessoas também podem começar a adotá-la umas das outras: você ouve outros falantes de turco usarem a palavra e adota esse comportamento dos outros. Pode-se dizer que nesse caso não se trata mais de um *codeswitch*, pois você não está adotando-o da outra língua, mas de outro falante da mesma língua: um caso clássico de difusão sociolinguística. Ao mesmo tempo, o falante em questão provavelmente sabe muito bem que é uma palavra holandesa de origem holandesa, e pode muito bem pretender usá-la por causa das conotações que vêm com ser uma palavra holandesa usada ao falar turco: pode transmitir o significado social de estar em contato com a vida na Holanda. Nesse caso, também seria uma mudança típica de código. Se a palavra continua a ser usada em enunciados turcos, torna-se parte regular do léxico turco de pelo menos aqueles que a usam, embora todos os falantes saibam que é de origem holandesa. Por esse tempo, é uma palavra de empréstimo estabelecida. Ao mesmo tempo, precisamente porque todo mundo sabe que é uma palavra holandesa, ela ainda pode ser usada com as intenções pragmáticas que a troca de código geralmente cumpre, como um apelo à modernidade ou ao poder. Presumivelmente, quanto mais estabelecida a palavra se torna, e quanto mais ela é simplesmente vista como a única ou a melhor palavra para o conceito que ela nomeia, menos potencial ela tem para esse uso pragmático. Ele simplesmente é ativado sempre que o conceito precisa ser verbalizado. Finalmente, se muitos falantes da comunidade usam rotineiramente a palavra, seria lógico se referir a ela como uma palavra turca estabelecida e, por ser de origem holandesa, uma palavra emprestada. Foi feito um empréstimo linguístico.

Nesse cenário, seria impossível dizer quando o *status* de palavra emprestada foi alcançado, a menos que impusesse uma definição arbitrária (por exemplo, número de usos por pessoa x número de enunciados, ou uso por uma certa porcentagem da população, medidas do que não seria absurdo, mas virtualmente impossível de implementar metodologicamente). O que faria mais sentido, penso eu, é aceitar alguma indeterminação e reservar o termo ‘*codeswitching*’ para a seleção sincrônica de um item de origem estrangeira e ‘empréstimo’ como o nome para o efeito diacrônico desses atos de seleção sincrônica.

Para transferência ou interferência gramatical, o mesmo cenário é válido. Um falante pode usar uma estrutura semelhante ao holandês ao falar turco pela primeira vez, como uma oração subordinada semelhante ao holandês, e isso seria um caso de interferência, e o mesmo pode acontecer com muitos outros falantes, pois todos eles também sabem holandês. Os falantes também serão expostos a outros falantes usando o recurso estrutural e adotá-lo deles. Ao contrário do uso de uma palavra holandesa, é improvável que tal interferência estrutural seja uma questão de intenção, razão pela qual a conotação com intencionalidade e indexicalidade social que é sempre uma parte possível do ato de *code-switching* está em grande parte ausente das discussões sobre interferência, como também está implícito no próprio termo ‘interferência’: acontece com você). À medida que o falante individual usa a estrutura repetidamente, ela se torna um aspecto convencional do turco desse falante cada vez mais (um empréstimo estrutural no nível individual), e como o mesmo acontece com outros falantes, pode ser dito com mais justificativa de que o turco emprestou a estrutura. Mais uma vez, não podemos dizer exatamente quando o empréstimo foi estabelecido, a menos que uma operacionalização arbitrária seja imposta. ‘Interferência’ (ou ‘transferência’) é um nome para a seleção sincrônica de uma estrutura reforçada por equivalente estrangeiro; ‘empréstimo’ é um nome para o efeito diacrônico desses atos de seleção sincrônica.

5. Implicações para a teoria

As semelhanças resumidas acima sugerem que uma descrição mais unificada dos fenômenos de contato deve ser possível, em vez de ter os subcampos relativamente separados de estudos de troca de código, por um lado, e relatos de mudança induzida por contato, por outro. Dado que os fenômenos empíricos tendem a ocorrer juntos, nos mesmos falantes e nas mesmas instâncias de uso da linguagem, talvez isso não seja uma má ideia para começar. No entanto, teria implicações em relação a pelo menos três tópicos.

Em primeiro lugar, a fronteira entre as palavras e a gramática pode não ser tão difícil quanto geralmente se supõe nas teorias linguísticas. Se empréstimos lexicais e estruturais ocorrem concomitantemente e, mais importante, seguem os mesmos mecanismos de enraizamento crescente em falantes e convencionalização em suas comunidades, pode ser útil investigar o que

eles compartilham em vez de apenas como eles diferem. A linguística baseada no uso apresenta todas as unidades linguísticas como unidades de significado de forma, e estas podem ser de qualquer tamanho e complexidade e variam no grau em que são lexicalmente fixas ou esquemáticas. Os itens lexicais típicos são simples e específicos; os padrões gramaticais típicos são complexos e esquemáticos. Uma segunda questão é como devemos ver a relação entre sincronia e diacronia. Se a mudança induzida pelo contato é resultado do uso bilíngue da língua, isso não deve, em princípio, ser diferente do que acontece com todo o uso da língua. Lembre-se de que, na descrição do processo de empréstimo acima, nem sempre fica claro se um falante individual recebe uma nova unidade ou um novo uso de falantes da língua estrangeira ou de outros falantes da língua que faz o empréstimo. Finalmente, devemos considerar o impacto da comunicação na competência linguística. O *locus* da mudança é a mente individual, mas o gatilho para isso está localizado nos eventos comunicativos em que o indivíduo se envolve.

Seguindo a lógica baseada no uso, seria útil aplicar esses pensamentos ao consenso sobre a mudança induzida pelo contato, que surgiu de décadas de pesquisa sobre bilinguismo a partir de perspectivas linguísticas, sociolinguísticas e psicolinguísticas. Geralmente, a pesquisa linguística descobriu que algumas coisas são mais emprestáveis do que outras e que existem certos princípios que governam o processo. Estudos sociolinguísticos estabeleceram que os bilíngues muitas vezes tiram coisas da língua que contém mais poder porque o empréstimo instancia um ato de identidade: usar algo da outra língua indexa uma associação com a cultura que produz essa língua. Estudos psicolinguísticos estabeleceram que, cognitivamente, algumas coisas são mais fáceis de ativar do que outras, ou mais difíceis de inibir, e que algumas coisas são mais fáceis de aprender do que outras, ou mais difíceis de esquecer. Como resultado, uma vez suficientemente frequentes, estruturas e palavras estrangeiras são ativadas quase automaticamente. A alegação maior por trás da linguística baseada no uso é que tudo o que chamamos de “linguística” pode se reduzir à combinação de perspectivas sociolinguísticas e psicolinguísticas. No que diz respeito ao empréstimo, isso deve significar que algo pode ser emprestado se for socialmente atraente ou cognitivamente fácil de ativar.

A combinação de considerações cognitivas e sociais talvez possa ajudar a entender a questão fundamental de porque, no mesmo par de línguas e na saída dos mesmos falantes, muitas vezes obtemos mudanças substanciais em um domínio gramatical e não em outro. Vimos que no contato turco-holandês, por exemplo, vemos muito mais evidências de mudança induzida pelo contato na subordinação do que no pro-drop ou marcação de casos. Uma lógica baseada no uso sugeriria que os domínios sintáticos centrais hospedam estruturas que são usadas o tempo todo: milhares de enunciados todos os dias as instanciarão. Como resultado, eles estarão profundamente enraizados para todos os falantes. Isso valeria, por exemplo, para pronomes nulos em enunciados pragmaticamente neutros em turco. Por outro lado, domínios sintáticos menos usados, por outro lado, hospedam estruturas menos bem enraizadas. Para falantes de turco na Holanda, isso pode

valer para orações subordinadas, pois elas podem ser mais típicas do discurso formal ou acadêmico do que para o vernáculo cotidiano. O turco formal não desempenha um grande papel na fala da maioria dos falantes de turco na Holanda: a maioria dos domínios nos quais eles usam registros mais formais são em holandês. Isso sugeriria que os pronomes nulos são facilmente ativados por esses falantes, mas a estrutura da oração subordinada não finita não é. A alta frequência protege contra variações e, portanto, contra mudanças; baixa frequência torna uma estrutura vulnerável. Tudo isso acontece mais ou menos automaticamente, pois as escolhas gramaticais são muito menos acessíveis à atenção consciente do que, por exemplo, itens lexicais. A mudança gramatical é principalmente uma questão cognitiva, embora em circunstâncias particulares a atenção possa ser focada em construções gramaticais individuais. Nesse caso, os oradores podem ser capazes de deter conscientemente a mudança, ou promovê-la.

De modo mais em geral, uma perspectiva baseada no uso geralmente requer respostas a perguntas básicas que se referem tanto à cognição quanto à sociabilidade. Se a mudança é uma característica de *design* da linguagem, e a linguagem realmente muda o tempo todo, então obviamente precisamos saber por que isso acontece, e muitas sugestões foram feitas ao longo da era moderna da teorização linguística, começando com a linguística histórica do século XIX. Quanto às mudanças induzidas pelo contato, novamente vemos uma série de explicações sugeridas que às vezes são puramente linguísticas (por exemplo, um desequilíbrio no sistema), puramente cognitivas (por exemplo, redução da carga cognitiva) e às vezes puramente sociolinguísticas (por exemplo, o poder de uma língua estrangeira). Uma visão baseada no uso nos compromete a tentar unificar essas diferentes explicações, uma vez que todas elas fazem sentido e todas elas foram demonstradas empiricamente. Se olharmos essas explicações um pouco mais de perto, começamos a ver semelhanças e pontos de contato entre elas. Um desequilíbrio no sistema refere-se a uma situação em que existem maneiras concorrentes de dizer a mesma coisa. Se a redução da carga cognitiva está impulsionando uma mudança, é porque, como falante, você quer manter as coisas simples e, portanto, usa o que é familiar: faz uso do processamento fácil que altos graus de enraizamento possibilitam (opta pelo menor esforço ‘erro de previsão’; CLARK, 2013). Finalmente, quando a influência estrangeira impulsiona a mudança, os falantes podem querer usar o que é semântica ou socialmente útil e o que os outros usam. Obviamente, é valioso estudar esses fatores em si mesmos, mas eles não são completamente independentes. O que é especialmente uma tarefa para a análise baseada no uso é ver se as explicações linguísticas se reduzem a psicolinguísticas, sociolinguísticas ou a uma combinação das duas.

6. Conclusão

O que podemos concluir do exposto é que nossos estudos sobre aquisição e socialização bilíngues sugerem que as fronteiras linguísticas são amplamente construídas socialmente. As crianças

bilíngues misturam suas línguas e param de fazê-lo quando aprendem que é apreciado por aqueles ao seu redor se mantiverem suas línguas separadas. Eles continuam fazendo isso se a troca de código não estiver marcada em seu ambiente sociolinguístico. Essa prática comunitária de misturar línguas pode ser o resultado inevitável dos fatores sociais e cognitivos que co-determinam o uso da língua. As mudanças lexicais e gramaticais são uma marca registrada dos idiomas observados nas configurações de contato. As mudanças lexicais são explicadas principalmente por fatores sociais; mudanças gramaticais por fatores cognitivos, mas essa oposição não é absoluta.

Juntos, esses fatores fornecem os contornos de uma estrutura baseada no uso para efeitos de contato. A cognição e a sociabilidade formam a base, assim como para todo o uso da linguagem. A sociabilidade garante a comunicação, que fornece informações para a cognição trabalhar e assim por diante em um ciclo de *feedback* sem fim. Esta forma de construção de modelos tem pelo menos duas implicações. Uma é que a necessidade da linguística, que limitou seu objeto de estudo em grande parte às unidades verbais da comunicação, de atender mais fundamentalmente aos aspectos sociais e cognitivos do comportamento humano exige uma discussão mais aprofundada sobre o que a disciplina deveria ser, qual é seu objeto de estudo. A segunda implicação é que o grau em que a linguística adota acriticamente a suposição, ou ideologia, de que comunidades particulares têm línguas particulares nos cega um pouco para o que a variação que vemos no uso da língua bilíngue está nos dizendo.

A separação da linguagem, argumentei, é o resultado lógico do *input* separado e da mistura de línguas do *input* misto. A co-ocorrência de duas ou mais unidades leva ao enraizamento da combinação em uma unidade maior, e isso gera uso adicional da mesma unidade. São mecanismos cognitivos. No entanto, o *input* separado e o uso separado de duas línguas também é resultado do uso da língua como marcador de pertencimento ao grupo, um fator social. A questão então é por que a língua funciona assim? Talvez tenha a ver com a pura frequência do uso da língua: usamos o dia todo, dando ampla chance para a cognição fazer seu trabalho, e construímos muitas unidades que podem ser reutilizadas com alto grau de automaticidade e baixo grau de atenção consciente. E talvez também tenha a ver com sua versatilidade: usamos a língua como ferramenta cultural para transmitir matizes ilimitados de significado, da semântica à indexicalidade social. A sociabilidade é o que impulsiona porque usamos a linguagem em primeiro lugar.

Em suma, argumentei que uma lente baseada no uso impõe à análise linguística que tanto os princípios cognitivos quanto os objetivos sociais sejam levados em consideração juntos. Os princípios cognitivos são importantes porque a comunicação é um comportamento orientado à rotina: o grau de enraizamento determina a facilidade de ativação e a facilidade de ativação co-determina quais unidades são selecionadas na fala. Os objetivos sociais são importantes porque a comunicação também é um comportamento orientado a objetivos: a seleção consciente determina qual é a melhor

opção para transmitir o significado pretendido. Explicações psicolinguísticas e sociolinguísticas são, portanto, necessárias, embora talvez nem sempre ao mesmo tempo. É possível, e até mesmo esperado de uma perspectiva baseada no uso, que a articulação adicional de tais análises torne desnecessário um nível linguístico separado de explicação, pois as explicações linguísticas sempre se reduzirão a uma combinação de fatores cognitivos e sociais.

Referências

AALBERSE, S.; BACKUS, A.; MUYSKEN, P. *Heritage Languages. A Language Contact Approach*. Studies in Bilingualism 58. Amsterdam/Philadelphia: John Benjamins Publishing Company, 2020.

AZAR, Z.; ÖZYÜREK, A.; BACKUS, A. Turkish-Dutch bilinguals maintain language-specific reference tracking strategies in elicited narratives. *International Journal of Bilingualism*. Advance online Publication. Published online 5 April 2019. DOI: <https://doi.org/10.1177/1367006919838375>

BACKUS, A. Turkish as an immigrant language in Europe. In: BHATIA, T. K.; WILLIAM, C. R. (ed.). *The Handbook of Bilingualism and Multilingualism*. 2nd edition. Blackwell Handbooks in Linguistics, 2013. p. 770-790.

BACKUS, A. Towards a usage-based account of language change: implications of contact linguistics for linguistic theory. In: NICOLAÏ, R. (ed.). *Questioning language contact. Limits of contact, contact at its limits*, 91-118. Leiden/Boston: Brill, 2014.

BARKING, M.; BACKUS, A.; MOS, M. (forthc.). Comparing Forward and Reverse Transfer from Dutch to German. *International Journal of Bilingualism*.

CHRISTIANSEN, M.; CHATER, N. The Now-or-Never Bottleneck: A Fundamental Constraint on Language. *Behavioral and Brain Studies*, v. 39, e62, 2016. DOI: <https://doi.org/10.1017/S0140525X1500031X,e62>

CLARK, A. Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*, v. 36, p. 181-253, 2013. DOI: <https://doi.org/10.1017/S0140525X12000477>

DEMIRÇAY, D. *Connected Languages: Effects of intensifying contact between Turkish and Dutch*. Utrecht: LOT, 2017.

DOĞRUÖZ, S.; BACKUS, A. Innovative constructions in Dutch Turkish: An assessment of on-going contact-induced change. *Bilingualism: Language & Cognition*, v. 12, n. 1, p. 41-63, 2009.

GASKINS, D.; BACKUS, A.; ENDEFELDER QUICK, A. Slot and frame schemas in the language of a Polish and English speaking child: the impact of usage patterns on the switch placement. *Languages*, 2019, 4, 8. DOI: <https://doi.org/10.3390/languages4010008>

HAKIMOV, N.; BACKUS, A. (forthc.). Usage-Based Contact Linguistics: Effects of Frequency and Similarity in Language Contact. *Journal of Language Contact*, v. 13, p. 459-481, 2021.

HEINE, B.; KUTEVA, T. *Language Contact and Grammatical Change* (Cambridge Approaches to Language Contact). Cambridge: Cambridge University Press, 2005.

MATRAS, Y. *Language contact*. Cambridge: Cambridge University Press, 2009.

MUYSKEN, P. *Bilingual Speech. A Typology of Code-Mixing*. Cambridge: Cambridge University Press, 2000.

MYERS-SCOTTON, C. *Social Motivations for Codeswitching: Evidence from Africa*. Oxford: Oxford University Press, 1993.

ONAR VALK, P. *Transformation in Dutch Turkish Subordination? Converging Evidence of Change Regarding Finiteness and Word Order in Complex Clauses*. PhD. Tilburg University, 2015.

PICKERING, M. J.; GARROD, S. Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences*, 27, 169-190; discussion 190-226, 2004.

QUICK, A. E.; LIEVEN, E.; BACKUS, A.; TOMASELLO, M. Constructively combining languages: The use of code-mixing in German-English bilingual child language acquisition. *Linguistic Approaches to Bilingualism*, v. 8, n. 3, p. 393-409, 2018. DOI: <https://doi.org/10.1075/lab.17008.qui>

SODACI, H.; BACKUS, A.; KOOTSTRA, G. J. (forthc.). Role of structural priming in contact-induced change: Subject pronoun expression in L1 Turkish by Turkish-Dutch bilinguals. *Journal of Cultural Cognitive Science*. Pre-print. DOI: <https://doi.org/10.31234/osf.io/2uxej>

TORRES CACOULLOS, R.; TRAVIS, C. *Bilingualism in the Community. Code-Switching and Grammars in Contact*. Cambridge: Cambridge University Press, 2018.

TREFFERS-DALLER, J. *Mixing two languages: French-Dutch contact in a comparative perspective*. Berlin: De Gruyter, 1994.

ZENNER, E.; BACKUS, A.; WINTER-FROEMEL, E. (ed.). *Cognitive Contact Linguistics. Placing usage, meaning and mind at the core of contact-induced variation and change*. *Cognitive Linguistics Research*, Volume 62. Berlin: De Gruyter Mouton. I-VI + 1-342, 2019.

Prosody and bilingualism

Chapter 5

**Resolving the
prosody paradox**

Vincent J. van Heuven

I. Introduction

In this tutorial paper, I will first explain what we mean by prosody in language and speech. I will then show that, generally, prosody is highly redundant, i.e., it can be omitted from the speech signal without causing speech to become unintelligible or incomprehensible. Yet, it has often been mentioned, also in recent publications, that getting the prosody right should have priority in the acquisition of the phonology of a foreign language. This seems a contradiction. Why should a generally redundant feature of a code deserve priority in teaching? It is the purpose of this paper to resolve this paradox.

I.1 What is prosody?

In the study of speech and language, it is customary to analyze a sentence as a sequence of discrete time-ordered units often referred to as “segments”. A spoken sentence is analyzed as a sequence of morphemes, i.e., minimal sequences of sound with an identifiable meaning associated with them, while each morpheme is decomposed into a sequence of vowels and consonants – which are seen as the shortest building blocks from which spoken sentences are constructed. At the same time, it is readily acknowledged that words and sentences cannot be described just as a sequence of segments. In fact, the same sequence of vowels and consonants may constitute different words with different meanings (and the same sequence of words may make up different sentences) depending on the prosody. Prosody is defined, then, as the ensemble of properties characterizing words and sentences that cannot be predicted from the mere sequence of smaller sound segments.⁹ For instance, Dutch has two different words with unrelated meanings that are composed of the same five sounds in the same order: /ka:nɔn/. If the first syllable is pronounced stronger than the second, the meaning is ‘list of saints’ or ‘round song’ (spelled *canon*), but with the emphasis on the second syllable, the word means ‘large gun’ (spelled *kanon*; for more discussion of the definition of prosody see VAN HEUVEN, 1994, 2008; VAN HEUVEN; SLUIJTER, 1996; NOOTEBOOM, 1997; RIETVELD; VAN HEUVEN, 2016, chapter 12).

In most textbooks on phonetics, authors seem satisfied with simply listing those properties of speech that can generally be considered to be prosodic in nature, such as:

- melodic organization: pitch, tone, intonation;
- temporal organization: length, duration, tempo and tempo variation, and pause;
- dynamic organization: loudness and loudness variation, word and sentence stress.

⁹ The word prosody derives from Greek *pros* ‘with’ and *oidein* ‘sing’, i.e., it refers to the musical accompaniment of sung words. Somewhat confusingly, the Latin loan translation *accent* (= *ad* + *cantus* ‘with the song’) does not refer to all aspects of prosody but is restricted to the phonetic realization of emphasis only.

Although such a taxonomy gets us a long way, matters get more complicated when we consider the fact that all individual vowels and consonants have intrinsic melodic, temporal, and dynamic properties. Generally, for instance, closed vowels such as [i] and [u] are pronounced – all else being equal – with a higher pitch than an open vowel [a] (e.g., LEHISTE; PETERSON, 1961, see Figure 1-a). Conversely, open vowels naturally have greater intensity than closed vowels (e.g., PETERSON; LEHISTE, 1959, see Figure 1-b), if only because the mouth radiates sound pressure more effectively from an open (horn-shaped) mouth than from a closed mouth (which functions as a funnel, stifling the sound). Similarly, open vowels tend to be longer than closed vowels (e.g., LEHISTE; PETERSON, 1960, see Figure 1-c), since opening the jaw widely (as for [a]) and closing it again takes up more time than just opening it a little and closing it again (as for an [i] or [u]). Yet, these intrinsic properties, by their very nature, are fully predictable from the identity of the segments and should therefore be excluded from the realm of prosody.

To belabor this point a little further, there is also a class of properties of segments that is referred to by the term “co-intrinsic”. This term subsumes those properties of a segment that are fully predictable from the identity of the neighboring segments. For example, a vowel will last longer when followed by a voiced obstruent than when it is followed – all else being equal – by a voiceless obstruent (SLIS; COHEN, 1969, LISKER; ABRAMSON, 1964). The fundamental frequency (pitch) of a vowel after a voiceless obstruent starts some 20 to 30 Hz higher than after a voiced obstruent (so-called co-intrinsic pitch, e.g., LEHISTE; PETERSON, 1961, see Figure 1-d). Again, this part of the melodic, temporal, and dynamic properties of sounds is predictable from the sequence of segments and is, therefore, no part of prosody. To uncover the true prosody of speech, some form of melodic, temporal, and dynamic decomposition is needed, which extracts the inherent segmental properties first and then factors out the co-intrinsic influences so that pure prosody remains.

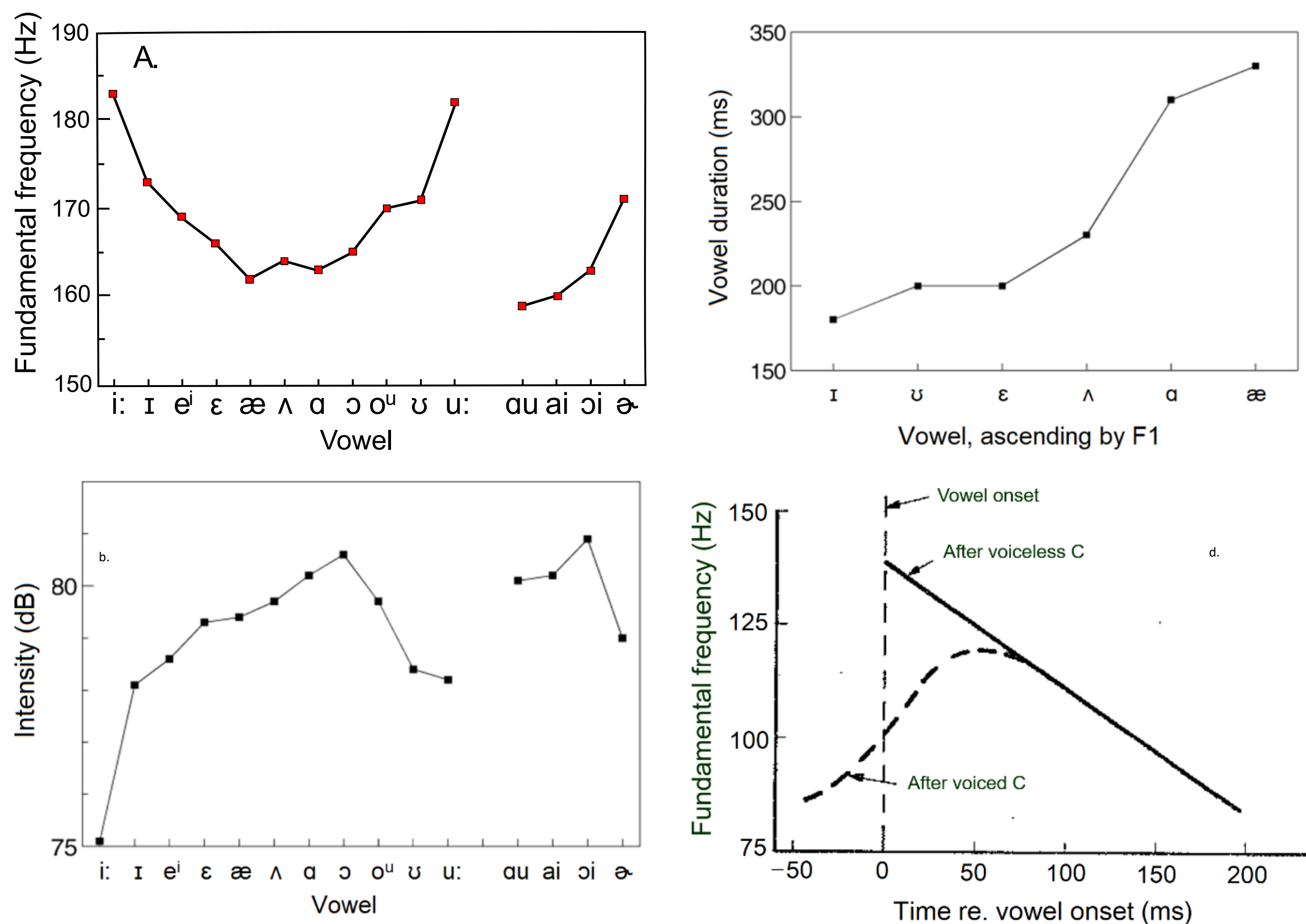


Figure 1: (a) Fundamental frequency (Hz) for 15 American English vowels (adapted from RIETVELD; VAN HEUVEN, 2001: 234, after LEHISTE; PETERSON, 1961). (b) Peak intensity (dB) for the same 15 vowels (adapted from RIETVELD; VAN HEUVEN, 2001: 233, after PETERSON; LEHISTE, 1959). (c) Duration (ms) of six phonologically lax vowels in American English ordered by degree of openness (adapted from RIETVELD; VAN HEUVEN, 2001, p. 235, after LEHISTE; PETERSON, 1960). (d) Idealized fundamental frequency (Hz) as a function of time (ms) after voiced and voiceless obstruents (adapted from NOOTEBOOM; COHEN, 1976, p. 85, after SLIS; COHEN, 1969).

1.2 Communicative functions of prosody

It is customary to differentiate between word prosody and sentence prosody. There are basically two types of word-prosodic systems in the languages of the world, i.e., stress languages and tone languages. The communicative function of tone would be to help differentiate between words

in the lexicon. For instance, in Mandarin Chinese, the same syllable /ma/ may signal five different and unrelated meanings, depending on the tone (e.g., VAN HEUVEN, 2008). Stress may also help to differentiate between words (as in *canon* ~ *kanon* above) but only in languages in which the location of the stressed syllable within the word is unpredictable. In languages with fixed stress, such as in Hungarian, where stress is always word-initial, stress may signal word boundaries.

The communicative functions of sentence prosody are many.

- *Phrasing*. By acceleration (initial strengthening), deceleration (pre-boundary lengthening), and pitch changes, the speaker may signal which sequence of words should be analyzed as a prosodic domain – with different markings depending on the domain’s hierarchical depth.
- *Clause typing*. Groups of words (phrases), whether they are a sentence or a clause within a sentence, can be qualified as a specific type. Well-known sentence types are statement, question, exclamation, command, and negation. On a lower level, there are at least two clause types, i.e., clauses that are sentence-final (finality) and non-final clauses, which typically contain a prosodic continuation marker.
- *Attentional function*. Within such prosodic domains, one word is typically presented as communicatively more important than the others through sentence stress – typically marked by a conspicuous change in vocal pitch on the syllable within the word that also has word stress (see e.g., VAN HEUVEN, 2018 for a cross-linguistic survey of findings).
- *Intentional function*. Different melodies may then further qualify the information status of referents relative to the background shared by the speaker and hearer. In Dutch, for instance, sentence stress marked by only a fall in pitch suggests that the speaker could easily guess the identity of the focused referent, whereas rise-fall stress indicates that the referent was both new and unexpected (KEIJSPER, 1984, CASPERS, 2000).
- *Affect*. Prosody also makes an important contribution to the expression of affect. This is an umbrella term covering attitudes (e.g., sincerity, irony, sarcasm) and emotions (e.g., happiness, sadness, anger, fear) on the part of the speaker. Since such distinctions are typically not made in a categorical but gradient fashion, it is often argued that they are not part of language (i.e., phonology) but should be considered paralinguistic properties.
- *Continuity*. The melody of a sentence unfolds rather slowly in time and obeys a highly regular and predictable pattern. This is an important source of information that allows a listener to keep track of a speaker amidst competing voices in a crowd.

1.3 Intelligibility vs. comprehensibility

When two persons communicate through spoken language, thoughts that enter the speaker’s mind must first be expressed in terms of the vocabulary and grammatical structures of the speaker’s language. The mental linguistic structures are then used to make the speaker’s vocal organs move so that they produce audible sounds. The sounds travel through the air (or some other medium – e.g., a telecommunication channel) and then impinge on the ear of the listener. The result is that the listener hears a stream of sounds. If the listener is familiar with the language, he or she will

recognize (the same) linguistic units (e.g., words) in the same order in which they left the speaker's mouth. This part of the communication process is what we call *speech recognition*. If a sufficient number of units have been correctly recognized in the correct sequential order, the listener will be able to reconstruct the original thoughts and intentions of the speaker. This last part of the process is what we call *speech understanding* or *comprehension*. The sequence of events sketched here is known as the *speech chain* (DENES; PINSON, 1963), and it has been the blueprint of Levelt's (1989) model of speech production and Cutler's (2012) model of native listening.

Different methods are required to assess a speaker's intelligibility than to assess the individual's comprehensibility. If the listener does not know the language the speaker uses, the speaker's intelligibility is (close to) zero – even if the utterance(s) would be perfectly intelligible to native listeners of the language. The *comprehensibility* of a speaker (or a spoken text) is the degree to which a listener can understand the speaker's meaning and intentions. Intelligibility, then, correlates with speech recognition and comprehensibility with speech understanding. In this view, speech understanding is a higher-order process than speech recognition. Different methods are required to assess a speaker's intelligibility and then assess the individual's comprehensibility. For instance, a functional intelligibility test would be to ask a listener to take down, by way of dictation, a series of nonsense utterances produced by the speaker. This is what the SUS test (SUS: semantically unpredictable sentences, Benoît et al., 1996) does with utterances as *The state sang through the whole week*. A functional speech comprehension test would, for example, require a listener to determine whether a spoken sentence embodies a truth or a falsehood. If a listener would think that *Most human babies are heavier than a full-grown elephant* is true, he (or she) obviously has not understood the sentence.¹⁰

It should be clear from the above explanation that word prosody, i.e., stress and lexical tone, make a potential contribution to the recognition of words. In fact, stress and tone are indispensable for word recognition when a language has lexical items consisting of identical segment strings that differ only in the location of the stress or in the type of lexical tone (see the examples of minimal stress pairs in Dutch and English, and the minimal tone quadruplets in Mandarin). Generally, however, the ambiguity that may arise due to such minimal word-prosodic contrasts is easily resolved based on top-down contextual information. It has been shown that stress and lexical tone are used in the online word recognition process. If a Dutch or Spanish listener hears that a spoken word begins with a stressed syllable, any word in the lexicon that does not have a stressed initial syllable is no longer a candidate eligible for recognition. This would be an example of how stress information can help

¹⁰ I spell this out in some detail because other disciplines and other researchers have used the terms differently from the way phoneticians do. For instance, applied linguists (MUNRO; DERWING, 1995; MUNRO et al., 2006) use the term *intelligibility* as the degree to which a speaker can be understood using functional tests, and the term *comprehensibility* for the listener's opinion as to how well a speaker (or utterance) can be understood. Our position is that intelligibility and comprehensibility address two different stages in the speech chain, and that each can be measured both by functional tests and by opinion tests.

narrow down the pool of recognition candidates in the early stage of the online word recognition process called lexical access. In other cases, word prosody can be used in the later stages of the recognition process, the verification stage, to check whether the perceived word prosody matches the stored pattern for the word that has been provisionally recognized on the basis of the sequence of segments. I will come back to this use of word prosody in later sections of this paper when we review some studies addressing this issue.

Another way in which word prosody may assist word recognition, and thereby contribute to speech intelligibility, is seen in languages with fixed word stress. It has been suggested that in such languages, hearing a stress informs the listener that s/he is at the beginning of a new word when stress is fixed-initial (as in Finnish or Hungarian) or that the very next syllable will be a word onset when stress is fixed-final (as in Turkish or Persian). In fact, the stress does not have to be fixed. In English and Dutch, most words begin with a stressed syllable; listeners assume that every time they hear a stress, they are at the beginning of a new word – even if this is not always true. It has even been found that infants use this segmentation strategy in the first year of their lives to detect words in the continuous stream of speech in their environment.

Sentence prosody makes very little contribution to intelligibility, as defined in this section. The communicative functions listed for sentence prosody (see section 1.2) are there to assist the listener in grasping the intentions the speaker wants to impart to the hearer by the sequence of words he has strung together. The phrasing, through acceleration and deceleration, pausing, and the use of boundary tones, tells the listener which words should be parsed together into cohesive thought units and where important breaks are in the linguistic structure of the spoken text. Sentence stress informs the hearer which words he should pay special attention to, while sentence melody is used to divulge the speaker's deeper intentions, attitudes, and emotions.¹¹

It is possible, to some extent, to remove all prosodic information from a spoken text and leave only the segmental information, i.e., the sequence of vowels and consonants, in place. This can be done by, for instance, using speech synthesis based exclusively on sounds excerpted from fully stressed syllables, omitting any pitch inflection or adjustment of segment durations, and reducing all grammatical pauses to zero. The resulting speech is spoken on a perfectly flat monotone, with a staccato type of rhythm (since all syllables are pronounced as if they were fully stressed, all words are given equal weight, and any suggestion of phrasing would be absent). The result would be comparable to a printed text without punctuation marks and paragraph indents. Even if we were to remove the spaces and hyphens that normally separate words, we would nevertheless be able,

¹¹ Van Wijngaarden (2001) measured the intelligibility (in terms of Speech Reception Threshold SRT, i.e., the amount of noise that the listener can tolerate and still achieve 50% correct sentence recognition) for redundant every day Dutch utterances in two prosodic conditions, i.e., with their original melody and with the melody replaced by a fixed constant pitch (through PSOLA resynthesis). The SRT for the monotonized utterance was 2 dB poorer for the monotonized versions (probably because the artificial monotonization causes a ringing effect that interferes with segmental intelligibility). In quiet, however, both versions were perfectly intelligible.

with some effort, to reconstruct the missing prosodic structures from the segmental information that is still coded in the printed words. To illustrate the point, I print here the opening paragraph of an article that appeared some 40 years ago (SMITH; RAFIQZAD, 1979).

in discussions of the use of English for cross-cultural communication it is often maintained that the educated native speaker is more likely to be intelligible to other than the educated non-native speaker. These people therefore claim that a native speaker variety of English should be the model for production as well as the standard with which to compare all other varieties. People of this opinion are also often concerned that the use of other models will lead to such a great diversity of non-native varieties of educated English that soon persons speaking English may not be intelligible to their listeners.

Very little research has been done on comprehending speech devoid of prosody. An exception is Keskin et al. (2019), which used a text-to-speech system for Turkish with all punctuation marks deleted from the input text and all prosodic rules disengaged as a control condition in an experiment that targeted the comprehensibility of speech with and without prosody. Four matched groups of Turkish primary school children listened to the same story in four different versions, i.e., spoken by a human speaker or by a text-to-speech system, with and without prosody. Listening comprehension was tested by six multiple-choice items on literal comprehension and six more on inferential comprehension. Results were as follows:

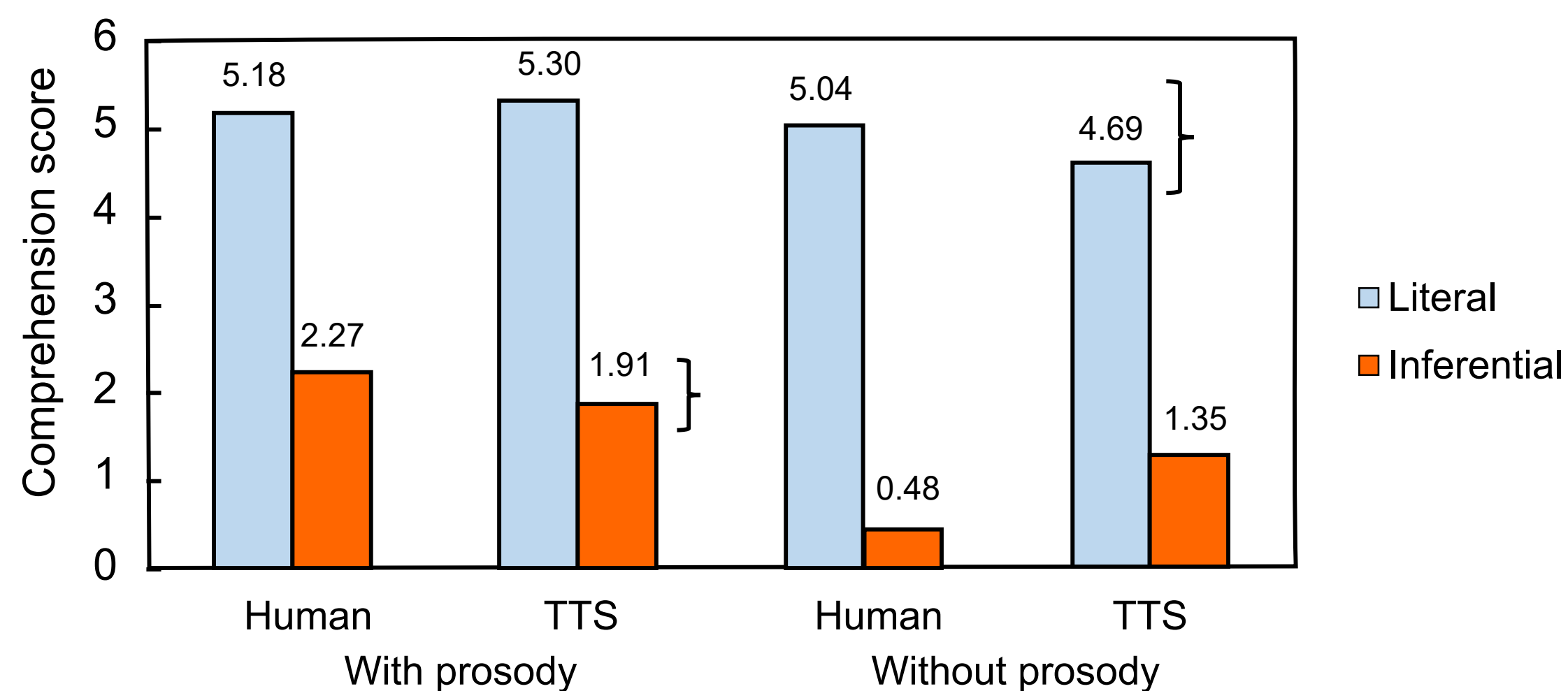


Figure 2: Listening comprehension score (maximum = 6) for four versions of the same text, spoken by a human or by a text-to-speech system (TTS), with and without prosody. Braces enclose means that do not differ significantly from one another. See text for details. Adapted from Keskin et al. (2019).

Although the versions without prosody obtained slightly poorer comprehension scores, the differences were not significant for the literal content questions. Only when the children had to make inferences from what they had heard in the spoken passage was comprehension negatively (and significantly) affected by the absence of prosody. This demonstrates that comprehension remains generally possible even in the absence of prosodic information. Still, there is some circumstantial evidence suggesting that removing prosody from spoken materials impacts comprehension. For one

thing, listeners tend to lose concentration and complain that they find it difficult to keep track of what the speaker is trying to say when prosody is (artificially) eliminated from spoken passages. And in the paralinguistic domain, the recognition of emotions may well be compromised in monotonized speech (e.g., VAN DE VELDE *et al.*, 2017).

The upshot of this section would be that prosody generally contributes redundant, non-vital information. Removing prosodic information renders the input speech unnatural (and unacceptable) but generally does not prevent the native listener from recognizing the words in sequence and comprehending the message.

1.4 Relative importance of segments vs. prosody for intelligibility and comprehensibility

There has been considerable debate on the relative importance of segmental structure versus prosody. In written language, prosody is hardly reflected, while most orthographies provide a fairly precise and exhaustive coding of the identity of the vowels and consonants that make up a sentence. Word and sentence stress is usually not marked in written language at all (although irregular word stress is marked in Spanish orthography, and high and low word tones are explicitly marked by acute and grave accents in the orthographies of some African tone languages), while only the deepest prosodic boundaries (corresponding approximately with syntactic paragraph, sentence and clause boundaries) are represented in spelling. Prosody is almost completely redundant given the segmental composition of the words in sequence. If we remove all useful prosodic information from a spoken sentence or paragraph, the result is still perfectly intelligible (although more difficult to comprehend and remember, see the preceding section). However, if we eliminate all segmental information from an utterance while fully maintaining prosodic information, e.g., through low-pass filtering at 300 Hz for a male speaker, no intelligibility remains.¹² Even in tone languages such as Mandarin, intelligibility remains close to the ceiling level when prosody (including word tones) is eliminated (e.g., QIN, 2007, VAN HEUVEN, 2008, WANG *et al.*, 2011).

2. Prosody and foreign language learning

In the preceding section, we have speculated on the relative importance of segmental vs. suprasegmental (or: prosodic) speech properties that allow a native listener to recognize the words in sequence (correlated with intelligibility) and reconstruct the speaker's meaning and intentions (correlated with comprehensibility). In this section, I will describe and discuss a few publications that have often been cited in support of the idea that prosody is more important for speech recognition and understanding than segmental properties when native speakers (of English) listen to a non-

¹² These effects are demonstrated here: <https://figshare.com/s/7ab0b2461206b42f53aa>

native speaker. If this claim is true, then a question crops up. How is it possible that elements of speech that are almost completely redundant in English, such as the word and sentence prosody, can be more important than segmental structure in the communication between a non-native talker and a native English listener? This is what I call the prosody paradox.

2.1 Anderson-Hsieh *et al.* (1992).

The first paper we should look at is Anderson-Hsieh *et al.* (1992). This study investigated the relationship between experienced raters' judgments of non-native pronunciation and actual deviance in segmentals, prosody, and syllable structure. Sixty speakers from 11 language groups recorded the same reading passage and were rated by expert raters on pronunciation and later analyzed impressionistically for deviance in segmentals, prosody, and syllable structure. The deviance found in each pronunciation area was then correlated with the pronunciation ratings. An analysis of the 60 speakers showed that deviance in segmentals, prosody, and syllable structure all showed a significant influence on the pronunciation ratings. However, the prosodic variable proved to have the strongest effect. When separate analyses were done on two language subgroups within the sample, prosody was always found to be significantly related to the global ratings, whereas this was not always true for the other variables investigated.

First, it should be realized that syllable structure per se is not an independent area of pronunciation. Incorrect pronunciation of English syllable structure is generally found when the learner's native language does not have the complex consonant clusters of English. They then either delete one or more consonants from the cluster or break up the cluster by inserting an epenthetic vowel before it somewhere within the cluster. I would argue that this reduces syllable structure errors to segmental errors. Segmental errors can be of three types: segments can be deleted, inserted, or replaced. There is no principled reason why insertions and deletions should be classified as basically different from substitutions. When all segmental errors are combined, they may well be more influential as a whole than prosodic errors.

In the study, the non-native speakers hailed from 11 different language backgrounds. Generally, we should find relatively high correlations between the various pronunciation components for EFL learners sharing the same L1. The pronunciation quality of the consonants, vowels, clusters, and prosody should develop along a parallel course. However, when the learners have different L1 backgrounds, it may well be the case that the pronunciation of the consonants, for instance, is better than that of the vowels and/or the prosody. The interfering factor here is the way the learner's L1 differs from English. It has been observed that EFL learners generally pronounce the (single) consonants quite well and that the communicative problems they experience with English reside in the pronunciation of the vowels. This discrepancy can be understood if we consider that English

has just an average-size consonant inventory, as determined, for instance, from the *World Atlas of Linguistic Structures* (COMRIE et al., 2005). This means that, statistically, at least 50% of the EFL learners have a native language with a consonant inventory that is equal to or larger in size than English. In such cases, it will be relatively easy for the learner to substitute L1 consonants in English which are close enough to one another that intelligibility is not overly compromised. However, when it comes to vowels, English belongs to the category of languages with large vowel inventories. The large majority of EFL learners, therefore, will be faced with the problem that their L1 inventory lacks some vowels that can be used as reasonable substitutes in English, so vowel pronunciation will compromise the learner's intelligibility until s/he has established new vowel sounds and shifted boundaries of existing L1 categories (a process that may take years).

Based on what was said above, it will generally be the case that EFL learners will show deviant realizations of the consonants and especially the vowels of English. As I will argue later in this chapter, the role of prosody is not constant but varies depending on the quality of the segmental structure. If the vowels and the consonants are poorly defined, then prosody assumes a more important role, while its influence remains modest (and unnoticeable) as long as the segmental quality of EFL is good (native).

2.2. Hahn (2004)

A single proficient Korean speaker read out a short English text (270 words) in three versions (harmonized post hoc for overall speed and pause duration). Version A: sentence stresses (primary stress) on [+F] (i.e., focused, communicatively important) words only; version B: sentence stress on one [-F] (non-focused, communicatively unimportant) word; version C: no sentence stress at all.¹³ Each of the three versions was played to a different group of 30 American native listeners with little or no experience with non-native Englishes. Three functional intelligibility/comprehensibility tests were administered, and a questionnaire was filled in after the functional tasks were completed.

- Test 1: Listeners had to click their mouse every time they heard a beep while trying to understand and remember the contents of the text. Accuracy and RT were measured.
- Test 2: Listeners had to retell the contexts as accurately and completely as they could. Correctly reported main points and minor points (details) were counted.
- Test 3: A short-answer 6-item open comprehension test was administered, capturing the main points of the text; each question yielded 0, 1, or 2 points, max = 12.

¹³ It seems impossible to me to produce a sentence without any sentence stress. Possibly, Hahn meant that all sentence stresses were of equal magnitude. It would also be very strange to have a text composed of many sentences, all of which contain just one sentence stress on one [-F] word. There are various other oddities in the description of the materials. If condition C has equal-sized pitch movements on all [+F] words, the text should be perfectly OK. The results, however, indicate that it is just as bad as version B. This renders the paper and its results difficult to understand.

- The opinion test afterwards asked listeners to rate the speaker on 11 scales having to do with intelligibility, comprehensibility, liveliness, and pleasantness of voice.

As for test 1, version A (134) was somewhat better than versions C (128) and B (125), but the effect was not significant.¹⁴ The scoring of the retelling data in test 2 is unclear as the maximum number of main points and details that could be remembered is not mentioned. Be this as may, A was retold more accurately than B and C (C was slightly better, which did not differ significantly from B). A count of main points was more informative than the count of details (no significant differences in the details), but the unweighted simple addition of main points and details discriminated best. The same goes for the comprehension quiz data in test 3: the three conditions are in the same order, but the effects were insignificant this time. The differences among the three means are within 1 point (on a scale from 0 to 12). The judgment data, finally, are in line with the retelling results: version A gets more favorable ratings than either B or C, with little difference between the latter two. What is strange is that the ratings for “better emphasis on important points” and (significantly) “lack of monotony” were more favorable for version B (with NO sentence stresses) than for C (with one primary stress per sentence on the wrong word). Figure 3 summarizes the results of the opinion test.

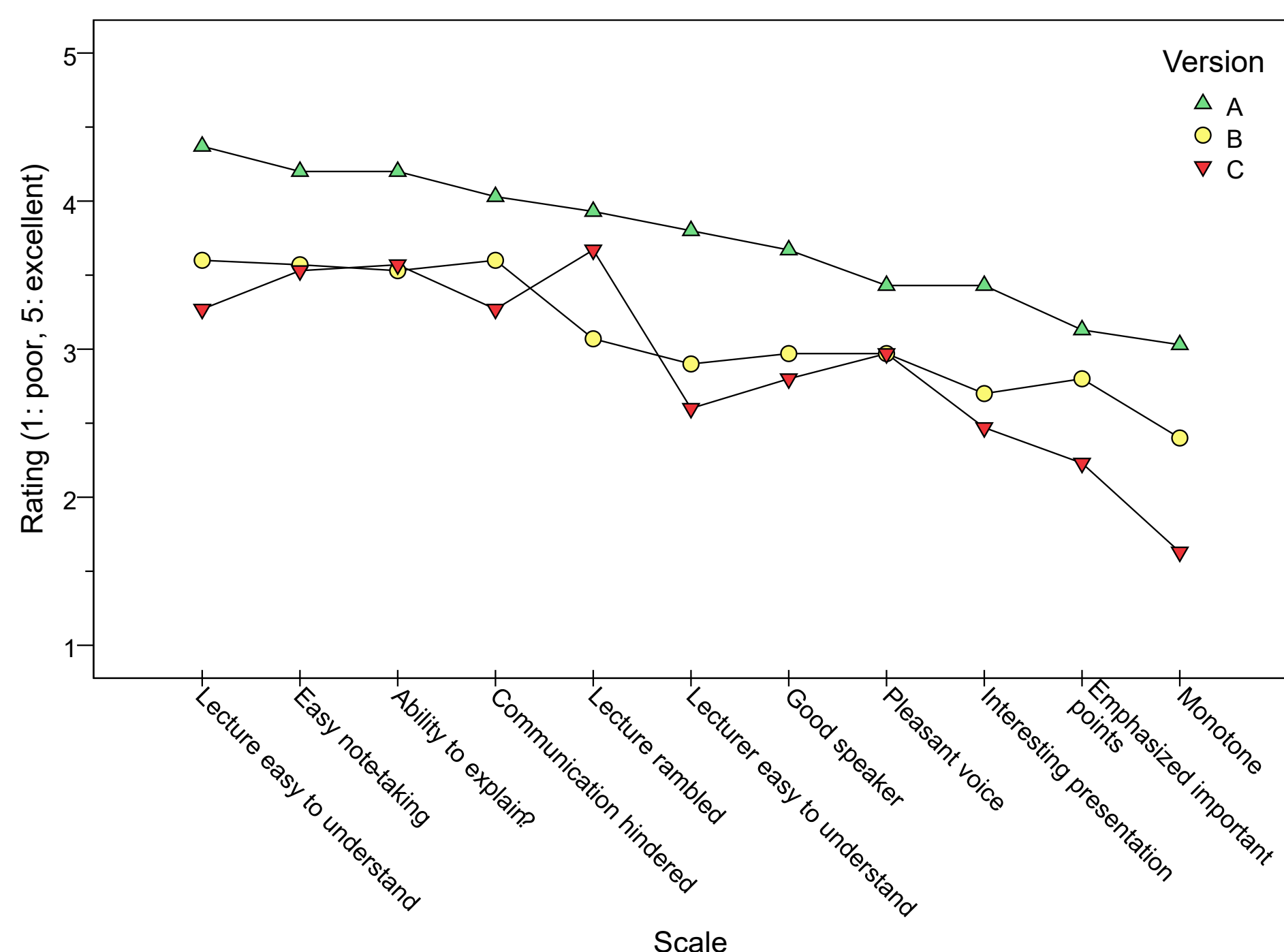


Figure 3: Judgments by American native listeners on eleven rating scales of speakers in three versions. A: Sentence stress on +Focus words only, B: One sentence stress on a –Focus word, C: No sentence stress.

The crucial problem with this study is that the Korean speaker, even if proficient in English, must have read out the text with a Korean accent, which will have affected his vowels and consonants. To

¹⁴ It is unclear what the numbers refer to. My guess is that it is the reciprocal of RT in ms multiplied by 10,000 – but Hahn does not say whether the RT was measured from the onset or the offset of the beep, nor does she explain how she dealt with missed beeps. Suppose the mean RT to a beep would be 500 ms, then a reasonable value would be $1/500 = .002 \times 10,000 = 20$. An RT of 1000 ms would yield a value of $1/1000 = .001 \times 10,000 = 10$. The value of 134 reported for version A would correspond to a mean RT of $1/(134/10000) = 75$ ms – which is impossibly fast (no one can respond faster than 200 ms). Probably the numbers in Table 1 are the reciprocals ($\times 10,000$) of RTs measured from the end of a beep (with a length between 250 and 500 ms).

be able to interpret Hahn's results, we need a control experiment with the same text spoken (in the same three versions) by a native speaker of American English. It is, in fact, not clear to me why the experiment was done with a non-native speaker in the first place. If the researcher was interested in the role of sentence stress in the processing of a spoken text, then a native speaker would have been a more obvious choice. Be this as it may, the results would seem to indicate that putting sentence stress on a communicatively unimportant (-Focus) word makes the text more difficult to understand for a native English listener. This confirms what we know from the literature on native speakers and listeners of English (e.g., FOWLER; HOUSUM, 1987) and other (related) languages such as Dutch (e.g. TERKEN; NOOTEBOOM, 1987; VAN DONSELAAR, 1995): putting sentence stress on a non-focused word confuses the listener, who then believes that he should look for a new referent; similarly, omitting sentence stress on a new referent prompts the English listener to incorrectly look for a given referent in the preceding context. My prediction would be that the effect of the wrong sentence stresses will be (more) modest when the speaker is native than when s/he is a non-native (whether Korean or any other language).

2.3. Field (2005)

This is an often-cited study in support of the claim that correct word stress in English is important both for native and non-native listeners. One (male) native speaker of English recorded 12 high-frequency two-syllable words with a Strong-weak (Sw) stress pattern and another twelve with a wS stress pattern (see Table 1).

Table 1: Materials used by Field (2005).

Quality change possible in unstressed V?	Lexical stress pattern							
	Sw				wS			
	Orthogr.	Default	Shift only	Shift+Q	Orthogr.	Default	Shift only	Shift+Q
Yes	<i>husband</i> <i>second</i> <i>different</i> <i>person</i> <i>Friday</i> <i>women</i>	'hʌzbənd 'sekənd 'dɪfrənt 'pɜ:sən 'fraɪdɪ 'wɪmən	hʌz' bə:nd se' kə:nd dɪ' frə:nt pə:' sən fraɪ' dɪ: wɪ' mə:n	hʌz' bænd se' kɒnd ¹⁵ dɪ' frent pə:' sɒn fraɪ' deɪ wɪ' men	<i>enjoy</i> <i>forget</i> <i>contain</i> <i>provide</i> <i>today</i>	ən' dʒɔɪ fə' get kən' teɪn prə' vaɪd tə' deɪ	' ə:ndʒɔɪ ' fə: get ' kə:nteɪn ' prə:vaɪd ' tə:deɪ	' ɛndʒɔɪ ' fɔ: get ' kɒnteɪn ' prəʊvaɪd ' tu:deɪ
No	<i>follow</i> <i>coffee</i> <i>water</i> <i>listen</i> <i>money</i> <i>notice</i>	'fɒləʊ 'kɒfɪ 'wɔ:tə 'lɪsən 'mʌni 'nəʊtɪs	fɒ' ləʊ kɒ' fɪ: wɔ:' tə: lɪ' sən mʌ' ni: nəʊ' tɪ:s		<i>begin</i> <i>decide</i> <i>discuss</i> <i>inform</i> <i>perhaps</i> <i>prevent</i> <i>return</i>	bɪ' gɪn dɪ' saɪd dɪs' kʌs ɪn' fɔ:m pə:' hæps prɪ' vent rɪ' tən	' bɪ: gɪn ' dɪ:saɪd ' dɪ:skʌs ' ɪ:nfɔ:m ' pə:hæps ' prɪ:vent ' rɪ:tən	

¹⁵ The three versions of the word *second* are the only examples given in Field (2005). He writes (p. 406): "Shift of stress with change to full vowel quality [...]; example: [se'kənd]. The field must have meant to write [se'kɒnd], given the existence of the word *secondant* [sə'kɒndənt]. The full vowel in the initial syllable suggests that full vowels were not reduced to schwa when the lexical stress was shifted away from them.

The basic idea of the experiment was to create for each of the 24 words a version with the stress (realized as sentence stress in a citation form) on the wrong syllable. As a result of the stress shift, the pitch movement normally associated with the lexically stressed syllable must shift to the incorrectly stressed syllable, which should be a sufficient (if not overriding) cue for the stress shift (Van Heuven, 2018, and references therein). Unfortunately, the description of the manipulations that were carried out does not allow me to reconstruct exactly what was done. The transcriptions in Table 1 are my best guess as to what the speaker did when asked to shift the stress to the wrong syllable. It would appear that the lexically stressed syllable was always pronounced with a full vowel, even when unstressed – although, presumably, it was shortened somewhat in initial syllables when the stress was shifted to the other syllable. In final syllables, no such shortening is expected due to final lengthening (Van Heuven & Turk, 2020, and references therein). In the words in the upper half of Table 1, the wrong stress was implemented in two different ways. The first was to simply lengthen the incorrectly stressed syllable, in which case a schwa or reduced vowel [ɪ] was lengthened by 50% or more (“shift only”). The second way to shift the stress was to change the vowel quality from reduced [ə] or [ɪ] to a full vowel (which should be longer and was probably additionally lengthened because it was stressed). The quality of the full vowel was derived from the way the syllable or morpheme can be pronounced in English if stressed.

In the lower half of Table 1, the stress shifts were implemented by only lengthening the non-lexically stressed vowel; Field claims that these lexically unstressed vowels are either full when unstressed or are reduced but cannot alternate with a full vowel.¹⁶

Words were offered to three groups of native English listeners and three groups of non-natives in three subsets A-B-C with equal division of types and versions over blocks. Listeners heard each stimulus word in only one version. Listeners were 82 15-year old natives (NL); 28 listened to subset A, 28 to B, and 26 to C. 76 (intermediate level) non-native listeners (same age) from a range of L1 backgrounds (see Figure 4) also listened to either A, B, or C. Subjects were told to expect English words, which they had to write down in ordinary spelling (even if they did not know the word).

The English listeners recognized the words with the correct stress with shifted-only stress and with shift plus vowel quality change in 96, 77, and 88%. This means that lengthening a reduced vowel compromises word recognition more than shifting stress with appropriate substitution of a full vowel for the reduced vowel, even when, in both situations, the pitch change moves to the lexically non-stressed syllable. The results further indicated that a rightward stress shift in an Sw word is more detrimental than a leftward shift in a wS word and that this effect is even greater when the stress is shifted by lengthening the reduced vowel than when the incorrectly stressed syllable is given a full vowel (Table 2).

¹⁶ This latter idea hinges on the assumption that the vowel in *sir* or *first* (which is often transcribed as /ɜ:/ in British English) is just a lengthened schwa /ə:/.

Table 2: Correct word recognition by native English listeners of correct and incorrectly stressed words.

Wrong stress marked by	Sw words			wS words		
	Correct stress	Right-shift	N	Correct stress	Left-shift	N
lengthening only	96	71	12	95	88	12
Substituting full V	98	83	5	97	95	6

Figure 4 presents the word recognition scores for the three types of stressing (correct, wrong stress by lengthening only, wrong stress by substitution of a full vowel) for all groups of listeners, both native and non-native.

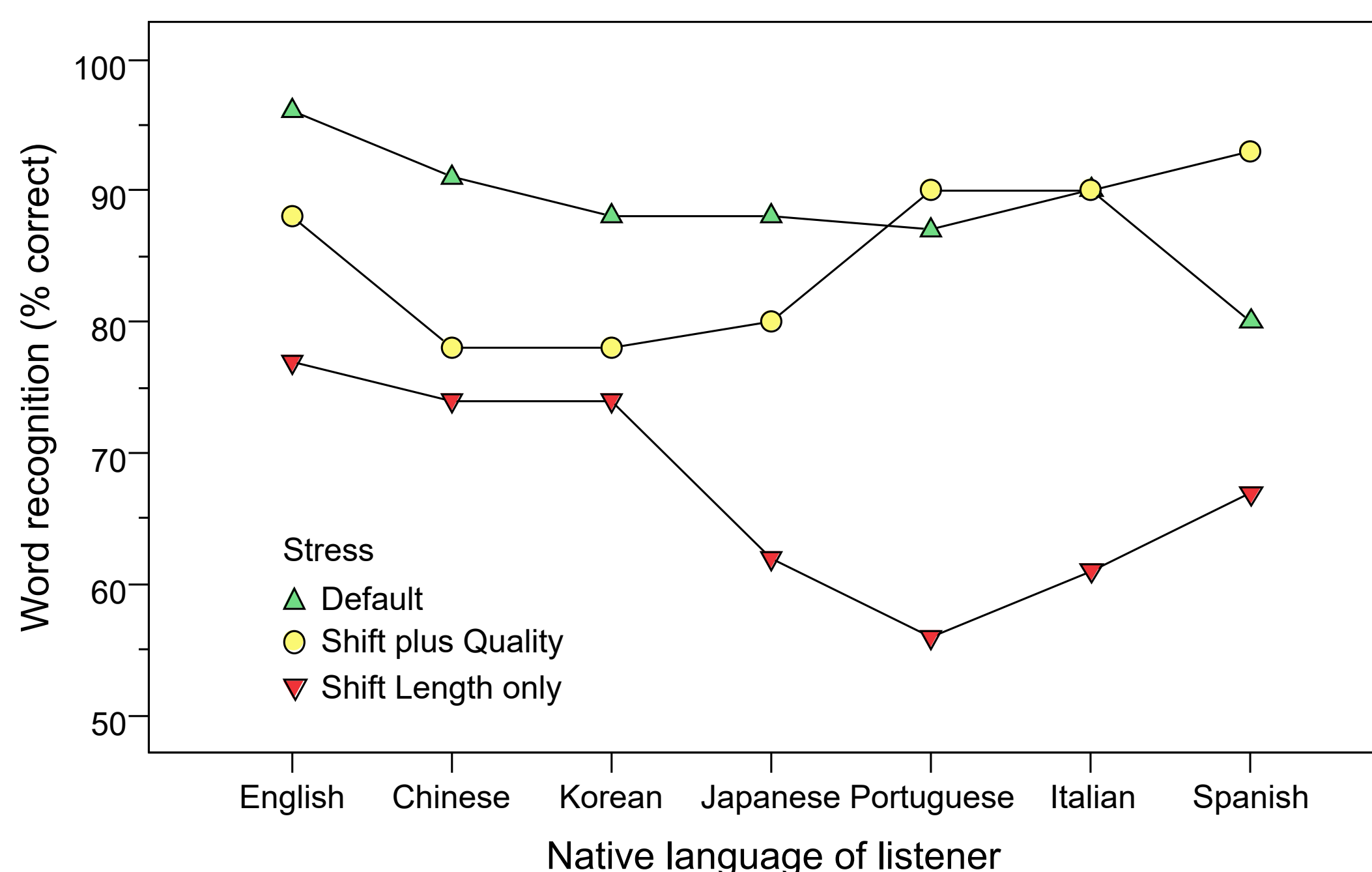


Figure 4: Percent of correct word recognition for seven groups of listeners differing in native language. English two-syllable words were either correctly stressed (Default) or were spoken with stress shifted to the wrong syllable, which retained its original vowel quality (change in “Length only”) or which was changed both in Length and in Vowel Quality (“plus Quality”). (AFTER FIELD, 2005).

The main effects are clear. Correct stress works best, then wrong stress is realized with natural vowel quality change, and worst for wrong stress with change in length only (unnatural sound, impossible in English – vowel has the length of a stressed syllable and a pitch movement, but the vowel quality is not adjusted). Generally, the effects of native language and stress type are additive, but there is one remarkable interaction: listeners with a Romance L1 respond to incorrectly stressed words (with vowel quality adjusted) as if these were correctly stressed. The Romance listeners only suffer if the incorrectly stressed syllable contains a long reduction vowel.

The paper reiterates what was known at the time, i.e., that English words are harder to recognize when they are (sentence) stressed on the wrong syllable and that wrong initial stress is more disruptive than wrong final (or later) stress. This effect, and the asymmetry, were shown earlier by, e.g., Cutler and Clifton (1984) for English and by Van Heuven (1984, 1985) for Dutch (see also next section). What is new is that the same effects were found for non-native listeners (disregarding the remarkable interaction found with the EFL listeners with a Romance L1).

The study shows that correct stress position (when marked by a stressed sentence, i.e., a pitch movement) is important for word recognition, more so for words with initial than with non-initial lexical stress. Not only do native English listeners need to hear the stress on the right syllable, but also non-native listeners have a better chance of recognizing English words if they have the stress on the correct syllable. That said, the paper does not show that correct stress is more important than hearing correct segments. In fact, the paper supports the idea that incorrect segments (lengthened reduction vowels in incorrectly stressed syllables) are especially detrimental to word recognition. Notice, however, that the effect of putting the stress on the wrong syllable is relatively minor, especially when the incorrect stress pattern uses normal English vowels (no unnatural lengthened reduction vowels): the recognition scores go down from 96% to 88% correct, at least in this off-line transcription task. This means that incorrect stress does not block word recognition: the listener is still able to detect and repair the incorrect stress pattern – due to lexical redundancy.

2.4. Summary

To sum up this part of the chapter, we have reviewed some papers which are often cited to support the claim that getting prosody right should be a more important learning goal for non-native speakers of English than getting the vowels and the consonants right. My conclusion at this point is that the evidence is not compelling. The studies suffer from procedural weaknesses and lack necessary control conditions, which would allow the reader to estimate the relative contribution of (incorrect) segmental and suprasegmental properties to the processing of non-native English.

In the next part of the chapter, I will nevertheless assume that correct prosody is an important learning goal for EFL speakers. The question then remains: how can prosody be redundant and relatively unimportant in the communication between native speakers and listeners in English and yet constitute a source of miscommunication when the speaker and/or listener is foreign? In order to understand how this paradox can be solved, we have to know more about the role of prosody when speech has to be recognized under adverse circumstances.

3. Towards solving the paradox

3.1 Information imparted by segments vs. prosody

Given a mean speaking rate of 5 syllables per second and a minimal syllable structure of one consonant followed by one vowel (CV structure), the delivery rate of segments has a lower boundary at approximately 10 segments per second. The delivery rate may go down when syllable types assume greater complexity and/or up when the speaker talks fast. Prosody changes slowly over time, as the minimal prosodic domain is typically the syllable. It is a syllable (rather than a segment) that carries a lexical tone, or that is stressed or unstressed. Prosody, therefore, develops more slowly in time

than segmental information. For this reason, it is commonly understood that segmental information by itself is sufficient to differentiate the word forms in the lexicon. This insight is captured by the orthographies of Western languages, in which the letters of the alphabet exhaustively express the identities of the vowels and consonants (although not always in a straightforward fashion), but prosodic information is largely absent. No information is given in Western orthographies on the location of word stress or sentence accents, nor on the speech melody (note that the question mark expresses an interrogative rather than a high pitch – wh-questions that often end in a low pitch are marked with ‘?’ (e.g., VAN HEUVEN; HAAN, 2000, for Dutch).¹⁷ Although word boundaries are generally expressed in writing (as spaces between words), many phrase boundaries are not marked as such.

In line with this insight, experiments with speech materials from which prosodic information was systematically eliminated have demonstrated that intelligibility was hardly compromised. For instance, Van Heuven (1984) generated polysyllabic Dutch words which should carry stress on either the first, second, or third syllable by concatenating diphones that were excerpted exclusively from syllables spoken with a nuclear sentence accent so that any differences between stressed and unstressed syllables were effectively eliminated from the stimuli. Nevertheless, these unnatural words, presented in a carrier sentence, were correctly recognized by native Dutch listeners in a gating experiment, be it that a longer onset portion of the target words was needed for the listeners to reach the recognition threshold. An informal listening test with a longer Dutch sentence that was entirely composed of fully accented diphones showed excellent intelligibility of all the words in the sentence. However, when segmental information is obliterated, and only prosodic information persists in the stimulus materials, intelligibility is practically reduced to zero. This has been shown for low-pass filtered speech (e.g., MILLER; NICELY, 1955) and for spectrally rotated speech (BLESSER, 1966). An error analysis of the responses collected for such segmentally degraded materials shows nevertheless that rhythmic properties such as the location of prosodic boundaries and stressed or accented syllables were often preserved (BLESSER, 1966, for English; SVENSSON, 1974, for Swedish; VAN HEUVEN, 1984; JONGENBURGER, 1996, for Dutch).

It has also been shown that prosodic information is highly resistant to noise, much more so than segmental information. When speech is presented in noise, segmental errors abound, but the prosodic structure is preserved in the error responses (e.g., SVENSSON, 1974, for Swedish; VAN HEUVEN, 1984, for Dutch). The reasons for the robustness of prosody are manifold. For one thing, important prosodic information, such as speech melody, syllable duration, and differences in loudness, are redundantly coded throughout the spectrum and are very strongly represented in the

¹⁷ Spanish orthography is an exception to this general practice. Not only does it mark questions and exclamations both at the end and at the beginning of a written sentence (where the punctuation mark at the beginning of the sentence is printed upside down), but it also explicitly marks the position of unpredictable word stresses by an acute accent mark over the vowel in the stressed syllable. No other European spellings have adopted this practice, nor does any European spelling attempt to code lexical word tone – even though lexical tones are distinctive properties of Norwegian, Swedish, Serbo-Croatian, Welsh, Lithuanian, and several more languages.

fundamental frequency. When speech is presented in noise, the low-frequency portion of the signal that contains the first few harmonics is not fully obliterated by the noise and therefore continues to weigh in, whereas the less intense higher frequencies that signal the segmental contrasts are made inaudible. For another thing, since prosody develops slowly in time, its time course is highly predictable, much more so than the segmental changes, so that the listener may easily keep track of prosodic features even if he occasionally misses out (see also CUTLER; MCQUEEN, 2014). This so-called prosodic continuity is also one of the reasons why we are able to successfully attend to one speaker amidst a crowd of competing speakers (so-called cocktail-party effect, e.g., BROKX; NOOTEBOOM, 1982).

The role of word prosody may well be different in languages that use prosody to signal lexical distinctions at the word level. The hallmark of stress is that only one syllable in a word can be stronger than all other syllables. This is the so-called culminative defining property (TRUBETSKOY, 1939; GARDE, 1968) that distinguishes word stress from the lexical tone. Hearing stress then allows the listener to predict that the next syllable will typically be unstressed (and verify the hypothesis that the preceding syllable was also unstressed). This is a direct consequence of the alternating stress pattern that is often characteristic of stress languages.¹⁸ Although stress in languages such as English and Dutch can be used to differentiate between otherwise identical segment strings, such as *trusty* ‘trustworthy’ (with initial stress) versus *trustee* ‘board member’ (with final stress), the number of such minimal stress pairs is limited – at least in Germanic languages. Therefore, it seems unlikely that the primary function of stress is in differentiating between words (CUTLER, 1986; FIELD, 2005).¹⁹

In a prototypical tone language, every syllable in a word may carry at least two different tones, e.g., a high one (H) or a low tone (L). Disyllabic words in such a language can then be of four prosodic types, i.e., HH, HL, LH, and LL. Crucially, it is not the case that either the H or the L tone is associated with greater strength or prosodic prominence: all the syllables carry equal weight. The primary function of lexical tone would appear to be to help differentiate between words in the lexicon, in much the same way as is done by vowels and consonants.

There seems to be a division of work between prosodic and segmental properties in speech communication. The fast-changing spectral properties are capable of differentiating between many vowels and consonants but are vulnerable to background noise and distortion. The more

18 The alternating stress pattern is formally accounted for in metrical phonological theory by the Obligatory Contour Principle (OCP), which is a constraint that disallows (or disfavors) immediate succession of the same structural property, such as the immediate succession of two strong syllables. Very often, if one word ends with a stressed syllable and the next word begins with one, the language resolves such a ‘stress clash’ by inverting the stress pattern in one of the words or by deleting one of the two stresses (LIBERMAN; PRINCE, 1977; HORNE, 1990; GUSSENHOVEN, 1991, for English; VAN HEUVEN, 1987; VAN HEUVEN; VAN DONZEL, 1995, for Dutch).

19 This is true, *a fortiori*, of languages with fixed stress, such as Finno-Ugric languages with fixed initial stress or French with fixed final stress. The contrastive function of stress may be more important in languages such as Russian and Greek, for which stress is largely unpredictable so that they contain a great many minimal stress pairs (e.g., JOURAVLEV; LUPKER, 2015, for Russian; TZAKOSTA, 2004, for Greek).

slowly varying prosodic properties (e.g., pitch and duration) contribute less to the differentiation of segments but are highly robust against noise and distortion. As a result, the relative importance of segmental and prosodic properties for speech intelligibility will vary depending on the communicative circumstances. If segments are poorly defined, e.g., due to noise in the communication channel (noise, electronic distortion, computer speech, or foreign accent), the importance of prosody will increase.

In the next few subsections, I will present some experiments I did in the 1980s, which illustrate the point I am trying to make. The experiments were done with Dutch materials and listeners, but I expect similar results would be obtained for other West-Germanic languages, including English.

3.2 Van Heuven (1984)

One male native speaker produced 26 Dutch words varying in length between two and five syllables, with lexical stress systematically equally divided over the maximally five positions. The target words were produced in two different sentences such that they were pronounced with sentence stress (i.e., in focus) on one occasion and without sentence stress (i.e., out of focus) on the other. The words were excerpted from their spoken context and presented for recognition in a gating task, either with or without ANSI noise added.²⁰ In the gating task, the listener first heard the initial CV; on each successive presentation, the audible portion of the target word was incremented with the next phoneme, until, at the last pass, the entire word was audible. After each pass ('gate'), listeners wrote down their best guess about the word they eventually heard. The target words had been chosen such that they could be uniquely distinguished from all other words in the Dutch lexicon based on roughly the first half of the word shape, so that (under good listening conditions) word recognition could take place before the end of the word had been made available. The materials were blocked over four matched groups of Dutch listeners, with the versions with and without sentence stress and with and without added noise, blocked over listeners so that a listener would hear only one version of each word. The results can be seen in Figure 5.

²⁰ ANSI (American National Standards Institute) noise roughly covers the long-term average speech spectrum, so it is an effective masker of speech even when used with modest signal-to-noise ratios.

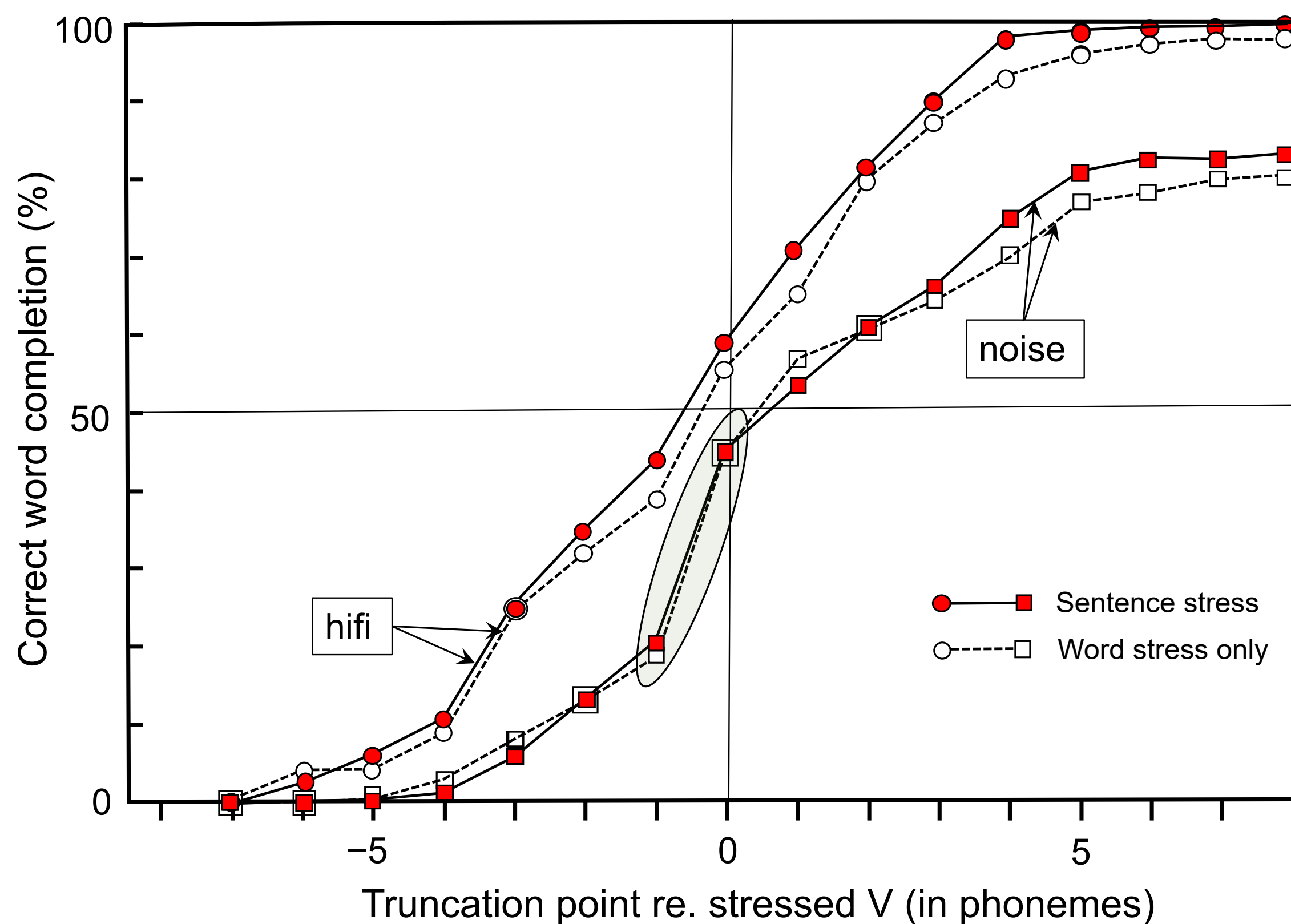


Figure 5: Cumulative correct word completion (%) in a gating task with 26 Dutch words differing in length and stress position. Words were produced with (+accent) or without (-accent) sentence stress and were presented either in quiet (hifi) or in noise. Each division along the X-axis represents the addition of one phoneme in the gating task. Words are lined up on the location of the stressed vowel (0 gate). Phonemes preceding the stressed vowel have negative gate numbers, while those following the stressed vowel have positive numbers.

Figure 5 shows that all target words presented in quiet were correctly completed (recognized) after the fourth phoneme following the stressed vowel if the word had been pronounced with sentence stress. When realized without sentence stress (so that the stressed vowel was not marked by a pitch movement but by length and loudness only) the completion scores are only marginally lower. Importantly, the sigmoid recognition curve shows steady increments over the range between -4 and +4 added segments around the stressed syllable. When the same targets were presented in noise, however, word recognition was delayed severely (curve rises slowly and later) and never exceeded 80% correct completion (where targets with sentence stress do slightly better than those without). Crucially, the cumulative completion score shows a quantum leap at the moment the stressed vowel is made audible (marked in Figure 5 by the green ellipse). The stressed syllable, which is longer and louder than the earlier (and later) unstressed syllables, forms an island of certainty in a sea of noise, which the listener jumps to.

The beneficial effect of hearing the stressed syllable can be observed (even) more clearly if we separate the results of the experiment into those that were obtained for words with early lexical stress positions against those with late stress positions. Early-stress words have the position of the stressed syllable at or before the lexical uniqueness point. In late-stress words, the stressed vowel occurs after the uniqueness point. In the English word *elephant*, for instance, the lexical uniqueness

point is at the /f/ in /'ɛləf/; there are no other words in the English lexicon that begin with /'ɛləf/ than *elephant* (and its derivations). The initial stress in this word precedes the uniqueness point, so that *elephant* qualifies as an early-stress word.

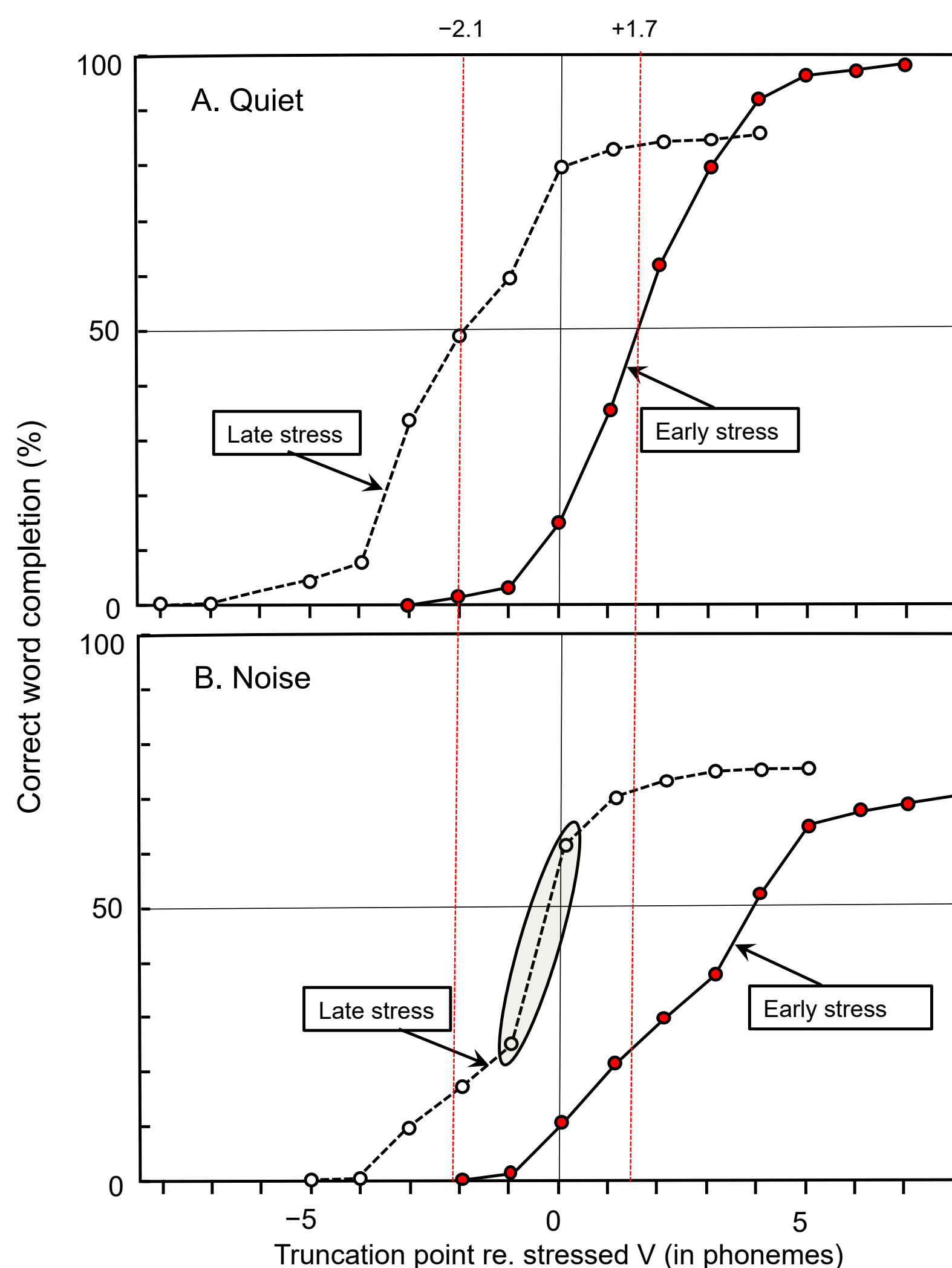


Figure 6: Cumulative correct word completion (%) as a function of the truncation point in a gating task. The upper panel shows the results for words presented in quiet, and the lower panel represents the same words presented in noise. In each panel, words with early and late stress (stress before vs. after the lexical uniqueness point) are plotted separately. The (mean) lexical uniqueness points are marked by a vertical line, at -2.1 phoneme for late stress and +1.7 phoneme for early stress.

When presented in quiet, successful word completion increments steadily as before in Figure 5. When presented in noise, the increments are roughly equal-sized as well, at least for words with early stress. However, when the stressed vowel occurs after the lexical uniqueness point, we see the quantum leap in correct completions (from 26 to 62% correct) at the moment the stressed vowel is made audible (marked by the green ellipse).

We may interpret the results of this experiment as follows. Hearing the stressed syllable (or the stressed vowel within it) provides no special information (relative to other syllables or segments) when word recognition takes place under favorable listening conditions (in quiet). When words have to be recognized under adverse listening conditions, as in noise, hearing the stressed syllable suddenly discloses a disproportionate amount of information as to the identity of the time-developing word. Earlier information falls into place once the listener realizes that the lexical stress is somewhere

later in the word, i.e., not in the “default” word-initial position. However, when the lexical stress is the early default (in Germanic languages typically on the first syllable), hearing the stress provides no information the listener did not already expect.

It is important to realize that the special contribution of stress to word recognition (in Dutch, English, and other Germanic languages resides with the non-default stresses, i.e., the non-initial stresses. Moreover, this special contribution of the non-default stress comes to the surface only in adverse listening conditions, when a redundant property that can normally be reconstructed from the sequence of segments that makes up the word is the only reliable clue left.

In a second experiment in the 1984 paper, I used trisyllabic CVCVCVC words only, ten with initial lexical stress, ten with medial stress, and another ten with final stress. The words were synthesized from diphones, i.e., speech synthesis building blocks consisting of roughly the second half of one phoneme followed seamlessly by the first half of the next phoneme. The diphones I used were excerpted from hyperarticulated stressed syllables in nonsense words with an H*L pitch configuration (i.e., sentence stress). The diphones were monotonized, and intensity normalized. As a result of this synthesis scheme, all syllables and segments within them were equally suggestive of stress – so that stress was not prosodically marked. In a second condition, I added prosody to the diphone words in a controlled manner by generating a fixed rise-fall pitch configuration on the stressed syllable and reducing the duration of the unstressed syllables by 20 percent. The 2 x 30 words were blocked and presented in a gating task to two groups of Dutch listeners in such a way that each listener heard every word once, with 15 words with and 15 words without explicit prosodic marking of stress. On successive passes, the audible portion of the words was incremented by one added diphone until, after the final pass, the words could be heard in their entirety. Figure 7 shows the results.

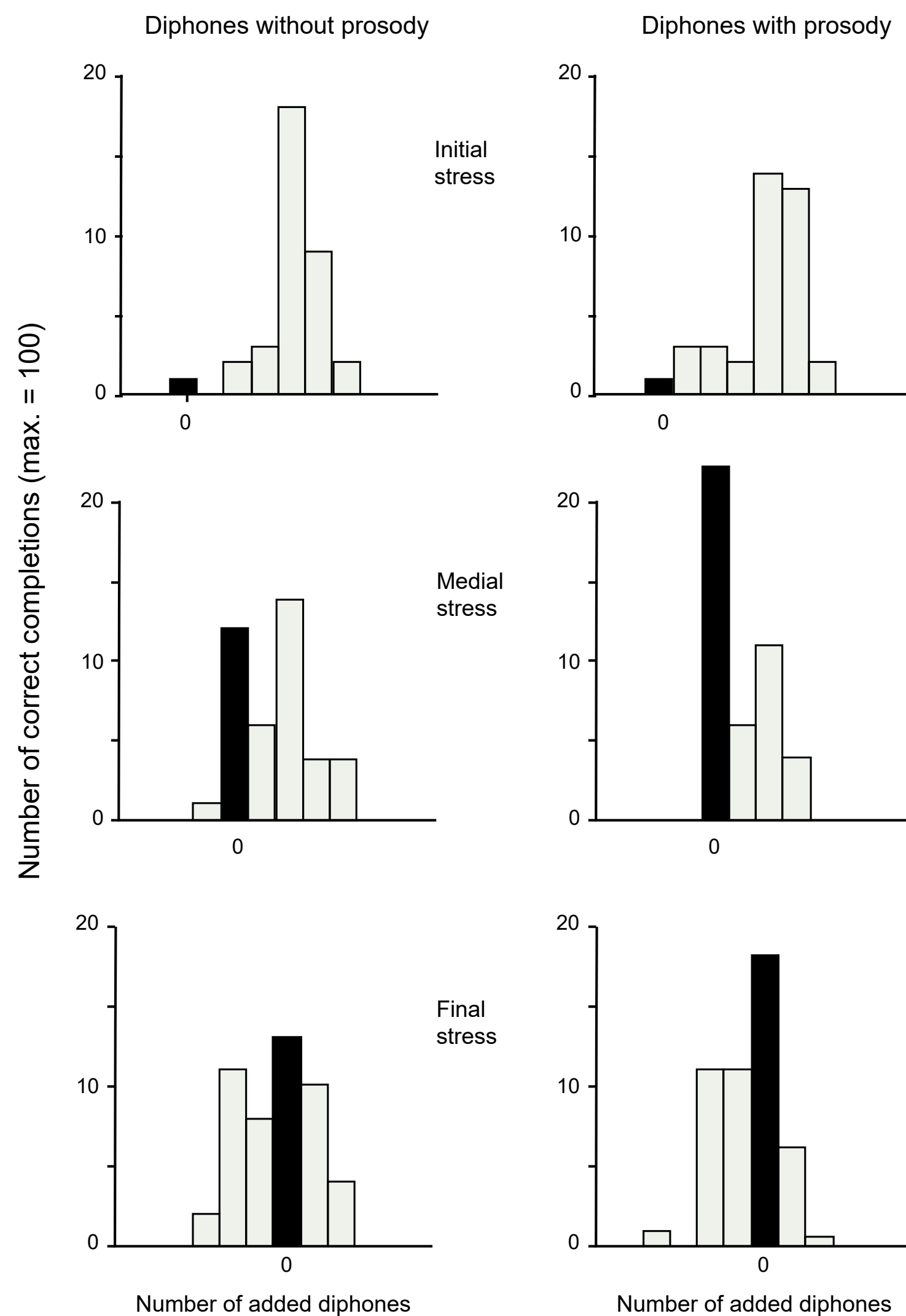


Figure 7: The number of correctly completed words as a function of the length of the diphone sequence made audible, broken down by type of speech (diphone synthesis with or without prosody) and by lexical stress position.

The results show no special role for hearing the stressed CV diphone in the initial (default) position. However, when the stressed CV diphone occurred in the medial or (even more so) in the final syllable, we see a disproportionate increment in correct completions at the point in time where the stressed CV diphone was made audible. This is seen when the stress was not prosodically marked (left column in Figure 7), and it shows up much more clearly when the non-initial stress is prosodically marked by the temporal organization and – especially – the H*L pitch configuration marking the sentence stress.

Again, we have seen that the Dutch listener assumes that words begin with a stressed syllable. However, as soon as the listener perceives a (sentence) stress in a non-default position (here medial, and even more strongly, final position), word recognition is completed. It is essential to realize that this special role of the non-initial stress (in Dutch) comes to the surface only when words have to be recognized in adverse listening conditions – such as (here) poor quality diphone synthesis.

3.3 Van Heuven (1985)

In the experiments described above, the prosodic marking of stress position was either weakened (word stress instead of sentence stress) or omitted altogether. A different way of investigating the importance of stress for word recognition is to deliberately produce stress on the wrong syllable. The more disruptive such misplaced stress in the word recognition process is, the more important stress must be. If an incorrectly placed stress goes unnoticed (as in Indonesian, see Van Zanten & Van Heuven, 1998, 2004), the conclusion would follow that stress plays no role in word recognition.

In this experiment, I synthesized 24 trisyllabic Dutch words from low-quality diphones. Eight words had lexical stress on the first syllable, eight on the medial, and eight on the final syllable. Each word was then synthesized in three different versions at the end of a fixed carrier meaning 'The next word is ...', with a prosodic marking (by an H*L pitch configuration) on the first, second or third syllable of the target word. For each word, then, there was one correctly stressed version and two incorrect versions. The materials were divided into three equal parts using a split-plot design (only one version of each target word would be heard by a listener) and presented to twelve Dutch listeners in a speeded repetition task. The participant was asked to repeat the word he had just heard while correcting possible erroneous stress. The correctness of the response was determined, as well as the time it took for the participant to start producing his response – measured from the offset of the stimulus. Typically, the result was that the participant recognized the word and corrected it on the fly or – alternatively – did not recognize the word and remained silent. The results are seen in Table 3.

When correctly stressed, the target words were repeated with a success rate between 63 and 88%, depending on the position of the lexical stress. This shows that the segmental quality of the diphone synthesis was poor. However, when the stress was shifted to a wrong syllable, the chances of recognizing the word were roughly halved. Correctly stressed targets were repeated successfully with latencies about 120 to 300 ms shorter than incorrectly stressed targets. As was seen in my earlier experiments, incorrectly stressing an initial syllable was clearly more detrimental than hearing wrong stress on a later syllable.

These results demonstrate, once more, that word recognition (in Dutch) is harmed especially when an unstressed initial syllable is incorrectly stressed. This suggests that Dutch listeners (and English listeners as well) have a default recognition strategy that assumes that words begin with a stressed syllable (word-initial stress bias). When they hear stress on an initial syllable, their expectation is confirmed so that words beginning with an unstressed syllable are no longer eligible candidates for recognition. Hearing incorrect stress on a later syllable will be less hurtful because the listener is already searching in the unmarked part of the lexicon with initially-stressed words.

Table 3: Correctly repeated words (upper panel) broken down by lexical and realized stress position. Correctly stressed words are in green cells along the main diagonal. Right-shifting of initial lexical stress (yellow cells) is less disruptive than (left-)shifting non-initial lexical stress (red cells). The bottom panel displays the repetition latency (in ms) for those words that were correctly repeated.

Lexical stress on syllable #	Stress synthesized on syllable #		
	1	2	3
1	66%	44%	56%
2	34%	81%	31%
3	34%	25%	63%

1	1500	1800	1650
2	1630	1510	1640
3	1700	1690	1390

The overall conclusion of this experiment, and the earlier experiments, is that correct stress (position) is an important prerequisite for successful word recognition when the segmental quality of the stimuli is poor, as in low-quality speech synthesis and non-native speech.

3.4 Wang *et al.* (2011)

I end this survey with a study on a different type of word-prosodic system, i.e., lexical tone rather than stress. Mandarin Chinese has four lexical tones by which identical strings of segments can still be differentiated as separate words. For instance, the syllable /ma/ means ‘mother’ when spoken with tone 1 (high level), ‘hemp’ with tone 2 (mid-rising), ‘horse’ with tone 3 (low dipping), and ‘scold’ with tone 4 (high-falling).

We knew from earlier studies that Mandarin speech can readily be understood when information about the lexical tones is eliminated from the utterance. For instance, Xin (2007) synthesized Mandarin utterances from prosodically neutral Dutch diphones and replaced the pitch contours on all content words with simple rise-fall pitch movements (which are not part of the Mandarin tone inventory). In comparison with the version in which the original Mandarin tones were left in place, intelligibility for native Mandarin listeners was reduced by merely 4 percent.

In Wang *et al.* (2011), we aimed to test the hypothesis that the contribution of prosody to word recognition is not fixed but assumes a more important role as the segmental quality deteriorates. Under normal listening circumstances, segments are most important, while prosody by itself does not afford word recognition. However, prosody will contribute substantially when segmental quality deteriorates.

We carried out an intelligibility test with words in meaningful sentences, using the Speech-Perception in-Noise test (SPIN test, Kalikow et al., 1977), which was adapted for Mandarin (and 14 other Chinese languages/dialects) by Tang & Van Heuven (2009). The test comprised 60 short, everyday sentences. After hearing a sentence, the listener has to write down the last word of the sentence (which is made rather predictable by the preceding context), as in *Keep your broken arm in a sling* (where *sling* is the predictable target word). The materials (see Tag, 2009 for details) were spoken by a male and a female speaker of Mandarin. We created six conditions by varying the segmental quality and the prosody. Segmental quality was varied in three steps: High quality (HQ, unfiltered), Medium Quality (MQ, low-pass filtered speech with a cut-off at 1000 Hz), and Low Quality (LQ, low-pass filtered with a 300-Hz cut-off frequency). Prosody was varied by leaving the original speech melody in place or monotonizing the speech by PSOLA analysis-resynthesis at a fixed pitch of 100 Hz for the male speaker and 200 Hz for the female.

The materials were blocked into six groups of Mandarin listeners (24 in total) in a Latin Square design, such that a listener heard each sentence only once and yet received an equal number of sentences in each of the six conditions. The results are shown in Figure 8.

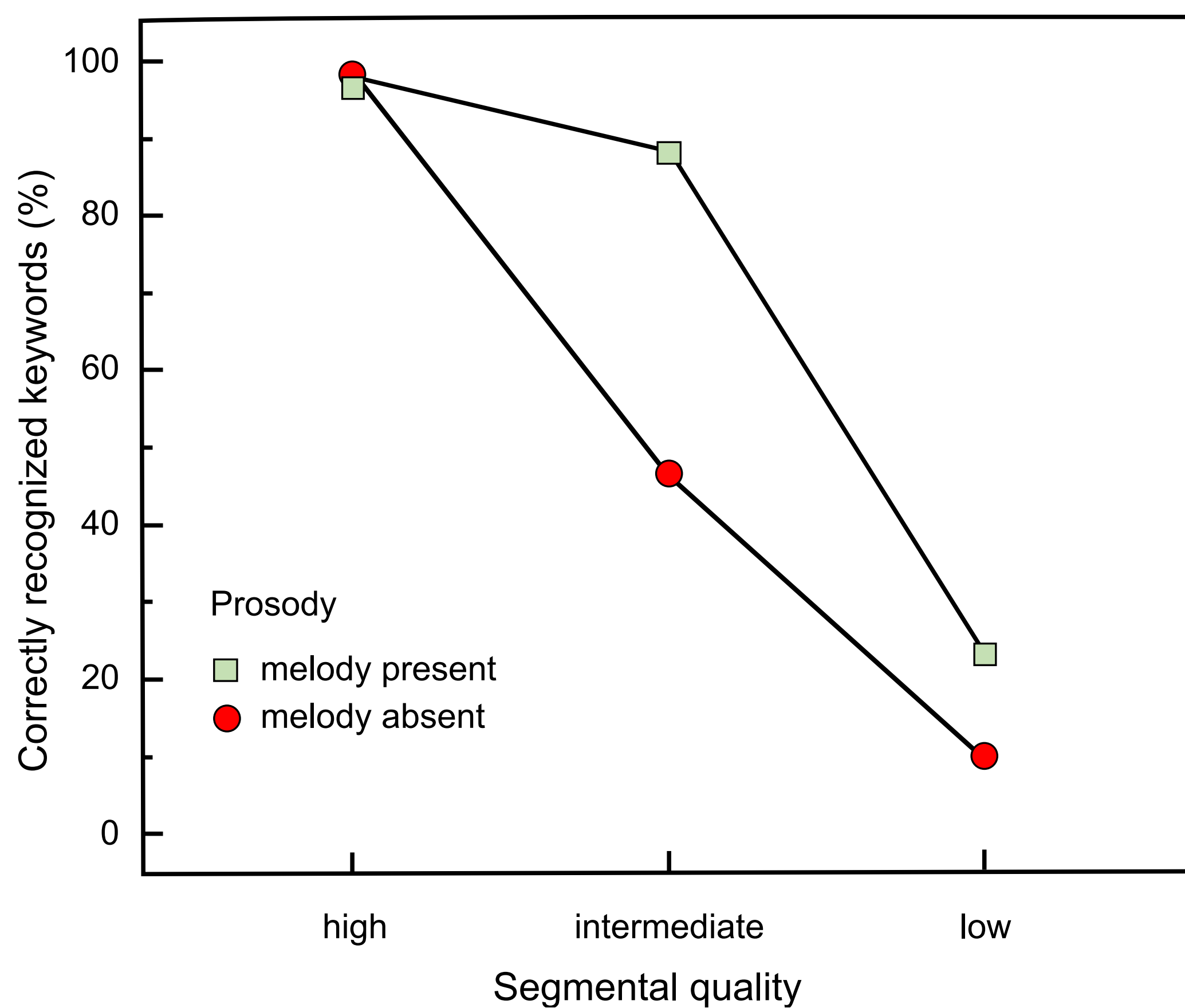


Figure 8: Correctly recognized keywords (%) in the Mandarin SPIN test as a function of segmental quality (high, medium, poor) and presence vs. absence of speech melody.

When the segmental quality is good, the intelligibility is close to the ceiling, whether the melody is present or absent. When the segmental quality is low, hardly any intelligibility remains. This finding mimics earlier results reported by Zhang et al. (1981). The important and novel part of the results is intelligibility for the medium-quality materials depending on the availability of prosody. When the

lexical tones are left in place, the intelligibility of the sentences remains very high (better than 90%). However, when the pitch is replaced by a fixed monotone, intelligibility is reduced to less than 50%.²¹

These results lend strong support to the idea that the contribution of prosody to speech intelligibility is not fixed but varies with the quality of the segmental information. Segmental quality is of overriding importance. If the segmental quality is poor, no word recognition is possible. If the segmental quality is perfect, there is no added value for tone. But tone really helps when the segmental quality is intermediate (and even insufficient by itself).

4. Why is incorrect stress so harmful?

What we have seen so far is that it would be incorrect to say in general that prosody is more important for speech intelligibility than segments (or vice versa). Segments take precedence over prosody when they are well defined. Conversely, prosody assumes greater importance as segmental information deteriorates. These basic effects were found both for stress and tone languages, which means we may have a language-universal principle here.

The probable cause for the success of word prosody in poor listening conditions is its robustness. The most important differences between segments are in the spectral distribution. The vowels of a language very often differ only in their spectral composition, i.e., the location and dynamic change (in the case of diphthongs and semivowels) of the formants, i.e., relatively narrow frequency bands that are amplified through resonance in the vocal tract, as determined by the degree of jaw opening (corresponding with the center frequency of the lowest resonance), the location of the constriction place (front-back dimension, corresponding with the second-lowest resonance, the degree of lip rounding, etc., see for instance LADEFOGED; DISNER, 2012 for details). Some languages also use duration to differentiate between long and short vowels. Since vowels generally have (much) more intensity than the surrounding consonants, it is not difficult for a listener to roughly determine the length of the loud vowel sound, even amid background noise and competing voices. Spectral differences (in the noise bursts) are also the primary cues to distinguish the place of articulation for stops, fricatives, and affricates, i.e., the obstruents or true consonants.

In poor listening conditions, e.g., when there is a lot of background noise, when there are competing voices (as when talking in a crowd of people), or through degraded speech quality due to filtering and harmonic distortion in tele-communication systems, the spectral information is practically obliterated, so that the identification of the consonants, and especially of the vowels, becomes largely a matter of guesswork.

²¹ In the experiment, the absence of tone was operationalized by monotonizing the speech melody, which is indeed the strongest correlate of tone. However, the tones of Mandarin are distinguished to some extent by other acoustic properties than pitch. For instance, tone 3 is about 50% longer than tones 2 and 4, whereas tone 4 is shorter. Also, the intensity contour of the syllables will impart information on the tone. The effect of the absence of tone would have been stronger still if we had eliminated all tone information.

Acoustic information about the sentence melody is probably the most important and noise-resistant aspect of speech prosody. The melody is carried by the repetition rate of the vocal fold vibration, which generates a complex harmonic source signal. Throughout the sentence, the pitch of this source signal goes up and down, marking boundaries and sentence stresses. The information about the pitch is omnipresent in the speech signal. The pitch can be determined in the time domain by estimating the time lapse between successive glottal pulses, in which events are loud enough to resist even severe background noise. The pitch can also be determined directly by locating the fundamental frequency of the glottal wave, but even when the fundamental frequency is absent from the signal (due to filtering, as in telephone speech), the pitch can easily be reconstructed by estimating the distance between adjacent harmonics somewhere in the spectrum (see LADEFOGED; JOHNSON, 2011 for more background information). Other prosodic properties that are relatively resistant to noise and distortion are duration and intensity differences, as was noted above.

Listeners know from experience that, under adverse listening circumstances, prosodic information is generally a more reliable source of information than segmental/spectral information. They also know that native speakers of their language make proper use of prosody by marking phrase and utterance boundaries, and stressing the correct syllables at the word level and the appropriate words at the sentence level. Therefore, it is reasonable to assume that when the communicative situation deteriorates, the listener will rely strongly on the prosodic information that survives and will reinterpret the available spectral information to come up with plausible word recognition candidates.

Listening to foreign-accented speech is a form of listening to speech in which the spectral information is severely distorted. Non-native speakers typically use the nearest available vowel or consonant in their native-language sound inventory as a substitute for the vowels and consonants in the foreign language (this is what we call a foreign accent). As a result of this substitution process, the native listener is never certain of which vowel or consonant the speaker intended, so segmental identification becomes a matter of guesswork, of a type similar to what we find when listening to speech in noise. In such circumstances, the listener will primarily rely on prosodic information, which then sets the constraints within which the available segmental information is optimally fitted (or “reinterpreted” to fit the prosody). This strategy works well when the speaker uses the same language as the listener, but it backfires when the speaker’s language (or foreign accent) does not match the listener’s language.

As a nice example of this process, Bansal (1966) showed that native listeners of British English reinterpret segments to fit the incorrectly perceived rhythmic pattern in English pronounced by Indian speakers. To the British-English ear, the stress in the words *character* and *written* is perceived on the second syllable when pronounced by an Indian speaker of English. Given that the segmental

pronunciation of Indian English strongly deviates from the British norm, it makes sense for the British listener to reinterpret the information as *director* and *retain*, respectively (errors reported by BANSAL, see also CUTLER, 1984).

Similarly, at the sentence level, removing prosodic information from utterances is largely inconsequential. However, when a conflict is created between the syntactic-semantic structure of the sentence and the prosody, listeners tend to attach more weight to the prosodic information if the speech quality is poor (e.g., WINGFIELD, 1975).

Van Heuven and Hagman (1988) cross-tabulated the number of (monomorphemic) Dutch words in terms of word length (expressed in the number of syllables) and position of the primary stress. This allowed us to subdivide the Dutch lexicon into 27 rhythmic types.

Table 4: Size of sublexical defined by rhythmic pattern in Dutch monomorphemic words (after VAN HEUVEN; HAGMAN, 1988). Rhythmic word patterns are defined by word length in syllables and the position of the primary stress.

Word length	Position of primary stress in (monomorphemic) word						Total
	1	2	3	4	5	6	
1	4,284						4,284
2	2,703	1,682					4,385
3	408	808	1,032				2,248
4	56	128	474	314			972
5	7	0	43	76	52		178
6	0	0	0	6	9	5	20
>6	0	0	0	0	2	0	2
Total	7,458	2,618	1,549	396	63	5	12,089

The counts show that the majority of Dutch words have stress on the first (or only) syllable. This will prompt the stress-initial bias by which listeners assume that hearing stress signals the beginning of a new word. The data also show that stress is always on one of the three last syllables of the word (the universal three-syllable window identified by HAYES, 1995).²² Obviously, if the listener knows the rhythmic type of the word he has just heard (or is in the process of hearing), the lexical search space is much smaller than the 12,000-item lexicon as a whole, so that words can still be recognized from increasingly less articulated and less clear segments. A profitable strategy for a Dutch listener would be to decide, as early as possible, whether the word begins with a stressed syllable or not. Given that the majority of Dutch words have initial stress, the listener will use the presence of a pitch movement on the initial syllable as clear confirmation that he should look for

²² The 63 cases of initial stress in 4- and 5-syllable words are exceptions that can be explained on historical grounds (these words used to be compounds but are no longer felt to be complex).

words that begin with a stressed syllable. An incorrect (sentence) stress on the beginning of a Dutch word (and English word as well) blocks access to the roughly 4,500 recognition candidates that begin with an unstressed syllable and will be extremely detrimental to successful recognition if the target word has non-initial lexical stress (as was shown in section 3.3). On the other hand, not hearing the sentence stress on the first syllable of a word provides no compelling negative information; it is always possible that the target word is communicatively unimportant so that the speaker omits the sentence stress from it. Access to the initially stressed part of the lexicon will not be blocked until the listener hears the sentence stress somewhere later in the word. In poor-quality speech, as in non-native speech, word recognition is generally delayed (as was shown in Figures 5 and 6 above). Often, it will be the case that the end of the word has been received by the listener, while word recognition is still in progress. In such cases, knowing whether the word stress is initial or not is a very useful cue to limit the search space or to check whether a provisionally recognized word fits the stress pattern that is part of its lexical representation.

We know that hearing the difference between stressed and unstressed word beginnings is used by Dutch listeners even in the early stages of the word recognition process, in the way described in the preceding paragraph (see VAN HEUVEN, 1988, for data in an online shadowing task). English listeners are not sensitive to the difference between stressed and unstressed full vowels (or syllables). They only consider an initial syllable unstressed if the vowel is completely reduced to schwa or [ɪ] (as in Table 1). This is why both recognition candidates 'trusty 'trustworthy' and *trus* 'tee 'board member' (or 'forebear 'ancestor' vs. *for* 'bear 'tolerate, endure') are compatible with hearing an initial stressed as well as an unstressed first syllable for English listeners (see CUTLER, 1986: Forbear is a homophone). After the initial stage of the word recognition process (so-called lexical access phase), however, the difference in initial and non-initial stress is also used profitably in English to distinguish alternatives (see CUTLER, 2005, for a comprehensive review of the difference in the processing of stress information in English and Dutch).

What was said here about the role of lexical stress in languages like English and Dutch (languages with a variety of stress patterns) should also apply to the role of lexical tone, as in Mandarin (and many other South-East Asian languages). Here, too, the information on the lexical tone will be highly noise-resistant and will take priority over segmental information in low-quality speech (or speech with a non-native accent).

6. Conclusion

In this chapter, I showed that, generally, prosody can be omitted entirely from the spoken text (as indeed it is in the orthographies of most languages) without causing any serious communication problem. Prosody is largely redundant relative to the information imparted by the vowels and

consonants that make up spoken sentences. Yet, foreign-language pedagogists insist that getting the prosody right in the target language is essential for the learner's intelligibility and comprehensibility and should accordingly be given the highest priority as a learning goal, higher than correctly pronouncing the vowels and the consonants. This seems a paradox.

The paradox can be resolved if we realize that prosody is redundant in high-quality native speech and can therefore be omitted with impunity. In fact, the constraining power of reliable, high-quality segmental information is so strong that prosodic errors can be detected and ignored by the native listener (although some processing delay may be incurred). However, in poor-quality speech, for instance, in noise, amid competing voices or over a defective telecommunication channel, or if the speaker is a non-native with a foreign accent, the native listener relies on prosody and reinterprets poorly defined segmental information to fit the prosodic pattern that was perceived. This processing strategy is formed by the listener's experience with the difference in reliability between segmental information and prosody. Whereas segmental information is easily destroyed and highly vulnerable to noise and acoustic distortion, prosodic information is robust and resistant to noise. Listening to non-native speech means listening to poorly defined segments. In such cases, prosody takes precedence over segmental information.

This leaves two avenues open in foreign-language teaching. One solution would be to give top priority to improving the learner's vowels and consonants so that these assume native-like quality. In that case, the prosody (stress, intonation) can remain defective without much communicative harm. The alternative solution would be to give priority to getting the prosody right. If at least the prosody is right, a modicum of segmental deviance can be tolerated by the listener, and communication will not break down. The second alternative is probably cheaper (requires less time and effort), which would explain the advice given by the foreign-language pedagogists.

References

ANDERSON-HSIEH, J.; JOHNSON, R.; KOEHLER, K. The relationship between native speaker judgments of nonnative pronunciation and deviance in segmentals, prosody, and syllable structure. *Language Learning*, v. 42, p. 529-555, 1992.

BANSAL, R. K. *The intelligibility of Indian English*. Ph.D. dissertation, London University, 1966.

BENOÎT, C.; GRICE, M.; HAZAN, V. The SUS test: A method for the assessment of text-to-speech synthesis intelligibility using semantically unpredictable sentences. *Speech Communication*, v. 18, p. 381-392, 1996.

BLESSER, B. *Perception of spectrally rotated speech*. PhD dissertation, Massachusetts Institute of Technology, 1966.

BROKX, J. P. L.; NOOTEBOOM, S. G. Intonation and the perceptual separation of simultaneous voices. *Journal of Phonetics*, v. 10, p. 23-36, 1982.

COMRIE, B.; DRYER, M. S.; HASPELMATH, M.; GIL, D. (ed.). *World Atlas of Language Structures*. Oxford: Oxford University Press, 2005.

CUTLER, A. Stress and accent in language production and understanding. In: GIBBON, D.; RICHTER, H. (ed.). *Intonation, accent and rhythm, studies in discourse phonology*. Berlin: De Gruyter, 1984. p. 77-90.

CUTLER, A. Forbear is a homophone: Lexical stress does not constrain lexical access. *Language and Speech*, v. 29, p. 201-220, 1986.

CUTLER, A. Lexical stress. In: PISONI, D. B.; REMEZ, R. E. (ed.). *The handbook of speech perception*. Oxford: Blackwell, 2005. p. 264-289.

CUTLER, A. *Native listening: Language experience and the recognition of spoken words*. Cambridge, MA: MIT Press, 2012.

CUTLER, A.; CLIFTON, C. E. The use of prosodic information in word recognition. In: BOUMA, H.; BOUWHUIS, D. G. (ed.). *Attention and Performance X*. Hillsdale, NJ: Erlbaum, 1984. p. 183-196.

CUTLER, A.; MCQUEEN, J. How prosody is both mandatory and optional. In: PACILLY, J.; SCHILLER, N. O. (ed.). *Above and beyond the segments. Experimental Linguistics and Phonetics*. Amsterdam: John Benjamins, 2014. p. 71-82.

DENES, P. B.; PINSON, E. N. *The Speech Chain. The Physics and Biology of Spoken Language*. Murray Hill, NJ: Bell Telephone Laboratories, 1963.

FIELD, J. Intelligibility and the listener: The role of lexical stress. *TESOL Quarterly*, v. 39, p. 399-423, 2005.

FOWLER, C.; HOUSUM, J. Talkers' signaling of "new" and "old" words in speech and listeners' perception and use of the distinction. *Journal of Memory and Language*, v. 26, p. 489-504, 1987.

GARDE, P. *L'accent*. Paris: Presses Universitaires de France, 1968.

GUSSENHOVEN, C. The English rhythm rule as an accent deletion rule. *Phonology*, v. 8, p. 1-35, 1991.

HAHN, L. D. Primary stress and intelligibility: Research to motivate the teaching of suprasegmentals. *TESOL Quarterly*, v. 38, p. 201-233, 2004.

HAYES, B. *Metrical stress theory: Principles and case studies*. Chicago IL: University of Chicago Press, 1995.

HORNE, M. Empirical evidence for a deletion formulation of the rhythm rule in English. *Linguistics*, v. 28, p. 959-981, 1990.

JONGENBURGER, W. *The role of lexical stress during spoken-word processing*. The Hague: Holland Academic Graphics, 1996.

JOURAVLEV, O.; LUPKER, S. J. Predicting stress patterns in an unpredictable stress language: The use of non-lexical sources of evidence for stress assignment in Russian. *Journal of Cognitive Psychology*, v. 27, p. 944-966, 2015.

KALIKOW, D. N.; STEVENS, K. N.; ELLIOTT, L. L. Development of a test of speech intelligibility in noise using sentence materials with controlled word predictability. *Journal of the Acoustical Society of America*, v. 61, p. 1337-1351, 1977.

KEIJSPER, C. E. Vorm en betekenis in Nederlandse toonhoogtecontouren. *Forum der Letteren*, v. 25, p. 20-37 e 113-126, 1984.

KESKIN, H. K.; ARI, G.; BAŞTUĞ, M. Role of prosodic reading in listening comprehension. *International Journal of Education and Literacy Studies*, v. 7, n. 1, p. 59-65. 2019. DOI: <https://doi.org/10.7575/aiac.ijels.v.7n.3p.59>

LADEFOGED, P.; DISNER, S. F. *Vowels and consonants*. Chichester: Wiley, 2012.

LADEFOGED, P.; JOHNSON, K. *A course in Phonetics*. Boston: Wadsworth, 2011.

LEHISTE, I.; PETERSON, G. E. Vowel amplitude and phonemic stress in American English. *Journal of the Acoustical Society of America*, v. 31, p. 428-435, 1959.

LEHISTE, I.; PETERSON, G. E. Some basic considerations in the analysis of intonation. *Journal of the Acoustical Society of America*, v. 33, p. 419-425, 1961.

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

LIBERMAN, M.; PRINCE, A. On stress and linguistic rhythm. *Linguistic Inquiry*, v. 8, p. 249-336, 1977.

LISKER, L.; ABRAMSON, A. S. A cross-language study of voicing in initial stops: Acoustical measurements. *Word*, v. 20, p. 384-422, 1964. DOI: <https://doi.org/10.1080/00437956.1964.11659830>

- MILLER, G. A.; NICELY, P. E. An analysis of perceptual confusions among some English consonants. *Journal of the Acoustical Society of America*, v. 27, p. 338-352, 1955.
- MUNRO, M. J.; DERWING, T. M. Foreign accent, comprehensibility, and intelligibility in the speech of second language learners. *Language Learning*, v. 45, p. 73-97, 1995.
- MUNRO, M. J.; DERWING, T. M.; MORTON, S. L. The mutual intelligibility of L2 speech. *Studies in Second Language Acquisition*, v. 28, p. 111-131, 2006.
- NOOTEBOOM, S. G. The prosody of speech: Melody and rhythm. In: HARDCASTLE, W.; LAVER, J. (ed.). *The Handbook of Phonetic Sciences*. Oxford: Blackwell, 1997. p. 640-673.
- NOOTEBOOM, S. G.; COHEN, A. *Spreken en verstaan, een inleiding tot de experimentele fonetiek*. Assen: Van Gorcum, 1984.
- PETERSON, G. E.; LEHISTE, I. Duration of syllable nuclei in English. *Journal of the Acoustical Society of America*, v. 32, p. 693-703, 1960.
- QIN, Q. *Understanding foreign-accented language*. Experiments on the relative contribution of segments and prosody in Dutch-accented Mandarin. M.Phil. Thesis, Leiden University Centre for Linguistics, 2007.
- RIETVELD, A. C. M.; VAN HEUVEN, V. J. *Algemene fonetiek [General phonetics]*. 2nd edition. Bussum: Van Gorcum, 2001.
- RIETVELD, A. C. M.; VAN HEUVEN, V. J. *Algemene fonetiek [General phonetics]*. 4th revised edition. Bussum: Van Gorcum, 2016.
- SLIS, I. H.; COHEN, A. On the complex regulating the voiced-voiceless distinction. *Language and Speech*, v. 12, p. 80102 e 137155, 1969.
- SMITH, L. E.; RAFIQZAD, K. English for cross-cultural communication: The question of intelligibility. *TESOL Quarterly*, v. 13, p. 371-380, 1979.
- SVENSSON, S.-G. *Prosody and grammar in speech perception*. MILUS 2. Stockholm: Institute of Linguistics, University of Stockholm, 1974.
- TANG, C. *Mutual intelligibility of Chinese dialects: An experimental approach*. LOT dissertation series, 228. Utrecht: LOT, 2009.
- TANG, C.; VAN HEUVEN, V. J. Mutual intelligibility of Chinese dialects experimentally tested. *Lingua*, v. 119, p. 709-732, 2009.

TERKEN, J. M. B.; NOOTEBOOM, S. G. Opposite effects of accentuation and deaccentuation on verification latencies for 'given' and 'new' information. *Language and Cognitive Processes*, v. 2, p. 145-163, 1987.

TRUBETSKOY, N. S. Grundzüge der Phonologie. Prague: *Travaux du Cercle Linguistique de Prague* 7. Reprinted as Baltaxe, C. A. M. (Trans. 1969). *Principles of Phonology*. Berkeley and Los Angeles: University of California Press, 1969.

TZAKOSTA, M. *Multiple parallel grammars in the acquisition of stress in Greek L1*. LOT dissertation series, 93. Utrecht: LOT, 2004.

VAN DONSELAAR, W. *Effects of accentuation and given/new information on word processing*. Ph. D. dissertation, Utrecht University, 1995.

VAN HEUVEN, V. J. Segmentele versus prosodische invloeden van klemtoon op de herkenning van gesproken woorden [Segmental versus prosodic influences of stress on the recognition of spoken words]. *Verlagen van de Nederlandse Vereniging voor Fonetische Wetenschappen*, v. 159/162, p. 22-38, 1984.

VAN HEUVEN, V. J. Perception of stress pattern and word recognition: recognition of Dutch words with incorrect stress position. *Journal of the Acoustical Society of America* 78, S21, 1985. Available from: www.academia.edu/11793960.

VAN HEUVEN, V. J. Stress patterns in Dutch (compound) adjectives: acoustic measurements and perception data. *Phonetica*, v. 44, p. 1-12, 1987. DOI: <https://doi.org/10.1159/000261775>

VAN HEUVEN, V. J. Effects of stress and accent on the human recognition of word fragments in spoken context: gating and shadowing. In: AINSWORTH, W. A.; HOLMES, J. N. (ed.). *Proceedings of the 7th FASE/Speech-88 Symposium*. Edinburgh The Institute of Acoustics, 1988. p. 811-818.

VAN HEUVEN, V. J. Introducing prosodic phonetics. In: ODÉ, C.; VAN HEUVEN, V. J. (ed.). *Phonetic studies of Indonesian prosody*. Semaian, 9, Vakgroep Talen en Culturen van Zuidoost-Azië en Oceanië, RU Leiden, 1994. p. 1-26.

VAN HEUVEN, V. J. Making sense of strange sounds: (Mutual) intelligibility of related language varieties. A review. *International Journal of Humanities and Arts Computing*, v. 2, p. 39-62, 2008. DOI: <https://doi.org/10.3366/E1753854809000305>

VAN HEUVEN, V. J. Prosody and sentence type in Dutch. *Nederlandse Taalkunde*, v. 22, p. 3-29 e 43-46, 2017.

VAN HEUVEN, V. J. Notes on the phonetics of word and sentence stress: A cross-linguistic (re-)view. In: VAN DER HULST, H.; HEINZ, J.; GOEDEMAN, R. (ed.). *The Study of Word Stress and Accent: Theories, methods and data*. Cambridge: Cambridge University Press, 2018. p. 13-59. DOI: <https://doi.org/10.1017/9781316683101.002>

VAN HEUVEN, V. J.; VAN DONZEL, M. E. Structural conditioning and phonetic implementation of Dutch stress retraction. *Proceedings of the 13th International Congress of Phonetic Sciences*, p. 638-641, 1995.

VAN HEUVEN, V. J.; HAAN, J. Phonetic correlates of statement versus question intonation in Dutch. In: BOTINIS, A. (ed.). *Intonation: Analysis, modelling and technology*. Dordrecht/Boston/London: Kluwer, 2000. p. 119-144.

VAN HEUVEN, V. J.; HAGMAN, P. J. Lexical statistics and spoken word recognition in Dutch. In: COOPMANS, P.; HULK, A. (ed.). *Linguistics in the Netherlands 1988*. Foris: Dordrecht, 1988. p. 59-68.

VAN HEUVEN, V. J.; SLUIJTER, A. M. C. Notes on the phonetics of word prosody. In: GOEDEMAN, R.; VAN DER HULST, H.; VISCH, E. (ed.). *Stress patterns of the world, Part 1: Background*. HIL Publications (volume 2), Holland Institute of Generative Linguistics, Leiden/Holland Academic Graphics, The Hague, p. 233-269, 1996.

VAN HEUVEN, V. J.; TURK, A. Phonetic correlates of word and sentence stress. In: GUSSENHOVEN, C.; CHEN, A. (ed.). *The Oxford handbook of language prosody*. Oxford: Oxford University Press, 2020. p. 150-165. DOI: <https://doi.org/10.1093/oxfordhb/9780198832232.013.8>

VAN WIJNGAARDEN, S. J. Intelligibility of native and non-native Dutch speech. *Speech Communication*, v. 35, p. 103-113, 2001.

VAN ZANTEN, E. A.; VAN HEUVEN, V. J. Word stress in Indonesian; its communicative relevance, *Journal of the Humanities and Social Sciences of Southeast Asia and Oceania [Bijdragen tot de Taal-, Land- and Volkenkunde]*, v. 154, p. 129-147, 1998.

VAN ZANTEN, E. A.; VAN HEUVEN, V. J. Word stress in Indonesian: fixed or free? *NUSA, Linguistic Studies on Indonesian and other Languages in Indonesia*, v. 53, p. 1-20, 2004.

WANG, H.; ZHU, L.; LI, X.; VAN HEUVEN, V. J. Relative importance of tone and segments for the intelligibility of Mandarin and Cantonese. In: LEE, W. S.; ZEE, E. (ed.). *Proceedings of the 17th International Congress of Phonetic Sciences*. Hong Kong: City University of Hong Kong, 2011. p. 2090-2093.

WINGFIELD, A. The intonation-syntax interaction: prosodic features in the perceptual processing of sentences. In: COHEN, A.; NOOTEBOOM, S. G. (ed.). *Structure and process in speech perception*. Berlin: Springer, 1975. p. 148-156.

ZHANG, J. L.; QI, S. Q.; SONG, M. Z.; LIU, Q. X. 汉语声调在言语可懂度中的重要作用 [On the important role of Chinese tones in speech intelligibility]. *声学学报 [Acta Acustica]*, v. 4, p. 237-241, 1981.

Resolvendo o paradoxo da prosódia

Vincent J. van Heuven

I. Introdução

Neste artigo tutorial, explicarei primeiro o que queremos dizer com prosódia na linguagem e na fala. Mostrarei então que, geralmente, a prosódia é altamente redundante, ou seja, pode ser omitida do sinal de fala sem tornar a fala ininteligível ou incompreensível. No entanto, muitas vezes tem sido mencionado, também em publicações recentes, que correção na prosódia deve ser prioridade na aquisição da fonologia de uma língua estrangeira. Isso parece uma contradição. Por que uma característica geralmente redundante de um código merece prioridade no ensino? O objetivo deste artigo é resolver esse paradoxo.

I.1 O que é prosódia?

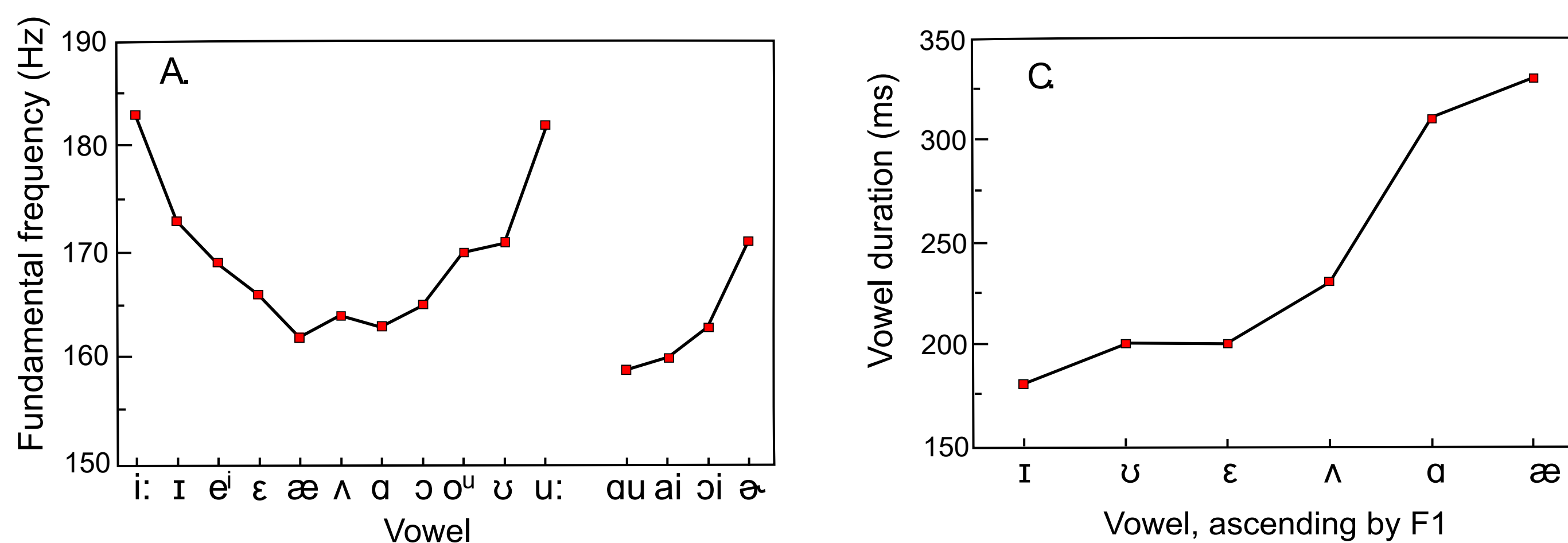
No estudo da fala e da linguagem, costuma-se analisar uma frase como uma sequência de unidades discretas ordenadas no tempo, muitas vezes chamadas de segmentos. Uma frase falada é analisada como uma sequência de morfemas, ou seja, sequências mínimas de som com um significado identificável associado a elas, enquanto cada morfema é decomposto em uma sequência de vogais e consoantes – que são vistas como os blocos de construção mais curtos a partir dos quais as frases faladas são construídas. Ao mesmo tempo, reconhece-se prontamente que palavras e frases não podem ser descritas apenas como uma sequência de segmentos. De fato, a mesma sequência de vogais e consoantes pode constituir palavras diferentes com significados diferentes (e a mesma sequência de palavras pode formar frases diferentes) dependendo da prosódia. A prosódia é definida, então, como o conjunto de propriedades que caracterizam palavras e frases que não podem ser previstas a partir da mera sequência de segmentos sonoros menores. Por exemplo, o holandês tem duas palavras diferentes com significados não relacionados que são compostos pelos mesmos cinco sons na mesma ordem: /ka:nɔn/. Se a primeira sílaba for pronunciada mais forte que a segunda, o significado é ‘lista de santos’ ou ‘estilo musical’ (escrito ‘canon’), mas com acento na segunda sílaba a palavra significa ‘arma grande’ (escrito ‘kanon’; para mais discussão da definição de prosódia ver VAN HEUVEN, 1994, 2008; VAN HEUVEN; SLUIJTER, 1996; NOOTEBOOM, 1997; RIETVELD; VAN HEUVEN 2016, capítulo 12).

Na maioria dos livros de fonética, os autores parecem satisfeitos em simplesmente listar as propriedades da fala que geralmente podem ser consideradas de natureza prosódica, como:

- Organização melódica: *pitch*, tom, entoação;
- Organização temporal: comprimento, duração, tempo e variação de tempo, e pausa;
- Organização dinâmica: intensidade e variação de intensidade, acento de palavra e de acento frasal.

Embora tal taxonomia seja bastante ampla, as considerações ficam mais confusas quando consideramos o fato de que todas as vogais e consoantes individuais têm propriedades melódicas, temporais e dinâmicas intrínsecas. Geralmente, por exemplo, vogais fechadas como [i] e [u] são pronunciadas – tudo o mais sendo igual – com um tom mais alto do que uma vogal aberta [a] (por exemplo, LEHISTE; PETERSON, 1961, ver Figura 1-a). Por outro lado, as vogais abertas naturalmente têm maior intensidade do que as vogais fechadas (porque funcionam como um funil, abafando o som). Da mesma forma, vogais abertas tendem a ser mais longas do que vogais fechadas (por exemplo, LEHISTE; PETERSON, 1960, veja Figura 1-c), uma vez que abrir a mandíbula amplamente (como para [a]) e fechá-la novamente leva mais tempo do que apenas abrí-la um pouco e fechá-la novamente (como para um [i] ou [u]). No entanto, essas propriedades intrínsecas, por sua própria natureza, são totalmente previsíveis a partir da identidade dos segmentos e, portanto, devem ser excluídas do domínio da prosódia.

Para aprofundar este ponto ainda mais, há também uma classe de propriedades de segmentos que é referida pelo termo “co-intrínseca”. Este termo inclui as propriedades de um segmento que são totalmente previsíveis a partir da identidade dos segmentos vizinhos. Por exemplo, uma vogal durará mais quando for seguida por uma obstruinte sonora do que quando for seguida – tudo o mais sendo igual – por uma obstruinte surda (SLIS; COHEN, 1969; LISKER; ABRAMSON, 1964). A frequência fundamental (tom) de uma vogal depois de uma obstruinte surda começa cerca de 20 a 30 Hz mais alta do que depois de uma obstruinte sonora (o chamado tom co-intrínseco, por exemplo, LEHISTE; PETERSON, 1961, ver Figura 1-d). Novamente, essa parte das propriedades melódicas, temporais e dinâmicas dos sons é previsível a partir da sequência de segmentos, portanto, não faz parte da prosódia. Para descobrir a verdadeira prosódia da fala, é necessária alguma forma de decomposição melódica, temporal e dinâmica que extraia primeiro as propriedades segmentais inerentes e, em seguida, fature as influências co-intrínsecas para que a prosódia pura permaneça.



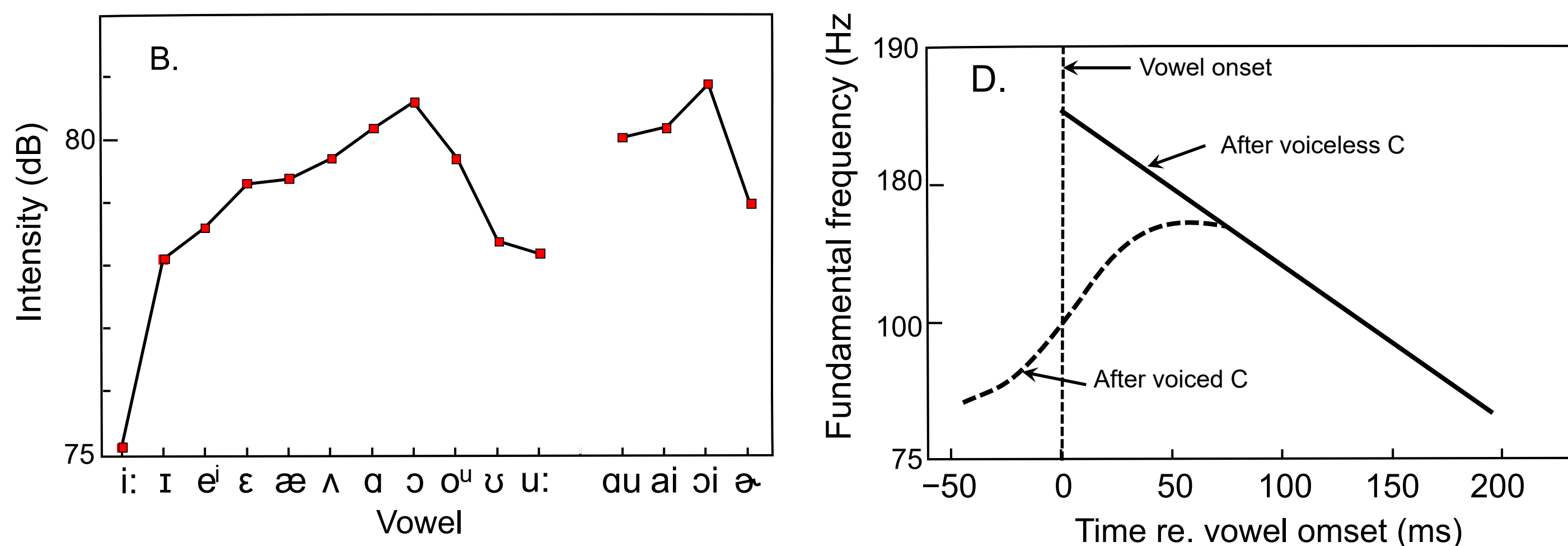


Figura 1: (a) Frequência fundamental (Hz) para 15 vogais do inglês americano (adaptado de RIETVELD; VAN HEUVEN, 2001, p. 234 apud LEHISTE; PETERSON, 1961). (b) Pico de intensidade (dB) para as mesmas 15 vogais (adaptado de RIETVELD; VAN HEUVEN, 2001, p. 233 apud PETERSON; LEHISTE, 1959). (c) Duração (ms) de seis vogais fonologicamente não-tensas no inglês americano ordenadas por grau de abertura (adaptado de RIETVELD; VAN HEUVEN, 2001, p. 235 apud LEHISTE; PETERSON, 1960). (d) Frequência fundamental idealizada (Hz) em função do tempo (ms) após obstruintes sonoras e surdas (adaptado de NOOTEBOOM; COHEN, 1976, p. 85 apud SLIS; COHEN, 1969).

1.2 Funções comunicativas da prosódia

Costuma-se diferenciar entre prosódia lexical e prosódia frasal. Existem basicamente dois tipos de sistemas prosódicos de palavras nas línguas do mundo, ou seja, línguas acentuais e línguas tonais. A função comunicativa do tom seria ajudar a diferenciar as palavras no léxico. Por exemplo, em mandarim chinês, a mesma sílaba /ma/ pode sinalizar cinco significados diferentes e não relacionados, dependendo do tom (por exemplo, VAN HEUVEN, 2008). O acento também pode ajudar a diferenciar as palavras (como em canon ~ kanon acima), mas apenas em línguas em que a localização da sílaba tônica dentro da palavra é imprevisível. Em línguas com acento fixo, como no húngaro, onde o acento é sempre inicial de palavra, o acento pode sinalizar limites de palavras.

As funções comunicativas da prosódia frasal são muitas.

- *Fraseamento.* Pela aceleração (reforço inicial) e desaceleração (alongamento pré-fronteira), e mudanças de tom, o falante pode sinalizar qual sequência de palavras deve ser analisada como um domínio prosódico – com marcações diferentes dependendo da profundidade hierárquica do domínio;
- *Digitação de orações.* Grupos de palavras (frases), sejam uma frase ou uma oração dentro de uma frase, podem ser qualificados como um tipo específico. Tipos de frases bem conhecidos são declaração, pergunta, exclamação, comando e negação. Em um nível inferior, existem pelo menos dois tipos de orações, ou seja, orações que são sentenças finais (finalidade) e orações não finais, que normalmente contêm um marcador de continuação prosódica;

- *Função de atenção.* Dentro desses domínios prosódicos, uma palavra é tipicamente apresentada como comunicativamente mais importante do que as outras por meio do acento frasal – tipicamente marcado por uma mudança saliente no tom vocal na sílaba dentro da palavra que também tem acento de palavra (veja, por exemplo, VAN HEUVEN, 2018 para uma pesquisa interlinguística dos resultados).
- *Função intencional.* Melodias diferentes podem então qualificar ainda mais o status da informação dos referentes em relação ao contexto compartilhado pelo falante e pelo ouvinte. Em holandês, por exemplo, um acento frasal marcado por apenas uma queda no tom sugere que o falante poderia facilmente adivinhar a identidade do referente focado, enquanto um acento de subida e queda indica que o referente era novo e inesperado (KEIJSPER, 1984).
- *Afeto.* A prosódia também dá uma importante contribuição para a expressão do afeto. Este é um termo abrangente que engloba atitudes (por exemplo, sinceridade, ironia, sarcasmo) e emoções (por exemplo, felicidade, tristeza, raiva, medo) por parte do falante. Como essas distinções geralmente não são feitas de forma categórica, mas em gradiente, argumenta-se frequentemente que elas não fazem parte da fonologia da língua, mas devem ser consideradas propriedades paralinguísticas.
- *Continuidade.* A melodia de uma frase se desenvolve lentamente no tempo e obedece a um padrão altamente regular e previsível. Esta é uma importante fonte de informação que permite que um ouvinte acompanhe um falante em meio a vozes concorrentes em uma multidão.

1.3 Inteligibilidade x compreensibilidade

Quando duas pessoas se comunicam por meio da linguagem falada, os pensamentos que entram na mente do falante, primeiro devem ser expressos em termos do vocabulário e das estruturas gramaticais da língua do falante. As estruturas linguísticas mentais são então usadas para fazer os órgãos vocais do falante se moverem para que produzam um som audível. Os sons viajam pelo ar (ou algum outro meio – por exemplo, um canal de telecomunicações) e depois atingem o ouvido do ouvinte. O resultado é que o ouvinte ouve um fluxo de sons. Se o ouvinte estiver familiarizado com a língua, ele reconhecerá (as mesmas) unidades linguísticas (por exemplo, palavras) na mesma ordem em que saíram da boca do falante. Essa parte do processo de comunicação é o que chamamos de reconhecimento de fala. Se um número suficiente de unidades foi reconhecido corretamente na ordem sequencial correta, o ouvinte será capaz de reconstruir os pensamentos e as intenções originais do falante. Esta última parte do processo é o que chamamos de compreensão ou compreensão da fala. A sequência de eventos esboçada aqui é conhecida como cadeia de fala (DENES; PINSON, 1963), e tem sido a fundamentação do modelo de produção de fala de Levelt (1989) e do modelo de escuta nativa de Cutler (2012).

A inteligibilidade de um falante, ou de um enunciado de fala, é o grau em que um ouvinte é capaz de reconhecer as unidades linguísticas no fluxo de sons e estabelecer a ordem em que foram ditas. Se o ouvinte não conhece a língua que o falante usa, a inteligibilidade do falante é (próxima de) zero – mesmo que o(s) enunciado(s) sejam perfeitamente inteligíveis para ouvintes nativos da língua. A compreensibilidade de um falante (ou de um texto falado) é o grau em que um ouvinte é capaz de entender o significado e as intenções do falante. A inteligibilidade, então, correlaciona-se com o reconhecimento da fala e a compreensibilidade com a compreensão da fala. Nessa visão, a compreensão da fala é um processo de ordem superior do que o reconhecimento da fala. São necessários métodos diferentes para avaliar a inteligibilidade de um falante do que para avaliar a compreensibilidade do indivíduo. Por exemplo, um teste de inteligibilidade funcional seria incumbir um ouvinte de anotar, por meio de um ditado, uma série de enunciados sem sentido produzidos pelo falante. É isso que o teste do SSI (SSI: sentenças semanticamente imprevisíveis, BENOÎT et al., 1996) faz com enunciados como “O estado cantou a semana inteira”. Um teste funcional de compreensão de fala exigiria, por exemplo, que um ouvinte determinasse se uma frase falada incorpora uma verdade ou uma falsidade. Se um ouvinte pensar que *a maioria dos bebês humanos é mais pesada do que um elefante adulto é verdade*, ele (ou ela) obviamente não entendeu a frase.²³

Deve ficar claro a partir da explicação acima que a prosódia de palavras, ou seja, acento e tom lexical, fazem uma contribuição potencial para o reconhecimento de palavras. De fato, acento e tom são indispensáveis para o reconhecimento de palavras quando um idioma possui itens lexicais consistindo em sequências de segmentos idênticas que diferem apenas na localização do acento ou no tipo de tom lexical (veja os exemplos de pares de acentos mínimos em holandês e inglês, e os quatro tons mínimos em mandarim). Geralmente, no entanto, a ambiguidade que pode surgir devido a esses contrastes mínimos palavra-prosódia é facilmente resolvida com base em informações contextuais *top-down*. Foi demonstrado que o acento e o tom lexical são usados no processo de reconhecimento de palavras *on-line*. Se um ouvinte holandês ou espanhol ouvir que uma palavra falada começa com uma sílaba tônica, qualquer palavra no léxico que não tenha uma sílaba inicial tônica não é mais candidata a reconhecimento. Este seria um exemplo de como as informações de acento podem ajudar a diminuir o grupo de candidatos a reconhecimento no estágio inicial do processo de reconhecimento de palavras *on-line*, chamado acesso lexical. Em outros casos, a prosódia da palavra pode ser usada nos estágios posteriores do processo de reconhecimento, o estágio de verificação, para verificar se a prosódia da palavra percebida corresponde ao padrão armazenado para a palavra que foi reconhecida provisoriamente com base na sequência de segmentos. Voltarei

23 Eu explico isso com alguns detalhes porque outras disciplinas e outros pesquisadores usaram os termos de maneira diferente da maneira como os fonéticos o fazem. Por exemplo, linguistas aplicados (MUNRO; DERWING, 1995; MUNRO et al., 2006) usam o termo *inteligibilidade* como o grau em que um falante pode ser entendido usando testes funcionais, e o termo *compreensibilidade* para a opinião do ouvinte sobre quão bem um falante (ou enunciado) pode ser entendido. Nossa posição é que inteligibilidade e compreensibilidade tratam de dois estágios diferentes na cadeia da fala, e que cada um pode ser medido tanto por testes funcionais quanto por testes de opinião.

a esse uso da palavra prosódia em seções posteriores deste capítulo, quando revisarmos vários estudos que abordam essa questão.

Outra maneira pela qual a prosódia lexical pode auxiliar no reconhecimento de palavras e, assim, contribuir para a inteligibilidade da fala, é vista em línguas com acento de palavra fixo. Tem sido sugerido que, em tais línguas, ouvir um acento informa ao ouvinte que ele está no início de uma nova palavra quando o acento é inicial fixo (como em finlandês e húngaro), ou que a próxima sílaba será um início de palavra quando o acento é fixo-final (como em turco e persa). Na verdade, o acento não precisa ser corrigido. Em inglês e holandês, a maioria das palavras começa com uma sílaba tônica; os ouvintes assumem que toda vez que ouvem um acento estão no início de uma nova palavra – mesmo que isso nem sempre seja verdade. Verificou-se até que os bebês usam essa estratégia de segmentação no primeiro ano de vida para detectar palavras no fluxo contínuo da fala em seu ambiente.

A prosódia frasal contribui muito pouco para a inteligibilidade, conforme definido nesta seção. As funções comunicativas listadas para a prosódia frasal (ver seção 1.2) existem para ajudar o ouvinte a compreender as intenções que o falante deseja transmitir ao ouvinte pela sequência de palavras que ele encadeou. O fraseamento, por meio de aceleração e desaceleração, pausa e uso de tons de fronteira, informa ao ouvinte quais palavras devem ser analisadas juntas em unidades de pensamento coesas e onde há quebras importantes na estrutura linguística do texto falado. O acento frasal informa ao ouvinte em quais palavras ele deve prestar atenção especial, enquanto a melodia frasal é usada para divulgar as intenções, atitudes e emoções mais profundas do falante.²⁴

É possível, até certo ponto, remover todas as informações prosódicas de um texto falado e deixar apenas as informações segmentais, ou seja, a sequência de vogais e consoantes. Isso pode ser feito, por exemplo, usando síntese de fala baseada exclusivamente em sons extraídos de sílabas totalmente acentuadas, omitindo qualquer inflexão de tom ou ajuste de durações de segmentos e reduzindo todas as pausas gramaticais a zero. A fala resultante é produzida em um tom monótono perfeitamente plano, com um tipo de ritmo *staccato*, uma vez que todas as sílabas são pronunciadas como se estivessem totalmente acentuadas, todas as palavras recebem o mesmo peso, e qualquer sugestão de fraseamento estaria ausente. O resultado seria comparável a um texto impresso sem sinais de pontuação e sem recuos de parágrafos. Mesmo que eliminemos os espaços e hifens que normalmente separam as palavras, poderíamos, no entanto, com algum esforço, reconstruir as estruturas prosódicas ausentes da informação segmentada que é ainda codificada nas palavras impressas. Para ilustrar o ponto, mostro aqui o parágrafo de abertura de um artigo que apareceu cerca de 40 anos atrás (SMITH; RAFIQZAD, 1979).

²⁴ Van Wijngaarden (2001) mediu a inteligibilidade (em termos de Speech Reception Threshold SRT, ou seja, a quantidade de barulho que o ouvinte pode tolerar e ainda atingir 50% de reconhecimento de frases corretas) para enunciados holandeses redundantes diários em duas condições prosódicas, ou seja, com sua melodia original e com a melodia substituída por um tom constante fixo (através da ressíntese PSOLA). O SRT para o enunciado monotonizado foi 2 dB pior para as versões monotonizadas (provavelmente porque a monotonização artificial causa um efeito de campanha que interfere na inteligibilidade segmental). Em silêncio, no entanto, ambas as versões eram perfeitamente inteligíveis.

em discussões sobre o uso do inglês para comunicação intercultural muitas vezes sustentase que o falante nativo educado é mais provável de ser inteligível para os outros do que o falante nativo educado assim como o padrão como qual comparar todas as outras variedades das pessoas com essa opinião também está frequentemente preocupada que o uso de outros modelos levará a uma diversidade de variedades na nativas de inglês educado que logo as pessoas que falam inglês podem não ser inteligíveis para seus ouvintes²⁵

Poucas pesquisas foram feitas sobre a compreensão da fala desprovida de prosódia. Uma exceção é Keskin, Ari e Baştuğ (2019), que utilizou um sistema de conversão de texto em fala para turco com todos os sinais de pontuação excluídos do texto de entrada e todas as regras prosódicas desativadas, como condição de controle em um experimento que visava a compreensibilidade da fala com e sem prosódia. Quatro grupos pareados de crianças turcas da escola primária ouviram a mesma história em quatro versões diferentes, ou seja, faladas por um falante humano ou por um sistema de conversão de texto em fala, com e sem prosódia. A compreensão auditiva foi testada por seis itens de múltipla escolha na compreensão literal e mais seis na compreensão inferencial. Os resultados foram os seguintes:

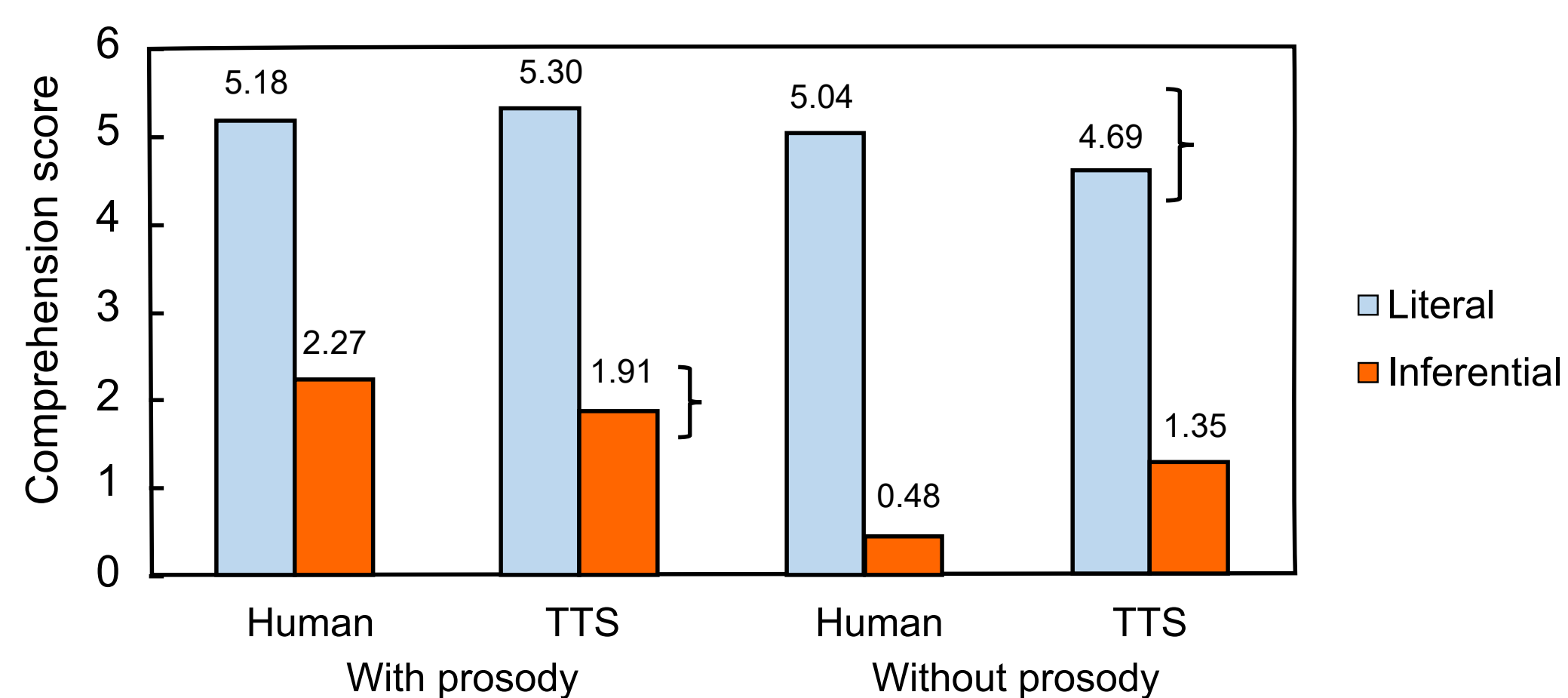


Figura 2: Pontuação de compreensão auditiva (máximo = 6) para quatro versões do mesmo texto, faladas por um humano ou por um sistema de conversão de texto em fala (TTS), com e sem prosódia. As chaves mostram que não diferem significativamente um do outro. Veja o texto para detalhes, adaptado de Keskin et al. (2019).

Embora as versões sem prosódia tenham obtido escores de compreensão ligeiramente inferiores, as diferenças não foram significativas para as questões de conteúdo literal. Quando as crianças tiveram que fazer inferências a partir do que ouviram na passagem falada, a compreensão foi afetada negativamente (e significativamente) pela ausência de prosódia. Isso demonstra que a compreensão é bem possível mesmo na ausência de informação prosódica. Há também evidências circunstanciais de que a remoção da prosódia de materiais falados afeta a compreensão. Por um lado, os ouvintes

25 Adaptação feita para o português com ausência de diacríticos que assinalariam o acento de palavra. Traduzido do original em inglês: “indiscussionsoftheuseofenglishforcrossculturalcommunicationitisoftenmaintainedthattheeducatednative speakerismorelikelytobeintelligibletoothersthantheeducatednonnativespeakerthesepeoplethereforeclaimthatanative speakervarietyofenglishshouldbethemodelforproductionaswellasthestandardwithwhichcompareallothervarietiespeopleofthisopinionarealsooftenconcernedthattheuseofothermodelswillleadtosuchagreatdiversityofnonnativevarietiesofeducatedenglishthatsoonpersonsspeakingenglishmaynotbeintelligibletotheirlisteners”

rapidamente perdem a concentração e reclamam que acham impossível acompanhar o que o falante está tentando dizer quando a prosódia é (artificialmente) eliminada das passagens faladas. O reconhecimento das emoções foi severamente comprometido na fala monotonizada.

O resultado desta seção seria que a prosódia geralmente contribui com informações redundantes e não vitais. A remoção de informações prosódicas torna o *input* da fala não natural (e inaceitável), mas não impede que o ouvinte nativo reconheça as palavras em sequência e compreenda a mensagem.

1.4 Importância relativa de segmentos *vs.* prosódia para inteligibilidade e compreensibilidade

Tem havido um debate considerável sobre a importância relativa da estrutura segmental *versus* a da prosódia. Na linguagem escrita, a prosódia dificilmente é refletida, enquanto a maioria das ortografias fornece uma codificação bastante precisa e exaustiva da identidade das vogais e consoantes que compõem uma frase. O acento de palavra e frase normalmente não é marcado na linguagem escrita (embora o acento de palavra irregular seja marcado na ortografia espanhola, e os tons de palavras altos e baixos sejam explicitamente marcados por acentos agudos e graves nas ortografias de algumas línguas de tons africanos), enquanto apenas os limites prosódicos mais profundos (correspondendo aproximadamente aos limites sintáticos de parágrafo, sentença e oração) são representados na ortografia. A prosódia é quase completamente redundante, dada a composição segmental das palavras em sequência. Se removermos todas as informações prosódicas úteis de uma frase ou parágrafo falado, o resultado ainda é perfeitamente inteligível (embora mais difícil de compreender e lembrar, veja a seção anterior). No entanto, se eliminarmos todas as informações segmentais de um enunciado, mantendo totalmente a informação prosódica, por exemplo, por meio de filtragem de baixa frequência em 300 Hz para um falante do sexo masculino, nenhuma inteligibilidade permanece. Mesmo em línguas tonais como o mandarim, a inteligibilidade permanece próxima do nível máximo quando a prosódia (incluindo tons de palavras) é eliminada (por exemplo, QIN, 2007; VAN HEUVEN, 2008, WANG *et al.*, 2011).

2. Aprendizagem de prosódia e línguas estrangeiras

Na seção anterior, especulamos a importância relativa das propriedades segmentais *versus* suprasegmentais (ou: prosódicas) da fala que permitem que um ouvinte nativo reconheça as palavras em sequência (correlacionadas com a inteligibilidade) e reconstrua o significado e as intenções do falante (correlacionada com a compreensibilidade). Nesta seção, descreverei e discutirei algumas publicações que têm sido frequentemente citadas em apoio à ideia de que a prosódia é mais importante para o reconhecimento e a compreensão da fala do que as propriedades segmentais quando falantes nativos (de inglês) ouvem um falante não nativo. Se esta afirmação for verdadeira, então surge uma pergunta. Como é possível que elementos da fala que são quase completamente

redundantes em inglês, como a prosódia de palavras e frases, possam ser mais importantes do que a estrutura segmental na comunicação entre um falante não nativo e um ouvinte nativo de inglês? Isso é o que chamo de paradoxo da prosódia.

2.1 Anderson-Hsieh *et al.* (1992)

O primeiro artigo que devemos olhar é o de Anderson-Hsieh *et al.* (1992). Este estudo investigou a relação entre os julgamentos de avaliadores experientes de pronúncia não nativa e desvio real da parte segmental, prosódia e estrutura silábica. Sessenta falantes de 11 grupos linguísticos gravaram a mesma passagem de leitura e foram avaliados por especialistas em pronúncia e posteriormente analisados impressionisticamente quanto a desvios em segmentos, prosódia e estrutura silábica. O desvio encontrado em cada área de pronúncia foi então correlacionado com as classificações de pronúncia. Uma análise dos 60 falantes mostrou que o desvio segmental, prosódia e estrutura silábica mostraram uma influência significativa nas classificações de pronúncia. No entanto, a variável prosódica provou ter o efeito mais forte. Quando análises separadas foram feitas em dois subgrupos de linguagem dentro da amostra, a prosódia sempre foi significativamente relacionada às classificações globais, enquanto isso nem sempre foi verdade para as outras variáveis investigadas.

Deve-se perceber, em primeiro lugar, que a estrutura silábica *per se* não é uma área independente de pronúncia. A pronúncia incorreta da estrutura silábica do inglês geralmente é encontrada quando a língua nativa do aluno não possui os complexos encontros consonantais do inglês. Eles, então, excluem uma ou mais consoantes do encontro, ou quebram o encontro consonantal inserindo uma vogal epentética antes em algum lugar dentro dele. Eu diria que isso reduz os erros de estrutura silábica a erros segmentais. Os erros segmentais podem ser de três tipos: segmentos podem ser excluídos, inseridos ou substituídos. Não há nenhuma razão de princípio para que inserções e exclusões sejam classificadas como basicamente diferentes de substituições. Quando todos os erros segmentais são combinados, eles podem ser mais influentes como um todo do que os erros prosódicos.

No estudo, os falantes não nativos vinham de 11 origens linguísticas diferentes. Geralmente, devemos encontrar correlações relativamente altas entre os vários componentes de pronúncia para alunos de por inglês língua estrangeira (ILE) que compartilham a mesma L1. A qualidade da pronúncia das consoantes, vogais, encontros e prosódia deve se desenvolver ao longo de um curso paralelo. No entanto, quando os aprendizes têm origens de L1 diferentes, pode ser que a pronúncia das consoantes, por exemplo, seja melhor do que a das vogais e/ou da prosódia. O fator interferente aqui é a forma como a L1 do aluno difere do inglês. Observou-se que os aprendizes de ILE geralmente pronunciam as consoantes (simples) bastante bem, e que os problemas de comunicação que eles experimentam com o inglês residem na pronúncia das vogais. Essa discrepância pode ser compreendida se considerarmos que o inglês possui apenas um inventário consonantal de tamanho

médio, conforme determinado, por exemplo, a partir do Atlas Mundial de Estruturas Linguísticas (COMRIE et al., 2005). Isso significa que, estatisticamente, pelo menos 50% dos alunos de ILE têm um idioma nativo com um inventário de consoantes igual ou maior em tamanho que o inglês. Nesses casos, será relativamente fácil para o aluno substituir as consoantes de L1 em inglês que estejam próximas o suficiente umas das outras para que a inteligibilidade não seja excessivamente comprometida. No entanto, quando se trata de vogais, o inglês pertence à categoria de idiomas com grandes inventários vocálicos. A grande maioria dos alunos de ILE, portanto, enfrentará o problema de que seu inventário de L1 carece de um número de vogais que possam ser usadas como substitutos razoáveis em inglês, de modo que a pronúncia das vogais comprometerá a inteligibilidade do aluno, desde que ele ou ela não configurou novos sons de vogais e mudou os limites das categorias de L1 existentes (um processo que pode levar anos).

Com base no que foi dito acima, geralmente será o caso que os aprendizes de ILE mostrarão realizações desviantes das consoantes e especialmente das vogais do inglês. Como argumentarei mais adiante neste capítulo, o papel da prosódia não é constante, mas varia de acordo com a qualidade da estrutura segmental. Se as vogais e as consoantes são mal definidas, então a prosódia assume um papel mais importante, enquanto sua influência permanece modesta (e imperceptível) desde que a qualidade segmental do ILE seja boa (fluente).

2.2. Hahn (2004)

Um único falante coreano proficiente leu um pequeno texto em inglês (270 palavras) em três versões (harmonizado *post hoc* para velocidade geral e duração de pausa). Versão A: acento frasal (acento primário) em palavras [+F] (ou seja, em foco, comunicativamente importante) apenas, B: acento frasal em uma palavra [-F] (não focada, comunicativamente sem importância), C: sem acento frasal em absoluto. Cada uma das três versões foi tocada para um grupo diferente de 30 ouvintes nativos americanos com pouca ou nenhuma experiência com ingleses não nativos. Três testes funcionais de inteligibilidade/compreensibilidade foram aplicados e um questionário foi preenchido após o término das tarefas funcionais.

- Teste 1: Os ouvintes tinham que clicar com o mouse toda vez que ouviam um bipe enquanto tentavam entender e lembrar o conteúdo do texto. Acurácia e tempos de reação foram medidos;
- Teste 2: Os ouvintes tiveram que recontar os contextos da forma mais precisa e completa possível. Pontos principais corretamente relatados e pontos menores (detalhes) foram contados;
- Teste 3: Foi aplicado um teste de compreensão aberta de 6 itens de resposta curta, capturando os pontos principais do texto; cada questão rendeu 0, 1 ou 2 pontos, max = 12;
- O teste de opinião posterior pediu aos ouvintes que classificassem o falante em 11 escalas relacionadas à inteligibilidade, compreensibilidade, vivacidade e agradabilidade da voz.

Quanto ao teste 1, a versão A (134) foi um pouco melhor que C (128) e do que B (125), mas o efeito não foi significativo. A pontuação dos dados de recontagem no teste 2 não é clara, pois o número máximo de pontos principais e detalhes que podem ser lembrados não é mencionado. Seja como for, A foi recontado com mais precisão do que B e C (C ligeiramente melhor, que não difere significativamente de B). Uma contagem de pontos principais foi mais informativa do que a contagem de detalhes (sem diferenças significativas nos detalhes), mas a adição simples não ponderada de pontos principais e detalhes discriminava melhor. O mesmo vale para os dados do teste de compreensão no teste 3: as três condições estão na mesma ordem, mas os efeitos foram insignificantes desta vez. As diferenças entre as três médias estão dentro de 1 ponto (em uma escala de 0 a 12). Os dados de julgamento, por fim, estão de acordo com os resultados da recontagem: a versão A obtém avaliações mais favoráveis do que B ou C, com pouca diferença entre estas duas últimas. O que é estranho é que as classificações para “melhor ênfase em pontos importantes” e (significativamente) “falta de monotonia” foram mais favoráveis para a versão B (sem acentos frasais) do que para C (com um acento primário por frase na palavra errada). A Figura 3 resume os resultados do teste de opinião.

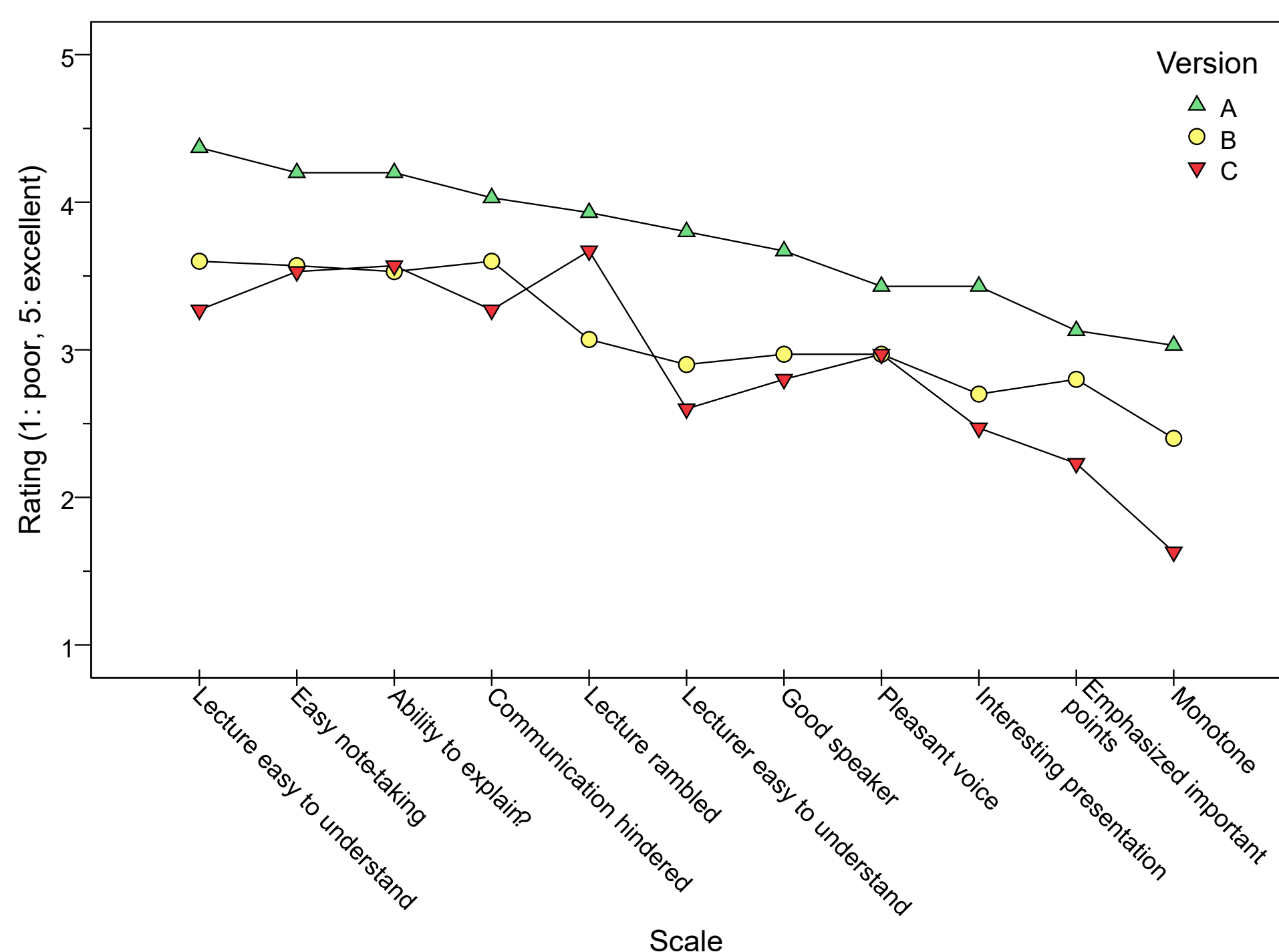


Figura 3: Julgamentos de ouvintes nativos americanos em onze escalas de classificação de falantes em três versões. A: Acento frasal apenas em palavras +Foco, B: Acento de uma frase em uma palavra –Foco, C: Sem acento frasal. Legenda: Eixo Y: Julgamentos (1: ruim, 5: excelente). Eixo X: Da esquerda para a direita: Fala fácil de entender; Fácil de tomar nota; Habilidade em explicar; Comunicação prejudicada; Fala com divagação; Falante fácil de compreender; Bom orador; Voz agradável; Apresentação interessante; Enfatizou pontos importantes; Monótono.

O problema crucial com este estudo é que o falante coreano, mesmo que seja proficiente em inglês, deve ter lido o texto com sotaque coreano, o que afetará suas vogais e consoantes. Para poder interpretar os resultados de Hahn, precisamos de um experimento de controle com o mesmo texto falado (nas mesmas três versões) por um falante nativo de inglês americano. Na verdade, não está claro para mim porque o experimento foi feito com um falante não nativo em primeiro lugar. Se o

pesquisador estivesse interessado no papel do acento frasal no processamento de um texto falado, então um falante nativo teria sido uma escolha mais óbvia. Seja como for, os resultados parecem indicar que colocar um acento frasal em uma palavra comunicativamente sem importância (-Foco) torna o texto mais difícil de entender para um ouvinte nativo de inglês. Isso confirma o que sabemos da literatura sobre falantes nativos e ouvintes de inglês (por exemplo, FOWLER; HOUSUM, 1987) e outras línguas (relacionadas) como o holandês (por exemplo, TERKEN; NOOTEBOOM 1987; VAN DONSELAAR, 1995): colocar um acento frasal em uma palavra não focalizada confunde o ouvinte, que então acredita que deve procurar um novo referente; da mesma forma, a omissão de um acento frasal em um novo referente leva o ouvinte em inglês a procurar incorretamente um determinado referente no contexto anterior. Minha previsão seria que o efeito do acento frasal errado será (mais) modesto quando o falante é nativo do que quando não é nativo (seja coreano ou qualquer outro idioma).

2.3. Field (2005)

Este é um estudo frequentemente citado em apoio à afirmação de que o acento correto da palavra em inglês é importante tanto para ouvintes nativos quanto para não nativos. Um (homem) falante nativo de inglês registrou 12 palavras de duas sílabas de alta frequência com um padrão de acento forte-fraco (Sw) e outras doze com um padrão de acento (fraco-forte) wS (ver Tabela 1).

Tabela 1. Materiais usados por Field (2005).

Mudança de qualidade em vogal átona ?	Padrões de acento lexical							
	Sw				wS			
	Ortogr.	Default	Mudança apenas	Mudança+ qualidade	Ortogr.	Default	Mudança apenas	Mudança+ qualidade
Sim	<i>husband</i> <i>second</i> ⁴ <i>different</i> <i>person</i> <i>Friday</i> <i>women</i>	'hʌzbənd 'sekənd 'dɪfrənt 'pɜ:sən 'fraɪdɪ 'wɪmən	hʌz'bə:nd se'kə:nd dɪ'frə:nt pə:'sə:n fraɪ'dɪ: wɪ'mə:n	hʌz'bænd se'kɒnd ²⁶ dɪ'frent pə:'sɒn fraɪ'deɪ wɪ'men	<i>enjoy</i> <i>forget</i> <i>contain</i> <i>provide</i> <i>today</i>	ən'dʒɔɪ fə'get kən'teɪn prə'vaɪd tə'deɪ	'ə:ndʒɔɪ 'fə:get 'kə:nteɪn 'prə:vaɪd 'tə:deɪ	'endʒɔɪ 'fɔ:get 'kɒnteɪn 'prəʊvaɪd 'tu:deɪ
Não	<i>follow</i> <i>coffee</i> <i>water</i> <i>listen</i> <i>money</i> <i>notice</i>	'fɒləʊ 'kɒfɪ 'wɔ:tə 'lɪsən 'mʌni 'nəʊtɪs	fɒ'ləʊ kɒ'fɪ: wɔ:'tə: lɪ'sə:n mʌ'ni: nəʊ'tɪ:s		<i>begin</i> <i>decide</i> <i>discuss</i> <i>inform</i> <i>perhaps</i> <i>prevent</i> <i>return</i>	bɪ'gɪn dɪ'saɪd dɪs'kʌs ɪn'fɔ:m pə:'hæps prɪ'vent rɪ'tə:n	'bɪ:gɪn 'dɪ:saɪd 'dɪ:skʌs 'ɪ:nfɔ:m 'pə:hæps 'prɪ:vənt 'rɪ:tə:n	

²⁶ As três versões da palavra *second* é o único exemplo dado em Field (2005). Ele escreve (p. 406): “Mudança de acento com mudança para qualidade de vogal plena [...]; exemplo: [se:kənd]. Field deve ter significado escrever [se:kɒnd] dada a existência da palavra *secondant* [sə'kɒndənt]. A vogal plena na sílaba inicial sugere que as vogais plenas não foram reduzidas a schwa quando o acento lexical foi desviado delas.

A ideia básica do experimento era criar para cada uma das 24 palavras uma versão com o acento (realizado como acento frasal em forma de citação) na sílaba errada. Como resultado da mudança de tonicidade, o movimento de *pitch* normalmente associado à sílaba tônica lexical deve mudar para a sílaba tônica incorreta, que por si só deve ser uma sugestão suficiente (se não predominante) para a mudança de tonicidade (VAN HEUVEN, 2018 e referências nele contidas). Infelizmente, a descrição das manipulações realizadas não me permite reconstruir exatamente o que foi feito. As transcrições na Tabela 1 são meu melhor palpite sobre o que o falante fez quando solicitado a mudar o acento para a sílaba errada⁴. Parece que a sílaba lexical tônica era sempre pronunciada com uma vogal plena, mesmo quando átona – embora, presumivelmente, fosse um pouco encurtada nas sílabas iniciais quando o acento era deslocado para a outra sílaba. Nas sílabas finais não se espera tal encurtamento, devido ao alongamento final (VAN HEUVEN; TURK, 2020 e referências neles). Nas palavras da metade superior da Tabela 1, acento errado foi implementado de duas maneiras diferentes. A primeira era simplesmente alongar a sílaba tônica incorretamente, caso em que um *schwa* ou vogal reduzida [ɪ] era alongado em 50% ou mais (“somente *shift*”). A segunda maneira de mudar o acento era mudar a qualidade da vogal de [ə] ou [ɪ] reduzida para uma vogal plena (que deveria ser mais longa e provavelmente foi alongada adicionalmente porque estava acentuada). A qualidade da vogal plena foi derivada da forma como a sílaba ou morfema pode ser pronunciado em inglês se for acentuado.

Na metade inferior da Tabela 1, as mudanças de acento foram implementadas apenas alongando a vogal não lexicalmente acentuada; Field afirma que essas vogais lexicalmente átonas são plenas quando átonas ou são reduzidas, mas não podem alternar com uma vogal plena.²⁷

As palavras foram oferecidas a três grupos de ouvintes nativos de inglês e a três grupos de não nativos, em três subconjuntos A-B-C com divisão igual de tipos e versões em blocos. Os ouvintes ouviram cada palavra-estímulo em apenas uma versão. Os ouvintes foram 82 nativos de 15 anos (NL); 28 ouviram o subconjunto A, 28 a B e 26 a C. 76 ouvintes não nativos (nível intermediário de mesma idade) de uma variedade de origens de L1 (ver Figura 4) também ouviram A, B ou C. Os indivíduos foram instruídos a esperar palavras em inglês, que eles tiveram que escrever na ortografia comum (mesmo que não conhecessem a palavra).

Os ouvintes ingleses reconheceram as palavras com acento correto, com acento apenas deslocado e com deslocamento mais alteração da qualidade da vogal em 96, 77 e 88%. Isso significa que o alongamento de uma vogal reduzida compromete mais o reconhecimento da palavra do que o deslocamento de um acento com a substituição adequada de uma vogal plena para uma vogal reduzida, mesmo quando em ambas as situações a mudança de *pitch* se move para a sílaba lexicalmente não acentuada. Os resultados indicaram ainda que um deslocamento de acento para

²⁷ Esta última ideia depende da suposição de que a vogal em *sir* ou *first* (que é frequentemente transcrita como /ɜ:/ no inglês britânico) é apenas um *schwa* alongado /ɜ:/.

a direita em uma palavra de padrão forte-fraco (Sw) é mais prejudicial do que um deslocamento para a esquerda em uma palavra de padrão fraco-forte (wS), e que esse efeito é ainda maior quando o acento é deslocado pelo alongamento da vogal reduzida, do que quando à sílaba tônica incorreta é dada uma vogal plena (Tabela 2).

Tabela 2. Reconhecimento correto de palavras por ouvintes nativos de inglês de palavras corretas e incorretas.

Acento incorreto marcado por	Palavras Sw			Palavras wS		
	Acento correto	Mudança-direita	N	Acento correto	Mudança-esquerda	N
Alongamento apenas	96	71	12	95	88	12
Substituição vogal plena	98	83	5	97	95	6

A Figura 4 apresenta os escores de reconhecimento de palavras para os três tipos de acentuação (correto, acento errado apenas por alongamento, acento errado por substituição de uma vogal plena) para todos os grupos de ouvintes, nativos e não nativos.

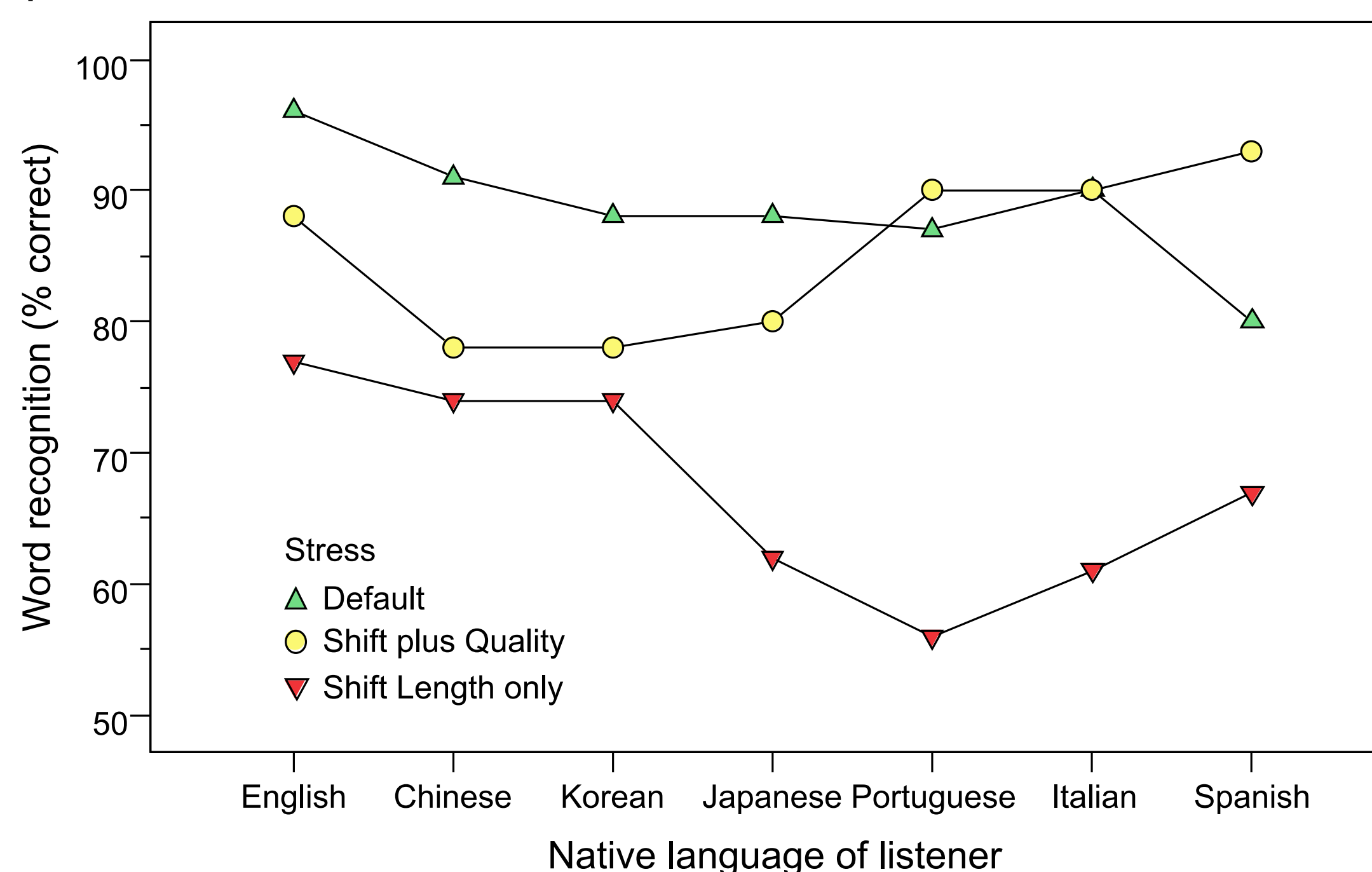


Figura 4: Percentual de reconhecimento de palavras corretas para sete grupos de ouvintes que diferem na língua nativa. Palavras inglesas de duas sílabas eram acentuadas corretamente (Padrão) ou foram faladas com acento deslocado para a sílaba errada, que manteve sua qualidade original da vogal (mudança em “Somente Comprimento”) ou que foi alterada tanto em Comprimento quanto em Qualidade Vogal (“mais Qualidade”). (Depois de Campo, 2005).
 Legenda: Eixo Y: Reconhecimento de palavra (% de acerto). Eixo X: Língua nativa do ouvinte; Inglês, Chinês, Coreano, Japonês, Português, Italiano, Espanhol.

Os principais efeitos são claros. Acento correto funciona melhor, então acento errado realizado com mudança natural da qualidade da vogal, e pior para acento errado com mudança apenas no comprimento (som não natural, impossível em inglês – a vogal tem o comprimento de uma sílaba tônica e um movimento de tom, mas a qualidade da vogal não é ajustada). Geralmente, os efeitos da língua nativa e do tipo de acento são aditivos, mas há uma interação notável: os ouvintes com

uma L1 românica respondem a palavras acentuadas incorretas (com a qualidade da vogal ajustada) como se estivessem corretamente acentuadas. Os ouvintes românicos só têm dificuldades se a sílaba acentuada incorreta contiver uma vogal reduzida longa.

O artigo reitera o que era conhecido na época, ou seja, que as palavras em inglês são mais difíceis de reconhecer quando são (frasalmente) acentuadas na sílaba errada, e que um acento inicial errado é mais perturbador do que um acento final (ou posterior) errado. Esse efeito e a assimetria foram mostrados anteriormente, por exemplo, por Cutler e Clifton (1984) para o inglês e por Van Heuven (1984, 1985) para o holandês (veja também a próxima seção). A novidade é que os mesmos efeitos foram encontrados para ouvintes não nativos (desconsiderando a notável interação encontrada com os ouvintes de ILE com uma L1 latina).

O estudo mostra que a posição correta do acento (quando marcada por um acento frasal, ou seja, um movimento de *pitch*) é importante para o reconhecimento de palavras, mais para palavras com acento lexical inicial do que com não inicial. Não apenas os ouvintes nativos de inglês precisam ouvir o acento na sílaba certa, mas também os ouvintes não nativos têm uma chance melhor de reconhecer palavras em inglês se tiverem o acento na sílaba correta. O artigo não mostra, no entanto, que o acento correto é mais importante do que ouvir os segmentos corretos. De fato, o artigo apoia a ideia de que segmentos incorretos (vogais curtas alongadas em sílabas acentuadas incorretamente) são especialmente prejudiciais ao reconhecimento de palavras. Observe, no entanto, que o efeito de colocar o acento na sílaba errada é relativamente menor, especialmente quando o padrão de acento incorreto usa vogais normais do inglês (sem vogais curtas alongadas não naturais): as pontuações de reconhecimento caem de 96% para 88% de acertos, ao menos nesta tarefa de transcrição *off-line*. Isso significa que o acento incorreto não bloqueia o reconhecimento de palavras: o ouvinte ainda é capaz de detectar e reparar o padrão incorreto de acento – devido à redundância lexical.

2.4. Resumo

Para resumir a análise, revisamos alguns artigos que são frequentemente citados em apoio à afirmação de que acertar a prosódia deveria ser uma meta de aprendizado mais importante para falantes não nativos de inglês do que acertar as vogais e as consoantes. Minha conclusão neste momento é que a evidência não é convincente. Os estudos sofrem de deficiências processuais e carecem de condições de controle necessárias que permitiriam ao leitor estimar a contribuição relativa de propriedades segmentais e suprasegmentais (incorretas) para o processamento de inglês não nativo.

Na próxima parte do capítulo, assumirei, no entanto, que a prosódia correta é um importante objetivo de aprendizado para falantes de ILE. A questão então permanece: como a prosódia pode ser redundante e relativamente sem importância na comunicação entre falantes nativos e ouvintes

em inglês, e ainda constituir uma fonte de falha de comunicação quando o falante e/ou ouvinte é estrangeiro? Para entender como esse paradoxo pode ser resolvido, precisamos saber mais sobre o papel da prosódia quando a fala deve ser reconhecida em circunstâncias adversas.

3.1 Rumo a resolver o paradoxo

3.1 Informações transmitidas por segmentos *vs.* prosódia

Dada uma taxa média de fala de cinco sílabas por segundo e uma estrutura de sílaba mínima de uma consoante seguida por uma vogal (estrutura CV), a taxa de elocução de segmentos tem um limite inferior em aproximadamente dez segmentos por segundo. A taxa de elocução pode aumentar quando os tipos de sílabas assumem maior complexidade e/ou quando o falante fala rápido. A prosódia muda mais lentamente ao longo do tempo, pois o domínio prosódico mínimo é tipicamente a sílaba. É uma sílaba (em vez de um segmento) que carrega um tom lexical, ou que é tônica ou átona. A prosódia, portanto, desenvolve-se mais lentamente no tempo do que a informação segmental. Por esta razão, é comumente entendido que a informação segmental por si só é suficiente para diferenciar as formas das palavras no léxico. Essa percepção é capturada pelas ortografias das línguas ocidentais, nas quais as letras do alfabeto expressam exhaustivamente as identidades das vogais e consoantes (embora nem sempre de maneira direta), mas a informação prosódica é amplamente ausente. Nenhuma informação é dada nas ortografias ocidentais sobre a localização do acento de palavra ou acento frasal, nenhuma informação é dada sobre a melodia da fala (note que o ponto de interrogação expressa a interrogatividade ao invés do tom alto – as perguntas que geralmente terminam em tom baixo são marcadas com ‘?’ (por exemplo, VAN HEUVEN; HAAN, 2000, para holandês).²⁸ Embora os limites das palavras sejam geralmente expressos por escrito (como espaços entre palavras), muitos limites de frases não são marcados como tal.

De acordo com esse *insight*, experimentos com materiais de fala dos quais a informação prosódica foi sistematicamente eliminada demonstraram que a inteligibilidade dificilmente foi comprometida. Por exemplo, Van Heuven (1984) gerou palavras holandesas polissilábicas que deveriam ter acento na primeira, segunda ou terceira sílaba, concatenando difones que foram extraídos exclusivamente de sílabas faladas com um acento frasal nuclear, de modo que quaisquer diferenças entre tônica e átona as sílabas foram efetivamente eliminadas dos estímulos. No entanto, essas palavras não naturais, apresentadas em uma frase-veículo, foram corretamente reconhecidas por ouvintes nativos de holandês em um experimento de *gating*, seja porque uma porção de início mais longa das palavras-alvo foi necessária para que os ouvintes atingissem o limiar de reconhecimento. Um

²⁸ A ortografia espanhola é uma exceção a esta prática geral. Não só marca perguntas e exclamações tanto no final como no início de uma frase escrita (onde o sinal de pontuação no início da frase é impresso de cabeça para baixo), também marca explicitamente a posição de acentos imprevisíveis de palavras por um agudo acento sobre a vogal na sílaba tônica. Nenhuma outra ortografia europeia adotou essa prática, nem qualquer ortografia europeia faz uma tentativa de codificar o tom da palavra lexical – embora os tons lexicais sejam propriedades distintas do norueguês, sueco, servo-croata, galês, lituano e vários outros idiomas.

teste de escuta informal com uma frase holandesa mais longa, inteiramente composta por difones totalmente acentuados, mostrou excelente inteligibilidade de todas as palavras da frase. No entanto, quando a informação segmental é obliterada e apenas a informação prosódica persiste nos materiais de estímulo, a inteligibilidade é praticamente reduzida a zero. Isso foi mostrado para fala filtrada passa-baixa (por exemplo, MILLER; NICELY, 1955) e para fala girada espectralmente (BLESSER, 1966). Uma análise de erro das respostas coletadas para esses materiais segmentados degradados mostra, no entanto, que as propriedades rítmicas, como a localização dos limites prosódicos e sílabas acentuadas ou acentuadas, foram frequentemente preservadas (BLESSER, 1966, para inglês; SVENSSON, 1974, para sueco; VAN HEUVEN, 1984; JONGENBURGER, 1996, para holandês).

Também foi demonstrado que a informação prosódica é altamente resistente ao barulho, muito mais do que a informação segmental. Quando a fala é apresentada no barulho, os erros segmentais são abundantes, mas a estrutura prosódica é preservada nas respostas de erro (por exemplo, SVENSSON, 1974, para sueco; VAN HEUVEN, 1984, para holandês). As razões para a robustez da prosódia são múltiplas. Por um lado, informações prosódicas importantes, como a melodia da fala, a duração da sílaba e as diferenças de intensidade, são codificadas de forma redundante em todo o espectro e são fortemente representadas na frequência fundamental. Quando a fala é apresentada em barulho, a porção de baixa frequência do sinal que contém os primeiros harmônicos não é totalmente obliterada pelo barulho e, portanto, continua a pesar, enquanto as frequências mais altas menos intensas que sinalizam os contrastes segmentais são inaudíveis. Por outro lado, como a prosódia se desenvolve lentamente no tempo, seu curso no tempo é altamente previsível, muito mais do que as mudanças segmentais, de modo que o ouvinte pode facilmente acompanhar as características prosódicas, mesmo que ocasionalmente perca (ver também CUTLER; MCQUEEN, 2014). Essa chamada continuidade prosódica também é uma das razões pelas quais somos capazes de entender com sucesso um falante em meio a uma multidão de falantes concorrentes (o chamado efeito coquetel-festa, por exemplo, BROKX; NOOTEBOOM, 1982).

O papel da prosódia da palavra pode ser diferente em línguas que usam a prosódia para sinalizar distinções lexicais no nível da palavra. A marca registrada do acento é que apenas uma sílaba em uma palavra pode ser mais forte do que todas as outras sílabas. Esta é a chamada propriedade definidora culminante (TRUBETSKOY, 1958; GARDE, 1968) que distingue acento de palavra de tom lexical. Ouvir um acento, então, permite ao ouvinte prever que a próxima sílaba será tipicamente átona (e verificar a hipótese de que a sílaba anterior também foi átona). Esta é uma consequência direta do padrão de acento alternado que muitas vezes é característico das línguas de acento lexical.²⁹ Embora o acento em idiomas como inglês e holandês possa ser usado para diferenciar sequências de

29 O padrão de acento alternado é formalmente explicado na teoria fonológica métrica pelo Princípio do Contorno Obrigatório (OCP), que é uma restrição que não permite (ou desfavorece) a sucessão imediata da mesma propriedade estrutural, como a sucessão imediata de duas sílabas fortes. Muitas vezes, se uma palavra termina com uma sílaba tônica e a próxima palavra começa com uma, a língua resolve esse “conflito de acento” invertendo o padrão de acento em uma das palavras ou excluindo um dos dois acentos (LIBERMAN; PRINCE, 1977; HORNE, 1990; GUSENHOVEN, 1991, para inglês; VAN HEUVEN, 1987; VAN HEUVEN; VAN DONZEL, 1995, para holandês).

segmentos idênticas, como *trusty* ‘trusty’ (com acento inicial) *versus* *trustee* ‘membro do conselho’ (com acento final), o número desses pares de acento mínimo é limitado – pelo menos nas línguas germânicas. Parece improvável, portanto, que a função primária do acento seja diferenciar entre palavras (CUTLER, 1986, FIELD, 2005).³⁰

Em uma língua tonal prototípica, cada sílaba em uma palavra pode carregar pelo menos dois tons diferentes, por exemplo, um tom alto (H) ou um tom baixo (L). Palavras dissilábicas em tal linguagem podem então ser de quatro tipos prosódicos, ou seja, HH, HL, LH e LL. Fundamentalmente, não é o caso de que o tom H ou L esteja associado a uma maior força ou proeminência prosódica: todas as sílabas têm o mesmo peso. A função primária do tom lexical parece ser ajudar a diferenciar as palavras no léxico, da mesma forma que é feito por vogais e consoantes.

Parece haver uma divisão de trabalho, então, entre propriedades prosódicas e segmentais na comunicação falada. As propriedades espectrais de rápida mudança são capazes de diferenciar entre muitas vogais e consoantes, mas são vulneráveis a barulho de fundo e distorção. As propriedades prosódicas de variação mais lenta (por exemplo, tom e duração) contribuem menos para a diferenciação de segmentos, mas são altamente robustas contra barulho e distorção. Como resultado, a importância relativa das propriedades segmentais e prosódicas para a inteligibilidade da fala irá variar dependendo das circunstâncias comunicativas. Se os segmentos forem mal definidos, por exemplo, devido a barulho no canal de comunicação (barulho, distorção eletrônica, fala de computador ou sotaque estrangeiro), a importância da prosódia aumentará. Nas próximas subseções, apresentarei alguns experimentos que fiz na década de 1980, que ilustram o ponto que estou tentando defender. Os experimentos foram feitos com materiais e ouvintes holandeses, mas espero que resultados semelhantes sejam obtidos para outras línguas germânicas ocidentais, incluindo o inglês.

3.2. Van Heuven (1984)

Um falante nativo do sexo masculino produziu 26 palavras holandesas variando em comprimento entre duas e cinco sílabas, com acento lexical sistematicamente dividido igualmente nas cinco posições máximas. As palavras-alvo foram produzidas em duas frases diferentes, de modo que foram pronunciadas com acento frasal (ou seja, em foco) em uma ocasião e sem acento frasal (ou seja, fora de foco) na outra. As palavras foram extraídas de seu contexto falado e apresentadas para reconhecimento em uma tarefa de *gating*, com ou sem adição de barulho ANSI. Na tarefa de *gating*, o ouvinte ouviu primeiro o CV inicial; a cada apresentação sucessiva, a porção audível da palavra alvo era incrementada com o próximo fonema, até que, na última passagem, a palavra inteira fosse audível.

³⁰ Isso é verdade, *a fortiori*, de línguas com acento fixo, como as línguas fino-úgricas com acento inicial fixo ou o francês com acento final fixo. A função contrastiva do acento pode ser mais importante em idiomas como russo e grego, para os quais o acento é amplamente imprevisível e contém muitos pares de acentos mínimos (por exemplo, JOURAVLEV; LUPKER, 2015, para russo; TZAKOSTA, 2004, para grego).

Após cada passagem (“portão”), os ouvintes escreviam seu melhor palpite sobre qual palavra eles ouviriam. As palavras-alvo foram escolhidas de modo que pudessem ser distinguidas exclusivamente de todas as outras palavras no léxico holandês com base aproximadamente na primeira metade da forma da palavra, de modo que (sob boas condições de audição) o reconhecimento da palavra pudesse ocorrer antes de o final da palavra ser disponibilizado. Os materiais foram bloqueados em quatro grupos pareados de ouvintes holandeses, com as versões com e sem acento frasal, e com e sem barulho adicionado, bloqueados sobre os ouvintes, de modo que um ouvinte ouvisse apenas uma versão de cada palavra. Os resultados podem ser vistos na Figura 5.

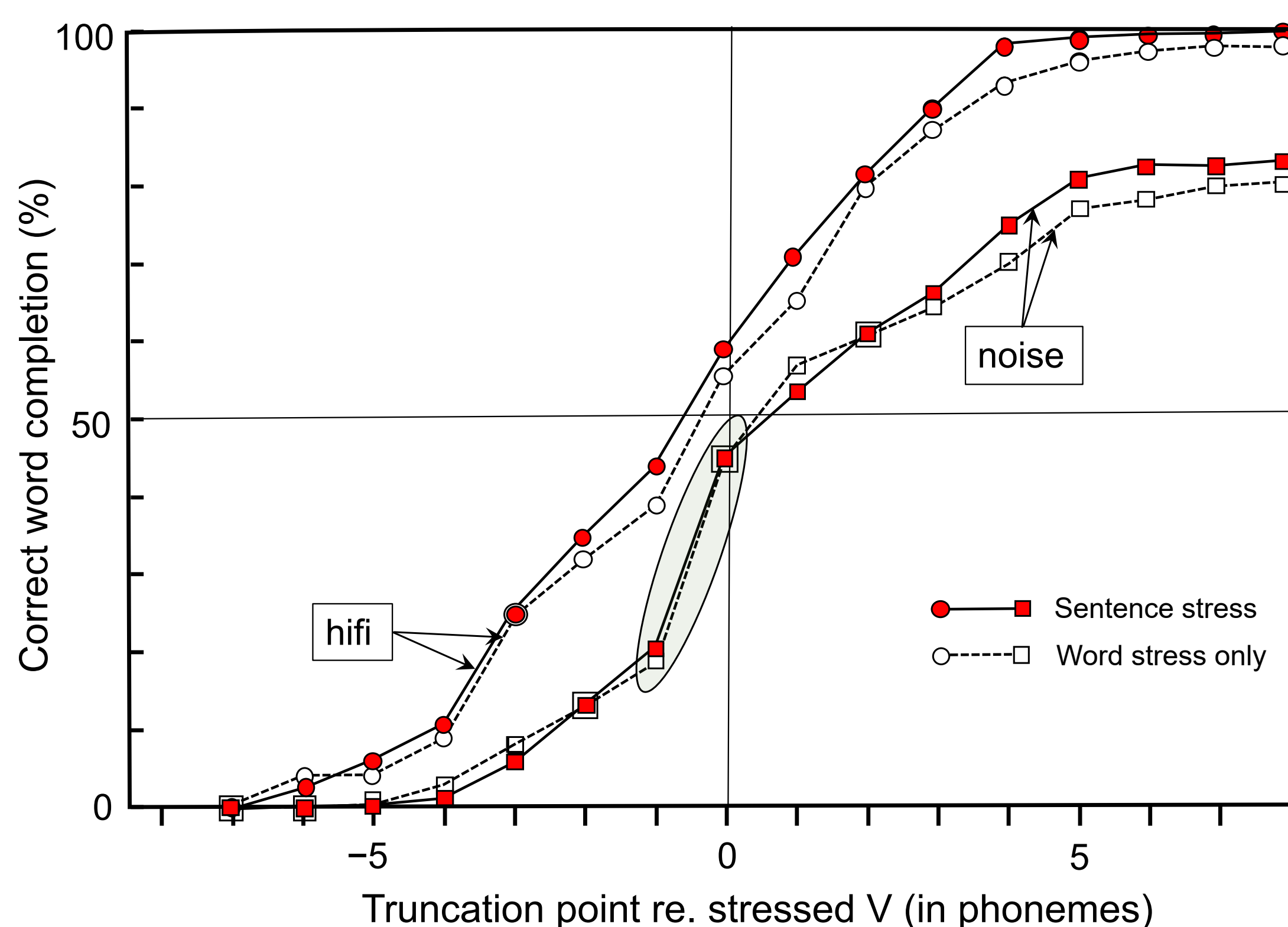


Figura 5: Preenchimento cumulativo de palavras corretas (%) em uma tarefa de *gating* com 26 palavras holandesas que diferem em comprimento e em posição de acento. As palavras foram produzidas com (+acento) ou sem (-acento) e foram apresentadas tanto no silêncio (*hifi*) quanto no barulho. Cada divisão ao longo do eixo X representa a adição de um fonema na tarefa de *gating*. As palavras são alinhadas na localização da vogal tônica (*gate* 0). Os fonemas que precedem a vogal tônica têm números de *gating* negativos, enquanto os que seguem a vogal tônica têm números positivos.

A Figura 5 mostra que todas as palavras-alvo apresentadas no silêncio foram preenchidas corretamente (reconhecidas) após o quarto fonema após a vogal tônica, caso a palavra tivesse sido pronunciada com acento frasal. Quando realizado sem acento frasal (de modo que a vogal tônica não foi marcada por um movimento de tom, mas apenas por comprimento e intensidade), as pontuações de conclusão são apenas marginalmente mais baixas. É importante ressaltar que a curva de reconhecimento sigmóide mostra incrementos constantes no intervalo entre -4 e +4 de segmentos adicionados ao redor da sílaba tônica. Quando os mesmos alvos foram apresentados no barulho, no entanto, o reconhecimento de palavras foi severamente atrasado (a curva sobe lentamente e mais tarde) e nunca excede 80% de conclusão correta (onde alvos com acento frasal

se saem um pouco melhor do que aqueles sem acento frasal). Fundamentalmente, a pontuação cumulativa de conclusão mostra um salto quântico no momento em que a vogal tônica se torna audível (marcada na Figura 1 pela elipse verde). A sílaba tônica, que é mais longa e mais audível que as sílabas átonas anteriores (e posteriores), forma uma ilha de certeza em um mar de barulhos para o qual o ouvinte salta.

O efeito benéfico de ouvir a sílaba tônica pode ser observado (ainda) mais claramente se separarmos os resultados do experimento daqueles que foram obtidos para palavras com acento lexical inicial contra aquelas com acento final. As palavras com acento inicial têm a posição da sílaba tônica no ou antes do ponto de singularidade lexical. Em palavras com acento final, a vogal tônica ocorre após o ponto de singularidade. Na palavra inglesa “elephant”, por exemplo, o ponto de singularidade lexical está no /f/ em /'eləf/; não há outras palavras no léxico inglês que comecem com /'eləf/ além de elefante (e suas derivações). O acento inicial nesta palavra precede o ponto de singularidade, de modo que elefante se qualifica como uma palavra de acento inicial.

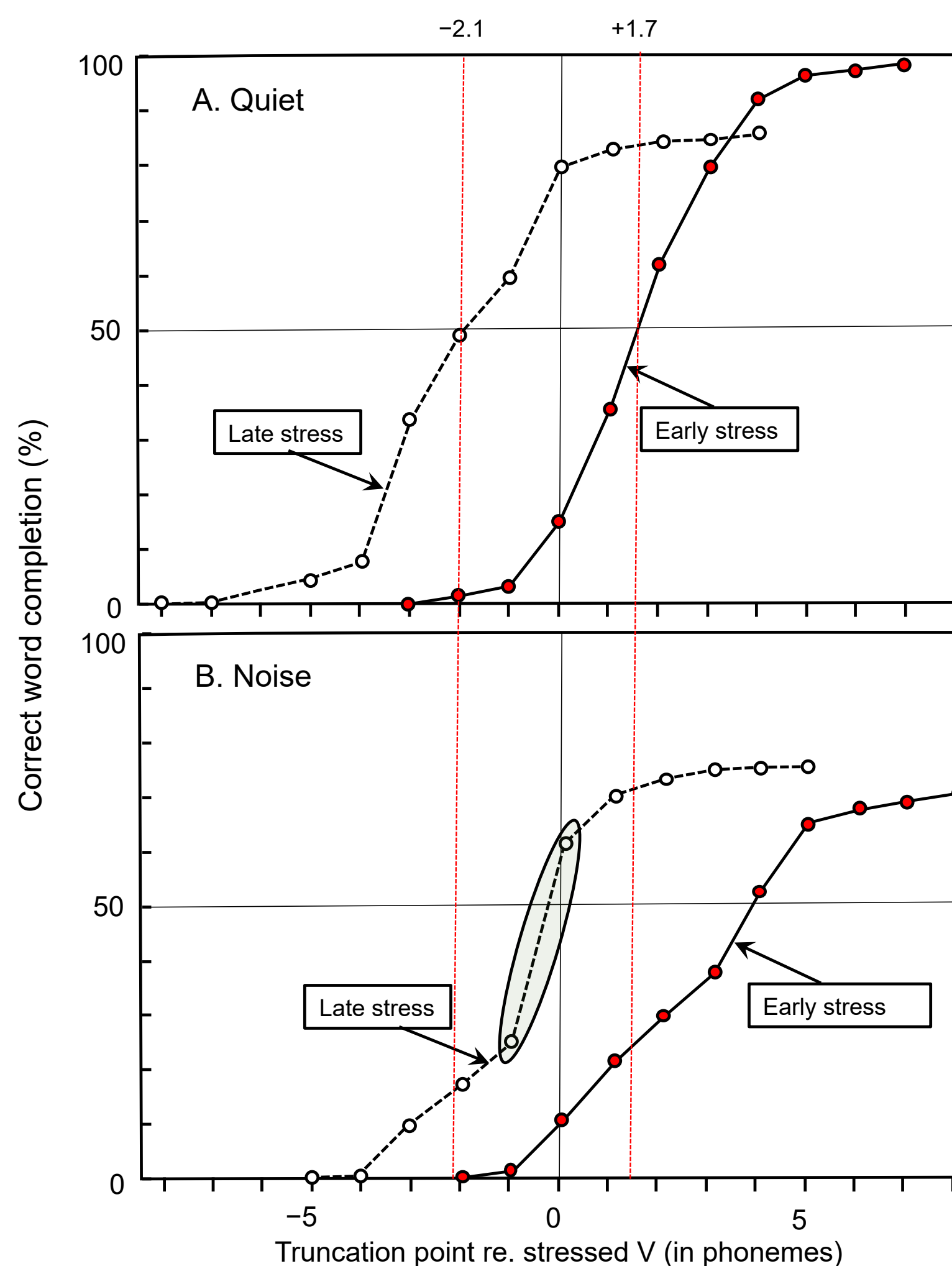


Figura 6: Preenchimento cumulativo de palavras corretas (%) em função do ponto de truncamento em uma tarefa de *gating*. O painel superior mostra os resultados para palavras apresentadas no silêncio, o painel inferior representa as mesmas palavras apresentadas no barulho. Em cada painel, as palavras com acento inicial e tardio (acento antes vs. depois do ponto de singularidade lexical) são plotadas separadamente. Os pontos de singularidade lexical (média) são marcados por uma linha vertical, em -2,1 fonema para acento tardio e em +1,7 fonema para acento precoce.

Quando apresentados em silêncio, há um aumento de acertos na tarefa de completar palavras de forma constante como antes na Figura 5. Quando apresentados em barulho, os incrementos também são aproximadamente do mesmo tamanho, pelo menos para palavras com acento inicial. No entanto, quando a vogal tônica ocorre após o ponto de unicidade lexical, vemos o salto quântico nas realizações corretas da tarefa de completar palavras (de 26% para 62% de acertos) no momento em que a vogal tônica se torna audível (marcada pela elipse verde no gráfico da Figura 6).

Podemos interpretar os resultados deste experimento da seguinte forma. Ouvir a sílaba tônica (ou a vogal tônica dentro dela) não fornece nenhuma informação especial (em relação a outras sílabas ou segmentos) quando o reconhecimento de palavras ocorre em condições favoráveis de escuta (no silêncio). Quando as palavras precisam ser reconhecidas em condições adversas de audição, como no barulho, ouvir a sílaba tônica de repente revela uma quantidade desproporcional de informações quanto à identidade da palavra que se desenvolve no tempo. As informações anteriores se encaixam quando o ouvinte percebe que o acento lexical está em algum lugar mais tarde na palavra, ou seja, não na posição inicial da palavra “padrão”. No entanto, quando o acento lexical é o padrão inicial (em línguas germânicas tipicamente na primeira sílaba), ouvir o acento não fornece nenhuma informação que o ouvinte já não esperava.

É importante perceber que a contribuição especial do acento para o reconhecimento de palavras (em holandês, inglês e outras línguas germânicas) reside nos acentos não padrão, ou seja, nos acentos não iniciais. O acento vem à tona apenas em condições adversas de escuta, quando uma propriedade redundante que normalmente pode ser reconstruída a partir da sequência de segmentos que compõe a palavra é a única pista confiável que resta.

Em um segundo experimento no artigo de 1984, usei apenas palavras trissilábicas CVCVCVC, dez com acento lexical inicial, dez com acento medial e outras dez com acento final. As palavras foram sintetizadas a partir de difonemas, ou seja, blocos de construção de síntese de fala que consistem em aproximadamente a segunda metade de um fonema seguido sem problemas pela primeira metade do próximo fonema. Os difones que usei foram extraídos de sílabas tônicas hiperarticuladas em palavras sem sentido com uma configuração de pitch H*L (ou seja, acento frasal). Os difones foram monotonizados e a intensidade normalizada. Como resultado desse esquema de síntese, todas as sílabas e todos os segmentos nelas eram igualmente sugestivos de acento – de modo que o acento não era marcado prosodicamente. Em uma segunda condição, adicionei prosódia às palavras difonéticas de maneira controlada, gerando uma configuração fixa de High e Low na sílaba tônica e reduzindo a duração das sílabas átonas em 20%. As 2 x 30 palavras foram distribuídas em blocos e apresentadas em uma tarefa de *gating* a dois grupos de ouvintes holandeses de tal forma que cada ouvinte ouviu cada palavra uma vez, sendo 15 palavras com e 15 palavras sem marcação prosódica explícita de acento. Nos filtros sucessivos, a porção audível das palavras era incrementada por um

difone adicionado, até que após o filtro final as palavras inteiras pudessem ser ouvidas. A Figura 7 mostra os resultados.

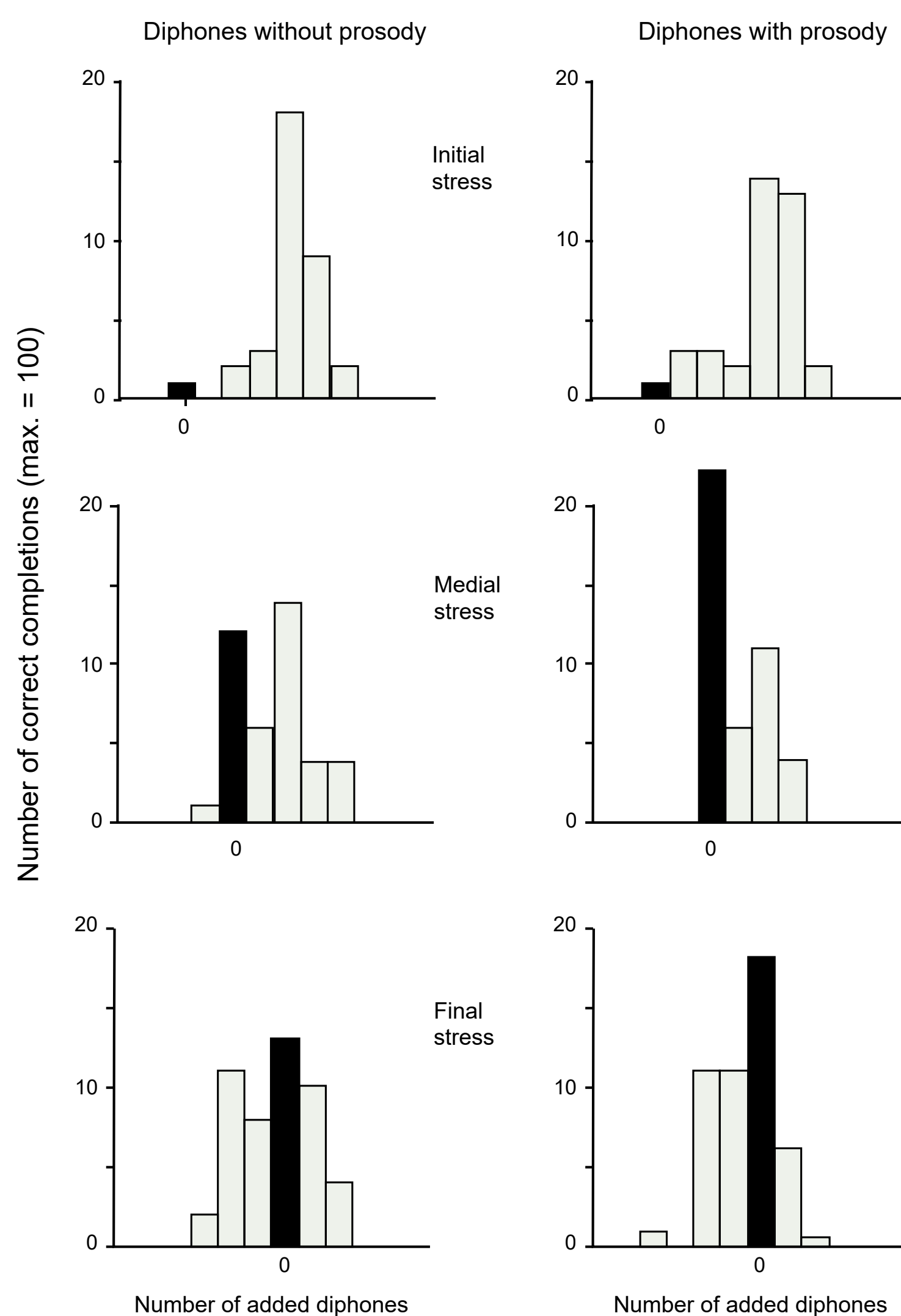


Figura 7: Número de palavras preenchidas corretamente em função da extensão da sequência difonética tornada audível, discriminada por tipo de fala (síntese difonética com ou sem prosódia) e por posição do acento lexical.

Os resultados não mostram nenhum papel especial em ouvir o difone CV acentuado na posição inicial (padrão). No entanto, quando o difone CV tônico ocorreu na sílaba medial ou (mais ainda) na sílaba final, vemos um incremento desproporcional nas compleções corretas no momento em que o difone CV tônico se tornou audível. Isso é visto quando o acento não foi marcado prosodicamente (coluna da esquerda na Figura 7), e fica muito mais claro quando o acento não inicial é marcado prosodicamente pela organização temporal e – especialmente – pela configuração do pitch H*L marcando a sentença acentuada.

Novamente, vimos que o ouvinte holandês assume que as palavras começam com uma sílaba tônica. No entanto, assim que o ouvinte percebe um acento (frasal) em uma posição não padrão (aqui na posição medial, e ainda mais fortemente, quando na posição final), o reconhecimento da palavra é concluído. É essencial perceber que esse papel especial do acento não inicial (em holandês) vem à tona apenas quando as palavras precisam ser reconhecidas em condições adversas de audição – como, neste caso, na síntese de difones com baixa qualidade segmental.

3.3 Van Heuven (1985)

Nos experimentos descritos acima, a marcação prosódica da posição do acento foi enfraquecida (acento de palavra em vez de acento frasal) ou omitida completamente. Uma maneira diferente de investigar a importância do acento para o reconhecimento de palavras é produzir deliberadamente o acento na sílaba errada. Quanto mais perturbador for esse acento mal atribuído no processo de reconhecimento de palavras, mais importante deve ser o acento. Se um acento atribuído incorretamente passa despercebido (como acontece em indonésio, veja VAN ZANTEN; VAN HEUVEN, 1998, 2004), então, a conclusão seria que o acento não desempenha nenhum papel no reconhecimento de palavras.

Neste experimento, eu sintetizei 24 palavras holandesas trissilábicas a partir de difones de baixa qualidade. Oito palavras apresentaram acento lexical na primeira sílaba, oito na posição medial e oito na última sílaba. Cada palavra foi então sintetizada em três versões diferentes no final da frase-veículo 'A próxima palavra é...', com uma marcação prosódica (uma configuração de tom H*L) na primeira, segunda ou terceira sílaba da palavra-alvo. Para cada palavra, então, havia uma versão acentuada correta e duas versões incorretas. Os materiais foram divididos em três partes iguais usando um desenho de parcelas divididas (apenas uma versão de cada palavra alvo seria ouvida por um ouvinte) e apresentados a doze ouvintes holandeses em uma tarefa de repetição rápida. O participante foi solicitado a repetir a palavra que acabara de ouvir enquanto corrigia um possível acento errado. A correção da resposta foi determinada, bem como o tempo que o participante levou para começar a produzir sua resposta – medido a partir do deslocamento do estímulo. Normalmente, o resultado era que o participante reconhecia a palavra e a corrigia na hora, ou – alternativamente – não reconhecia a palavra e permanecia em silêncio. Os resultados são vistos na Tabela 3.

Quando acentuadas corretamente, as palavras-alvo foram repetidas com uma taxa de sucesso entre 63 e 88%, dependendo da posição do acento lexical. Isso mostra que a qualidade segmental da síntese do difone era pobre. No entanto, quando o acento foi deslocado para uma sílaba errada, as chances de reconhecer a palavra foram reduzidas pela metade. Os alvos acentuados corretamente foram repetidos com sucesso com latências cerca de 120 a 300 ms mais curtas do que os alvos acentuados incorretamente. Como foi visto em meus experimentos anteriores, enfatizar incorretamente uma sílaba inicial era claramente mais prejudicial do que ouvir um acento errado em uma sílaba posterior.

Esses resultados demonstram, mais uma vez, que o reconhecimento de palavras (em holandês) é prejudicado principalmente quando uma sílaba inicial átona é acentuada incorretamente. Isso sugere que os ouvintes holandeses (e também os ouvintes ingleses) têm uma estratégia de reconhecimento padrão que assume que as palavras começam com uma sílaba tônica (viés de acento de início de

palavra). Quando ouvem um acento em uma sílaba inicial, sua expectativa é confirmada, de modo que palavras que começam com uma sílaba átona não são mais candidatas a reconhecimento.

Tabela 3: Palavras repetidas corretamente (%) discriminadas por posição lexical e acento percebido. As palavras acentuadas corretamente estão em células verdes ao longo da diagonal principal. O deslocamento para a direita de um acento lexical inicial é menos perturbador do que o deslocamento (para a esquerda) de um acento lexical não inicial. O painel inferior exibe a latência de repetição (em ms) para as palavras que foram repetidas corretamente.

lexical stress on syll. #	stress synthesized on syll #		
	1	2	3
1	66%	44%	56%
2	34%	81%	31%
3	34%	25%	63%

lexical stress on syll. #	stress synthesized on syll #		
	1	2	3
1	1500	1800	1650
2	1630	1510	1640
3	1700	1690	1390

A conclusão geral deste experimento, e dos experimentos anteriores, é que o acento correto (posição) é um pré-requisito importante para o reconhecimento de palavras bem-sucedido quando a qualidade segmental dos estímulos é ruim, como em síntese de fala de baixa qualidade e em fala não-nativa.

3.4. Wang *et al.* (2011)

Termino esta pesquisa com um estudo sobre um tipo diferente de sistema prosódico de palavras, ou seja, tom lexical em vez de acento. O mandarim chinês tem quatro tons lexicais pelos quais sequências idênticas de segmentos ainda podem ser diferenciadas como palavras separadas. Por exemplo, a sílaba /ma/ significa ‘mãe’ quando falada com tom 1 (High), ‘cânhamo’ com tom 2 (Mid-High), ‘cavalo’ com tom 3 (Low dip) e ‘repreensão’ com tom 4 (High-Fall).

Sabíamos de estudos anteriores que a fala em mandarim pode ser facilmente compreendida quando as informações sobre os tons lexicais são eliminadas do enunciado. Por exemplo, Xin (2007) sintetizou enunciados em mandarim a partir de difones holandeses prosodicamente neutros e substituiu os contornos de tom em todas as palavras de conteúdo por movimentos simples de tom ascendente e descendente (que não fazem parte do inventário de tons do mandarim). Em comparação com a versão em que os tons originais do mandarim foram mantidos, a inteligibilidade para os ouvintes nativos do mandarim foi reduzida em apenas 4%.

Em Wang *et al.* (2011), pretendemos testar a hipótese de que a contribuição da prosódia para o reconhecimento de palavras não é fixa, mas assume um papel mais importante à medida que a qualidade segmental se deteriora. Em circunstâncias normais de escuta, os segmentos são mais

importantes, enquanto a prosódia por si só não permite o reconhecimento de palavras. No entanto, a prosódia contribuirá substancialmente quando a qualidade segmental se deteriorar.

Realizou-se um teste de inteligibilidade com palavras em frases significativas, utilizando o teste Speech-Perception in-Noise (teste SPIN, KALIKOW et al., 1977), que foi adaptado para mandarim (e 14 outras línguas/dialetos chineses) por Tang e Van Heuven (2009). O teste compreendeu 60 frases curtas e cotidianas. Depois de ouvir uma frase, o ouvinte tem que escrever a última palavra da frase (que é bastante previsível pelo contexto anterior), como em 'Mantenha seu braço quebrado em uma tipóia' (onde tipóia é a palavra alvo previsível). Os materiais (ver TAG, 2009 para detalhes) foram falados por um homem e uma mulher que falavam mandarim como L1. Criamos seis condições variando a qualidade segmental e a prosódia. A qualidade segmental foi variada em três etapas: alta qualidade (HQ, não filtrada), qualidade média (MQ, fala filtrada com filtro-baixo com corte em 1000 Hz) e qualidade baixa (LQ, filtrada filtro-baixo com frequência de corte de 300 Hz). A prosódia foi variada deixando a melodia da fala original no lugar, ou monotonizando a fala pela análise-ressíntese PSOLA em um tom fixo de 100 Hz para a fala masculina e 200 Hz para a fala feminina.

Os materiais foram distribuídos em blocos de seis grupos de ouvintes de mandarim (24 ao todo) em um desenho de quadrado latino, de modo que um ouvinte ouvisse cada sentença apenas uma vez e ainda recebesse um número igual de frases em cada uma das seis condições. Os resultados são mostrados na Figura 8.

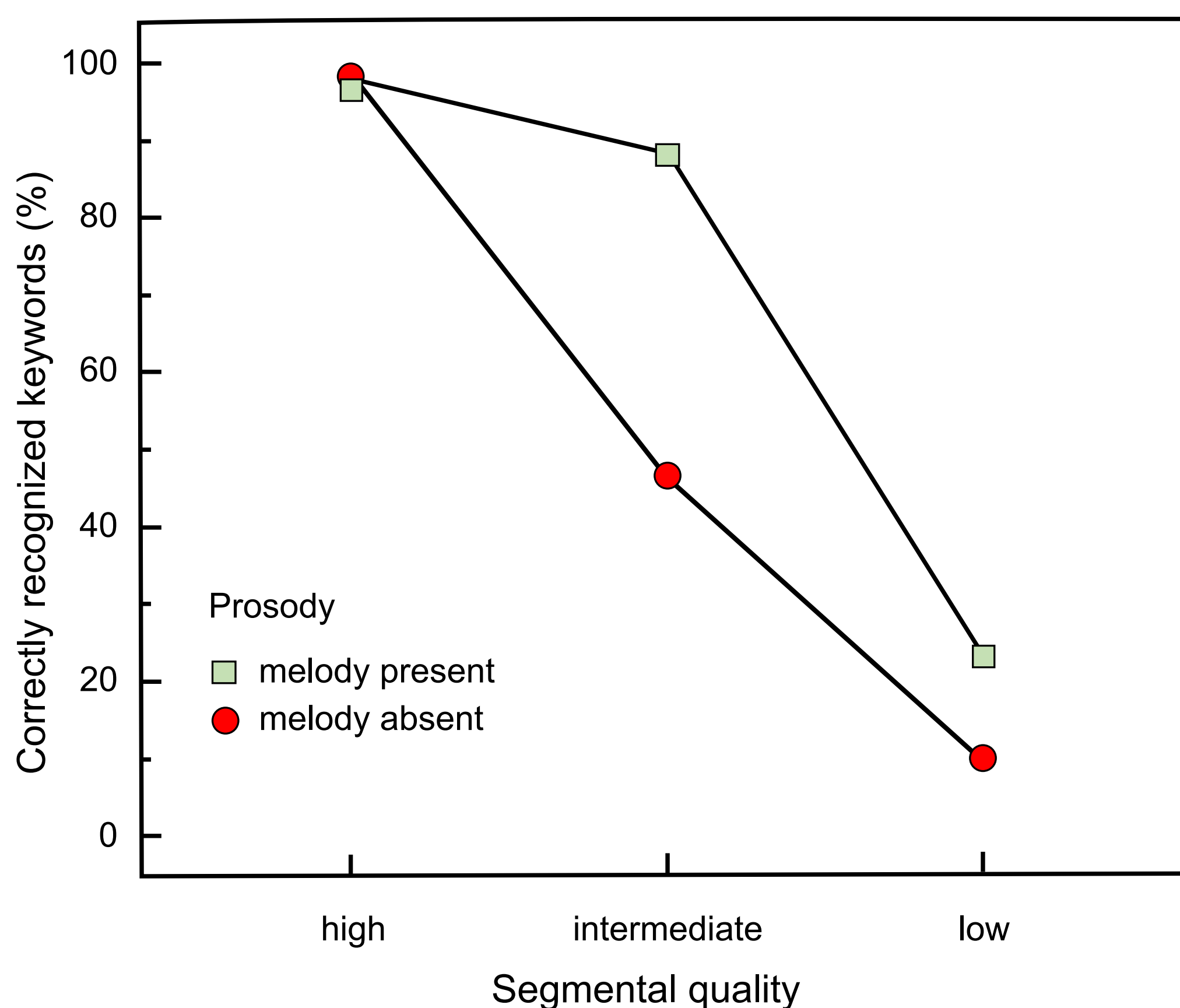


Figura 8: Palavras-chave corretamente reconhecidas (%) no teste SPIN de mandarim, em função da qualidade segmental (alta, média, baixa) e presença vs. ausência de melodia de fala.

Quando a qualidade segmental é boa, a inteligibilidade está próxima do teto, esteja a melodia presente ou ausente. Quando a qualidade segmental é baixa, quase nenhuma inteligibilidade permanece. Esse achado imita resultados anteriores relatados por Zhang et al. (1981). A parte importante e nova dos resultados é que a inteligibilidade dos materiais de qualidade média depende da disponibilidade da prosódia. Quando os tons lexicais são mantidos, a inteligibilidade das frases permanece muito alta (melhor que 90%). No entanto, quando o tom é substituído por um tom monótono fixo, a inteligibilidade é reduzida para menos de 50%.³¹

Esses resultados dão forte suporte à ideia de que a contribuição da prosódia para a inteligibilidade da fala não é fixa, mas varia com a qualidade da informação segmentada. A qualidade segmental é de importância primordial. Se a qualidade segmental for ruim, não será possível o reconhecimento de palavras. Se a qualidade segmental for perfeita, não há valor agregado para o pitch. Mas o pitch realmente ajuda quando a qualidade segmental é intermediária (e mesmo insuficiente por si só).

4. Por que a acentuação incorreta é tão prejudicial?

O que vimos até agora é que seria incorreto dizer em geral que a prosódia é mais importante para a inteligibilidade da fala do que os segmentos (ou vice-versa). Os segmentos têm precedência sobre a prosódia quando são bem definidos. Por outro lado, a prosódia assume maior importância à medida que a informação segmental se deteriora. Esses efeitos básicos foram encontrados tanto para as línguas tonais quanto para acentuais, de modo que podemos ter aqui um princípio universal da linguagem.

A causa provável para o sucesso da prosódia de palavras em más condições de escuta é sua robustez. As diferenças mais importantes entre os segmentos estão na distribuição espectral. As vogais de uma língua, muitas vezes, diferem apenas em sua composição espectral, ou seja, a localização e a mudança dinâmica (no caso de ditongos e semivogais) dos formantes, as bandas de frequência relativamente estreitas que são amplificadas por ressonância no trato vocal, conforme determinado pelo grau de abertura da mandíbula (correspondente à frequência central da ressonância mais baixa), a localização do local de constricção (dimensão frente-trás, correspondente à segunda ressonância mais baixa, o grau de arredondamento dos lábios, etc. (consulte LADEFOGED; DISNER, 2012 para detalhes). Algumas línguas também usam duração para diferenciar entre vogais longas e curtas. Como as vogais geralmente têm (muito) mais intensidade do que as consoantes circundantes, não é difícil para um ouvinte determinar aproximadamente a duração do som alto da vogal, mesmo em meio ao barulho de fundo e vozes concorrentes. Diferenças espectrais (nas rajadas de barulho) também são as principais pistas para distinguir o local de articulação para oclusivas, fricativas e africadas, ou seja, as obstruintes ou consoantes verdadeiras.

³¹ No experimento, a ausência de tom foi operacionalizada pela monotonização da melodia da fala, que é de fato o correlato mais forte do tom. No entanto, os tons do mandarim distinguem-se até certo ponto por outras propriedades acústicas que não o tom. Por exemplo, o tom 3 é cerca de 50% mais longo do que os tons 2 e 4, enquanto o tom 4 é mais curto. Além disso, o contorno de intensidade das sílabas dará informações sobre o tom. O efeito da ausência de tom teria sido ainda mais forte se tivéssemos eliminado todas as informações de pitch.

Más condições de escuta, quando há muito barulho de fundo, quando há vozes concorrentes (como quando se fala em uma multidão), qualidade de fala degradada devido à filtragem e distorção harmônica em sistemas de telecomunicações, a informação espectral é praticamente obliterada, de modo que a identificação das consoantes, e especialmente das vogais, torna-se em grande parte uma questão de adivinhação.

A informação acústica sobre a melodia da frase é provavelmente o aspecto mais importante e resistente ao barulho da prosódia da fala. A melodia é transportada pela taxa de repetição da vibração das pregas vocais, que gera um sinal de fonte harmônica complexa. Ao longo da frase, o *pitch* desse sinal de fonte sobe e desce, marcando limites e acentos frasais. A informação sobre o *pitch* é onipresente no sinal de fala. A altura pode ser determinada no domínio do tempo, estimando o lapso de tempo entre os pulsos glóticos sucessivos, cujos eventos são altos o suficiente para resistir a barulhos de fundo mesmo severos. O *pitch* também pode ser determinado diretamente pela localização da frequência fundamental da onda glótica, mas mesmo quando a frequência fundamental está ausente do sinal (devido à filtragem, como na fala telefônica), o *pitch* pode ser facilmente reconstituído estimando a distância entre harmônicos adjacentes em algum lugar no espectro (veja LADEFOGED; JOHNSON, 2011 para mais informações básicas). Outras propriedades prosódicas que são relativamente resistentes a barulhos e distorções são as diferenças de duração e intensidade, como observado acima.

Os ouvintes sabem por experiência que, em circunstâncias adversas de escuta, a informação prosódica é geralmente uma fonte de informação mais confiável do que a informação segmental/espectral. Eles também sabem que os falantes nativos de sua língua fazem uso adequado da prosódia, marcando os limites da frase e do enunciado e enfatizando as sílabas corretas no nível da palavra e as palavras apropriadas no nível da frase. É razoável, portanto, supor que, quando a situação comunicativa se deteriorar, o ouvinte confiará fortemente na informação prosódica que sobrevive e reinterpretará a informação espectral disponível para encontrar candidatos plausíveis ao reconhecimento de palavras.

Ouvir fala com sotaque estrangeiro é uma forma de ouvir fala na qual a informação espectral é severamente distorcida. Falantes não nativos normalmente usam a vogal ou consoante disponível mais próxima em seu inventário de sons da língua nativa como um substituto para as vogais e consoantes na língua estrangeira (isso é o que chamamos de sotaque estrangeiro). Como resultado desse processo de substituição, o ouvinte nativo nunca tem certeza de qual vogal ou consoante o falante pretendia, de modo que a identificação segmental se torna uma questão de adivinhação, de um tipo semelhante ao que encontramos ao ouvir a fala no barulho. Em tais circunstâncias, o ouvinte confiará principalmente na informação prosódica, que então define as restrições dentro das quais a informação segmental disponível é ajustada de maneira otimizada (ou “reinterpretada”

para se adequar à prosódia). Essa estratégia funciona bem quando o falante usa o mesmo idioma do ouvinte, mas sai pela culatra quando o idioma do falante (ou sotaque estrangeiro) não corresponde ao idioma do ouvinte.

Como um bom exemplo desse processo, Bansal (1966) mostrou que ouvintes nativos do inglês britânico reinterpretem segmentos de modo a se adequar ao padrão rítmico percebido incorretamente em inglês pronunciado por falantes indianos. Para o ouvido inglês-britânico, o acento nas palavras *caractere* e *escrita* é percebido na segunda sílaba quando pronunciada por um falante indiano de inglês. Dado que a pronúncia segmentada do inglês indiano se desvia fortemente da norma britânica, faz sentido para o ouvinte britânico reinterpretar a informação como “director” e “retain”, respectivamente (erros relatados por Bansal, ver também CUTLER, 1984).

Da mesma forma, no nível da sentença, remover informações prosódicas dos enunciados é amplamente inconsequente. No entanto, quando um conflito é criado entre a estrutura sintático-semântica da frase e a prosódia, os ouvintes tendem a atribuir mais peso à informação prosódica se a qualidade da fala for ruim (por exemplo, WINGFIELD, 1975).

Van Heuven e Hagman (1988) tabularam o número de palavras holandesas (monomórficas) em termos de comprimento de palavra (expresso em número de sílabas) e posição do acento primário. Isso nos permitiu subdividir o léxico holandês em 26 tipos rítmicos.

Tabela 4: Tamanho da sublexica definida pelo padrão rítmico em palavras monomorfêmicas holandesas (após VAN HEUVEN; HAGMAN, 1988). Os padrões de palavras rítmicas são definidos pelo comprimento da palavra em sílabas e pela posição do acento primário.

Tamanho de palavra	Posição do acento primário em palavras (monomorfêmicas)						Total
	1	2	3	4	5	6	
1	4,284						4,284
2	2,703	1,682					4,385
3	408	808	1,032				2,248
4	56	128	474	314			972
5	7	0	43	76	52		178
6	0	0	0	6	9	5	20
>6	0	0	0	0	2	0	2
Total	7,458	2,618	1,549	396	63	5	12,089

As contagens mostram que a maioria das palavras holandesas tem acento na primeira (ou única) sílaba. Isso provocará o viés de acento inicial pelo qual os ouvintes assumem que ouvir um acento sinaliza o início de uma nova palavra. Os dados também mostram que o acento está sempre em uma das três últimas sílabas da palavra (a janela universal de três sílabas identificada por HAYES, 1995). Obviamente, se o ouvinte conhece o tipo rítmico da palavra que acabou de ouvir (ou está

em processo de audição), o espaço de busca lexical é muito menor do que o léxico de 12.000 itens como um todo, de modo que as palavras ainda podem ser reconhecidas de segmentos cada vez menos claramente articulados. Uma estratégia lucrativa para um ouvinte holandês seria decidir, o quanto antes, se a palavra começa com uma sílaba tônica ou não. Dado que a maioria das palavras holandesas tem acento inicial, o ouvinte usará a presença de um movimento de *pitch* na sílaba inicial como uma confirmação clara de que deve procurar palavras que começam com uma sílaba tônica. Um acento incorreto (frase) no início de uma palavra holandesa (e da palavra inglesa também), bloqueia o acesso aos cerca de 4.500 candidatos a reconhecimento que começam com uma sílaba átona e será extremamente prejudicial para o reconhecimento bem-sucedido se a palavra alvo não tiver acento lexical inicial (como foi mostrado na seção 3.3). Por outro lado, não ouvir uma frase acentuada na primeira sílaba de uma palavra não fornece nenhuma informação negativa convincente; é sempre possível que a palavra alvo seja comunicativamente sem importância, de modo que o falante omita o acento frasal dela. O acesso à parte tônica inicial do léxico não será bloqueado até que o ouvinte ouça uma frase acentuada em algum lugar mais adiante na palavra. Na fala de baixa qualidade, como na fala não nativa, o reconhecimento de palavras geralmente é atrasado (como mostrado nas Figuras 5 e 6 acima). Muitas vezes, o final da palavra foi recebido pelo ouvinte, enquanto o reconhecimento da palavra ainda está em andamento. Nesses casos, saber se o acento da palavra é ou não inicial é uma dica muito útil para limitar o espaço de busca, ou para verificar se o reconhecimento provisório se encaixa no padrão de acento que faz parte da representação lexical da palavra candidata.

Sabemos que ouvir a diferença entre início de palavras tônicas e átonas é usado por ouvintes holandeses mesmo nos estágios iniciais do processo de reconhecimento de palavras, da maneira descrita no parágrafo anterior (veja VAN HEUVEN, 1988 para dados em uma tarefa de *on-line shading*). Os ouvintes em inglês não são sensíveis à diferença entre vogais plenas tônicas e átonas (ou sílabas). Eles só consideram uma sílaba inicial átona se a vogal estiver completa reduzida a schwa ou [ɪ] (como na Tabela 1). É por isso que ambos os candidatos a reconhecimento 'trusty' 'confiável' e trustee 'board member' (ou 'forebear' 'ancestral' vs. for 'bear' 'tolerar, suportar') são compatíveis com a audição de uma primeira sílaba tônica e átona para ouvintes de inglês (ver CUTLER, 1986: Forbear é um homófono). Após o estágio inicial do processo de reconhecimento de palavras (chamada fase de acesso lexical), no entanto, a diferença de acento inicial e não inicial também é usada de forma lucrativa em inglês para distinguir alternativas (veja CUTLER, 2005 para uma revisão abrangente da diferença do processamento de informações de acento em inglês e holandês).

O que foi dito aqui sobre o papel do acento lexical em línguas como o inglês e o holandês (línguas com uma variedade de padrões de acento) também deve se aplicar ao papel do *pitch* lexical, como no mandarim (e em muitas outras línguas do Sudeste Asiático). Aqui, também, as informações no tom do léxico serão altamente resistentes a barulhos e terão prioridade sobre as informações segmentais em fala de baixa qualidade (ou fala com sotaque não nativo).

5. Conclusão

Neste capítulo, mostrei que, geralmente, a prosódia pode ser omitida inteiramente do texto falado (como de fato é nas ortografias da maioria das línguas) sem nunca causar nenhum problema sério de comunicação. A prosódia é amplamente redundante em relação à informação transmitida pelas vogais e consoantes que compõem as frases faladas. No entanto, os pedagogos de língua estrangeira insistem que acertar a prosódia na língua-alvo é essencial para a inteligibilidade e compreensibilidade do aluno e, portanto, deve receber a mais alta prioridade como objetivo de aprendizagem, mais do que pronunciar corretamente as vogais e as consoantes. Isso parece um paradoxo.

O paradoxo pode ser resolvido se percebermos que a prosódia é redundante na fala nativa de alta qualidade e, portanto, pode ser omitida impunemente. De fato, o poder restritivo de informações segmentadas confiáveis e de alta qualidade é tão forte que erros prosódicos podem ser detectados e ignorados pelo ouvinte nativo (embora possa ocorrer algum atraso de processamento). No entanto, na fala de má qualidade, por exemplo no barulho, no meio de vozes concorrentes ou em um canal de telecomunicações defeituoso, ou se o falante não é nativo com sotaque estrangeiro, o ouvinte nativo confia na prosódia e reinterpreta segmentos mal definidos como informações de modo a se adequar ao padrão prosódico que foi percebido. Essa estratégia de processamento é formada pela experiência do ouvinte com a diferença de confiabilidade entre informação segmental e prosódica. Enquanto a informação segmental é facilmente destruída e altamente vulnerável ao barulho e à distorção acústica, a informação prosódica é robusta e resistente ao barulho. Ouvir a fala não nativa significa ouvir segmentos mal definidos. Nesses casos, a prosódia tem precedência sobre a informação segmental.

Isso deixa duas avenidas abertas no ensino de línguas estrangeiras. Uma solução seria dar prioridade máxima para melhorar as vogais e consoantes do aluno para que elas assumissem uma qualidade nativa. Nesse caso, a prosódia (acento, entonação) pode permanecer defeituosa sem muito dano comunicativo. A solução alternativa seria dar prioridade para acertar a prosódia. Se pelo menos a prosódia estiver correta, um mínimo de desvio segmental pode ser tolerado pelo ouvinte e a comunicação não será interrompida. A segunda alternativa é provavelmente mais barata (requer menos tempo e esforço), o que explicaria o conselho dado pelos educadores de língua estrangeira.

Referências

ANDERSON-HSIEH, J.; JOHNSON, R.; KOEHLER, K. The relationship between native speaker judgments of nonnative pronunciation and deviance in segmentals, prosody, and syllable structure. *Language Learning*, v. 42, p. 529-555, 1992.

BANSAL, R. K. *The intelligibility of Indian English*. Ph.D. dissertation, London University, 1966.

BENOÎT, C.; GRICE, M.; HAZAN, V. The SUS test: A method for the assessment of text-to-speech synthesis intelligibility using semantically unpredictable sentences. *Speech Communication*, v. 18, p. 381-392, 1996.

BLESSER, B. *Perception of spectrally rotated speech*. PhD dissertation, Massachusetts Institute of Technology, 1966.

BROKX, J. P. L.; NOOTEBOOM, S. G. Intonation and the perceptual separation of simultaneous voices. *Journal of Phonetics*, v. 10, p. 23-36, 1982.

COMRIE, B.; DRYER, M. S.; HASPELMATH, M.; GIL, D. (ed.). *World Atlas of Language Structures*. Oxford: Oxford University Press, 2005.

CUTLER, A. Stress and accent in language production and understanding. In: GIBBON, D.; RICHTER, H. (ed.). *Intonation, accent and rhythm, studies in discourse phonology*. Berlin: De Gruyter, 1984. p. 77-90.

CUTLER, A. Forbear is a homophone: Lexical stress does not constrain lexical access. *Language and Speech*, v. 29, p. 201-220, 1986.

CUTLER, A. Lexical stress. In: PISONI, D. B.; REMEZ, R. E. (ed.). *The handbook of speech perception*. Oxford: Blackwell, 2005. p. 264-289.

CUTLER, A. *Native listening: Language experience and the recognition of spoken words*. Cambridge, MA: MIT Press, 2012.

CUTLER, A.; CLIFTON, C. E. The use of prosodic information in word recognition. In: BOUMA, H.; BOUWHUIS, D. G. (ed.). *Attention and Performance X*. Hillsdale, NJ: Erlbaum, 1984. p. 183-196.

CUTLER, A.; MCQUEEN, J. How prosody is both mandatory and optional. In: CASPERS, J.; CHEN, Y.; HEEREN, W. F. L. J.; PACILLY, N.; SCHILLER, O.; VAN ZANTEN, E. (ed.). *Above and beyond the segments. Experimental Linguistics and Phonetics*. Amsterdam: John Benjamins, 2014. p. 71-82.

DENES, P. B.; PINSON, E. N. *The Speech Chain. The Physics and Biology of Spoken Language*. Murray Hill, NJ: Bell Telephone Laboratories, 1963.

FIELD, J. Intelligibility and the listener: The role of lexical stress. *TESOL Quarterly*, v. 39, p. 399-423, 2005.

FOWLER, C.; HOUSUM, J. Talkers' signaling of "new" and "old" words in speech and listeners' perception and use of the distinction. *Journal of Memory and Language*, v. 26, p. 489-504, 1987.

GARDE, P. *L'accent*. Paris: Presses Universitaires de France, 1968.

GUSSENHOVEN, C. The English rhythm rule as an accent deletion rule. *Phonology*, v. 8, p. 1-35, 1991.

HAHN, L. D. Primary stress and intelligibility: Research to motivate the teaching of suprasegmentals. *TESOL Quarterly*, v. 38, p. 201-233, 2004.

HAYES, B. *Metrical stress theory: Principles and case studies*. Chicago IL: University of Chicago Press, 1995.

HORNE, M. Empirical evidence for a deletion formulation of the rhythm rule in English. *Linguistics*, v. 28, p. 959-981, 1990.

JONGENBURGER, W. *The role of lexical stress during spoken-word processing*. The Hague: Holland Academic Graphics, 1996.

JOURAVLEV, O.; LUPKER, S. J. Predicting stress patterns in an unpredictable stress language: The use of non-lexical sources of evidence for stress assignment in Russian. *Journal of Cognitive Psychology*, v. 27, p. 944-966, 2015.

KALIKOW, D. N.; STEVENS, K. N.; ELLIOTT, L. L. Development of a test of speech intelligibility in noise using sentence materials with controlled word predictability. *Journal of the Acoustical Society of America*, v. 61, p. 1337-1351, 1977.

KEIJSPER, C. E. Vorm en betekenis in Nederlandse toonhoogtecontouren. *Forum der Letteren*, v. 25, p. 20-37 e 113-126, 1984.

KESKIN, H. K.; ARI, G.; BAŞTUĞ, M. Role of prosodic reading in listening comprehension. *International Journal of Education and Literacy Studies*, v. 7, n. 1, p. 59-65. 2019. DOI: <https://doi.org/10.7575/aiac.ijels.v.7n.3p.59>

LADEFOGED, P.; DISNER, S. F. *Vowels and consonants*. Chichester: Wiley, 2012.

LADEFOGED, P.; JOHNSON, K. *A course in Phonetics*. Boston: Wadsworth, 2011.

LEHISTE, I.; PETERSON, G. E. Vowel amplitude and phonemic stress in American English. *Journal of the Acoustical Society of America*, v. 31, p. 428-435, 1959.

LEHISTE, I.; PETERSON, G. E. Some basic considerations in the analysis of intonation. *Journal of the Acoustical Society of America*, v. 33, p. 419-425, 1961.

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

- LIBERMAN, M.; PRINCE, A. On stress and linguistic rhythm. *Linguistic Inquiry*, v. 8, p. 249-336, 1977.
- LISKER, L.; ABRAMSON, A. S. A cross-language study of voicing in initial stops: Acoustical measurements. *Word*, v. 20, p. 384-422, 1964. DOI: <https://doi.org/10.1080/00437956.1964.11659830>
- MILLER, G. A.; NICELY, P. E. An analysis of perceptual confusions among some English consonants. *Journal of the Acoustical Society of America*, v. 27, p. 338-352, 1955.
- MUNRO, M. J.; DERWING, T. M. Foreign accent, comprehensibility, and intelligibility in the speech of second language learners. *Language Learning*, v. 45, p. 73-97, 1995.
- MUNRO, M. J.; DERWING, T. M.; MORTON, S. L. The mutual intelligibility of L2 speech. *Studies in Second Language Acquisition*, v. 28, p. 111-131, 2006.
- NOOTEBOOM, S. G. The prosody of speech: Melody and rhythm. In: HARDCASTLE, W.; LAVER, J. (ed.). *The Handbook of Phonetic Sciences*. Oxford: Blackwell, 1997. p. 640-673.
- NOOTEBOOM, S. G.; COHEN, A. *Spreken en verstaan, een inleiding tot de experimentele fonetiek*. Assen: Van Gorcum, 1984.
- PETERSON, G. E.; LEHISTE, I. Duration of syllable nuclei in English. *Journal of the Acoustical Society of America*, v. 32, p. 693-703, 1960.
- QIN, Q. *Understanding foreign-accented language*. Experiments on the relative contribution of segments and prosody in Dutch-accented Mandarin. M.Phil. Thesis, Leiden University Centre for Linguistics, 2007.
- RIETVELD, A. C. M.; VAN HEUVEN, V. J. *Algemene fonetiek [General phonetics]*. 2nd edition. Bussum: Van Gorcum, 2001.
- RIETVELD, A. C. M.; VAN HEUVEN, V. J. *Algemene fonetiek [General phonetics]*. 4th revised edition. Bussum: Van Gorcum, 2016.
- SLIS, I. H.; COHEN, A. On the complex regulating the voiced-voiceless distinction. *Language and Speech*, v. 12, p. 80102 e 137155, 1969.
- SMITH, L. E.; RAFIQZAD, K. English for cross-cultural communication: The question of intelligibility. *TESOL Quarterly*, v. 13, p. 371-380, 1979.
- SVENSSON, S.-G. *Prosody and grammar in speech perception*. MILUS 2. Stockholm: Institute of Linguistics, University of Stockholm, 1974.

TANG, C. *Mutual intelligibility of Chinese dialects: An experimental approach*. LOT dissertation series, 228. Utrecht: LOT, 2009.

TANG, C.; VAN HEUVEN, V. J. Mutual intelligibility of Chinese dialects experimentally tested. *Lingua*, v. 119, p. 709-732, 2009.

TERKEN, J. M. B.; NOOTEBOOM, S. G. Opposite effects of accentuation and deaccentuation on verification latencies for 'given' and 'new' information. *Language and Cognitive Processes*, v. 2, p. 145-163, 1987.

TRUBETSKOY, N. S. *Grundzüge der Phonologie*. Prague: Travaux du Cercle Linguistique de Prague 7. Reprinted as Baltaxe, C. A. M. (Trans. 1969). *Principles of Phonology*. Berkeley and Los Angeles: University of California Press, 1939.

TZAKOSTA, M. *Multiple parallel grammars in the acquisition of stress in Greek L1*. LOT dissertation series, 93. Utrecht: LOT, 2004.

VAN DONSELAAR, W. *Effects of accentuation and given/new information on word processing*. Ph. D. dissertation, Utrecht University, 1995.

VAN HEUVEN, V. J. Segmentele versus prosodische invloeden van klemtoon op de herkenning van gesproken woorden [Segmental versus prosodic influences of stress on the recognition of spoken words]. *Verslagen van de Nederlandse Vereniging voor Fonetische Wetenschappen*, v. 159/162, p. 22-38, 1984.

VAN HEUVEN, V. J. Perception of stress pattern and word recognition: recognition of Dutch words with incorrect stress position. *Journal of the Acoustical Society of America* 78, S21, 1985. Available from: www.academia.edu/11793960.

VAN HEUVEN, V. J. Stress patterns in Dutch (compound) adjectives: acoustic measurements and perception data. *Phonetica*, v. 44, p. 1-12, 1987. DOI: <https://doi.org/10.1159/000261775>

VAN HEUVEN, V. J. Effects of stress and accent on the human recognition of word fragments in spoken context: gating and shadowing. In: AINSWORTH, W. A.; HOLMES, J. N. (ed.). *Proceedings of the 7th FASE/Speech-88 Symposium*. Edinburgh The Institute of Acoustics, 1988. p. 811-818.

VAN HEUVEN, V. J. Introducing prosodic phonetics. In: ODÉ, C.; VAN HEUVEN, V. J. (ed.). *Phonetic studies of Indonesian prosody*. Semaian, 9, Vakgroep Talen en Culturen van Zuidoost-Azië en Oceanië, RU Leiden, 1994. p. 1-26.

VAN HEUVEN, V. J. Making sense of strange sounds: (Mutual) intelligibility of related language varieties. A review. *International Journal of Humanities and Arts Computing*, v. 2, p. 39-62, 2008. DOI: <https://doi.org/10.3366/E1753854809000305>

VAN HEUVEN, V. J. Prosody and sentence type in Dutch. *Nederlandse Taalkunde*, v. 22, p. 3-29 e 43-46, 2017.

VAN HEUVEN, V. J. Notes on the phonetics of word and sentence stress: A cross-linguistic (re-)view. In: VAN DER HULST, H.; HEINZ, J.; GOEDEMANS, R. (ed.). *The Study of Word Stress and Accent: Theories, methods and data*. Cambridge: Cambridge University Press, 2018. p. 13-59. DOI: <https://doi.org/10.1017/9781316683101.002>

VAN HEUVEN, V. J.; VAN DONZEL, M. E. Structural conditioning and phonetic implementation of Dutch stress retraction. *Proceedings of the 13th International Congress of Phonetic Sciences*, p. 638-641, 1995.

VAN HEUVEN, V. J.; HAAN, J. Phonetic correlates of statement versus question intonation in Dutch. In: BOTINIS, A. (ed.). *Intonation: Analysis, modelling and technology*. Dordrecht/Boston/London: Kluwer, 2000. p. 119-144.

VAN HEUVEN, V. J.; HAGMAN, P. J. Lexical statistics and spoken word recognition in Dutch. In: COOPMANS, P.; HULK, A. (ed.). *Linguistics in the Netherlands 1988*. Foris: Dordrecht, 1988. p. 59-68.

VAN HEUVEN, V. J.; SLUIJTER, A. M. C. Notes on the phonetics of word prosody. In: GOEDEMANS, R.; VAN DER HULST, H.; VISCH, E. (ed.). *Stress patterns of the world, Part 1: Background*. HIL Publications (volume 2), Holland Institute of Generative Linguistics, Leiden/Holland Academic Graphics, The Hague, p. 233-269, 1996.

VAN HEUVEN, V. J.; TURK, A. Phonetic correlates of word and sentence stress. In: GUSSENHOVEN, C.; CHEN, A. (ed.). *The Oxford handbook of language prosody*. Oxford: Oxford University Press, 2020. p. 150-165. DOI: <https://doi.org/10.1093/oxfordhb/9780198832232.013.8>

VAN WIJNGAARDEN, S. J. Intelligibility of native and non-native Dutch speech. *Speech Communication*, v. 35, p. 103-113, 2001.

VAN ZANTEN, E. A.; VAN HEUVEN, V. J. Word stress in Indonesian; its communicative relevance, *Journal of the Humanities and Social Sciences of Southeast Asia and Oceania [Bijdragen tot de Taal-, Land- and Volkenkunde]*, v. 154, p. 129-147, 1998.

VAN ZANTEN, E. A.; VAN HEUVEN, V. J. Word stress in Indonesian: fixed or free? *NUSA, Linguistic Studies on Indonesian and other Languages in Indonesia*, v. 53, p. 1-20, 2004.

WANG, H.; ZHU, L.; LI, X.; VAN HEUVEN, V. J. Relative importance of tone and segments for the intelligibility of Mandarin and Cantonese. In: LEE, W. S.; ZEE, E. (ed.). *Proceedings of the 17th International Congress of Phonetic Sciences*. Hong Kong: City University of Hong Kong, 2011. p. 2090-2093.

WINGFIELD, A. The intonation-syntax interaction: prosodic features in the perceptual processing of sentences. In: COHEN, A.; NOOTEBOOM, S. G. (ed.). *Structure and process in speech perception*. Berlin: Springer, 1975. p. 148-156.

ZHANG, J. L.; QI, S. Q.; SONG, M. Z.; LIU, Q. X. 汉语声调在言语可懂度中的重要作用 [On the important role of Chinese tones in speech intelligibility]. *声学学报 [Acta Acustica]*, v. 4, p. 237-241, 1981.

Chapter 6

O acento primário em modelos psicolinguísticos de produção de palavra: propostas para L1 e L2

Amanda Post da Silveira

Introdução

Neste capítulo, reviso modelos psicolinguísticos que simulam como os monolíngues e bilíngues fazem uso de informações de acento de palavra durante o acesso lexical. Especificamente, foco nos mecanismos cognitivos envolvidos na atribuição de acento de palavra na produção de palavras a partir do conceito para fala e na leitura em voz alta da palavra impressa para a fala. Esclareço esses mecanismos em vários modelos de fundamentação empírico-experimental e chego ao entendimento de que, apesar de precisarem fazer implementações sobre a produção de palavras em L2 e acento de L2, suas descobertas contribuem para a formulação de um relato psicolinguístico a respeito da atribuição de acento de palavra no processamento de palavras monolíngues e bilíngues e sobre o papel que o acento de palavra desempenha na codificação de palavras.

Acento primário em modelos de produção de palavras de L1

Durante a fala, o usuário de uma língua converte uma estrutura de evento conceptual em uma palavra ou frase a ser enunciada. Quanto à produção de palavras, uma representação léxico-fonológica é especificada nas fases posteriores deste processo. Em contraste com o reconhecimento de palavras faladas, existem vários modelos de produção de palavras que explicam as propriedades gerais do acento primário das palavras como parte deste processo de codificação fonológica. Na seção seguinte, discutiremos três destes modelos: os Modelos de Dell de 1986 e 1988; o modelo WEAVER de Roelofs (1992) e modelo WEAVER++ (ROELOFS; MEYER, 1998; ROELOFS, 2003); e o modelo de Perry, Ziegler e Zorzi (2010), o “Connectionist Dual Processing” (CDP++).

Dell (1986, 1988) propôs uma estrutura de ativação interativa em duas etapas (de *bottom-up* e *top-down*) para a produção de palavras que, em sua forma básica, contém três camadas de nós de representação. Em cima, há uma camada semântica com nós de características semânticas. A camada central de representações consiste em representações de palavras ou morfêmicas³². A camada inferior representa o início, o núcleo e os fonemas de coda (que correspondem apenas a vogais e consoantes), ver Figura 1, a seguir.

32 As representações morfêmicas são às vezes chamadas de lemas, embora, a rigor, este termo tenha um significado um pouco diferente.

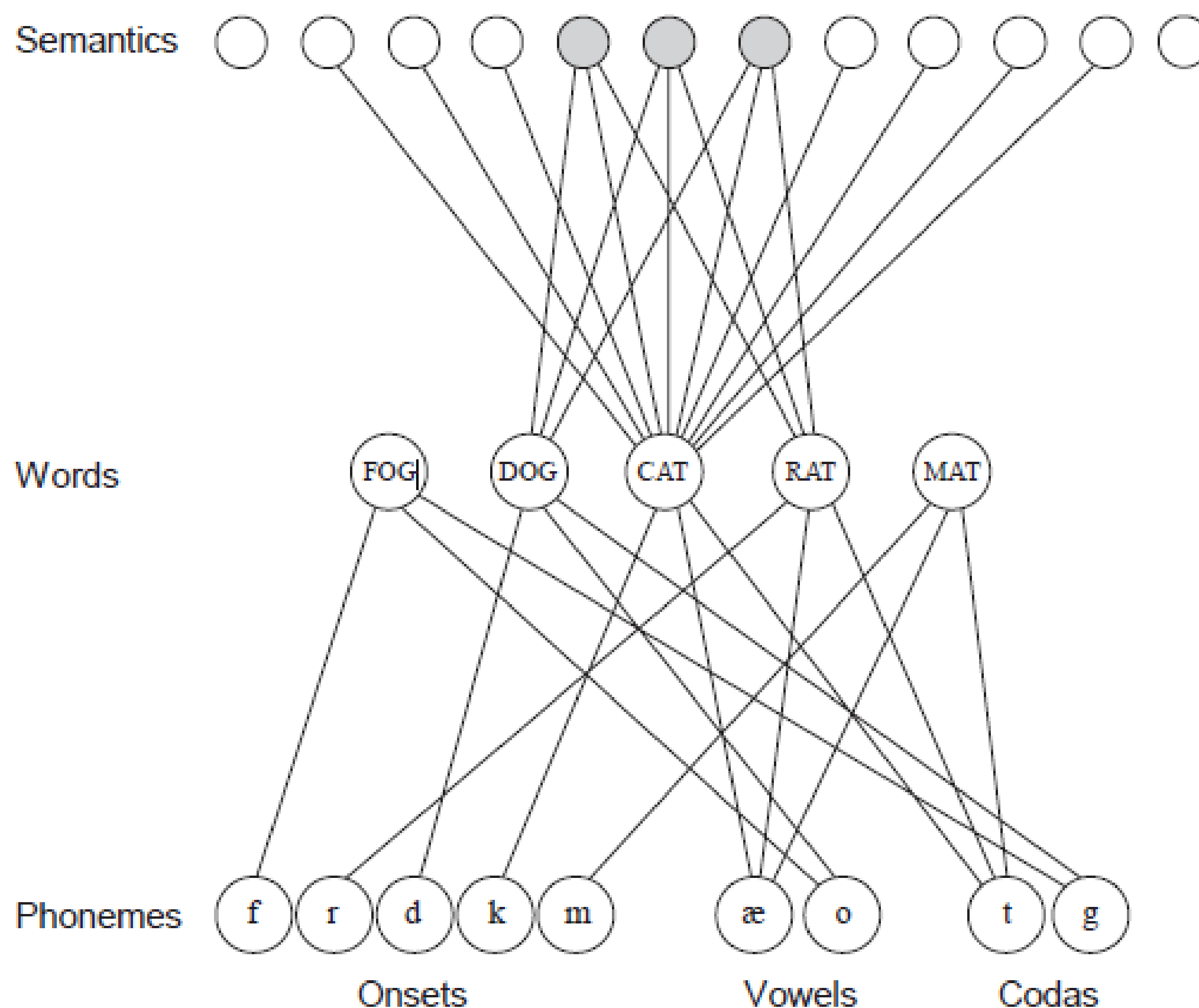


Figura 1: Fragmento da rede léxica interativa de Dell
 Fonte: Adaptação de Dell et al. (1997 feita por LEVELT, 1999).

Os nós na camada superior representam características semânticas. A camada média representa palavras ou lemas. Os nós na camada inferior representam fonemas em posições de onset, núcleo e coda (em particular as consoantes e vogais). Todas as conexões são bidirecionais e há apenas conexões facilitadoras, não inibitórias. A ativação se espalha em toda a rede sem restrições; há uma cascata total. É sempre a palavra ou nó de lema mais altamente ativado que recebe selecionados. O momento da seleção é determinado externamente pelo desenvolvimento do quadro sintático do enunciado. Após a seleção, o nó recebe uma sacudida extra de ativação, que aciona sua codificação fonológica. O modelo computacional tem muito mais características do que as representadas na figura atual. Há uma outra camada representando características fonológicas (tais como “vocalizado” ou “nasal”) e há versões do modelo com uma camada de nós de sílabas.

No modelo de Dell, todas as conexões entre unidades representacionais são bidirecionais e nunca inibitórias. Durante a fala, no curso do conceito à ativação fonológica, os nós fonológicos só são ativados indiretamente, depois que as representações semânticas e os lemas são ativados. No entanto, em estágios posteriores, os fonemas podem alimentar a ativação das representações léxico-fonológicas das quais fazem parte, devido à presença de *feedback* de ativação entre camadas.

No modelo de Dell (1986), as sílabas são CVC por *default*. Tanto a sequência fonotática quanto a codificação de acento das palavras são montadas após a seleção sucessiva dos nós de sílabas e segmentos mais ativos. A seleção e integração dos nós de segmento em quadros de sílabas é feita

por meio de regras. A ordenação dos segmentos em sílabas é feita através de uma verificação de *feedback* envolvendo os nós morfológicos. Dell (1988) propõe uma separação da estrutura fonológica das palavras em duas partes principais: um quadro de sílabas para as sequências sílabas abstratas e fonotáticas em uma palavra, como a estrutura CVC; e a representação segmental dos sons reais das palavras que se encaixam no quadro. Contudo, as simulações feitas com o modelo de Dell (1988) foram restritas a palavras monossílabas e o mecanismo que implementa os segmentos no proposto quadro silábico não é descrito (ROELOFS; MEYER, 1998).

Dell (1988) prevê que o acento das palavras seja ativado exclusivamente através do nível morfêmico, o que sugere que os componentes fonológicos pré-lexicais das palavras, isto é, os quadros silábicos e fonotáticos e as propriedades acústicas dos segmentos não desempenham um papel na codificação do acento de palavra durante a produção da fala. Os modelos de Dell também não contabilizam os processos de produção da fala após a montagem fonológica de uma palavra; eles pressupõem implicitamente que o *buffer* e a articulação fonológica não são propensos a erros (DELL, 1988).

O modelo original WEAVER (ROELOFS, 1992) é outro modelo de produção de palavras, mas que propõe a formação destas em duas etapas. Ele pressupõe a existência de três estratos de processamento: um estrato conceitual, um estrato de lema e um estrato de forma de palavra. A camada superior, do estrato conceitual, representa conceitos lexicais inteiros. Os conceitos léxicos, uma vez selecionados, ativam os lemas. Quando os lemas são selecionados, eles ativam uma representação fonológica no nível da forma. O nível da forma representa morfemas e sua conexão com os nós métricos e fonêmicos. A ativação entre os estratos no modelo WEAVER ocorre apenas em uma direção de avanço do nível conceitual para o nível da forma. No nível da forma, dois processos ocorrem independentemente e em paralelo: o sequenciamento de fonemas para compor as palavras e a metrificação de palavras em moldes de sílabas acentuadas e não acentuadas.

Uma árvore métrica é utilizada para esquematizar a distribuição de sílabas em termos de sequências fonêmicas. As árvores métricas descrevem como as sequências segmentares fazem parte das unidades silábicas e os quadros métricos descrevem como as subsequentes unidades silábicas são conectadas. No modelo WEAVER, a implementação do acento lexical à palavra ocorre antes da articulação, mas após um conceito de palavra já ter sido selecionado.

O modelo WEAVER++ (por exemplo, ROELOFS; MEYER, 1998; ROELOFS; VERHOEF, 2006) é uma extensão do modelo WEAVER que propõe três níveis consecutivos de codificação fonética: morfêmica, fonológica e fonética. Quando um morfema é ativado e selecionado no nível morfológico, seus segmentos fonêmicos correspondentes são dispostos em paralelo. Em seguida, a silabificação e o acento das palavras (prosodificação) são atribuídos sequencialmente no nível fonológico. No nível de codificação fonética, as representações abstratas fonológicas são então especificadas foneticamente para articulação.

O WEAVER++ assume que a codificação fonética inclua o acesso a um programa de sílabas, que é uma memória de programas motores prontos para formação de sílabas. Segmentos fonêmicos entre línguas, que variam pouco em sua realização articulatória, tais como /m/, /t/, /p/, e /k/, constituem unidades únicas no inventário sonoro dos falantes bilíngues que são compartilhados entre as línguas (ROELOFS, 2003). O mesmo mecanismo que armazena um programa motor silábico em um silabário que é acessado para codificação fonológica diz respeito à codificação prosódica ou prosodificação de uma palavra. A prosodificação deve ser específica de cada língua, pois cada língua tem suas próprias regras para a distribuição do padrão de acento de palavra e relações específicas entre o acento de palavra e a codificação silábica e segmental.

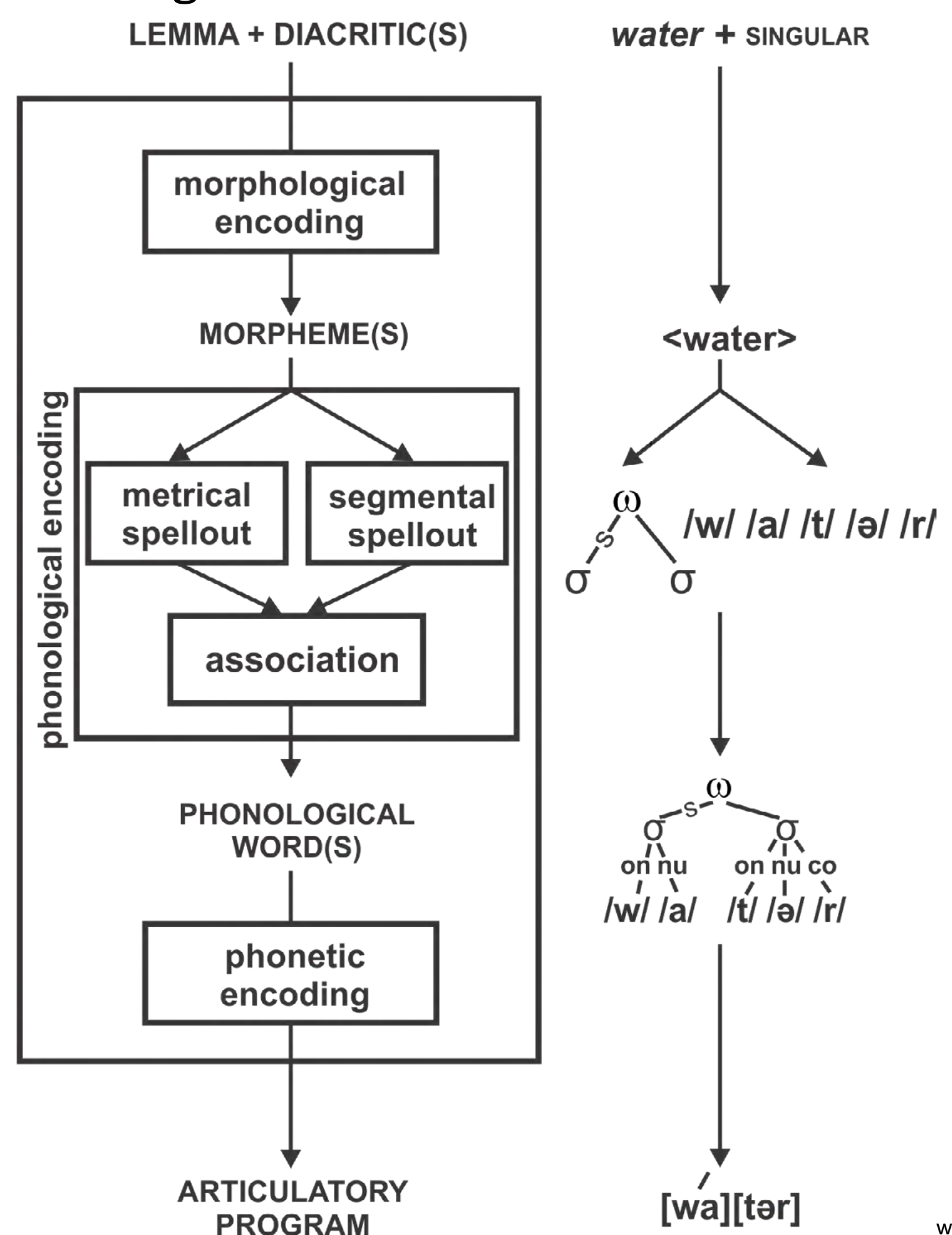


Figura 2: Esquema adaptado de WEAVER++ (Levelt, Roelofs, & Meyer, 1999) mostrando estágios de codificação na formação de palavras. ω = palavra fonológica; σ = sílaba; s = acento; on= onset; nu = núcleo; co = coda.

No modelo WEAVER++, o acento de palavra é atribuído unidirecionalmente do estrato de lema para o estrato de forma. Assim, a atribuição do acento de palavra é sempre feita de modo *top-down*. É importante notar que a distribuição fonotática das sílabas, ou mesmo estrutura de sílabas, não é levada em conta pelo modelo em relação à posição silábica e atribuição de acento lexical. O modelo relata detalhadamente as regras fonêmicas relativas ao início, núcleo e componentes da coda (tais como regras de adjacência do segmento, especialmente princípios de sequenciamento de sonoridade), mas nenhuma discussão sobre como as distribuições fonotáticas silábicas se relacionam com a atribuição do acento de palavra.

No que se refere à atribuição do acento de palavra em uma L2, o modelo WEAVER++ introduz a noção de regras de ação-condição que são específicas de língua: diferentes tipos de informações da L1 e da L2 são coativadas, mas as regras de ação-condição só selecionam lemas da L1 ou da L2, dependendo da tarefa em questão. As regras de ação-condição implicam, por exemplo, que quando os bilíngues executam uma tarefa como a nomeação de imagens em inglês como L2, tanto os lemas da L1 como da L2 e suas representações fonológicas são ativados, mas, então, apenas os nós de lemas da língua inglesa e seus correspondentes nós fonológicos em inglês são selecionados. Roelofs e Verhoef (2006, p. 169) formulam sua posição da seguinte forma (a “suposição de discricção” foi mencionada anteriormente em Levelt et al., 1999):

O modelo WEAVER++ implementa a hipótese de que somente as representações fonológicas correspondentes a um lema selecionado são ativadas. A suposição discreta implica que a ativação no nível de lema não se processa em cascata com as formas de tradução equivalentes. Entretanto, a discricção não exclui a ativação fonológica da língua não intencional. Em particular, as representações fonológicas do idioma não intencional são ativadas na medida em que as representações são compartilhadas entre os idiomas na memória. Além disso, as regras fonológicas da outra língua, tais como a regra de desvozeamento do holandês ao falar inglês, serão acionadas através destas representações compartilhadas. (ROELOFS; VERHOEF, 2006, p. 169).

Em suma, a ativação da forma, ou ativação pré-lexical, no WEAVER++ não está restrita à língua alvo devido à presença de representações fonológicas compartilhadas. Finalmente, conforme o modelo, as representações fonológicas compartilhadas em uma língua são suscetíveis de ativar programas motores de sílabas também no outro idioma (ROELOFS; VERHOEF, 2006).

Acento lexical na produção de palavras de L2

Entre os modelos de produção de palavras disponíveis, apenas o WEAVER++ distingue um nível fonêmico e de prosodificação da produção de palavras de L2. Entretanto, as simulações feitas com o modelo até agora não abordaram especificamente a produção de palavras de L2. Uma razão para isto pode ser que ainda não estão disponíveis dados substanciais de produção de palavras produzidas por bilíngues e/ou dados relevantes de produção de acento de palavra de L2.

De acordo com o modelo WEAVER++, a atribuição de acento de palavra procede de modo top-down (do estrato conceitual ao estrato de formas pré-lexicais). Na produção de palavras em L2, lemas da L1 e da L2 são ativados, mas somente lemas da L2 competem pela seleção se a tarefa envolver somente a produção de palavras da L2 (COSTA, 2005; COSTA; CARAMAZZA; SEBASTIAN-GALLES, 2000; ROELOFS; MEYER, 1998; ROELOFS, 2003), o que faz desse um modelo de ativação lexical seletiva.

Seguindo a proposta do WEAVER++ (ROELOFS, 2003), a parte que descreve o mecanismo de seleção de um dos léxicos de um bilíngue não é clara: se a língua é selecionada a partir do estrato conceitual e a direção de ativação acontece apenas *forward*, a implementação do molde segmental, do molde de prosodificação e o molde de silabificação não deveriam ser compartilhados entre L1 e L2, e deveriam seguir apenas as regras do léxico selecionado de modo *top-down*, e, portanto, os bilíngues não deveriam produzir desvios na produção de palavras que se relacionassem aos padrões da sua L1. Porém, não é o que se encontra normalmente em dados de produção de acento de palavra. Por exemplo, o francês tem como padrão acentuar a última sílaba das palavras e, portanto, de acordo com o mecanismo de descrição, um bilíngue francês-inglês não deveria usar a regra acentual de sua L1 na construção de palavras fonológicas em inglês, mas é até mesmo um exemplo anedótico que falantes nativos de francês geralmente não o fazem. Além disso, as aplicações de regras específicas do léxico não selecionado são previstas no modelo no nível fonêmico. O exemplo dado em Roelofs e Verhoef (2006) da regra de desvozeamento final do /d/, por exemplo, em holandês que é transferido para a produção de palavras em inglês, também não deveriam ocorrer tendo em conta o mesmo argumento acima.

Em resumo, WEAVER++ prevê uma ativação seletiva de acento lexical, segundo a qual somente as regras de acento de palavras específicas L2 devem ser ativadas na codificação de palavras de L2. Além disso, o acento de palavra de L2 é especificado no estrato de lema, sendo codificado à palavra de modo unidirecional na direção *top-down*. Finalmente, nenhum efeito do léxico de L1 deve estar presente na formação de palavras de L2 e vice-versa.

Acento lexical na nomeação de palavras impressas de L1

Na nomeação de palavras impressas, os leitores devem recuperar a representação fonológica associada a uma palavra impressa a fim de articulá-la. Em uma situação experimental, eles são solicitados a fazer isto da forma mais precisa e rápida possível. Junto com a decisão lexical, a nomeação de palavras está entre as tarefas mais frequentemente utilizadas nas pesquisas em psicolinguística (por exemplo, de GROOT, 1985; SEIDENBERG; MCCLELLAND, 1989; GRAINGER, 1990; VAN HEUVEN; DIJKSTRA; GRAINGER, 1998, LEVELT *et al.*, 1999; DIJKSTRA; VAN HEUVEN, 2002; BALOTA *et al.*, 2002). A nomeação de palavras impressas é um processo muito complexo, pois associa processos de reconhecimento visual de palavras a processos de produção fonológica, seja diretamente por conversão de traços em grafemas e da conversão de grafemas em fonemas, ou indiretamente através de uma representação de palavras inteiras ou uma representação de significado derivado. Nesta seção, descrevemos este processo com mais detalhes, para considerar onde e como no processo de nomeação de palavras impressas o acento de palavra pode desempenhar um papel. Começamos considerando a sistemática no mapeamento desde a ortografia até a fonologia.

O sistema de escrita alfabética, como o alfabeto romano em línguas da Europa Ocidental, permite um alto grau de correspondência entre letras e sons, ou mais tecnicamente dito, entre grafemas e fonemas (van LEERDAM, 2005). Entretanto, um desafio para o sistema cognitivo é que em muitos sistemas de escrita, ortografias e pronúncias nem sempre têm relações de mapeamento um-a-um (van ORDEN, 1990; van LEERDAM, 2005). Um exemplo proeminente são as inconsistentes correspondências entre grafemas e fonemas das vogais em palavras de língua inglesa. Por exemplo, a rima de palavras como BOUGH, COUGH, THOUGH e TOUGH é pronunciada de forma diferente, apesar de sua sequência de letras compartilhada OUGH. A consequência de ter múltiplas correspondências fonológicas para o mesmo grafema ou corpo ortográfico resulta em maior competição entre as palavras. As palavras grafadas consistentemente são semelhantes a muitas outras palavras, ou seja, elas têm muitos amigos lexicais. Amigos são palavras que compartilham as mesmas correspondências ortográficas; eles ajudam a ativar o item alvo lexical entre uma nuvem de candidatos lexicais. Por exemplo, a palavra inglesa *cat* ajuda a ativar *cap*, pois compartilha duas de três letras, que recodificam em dois fonemas idênticos aos da palavra alvo: /k/ e /æ/. Em contraste, as palavras grafadas de forma inconsistente têm muitos inimigos léxicos. Inimigos são palavras que compartilham uma porção da ortografia com a palavra alvo, mas diferem em relação às suas correspondências fonológicas; eles ativam candidatos lexicais concorrentes ambíguos, tornando assim mais difícil a seleção do candidato apropriado. Como exemplo, as palavras inglesas *pint* e *mint* são inimigas, pois embora compartilhem três de quatro letras, uma delas, a letra <i> recodifica em dois fonemas diferentes, a saber /aɪ/ vs. /ɪ/.

As línguas utilizam diferentes *scripts* e sistemas de escrita para representar correspondências entre impressões e sons. Ao redor do mundo, os *scripts* envolvem *abjads* (alfabetos de consoantes, como em hebraico) ou alfabetos (por exemplo, alfabeto romano, que incluem vogais e consoantes), abugidas (alfabetos silábicos) ou silabários (por exemplo, japonês Hiragana), bem como sistemas logográficos (por exemplo, mandarim). Diferentes *scripts* e sistemas de escrita pedem estratégias de leitura distintas em L1 e apresentam diferentes desafios para os aprendizes de L2 (ver, por exemplo, PERFETTI; LIU, 2005; MIWA *et al.*, 2014).

A hipótese de profundidade ortográfica (por exemplo, FROST *et al.*, 1987; KATZ; FELDMAN, 1983) postula que os leitores adotem estratégias diferentes, dependendo do tipo de seu sistema ortográfico de L1. Um sistema ortográfico é chamado de superficial (ou transparente) se grafemas e fonemas estiverem relacionados de forma simples e sistemática (por exemplo, como em finlandês ou italiano) e é chamado de profundo (ou opaco) se o mapeamento entre grafemas e fonemas for complexo (por exemplo, como em inglês e em francês). Em um sistema ortográfico pouco profundo, os leitores de L1 aprendem a usar a ortografia sublexical como um meio confiável de leitura de palavras. Em contraste, em um sistema ortográfico profundo, os leitores podem usar uma porção maior da palavra impressa ou mesmo a leitura de palavras inteiras para ativar representações

fonológicas de palavras (ZIEGLER; GOSWAMI, 2005; PERFETTI; DUNLAP, 2008). Modelos impressos para sons, como o Modelo de Dupla Rota (COLTHEART et al., 2001), argumentam que os leitores podem usar dois caminhos para acessar a representação lexical de uma palavra. Primeiro, há um caminho sublexical, indireto ou fonológico que envolve mapear grafemas em fonemas (ou seja, envolvendo a decodificação de cadeia de letras da esquerda para a direita) e depois recuperar a fonologia de palavras inteiras. Em segundo lugar, existe um caminho lexical, direto ou ortográfico no qual a representação ortográfica de uma palavra como um todo ativa tanto seu significado quanto sua representação léxico-fonológica associada (ou seja, envolvendo o reconhecimento de palavras inteiras). As estratégias de leitura entre idiomas podem diferir na predominância de uma ou outra via, porque os leitores podem se adaptar às demandas do(s) tipo(s) de *script*(s) envolvidos (ZIEGLER et al., 2001).

A (in)consistência das línguas com relação ao mapeamento ortográfico-fonológico também pode ser contabilizada em termos de representações de acento de palavra. Os usuários de línguas ditas de acento lexical (com posição de acento das palavras imprevisível por regras de atribuição) às vezes aplicam estratégias ortográficas de palavras inteiras para marcar a sílaba acentuada de uma palavra. Em grego, as sílabas acentuadas são sempre marcadas por um diacrítico ortográfico (PROTOPAPAS, 2006; PROTOPAPAS; GERAKAKI; ALEXANDRI, 2006). Entretanto, a maioria das línguas de acento lexical, incluindo alemão, holandês, inglês e russo, não faz uso de diacríticos para marcar sílabas acentuadas. Isto as torna ortografias relativamente profundas no que diz respeito às correspondências ortográficas de acento em comparação com as ortografias que sinalizam graficamente o acento de palavra. No entanto, falantes nativos e não nativos de ortografias profundas para o acento de palavras podem e atribuem o acento de palavras em tarefas de leitura.

Além dos diacríticos, pode haver outras características ortográficas que dão pistas sobre a posição do acento em uma palavra. Há evidências de que certas sequências de letras são mais facilmente identificadas como sílabas por leitores monolíngues, porque tais sequências ortotáticas em questão correspondem consistentemente a sequências fonotáticas que, por sua vez, correspondem a informações sobre o acento de palavra (JOURALEV; LUPKER, 2015). Isto implica a existência de um mecanismo cognitivo que permite aos leitores derivar informações de acento das palavras a partir das suas representações ortográficas.

A decodificação do acento de palavras da L1 baseada em ortografia pode acontecer através da rota lexical. Uma vez que as palavras são reconhecidas por mecanismos *bottom-up*, os nós de letras ativam tanto o léxico ortográfico quanto o fonológico, de modo que a evidência de acento da palavra do léxico fonológico possa ser alimentada aos nós de *output* de acento de palavra (PERRY et al., 2010). Além disso, os usuários das línguas podem aplicar ‘meta-informação’, por exemplo, nas distribuições de regularidade e consistência sobre os padrões de acento de palavras. Por exemplo,

há indicações nos estudos de nomeação de palavras de L1 de que as palavras são produzidas mais rapidamente quando seu padrão de acento específico tem uma alta frequência no léxico da L1. Sabe-se que as regularidades de acento de palavra aceleram a nomeação das palavras, especialmente, para palavras de baixa frequência (COLOMBO, 1992; RASTLE; COLTHEART, 2000).

Perry et al. (2010) propuseram o modelo de duplo processamento Connectionist (CDP++) implementado para nomear palavras impressas. Este modelo de rota dupla simula como a tensão das palavras é derivada da entrada ortográfica e é codificada na fala através de rotas sublexicais e lexicais. A arquitetura do CDP++ é mostrada na Figura 3.

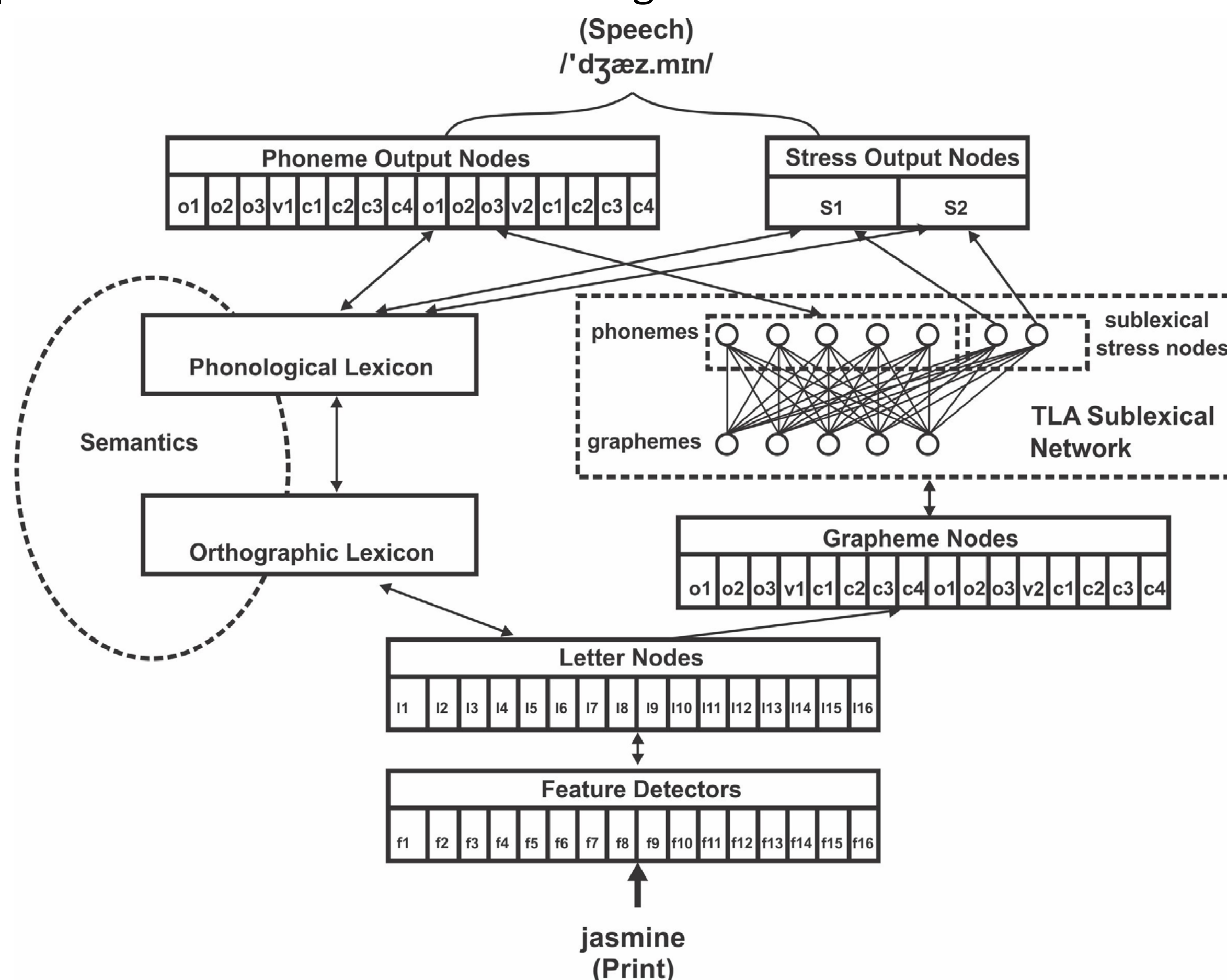


Figure 3: A arquitetura geral da percepção visual das palavras e da produção da fala na proposta do Processamento Conexionalista Duplo (CDP++) de Perry et al. (2010) (figura adaptada). Os números mostrados no interior das camadas indexam a posição do slot (molde), enquanto as letras indicam o tipo de representação (f = característica, l = letra, o = onset, v = vogal, c = coda). S1 = acento da primeira sílaba; S2 = acento da segunda sílaba.

A rota sublexical é um caminho em uma rede conexionalista que inclui um *buffer* grafêmico e uma rede de duas camadas de montagem fonológica. Os mapeamentos de grafemas para fonemas ocorrem independentemente da recuperação de informações lexicais. Uma rede de duas camadas de montagem fonológica (TLA; ver ZORZI et al., 1998a; ZORZI et al., 1998b) estrutura um modelo gráfico CCCVCCCCC com as transcrições grafema a fonema mais confiáveis. Tanto os nós ortográficos quanto fonológicos são posicionados em modelos de sílabas. A organização geral dos grafemas é baseada na frequência lexical das sequências de grafemas a fonemas que aparecem com maior frequência em posições de onset e coda de sílabas. O modelo propõe que as representações fonológicas de palavra inteira sejam ativadas pelos nós de letras. Assim como em outros modelos

conexionistas de leitura de palavras (por exemplo, DELL, 1986, 1988), as correspondências grafema a fonema são aprendidas por meio de uma busca recorrente das representações fonológicas lexicais, que estão sempre disponíveis durante o aprendizado das relações ortográficas. Na Figura 3 acima, a aprendizagem de grafema a fonema é indicada na parte direita da Figura, enquanto a fonologia lexical pode ser encontrada na parte esquerda.

Além disso, as correspondências de grafema a fonema e grafema a acento são aprendidas em paralelo. Similar ao aprendizado de grafema a fonema, o acento de palavra é aprendido com base tanto em informações sublexicais quanto lexicais. Na rota sublexical, há um quadro de nós silábicos separados dos nós fonêmicos, mas totalmente interconectados. Com base nas regularidades de acento, o modelo aprende a que nó silábico o acento de palavra é mais provável de ser atribuído. No nível sublexical, o modelo também aprende quais sequências fonêmicas (padrões fonotáticos) mapeiam melhor as sílabas acentuadas e não acentuadas, com base nas regularidades fonotáticas de acento. Os nós de acento no nível sublexical ativam os nós de acento do *buffer* de *output* fonológico.

A rota lexical é parte de uma rede totalmente interativa que envolve os léxicos fonológicos e ortográficos. Os nós de letras ativam os nós grafêmicos sobrepostos, que no léxico fonológico ativam palavras inteiras, enquanto as entradas de letra não sobrepostas são inibidas (ver figura 3). O *feedback* do léxico é possível porque todos os níveis estão totalmente interligados no CDP++, de modo que a ativação das unidades de acento ou unidades fonêmicas do *buffer* de *output* fonológico podem ativar tanto o léxico fonológico quanto as representações léxicas fonológicas.

Padrões de simulação obtidos com o CDP++ altamente correlacionados com padrões de nomeação de palavras de L1 para línguas de acento lexical (por exemplo, Perry et al., 2010, para o inglês; Perry et al., 2014, para o italiano). Entretanto, até agora o CDP+++ testou apenas a produção de palavras de L1 baseada na entrada ortográfica de palavras de mesmo comprimento em número de sílabas. Ainda não há nenhuma variante do modelo que também contabilize a nomeação de palavras de L2 e palavras que variam no comprimento em número de sílabas.

Acento de palavra na nomeação de palavras impressas de L2

Quando falantes bilíngues tardios nomeiam palavras impressas em sua segunda língua, eles devem aplicar as regras de conversão grafema a fonema de sua L2, e não de sua L1. Quando o roteiro para L2 tem uma profundidade ortográfica diferente daquela para L1, os bilíngues podem experimentar conflitos entre os mapeamentos de grafema para fonema na L1 dominante e os da L2 mais fraca. Nesse caso, eles podem recorrer a uma estratégia de leitura em L2 que seja análoga ou influenciada pela linguagem dominante. Entretanto, esta é uma hipótese não testada, pois apenas alguns poucos estudos consideraram as estratégias de reconhecimento e processamento que os leitores de L2 utilizam (por exemplo, JARED; SZUCS, 2002).

Quando os bilíngues se deparam com homógrafos interlinguísticos em leitura de L2, é provável que sejam feitas múltiplas conversões interlíngues de ortografias para fonemas (van LEERDAM, 2005). Quando os bilíngues percebem palavras impressas de L2, não somente a fonologia de L2, mas também a fonologia de L1 desempenha um papel no processo de decodificação. Van Leerdam (2005) conduziu um experimento de nomeação de palavras testando os efeitos de consistência entre palavras como *humor* e *blood* do inglês e *lood* do holandês por falantes monolíngues do inglês e bilíngues do holandês (L1) – bilíngues do inglês (L2). Suas descobertas indicam que o processo de reconhecimento de palavra é mais lento em L2 do que em L1, especialmente quando as sequências homográficas nos dois léxicos correspondem a representações fonológicas diferentes, por causa da competição lexical dos inimigos lexicais de L1. (Como definido anteriormente, inimigos são palavras com correspondências letra a letra que não combinam com os fonemas e que ativam candidatos lexicais diferentes). Como exemplo, van Leerdam dá a sequência ortográfica -oo em inglês que corresponde tanto a /u/, como em *humor*, e a /ʌ/, como em *blood*. Aqui /u/ é a correspondência mais consistente no léxico inglês, enquanto em holandês, -oo corresponde ao fonema /o/, como em *lood* ('chumbo' metálico). Van Leerdam descobriu que quando os bilíngues holandês-ingleses leram a palavra *blood* em inglês, três correspondências fonêmicas (duas de L2 e uma de L1) associadas à sequência homográfica -oo foram ativadas, assim como inimigos lexicais de língua cruzada, tais como *lood* (holandês L1) e *mood* (inglês L2), indicando o quão rápido o sistema bilíngue foi capaz de desambiguar o alvo dos candidatos inimigos correlacionando com a consistência das representações locais (VAN LEERDAM, 2005).

A concepção de Van Leerdam (2005) das categorias combinadas de L1-L2 refere-se explicitamente ao mapeamento de uma unidade visual na L2 (uma letra ou sequência de letra) para duas ou mais categorias fonológicas na L1 e na L2. Entretanto, o modelo não prevê as categorias fonológicas combinadas L1-L2 para sons que não fazem parte do inventário fonêmico de L1 dos bilíngues (ver estudos sobre a assimilação ou categorização da categoria L2, respectivamente por BEST, 1995 e FLEGE, 1995). Para o exemplo -oo de van Leerdam (2005), o inventário sonoro holandês não inclui a vogal inglesa /ʌ/, portanto esta vogal provavelmente se fundiu com uma categoria de vogal holandesa próxima a ela no espaço espectral vocálico, tal como o fonema /o/ do holandês. Como mencionado acima, as letras -oo são pronunciadas na palavra holandesa *lood* [chumbo] como /lot/, e -ood na palavra inglesa *blood* como /blʌd/. O efeito desta combinação vocálica de L1-L2 a nível fonético-fonológico na leitura é o aumento da competição por aquelas palavras do holandês como L1 contendo /o/ quando a conversão grafema a fonema do inglês como L2 corresponda a /ʌ/ como em *blood* e a baixa competição da L1 quando a grafia corresponde a /u/ como em *humor*. Em suma, na leitura de L2, os inimigos lexicais de L1 podem ser considerados como amigos na nuvem de candidatos locais, mas que podem ser falsos amigos quando os grafemas e corpos ortográficos correspondem a uma categoria fonêmica combinada de L1-L2, como a do exemplo.

Acento de palavra e modelos psicolinguísticos de produção de palavras: uma conclusão

Para resumir, uma consideração dos modelos disponíveis para produção de palavras faladas e nomeação de palavras impressas, mostra que especialmente os dois últimos tipos de modelos já deram um primeiro passo para especificar o papel do acento lexical na produção de palavras. Neste capítulo, meu objetivo foi o de fornecer um relato mais completo sobre a atribuição do acento de palavra em produção lexical de L1 e de L2. Em última análise, isto deve levar a adaptações dos modelos atuais de produção e nomeação de palavras impressas que incluam o acento de palavras.

Referências

COLOMBO, L. Lexical stress effect and its interaction with frequency in word pronunciation.

Journal of Experimental Psychology: Human Perception and Performance, v. 18, n. 4, p. 987-1003, 1992.

COSTA, A.; SANTESTEBAN, M.; CAÑO, A. On the facilitatory effects of cognate words in bilingual speech production. *Brain and Language*, v. 94, p. 94-103, 2005.

COSTA, A.; CARAMAZZA, A.; SEBASTIAN-GALLES, N. The cognate facilitation effect: Implications for models of lexical access. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, v. 26, p. 1283-1296, 2000.

DELL, G. S. A spreading activation theory of retrieval in language production. *Psychological Review*, v. 93, p. 283-321, 1986.

DELL, G. The retrieval of phonological forms in production: Tests of predictions from a connectionist model. *Journal of Memory and Language*, v. 27, p. 124-142, 1988.

DELL, G. S. et al. Lexical access in aphasic and non-aphasic speech. *Psychology Review*, v. 104, p. 801-837, 1997.

FLEGE, J. Second-language Speech Learning: Theory. Findings and Problems. In: STRANGE, W. (ed.). *Speech Perception and Linguistic Experience: Issues in Cross-language research*. Timonium, MD: York Press, 1995. p. 229-273.

FROST, R.; KATZ, L.; BENTIN, S. Strategies for visual word recognition and orthographical depth: A multilingual comparison. *Journal of Experimental Psychology: Human Perception and Performance*, v. 13, p. 104-115, 1987.

- JARED, D.; SZUCS, C. Phonological activation in bilinguals: Evidence from interlingual homograph naming. *Bilingualism. Language & Cognition*, v. 5, p. 225-239, 2002.
- KATZ, L.; FELDMAN, L. B. Relation between pronunciation and recognition of printed words in deep and shallow orthographies. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, v. 9, p. 157-166, 1983.
- LEVELT, W. J. M.; ROELOFS, A.; MEYER, A. S. Multiple perspectives on lexical access. Reply to commentaries. *Behavioral and Brain Sciences*, v. 22, p. 61-72, 1999.
- MIWA, K.; LIBBEN, G.; DIJKSTRA, T.; BAAYEN, R. H. The time-course of lexical activation in Japanese morphographic word recognition: Evidence for a character-driven processing model. *Quarterly Journal of Experimental Psychology*, v. 67, n. 1, p. 79-113, 2014.
- PERFETTI, C. A.; LIU, Y. Orthography to phonology and meaning: Comparisons across and within writing systems. *Reading and Writing*, v. 18, n. 3, p. 193-210, 2005.
- PERFETTI, C. A.; DUNLAP, S. Learning to read: General principles and writing system variations. In: KODA, K.; ZEHLER, A. (ed.). *Learning to read across languages*. Mahwah, NJ: Erlbaum, 2008. p. 13-38.
- PERRY, C.; ZIEGLER J. C.; ZORZI, M. Beyond single syllables: large-scale modeling of reading aloud with the Connectionist Dual Process (CPD++) model. *Cognitive Psychology*, v. 61, p. 106-115, 2010.
- PROTOPAPAS, A. On the Use and Usefulness of Stress Diacritics in Reading Greek. *Reading and Writing*, v. 19, p. 171-198, 2006.
- PROTOPAPAS, A.; GERAKAKI, S.; ALEXANDRI, S. Lexical and default stress assignment in reading Greek. *Journal of Research in Reading*, v. 29, n. 4, p. 418-432, 2006.
- RASTLE, K.; COLTHEART, M. Lexical and nonlexical print-to-sound translation of disyllabic words and nonwords. *Journal of Memory and Language*, v. 42, n. 3, p. 342-364, 2000.
- ROELOFS, A. A spreading-activation theory of lemma retrieval in speaking. *Cognition*, v. 42, p. 107-142, 1992.
- ROELOFS, A.; MEYER, A. Metrical Structure in Planning the Production of Spoken Words. *Journal of Experimental Psychology: Learning, Memory and Cognition*, v. 24, n. 4, p. 922-939, 1998.
- ROELOFS, A. Goal-referenced selection of verbal action: Modeling attentional control in the Stroop task. *Psychological Review*, v. 110, p. 88-125, 2003.

ROELOFS, A.; VERHOEF, K. Modeling the control of phonological encoding in bilingual speakers. *Bilingualism: Language and Cognition*, v. 9, p. 167-176, 2006.

VAN LEERDAM, M. *Dynamics of Phonological Coding in Bilingual Visual Word Perception*. Doctoral Thesis, Universiteit van Amsterdam, Amsterdam, 2005.

VAN ORDEN, G. C.; PENNINGTON, B. F.; STONE, G. O. Word identification in reading and the promise of subsymbolic psycholinguistics. *Psychological Review*, v. 97, p. 488-522, 1990.

VROOMEN, J.; DE GELDER, B. Metrical segmentation and lexical inhibition in spoken word recognition. *Journal of Experimental Psychology: Human Perception and Performance*, v. 21, n. 1, p. 98-108, 1995.

ZIEGLER, J. C.; GOSWAMI, U. C. Reading acquisition, developmental dyslexia and skilled reading across languages: A psycholinguistic grain size theory. *Psychological Bulletin*, v. 131, n. 1, p. 3-29, 2005.

Chapter 7

Análise quantitativa da entoação de interrogativas pronominais: evidências de aprendizagem e atrito na produção de falantes de espanhol l2 brasileiros

**Cristiane Conceição Silva
Pablo Arantes**

I. Introdução

Apresentamos neste capítulo resultados de um estudo experimental que procura obter evidência a respeito da ocorrência dos processos de aprendizagem e de atrito linguístico na produção de contornos entoacionais de brasileiros que são falantes de espanhol como L2 em contexto de imigração. O foco do estudo são produções paralelas de perguntas pronominais feitas pelos bilíngues em português e em espanhol L2, que são comparadas à produção de falantes monolíngues do português brasileiro e do espanhol. Uma novidade do trabalho é o uso da técnica *dynamic time warping* (DTW) para produzir uma estimativa objetiva e quantitativa da distância entre pares de contornos sem a necessidade de fazer referência a uma transcrição entoacional prévia.

O capítulo está organizado da seguinte forma: a seção 1.1 apresenta um quadro de referência teórico recente que entende os fenômenos da aprendizagem e do atrito como ocorrências comuns ao processo de bilinguismo e algumas das principais referências a respeito do estudo dos dois fenômenos no nível segmental e prosódico. As seções 1.2 e 1.3 apresentam alguns estudos que descrevem a entoação das interrogativas pronominais, respectivamente, no espanhol e no português como L1. A seção 1.4 trata brevemente dos estudos prévios que descrevem a produção de interrogativas pronominais por parte de brasileiros falantes de espanhol como L2. A seção 2 descreve os aspectos metodológicos do presente estudo, como a caracterização dos participantes, a descrição dos materiais linguísticos usados e os procedimentos de análise linguística e estatística. Na seção 3 são apresentados os resultados e, finalmente, na seção 4, desenvolvemos uma discussão sobre eles.

I.1 Bilinguismo e atrito

Diversos estudos sobre o bilinguismo³³ têm chamado a atenção para a interação de natureza bidirecional que existe entre a L1 e a L2 (FLEGE, 1995; FLEGE; BOHN, 2021; MENNEN, 2004, 2015; SCHMID; MONIKA, 2011; SCHMID; KÖPKE, 2017). Flege e Bohn (2021), por exemplo, apresentam o *Speech Learning Model-revised* (SLM-r), no âmbito do qual propõem que as categorias dos sons da L1 e da L2 do bilíngue compartilham um espaço fonético comum e, conseqüentemente, se influenciam mutuamente. Sendo assim, para compreender o processo de formação ou não de uma nova categoria da L2, por exemplo, um contraste entre duas classes de sons, é preciso levar em conta essa interação. Segundo os autores, quando uma nova categoria de L2 se forma, ela pode sofrer dissimilação, ou seja, pode se afastar de uma categoria vizinha da L1 para que, assim, seja mantido o contraste fonético entre esse par de sons. Da mesma forma, na ausência da formação de uma categoria de som para a L2, pode haver a fusão das propriedades fonéticas do som da L1 e do som da L2 ao qual

³³ Consideramos como falante bilíngue a pessoa que já adquiriu uma L1 antes de iniciar a aprendizagem de uma L2. Nesse contexto, o processo de aprendizagem da L2 não é simultâneo, mas sequencial. Seguimos, assim, a definição de bilinguismo proposta por autores como Flege (1995), Flege e Bohn (2021) e Schmid e Köpke (2017).

a categoria está ligada do ponto de vista da percepção, o que pode também deslocar o som da L1 em direção ao som da L2.

Da mesma maneira, Schmid e Köpke (2017, p. 637, tradução nossa) destacam que as pesquisas sobre aquisição de L2 têm sugerido fortemente que “as diferentes línguas não existem de forma isolada, ou seja, que elas estão interligadas e dependem uma das outras e que, além disso, há também interação constante entre esses diferentes sistemas de conhecimento”³⁴. Apesar disso, as autoras apontam que grande parte dos estudos sobre aquisição de L2 têm analisado o fenômeno apenas a partir de uma das direções, ou seja, se concentram, principalmente, na análise da transferência de aspectos da L1 para a L2. Essa influência, porém, pode ocorrer também na direção oposta. Consolidou-se na literatura a expressão “atrito linguístico” para denominar os casos nos quais o fenômeno de transferência ocorre da L2 para a L1. Ainda segundo as autoras, o atrito linguístico pode ser definido como “qualquer fenômeno que surge em uma língua nativa de um bilíngue sequencial como consequência da co-ativação de línguas, transferência interlinguística ou desuso, em qualquer estágio de desenvolvimento e uso da L2” (SCHMID; KÖPKE, 2017, p. 637, tradução nossa³⁵).

Seguindo o modelo proposto por Flege (1995), Mennen (2015) destaca também que essa interação entre L1 e L2 pode ocorrer por meio de fenômenos como a assimilação ou *merging* das propriedades tanto da L1 quanto da L2, já que a produção oral de aprendizes de L2, segundo a autora, tende a apresentar características intermediárias entre L1 e L2. Tal influência bidirecional entre as línguas foi demonstrada, inicialmente, em estudos a respeito do nível segmental (FLEGE, 1987; KUPSKE, 2016; MAJOR, 1992; SANCIER; FOWLER, 1997) e posteriormente passaram a incluir o nível prosódico (DE LEEUW; MENNEN; SCOBIE, 2012; MENNEN, 2004; SILVA; ARANTES, 2021; SILVA; ARANTES, submetido).

Neste estudo, buscamos evidências dessas influências bidirecionais no par linguístico português brasileiro (PB) como L1 e espanhol como L2, analisando, com esse fim, a produção de interrogativas pronominais nas duas línguas. Os primeiros estudos sobre a entoação de interrogativas pronominais em espanhol L2 por falantes de PB L1 analisaram a produção dos bilíngues somente a partir de uma direção, ou seja, destacaram, principalmente, aspectos relacionados à transferência de L1 para L2 (OLIVEIRA, 2013; SÁ, 2008; SILVA, 2016). Já os trabalhos de Silva e Arantes (2021) e Silva e Arantes (submetido), analisaram a produção dos bilíngues nas duas direções, ou seja, não apenas discutiram a aprendizagem da L2 e a transferência da L1 para L2, mas também observaram a ocorrência de atrito linguístico, isto é, a transferência da L2 para a L1.

34 No original: “[...] the different languages do not exist in isolation: they are closely linked, dependent on each other, and there is constant interaction between these different knowledge systems” (SCHMID; KÖPKE, 2017, p. 637).

35 No original: “[...] to any of the phenomena that arise in the native language of a sequential bilingual as the consequence of the co-activation of languages, crosslinguistic transfer or disuse, at any stage of second language (L2) development and use, as language attrition” (SCHMID; KÖPKE, 2017, p. 637).

Os resultados obtidos por Silva e Arantes (2021) apontaram para uma grande variabilidade na produção das interrogativas pronominais dos bilíngues, tanto em espanhol L2 como em português L1. A partir de uma análise quantitativa em um conjunto grande de enunciados em três modalidades (declarativas, interrogativas totais e pronominais), os autores discutiram em que medida essa variabilidade afeta a produção dos contornos de forma geral. Seu estudo revelou que os contornos de f_0 dos bilíngues brasileiros, tanto em PB L1 quanto em espanhol L2, são significativamente mais variáveis tanto do ponto de vista quantitativo quanto qualitativo quando comparados aos contornos dos informantes monolíngues. O estudo mostrou também evidências tanto de aprendizagem de espanhol L2 por parte dos bilíngues quanto de atrito, na forma de transferência de padrões do espanhol para sua produção em PB e, finalmente, apontou também para diferenças no nível de aprendizagem e de atrito dependendo da modalidade de enunciado. Já o estudo de Silva e Arantes (submetido) se concentrou em uma análise qualitativa, baseada na observação visual de padrões em contornos de f_0 normalizados temporalmente de um subconjunto dos dados analisados em Silva e Arantes (2021). O estudo mostrou evidência tanto de aprendizagem de espanhol L2, já que foram observados padrões semelhantes aos produzidos pelos falantes espanhóis nativos, quanto de atrito linguístico, pois foram observados contornos produzidos pelos bilíngues em PB L1 que são semelhantes aos produzidos pelos espanhóis nativos. Além disso, o estudo demonstrou também que os dois fenômenos manifestam-se de forma gradiente e variável entre os informantes e que podem estar relacionados a variáveis como a experiência dos falantes com a L2 e quantidade de uso de L1. No presente trabalho, combinamos o aspecto qualitativo desenvolvido em Silva e Arantes (submetido) com a abordagem quantitativa introduzida em Silva e Arantes (2021), que usa a técnica de análise de sinais *dynamic time warping* (DTW) para determinar a distância entre pares de contornos de f_0 de modo a servir como um estimador de semelhança entre eles e aplicamos essa metodologia combinada à análise de interrogativas pronominais no par linguístico PB e espanhol.

1.2 Entoação das interrogativas pronominais em espanhol L1

Há diversos estudos que descrevem a entoação das distintas modalidades de enunciado tanto do espanhol L1 quanto do português L1 (ESTEBAS-VILAPLANAS; PRIETO, 2010; FACE, 2005; NAVARRO TOMÁS, 1999; QUILIS; FERNÁNDEZ, 2003; SOSA, 1999) para espanhol L1 (FROTA; MORAES, 2016; LUCENTE, 2008, 2012; MORAES, 1998; MORAES, 2008; TENANI, 2002; TRUCKENBRODT; SANDALO; ABAURRE, 2008) para português europeu e brasileiro L1. Nesses estudos, estão incluídas descrições dos contornos das interrogativas pronominais em ambas as línguas.

Se, por um lado, há muitos estudos que descrevem os contornos das principais modalidades em espanhol e português (com ênfase para as declarativas e interrogativas totais), ainda há poucos estudos que se concentram exclusivamente na análise da entoação de interrogativas pronominais. Em espanhol, por exemplo, Henriksen (2009) destaca que uma das possíveis razões que poderia

explicar a menor quantidade de estudos sobre essa modalidade de enunciado estaria relacionada com a assunção de autores como Quilis (1993), por exemplo, de que a presença do pronome interrogativo seria suficiente para sinalizar esse tipo de modalidade de enunciado.

Apesar de haver poucos trabalhos sobre as interrogativas pronominais em espanhol, Sosa (2003, p. 230, tradução nossa³⁶) destaca que a entoação dessas interrogativas é marcada por grande variabilidade: “Essa variabilidade é consequência tanto dos diferentes tipos de funções pragmáticas que elas atendem no discurso quanto por diferenças dialetais reportadas em configurações de formas não marcadas”. A variabilidade das interrogativas pronominais em espanhol peninsular foi descrita por diversos autores (ESTEBAS-VILAPLANAS; PRIETO, 2010; HENRIKSEN, 2009, 2013; NAVARRO TOMÁS, 1999; QUILIS; FERNÁNDEZ, 2003).

Quilis e Fernández (2003) descrevem o padrão das interrogativas pronominais ou pronominais como um contorno final descendente que as torna semelhantes às declarativas, a presença do pronome interrogativo sendo o elemento que diferencia as duas modalidades. Os autores descrevem também contornos finais ascendentes que, para eles, poderiam expressar atitudes pragmáticas como polidez (QUILIS; FERNÁNDEZ, 2003). Além dos dois contornos descritos por Quilis e Fernández (2003), Navarro Tomás (1999, p. 230) descreveu um terceiro contorno, chamado por ele de “circunflexo final” ou contorno ascendente-descendente, caracterizado por um padrão nuclear descendente. Nele, há um pico no pronome interrogativo, seguido de um platô que termina a sentença com um movimento descendente na última sílaba tônica do enunciado. Segundo o autor, esse tipo de contorno seria mais comum em fala mais espontânea.

Os trabalhos de Henriksen (2009, 2013) analisaram experimentalmente as interrogativas pronominais em espanhol peninsular. O objetivo do estudo de Henriksen (2009) foi identificar as diferentes realizações dos contornos de f_0 das interrogativas pronominais e, para isso, controlou o contexto de produção, ou seja, analisou as produções a partir da leitura em voz alta de enunciados e também de um diálogo construído com o que o autor denominou de “tarefa de identificação”. Participaram do experimento seis informantes da região de León e seus resultados mostraram a realização de quatro contornos de f_0 diferentes para as interrogativas pronominais. O primeiro intitulado pelo autor como ascendente final (nossa tradução para a expressão *final rise*, bastante usada na literatura para descrever esse padrão) apresenta um pico de f_0 no final da primeira sílaba tônica da sentença, em seguida há uma queda gradual em direção à última sílaba tônica do enunciado e um vale na posição nuclear seguido de uma fronteira final ascendente. O segundo é chamado de circunflexo nuclear (*nuclear circumflex*) e começa com um platô que se inicia no início do enunciado e se mantém até o começo da sílaba tônica final. O autor observa também que pode haver um início

36 No original: “This variability is a consequence of both the different kinds of pragmatic functions they serve in natural discourse as well as the reported dialectal differences in the configurations of the unmarked forms”.

com tom ligeiramente mais alto e em seguida um movimento descendente até a posição final. Em ambos os contornos, há um movimento de subida e descida alinhado com a sílaba tônica da última palavra de conteúdo do enunciado. O terceiro contorno, descendente global (*global falling*), apresenta o pico de f_0 mais alto alinhado ao início do pronome interrogativo, seguido por uma descida ao longo da sentença até a sílaba final. Finalmente, o quarto contorno, descendente nuclear (*nuclear falling*), apresenta uma subida em direção ao pico de f_0 no pronome interrogativo, seguida de um platô que se estende até a sílaba acentuada nuclear, terminando em um movimento descendente e um vale tonal no final do enunciado. O estudo de Henriksen (2009), além de demonstrar a grande variabilidade na produção dos contornos de f_0 das interrogativas pronominais, também identificou uma diferença significativa entre os estilos de fala (leitura e fala mais espontânea) e os tipos de contornos empregados pelos informantes. Na tarefa de leitura, houve 70% de ocorrências do contorno ascendente final, 13% de descendente global, 10% de circunflexo nuclear e 7% de descendente nuclear. Já na tarefa de identificação de pessoas, a variabilidade foi menor, já que 57% dos contornos foram ascendentes finais e 43% circunflexos nucleares.

Henriksen (2013) analisou a entoação de interrogativas pronominais, totais e declarativas produzidas por 16 espanhóis que residiam na zona rural no centro da Espanha. Assim como no estudo anterior, o autor buscou compreender a interação entre o estilo de fala (leitura e diálogo) e também analisou o papel do gênero na produção dos informantes. Com relação às interrogativas pronominais, diferentemente do seu estudo anterior, Henriksen (2013) identificou a produção de apenas três contornos distintos: (1) *nuclear falling* que o autor denominou *early fall* em seu novo estudo; (2) *nuclear circumflex*, denominado *late fall* e, finalmente, (3) *final rise*. Sendo assim, o autor não identificou em seu novo estudo a produção do contorno *global falling*.

Seus resultados demonstraram novamente a variabilidade não apenas por meio da realização de três contornos diferentes para a mesma modalidade de enunciado, mas também a variabilidade individual, por estilo de elocução e por gênero do falante. Na tarefa de leitura, sete dos 16 informantes preferiram o contorno final ascendente, cinco preferiram o contorno *nuclear falling* e quatro informantes o *nuclear circumflex*. Já na tarefa de fala mais espontânea, 10 informantes preferiram o contorno *nuclear falling*, quatro o *nuclear circumflex* e apenas dois o *final rise*.

Ao analisar a distribuição global dos contornos em todo o corpus por estilo de elocução, o autor observou uma preferência geral pelo contorno *final rise* na tarefa de leitura. Já na tarefa de fala mais espontânea, a preferência foi pelo contorno *nuclear falling*, enquanto o contorno *nuclear circumflex* foi usado de forma similar nos dois estilos. Para o autor, a predominância de contornos ascendentes na leitura poderia ser interpretada como sendo mais formal e mais próxima, portanto, da variante padrão do espanhol Manchego enquanto o contorno *nuclear falling* seria uma variante menos formal e mais próxima do vernáculo dessa variante. Com relação ao contorno *nuclear circumflex*, o autor o

considerou um padrão intermediário que está entre o contorno ascendente mais formal e o *nuclear falling* menos formal.

Sendo assim, para o autor, o contorno ascendente estaria associado a um maior grau de formalidade como haviam observado autores como Navarro Tomás (1999) e Quilis (1993), por exemplo, ao fazer referência ao contorno ascendente das interrogativas pronominais como mais polidos. Além disso, Henriksen (2013) destacou que seu estudo, assim como o de Savino (2012) com dados do italiano, evidencia a importância da manipulação do estilo de fala para a compreensão do papel do efeito da leitura na produção dos informantes. Savino (2012) encontrou também uma maior ocorrência de fronteiras ascendentes em interrogativas se comparadas com a produção mais espontânea (*map task*). Finalmente, o autor destaca uma diferença de gênero na produção de tais contornos. Na tarefa de leitura, as mulheres empregaram o contorno final ascendente (mais formal) com maior frequência enquanto os homens preferiram o contorno *nuclear falling* (menos formal) no mesmo estilo. Na tarefa de identificação, as mulheres utilizaram o contorno circunflexo nuclear e descendente nuclear na mesma proporção e os homens preferiram novamente o contorno descendente nuclear.

1.3 Entoação das interrogativas pronominais em PB LI

Assim como em espanhol, Frota e Moraes (2016) destacam que as interrogativas pronominais em português são sintaticamente e lexicalmente marcadas, já que o pronome interrogativo identifica que o enunciado é uma pergunta. Em seu estudo, os autores analisaram a entoação de diversas modalidades de enunciado, dentre elas as interrogativas pronominais, tanto em PB quanto em português europeu (PE). Em PB, os autores descreveram apenas um contorno de f_0 possível para tais interrogativas que, em alguns aspectos, seria semelhante ao contorno *global falling* descrito para o espanhol. Nesse contorno em PB, há um pico de f_0 inicial extra-alto que está alinhado com o pronome interrogativo. Em seguida, há um movimento descendente gradual ao longo das sílabas seguintes até a última sílaba tônica no enunciado e a fronteira final é baixa.

Miranda (2015) analisa a entoação de enunciados em PB produzidos por informantes do Rio de Janeiro. Dentre as modalidades de enunciado estudadas, descreve as produções de interrogativas pronominais. Em seu estudo, a autora encontrou um padrão semelhante ao descrito por Frota e Moraes (2016). Nesse contorno, “há uma queda gradual que começa em um nível de referência alto no registro e se estende ao longo de todo o enunciado, terminando em um nível de referência baixo no registro” (MIRANDA, 2015, p. 52). Além disso, para três dos quatro informantes analisados, a autora observou um platô de curta duração no pronome interrogativo. O quarto informante, porém, realizou um movimento ascendente seguido de um platô alto. Este contorno se assemelha ao contorno *nuclear falling* descrito por Henriksen (2009).

Diferentemente dos trabalhos de Henriksen (2009, 2013) que analisaram experimentalmente a produção de interrogativas pronominais em espanhol, não encontramos até o momento nenhum estudo experimental voltado exclusivamente para análise das produções de interrogativas pronominais em PB. O estudo de Miranda *et al.* (2019) analisa experimentalmente interrogativas pronominais neutras e com valor exclamativo, porém a partir de uma perspectiva da percepção audiovisual de tal modalidade de enunciado.

1.4 Entoação das interrogativas pronominais em espanhol L2

Finalmente, com relação aos estudos sobre as interrogativas pronominais em espanhol L2 por falantes de PB L1, há diversos estudos que analisaram essa modalidade de enunciado, incluindo a análise de outras modalidades de enunciado como declarativas e interrogativas totais (OLIVEIRA, 2013; SÁ, 2008; SILVA, 2016). Com relação às interrogativas pronominais, os trabalhos iniciais realizados, principalmente, com corpus de leitura de informantes brasileiros que aprenderam espanhol no Brasil, apontaram para uma transferência do contorno descendente do PB para o espanhol L2 (SÁ, 2008). Já Oliveira (2013) analisou a entoação em espanhol L2 de brasileiros residentes em Barcelona a partir de uma entrevista na qual os informantes falaram sobre sua experiência com a língua e a cultura espanholas. Em seu estudo, Oliveira encontrou a produção de três padrões entoacionais diferentes para as interrogativas pronominais em espanhol L2 em fala espontânea: 52% de contornos enfáticos (tipo circunflexo), 33% de contornos suspensos (descida suave) e 20% de contornos neutros (descida abrupta). Assim como Oliveira (2013), Silva (2016) também aponta para a variabilidade na entoação em espanhol L2 de informantes brasileiros. A autora estudou também a entoação de declarativas e interrogativas totais e pronominais de informantes brasileiros residentes em Madri em leitura de enunciados isolados e inseridos em uma narrativa. A autora encontrou evidências de aprendizagem da entoação de interrogativas pronominais, já que alguns dos informantes foram capazes de reproduzir os contornos de f_0 em espanhol L1. Além disso, encontrou também evidências de transferência de contornos descendentes do PB para espanhol L2. Silva e Arantes (2021) analisaram a entoação de declarativas e interrogativas totais e pronominais a partir da produção em espanhol L1 de informantes espanhóis e espanhol L2 e PB L1 de informantes brasileiros bilíngues. Nesse novo estudo, os autores acrescentaram em suas análises a produção em PB L1 de brasileiros monolíngues como grupo de referência. Seu estudo revelou grande variabilidade na produção das interrogativas pronominais em espanhol L1 e L2 e PB L1 dos informantes bilíngues. Além disso, o estudo mostrou evidências de aprendizagem de L2 por parte dos bilíngues, além de transferências de contornos da L1 para a L2 assim como demonstrou evidências de atrito linguístico nas produções dos informantes bilíngues em PB L1.

2. Metodologia

2.1 Informantes

Neste estudo, analisamos 15 informantes brasileiros bilíngues, falantes de espanhol L2 e PB L1 e um grupo de falantes monolíngues³⁷ formado por cinco informantes espanhóis e cinco brasileiros. O grupo dos informantes bilíngues foi composto por 10 mulheres e cinco homens, naturais do estado de São Paulo, com formação universitária e idades entre 27 e 48 anos. Todos aprenderam espanhol depois dos 18 anos e moravam em Madri no momento das gravações. O tempo de residência variou entre seis meses e 16 anos (média de 6 anos). Esse grupo contribuiu com dois conjuntos de gravações: espanhol L2 (doravante referido como BL2) e PB L1 (doravante BL1). Já os grupos dos falantes monolíngues de espanhol e português foram compostos por 3 mulheres e 2 homens cada. Os espanhóis eram nativos de Madrid, Segovia e Ciudad Real, com idades entre 22 e 33 anos e os brasileiros eram nativos do estado de São Paulo, com idades entre 19 e 42 anos. Nesse grupo, todos tinham formação universitária e nunca estudaram português L2 ou espanhol L2, respectivamente. Os espanhóis contribuíram com as gravações em espanhol L1 (doravante EM) e os brasileiros monolíngues com as gravações em português L1 (doravante PM). O perfil detalhado dos informantes pode ser consultado em Silva (2016) e Silva e Arantes (2021).

2.2 Corpus

Os participantes leram um excerto de uma adaptação de *Don Quijote de la Mancha* para adolescentes (SÁNCHEZ AGUILAR, 2004). Trata-se do começo do capítulo intitulado “Gigantes con aspas”, que foi posteriormente modificado pela primeira autora para garantir que o excerto incluísse enunciados em três modalidades: declarativas, interrogativas totais e pronominais. O material em espanhol foi lido pelos grupos EM e BL2; os grupos PM e BL1 leram a versão do excerto traduzida para o português pela primeira autora. Finalizada a leitura dos excertos (2 repetições), os participantes leram 39 enunciados isolados (15 declarativas, 12 interrogativas totais e 12 interrogativas pronominais). Cada enunciado isolado foi lido três vezes. Maiores detalhes sobre a organização do corpus e os procedimentos de gravação podem ser consultados em Silva (2016) e Silva e Arantes (2021).

Neste estudo, analisaremos apenas um subconjunto desse corpus composto pelas cinco repetições (2 repetições da frase no excerto e 3 repetições da frase isolada) da interrogativa parcial em espanhol “¿Dónde están mis libros?” e “Onde estão meus livros?” em português.

Como mencionado anteriormente, os falantes monolíngues produziram os enunciados apenas em sua língua nativa, ou seja, os espanhóis leram as frases em espanhol e os brasileiros em português.

³⁷ Usamos o termo “monolíngue” nesse contexto para dizer que os falantes espanhóis não tinham nenhum conhecimento de português e os falantes brasileiros nenhum de espanhol. Não se controlou, no entanto, se tinham conhecimento ou estudo formal de outra língua.

Já os informantes bilíngues leram as frases nas duas línguas (espanhol e português). O número total de enunciados analisados foi de 200, sendo 25 enunciados (1 enunciado X 5 repetições X 5 informantes) para cada grupo de monolíngues (EM e PM) e 75 (1 enunciado X 5 repetições X 15 informantes) para cada grupo de bilíngues composto pelos grupos BL2 e BL1.

2.3 Procedimentos de análise

Os contornos de frequência fundamental foram extraídos para cada amostra de áudio do conjunto de dados por meio de um *script* para o programa de análise acústica Praat (ARANTES, 2019). Os contornos de f_0 extraídos com a ajuda do *script* foram posteriormente suavizados usando 2 Hz como parâmetro de largura de banda da função *Smooth* do Praat e, finalmente, os valores em Hz foram convertidos em semitons em relação ao valor de f_0 mínimo de cada contorno individual. Numa segunda etapa, os contornos foram normalizados temporalmente usando a sílaba como intervalo de normalização, com cinco amostras de f_0 para cada intervalo. O procedimento foi realizado com o auxílio de um *script* do Praat criado por Arantes (2021). Para uma descrição detalhada da técnica de normalização temporal, consultar Arantes (2015).

Os contornos normalizados temporalmente foram usados para gerar visualizações dos contornos em todas as condições (EM, PM, BL2 e BL1) que serviram de base para a avaliação qualitativa dos contornos. No caso do grupo PM, não se esperava variabilidade qualitativa na implementação dos contornos, dadas as descrições prévias na literatura, que sugerem a realização sistemática de contornos do tipo descendente global para perguntas pronominais. No grupo EM seria possível esperar o aparecimento de variabilidade qualitativa, isto é, a realização de contornos de tipos diferentes, que descrevemos na seção 1.2. No presente estudo, observamos a produção de três contornos distintos (ascendente final, circunflexo nuclear e duplo circunflexo). Na seção 3.1, apresentamos a quantificação dessas ocorrências.

Dessa forma, para realizarmos a classificação, observamos, primeiramente, cada um dos contornos de forma individual e buscamos classificá-los de forma global em uma das três versões possíveis presentes nos dados em espanhol monolíngue ou na versão única produzida pelos monolíngues brasileiros. No caso dos monolíngues de ambos os grupos, não houve dificuldade para realizar tal classificação. Por outro lado, a classificação de alguns contornos produzidos pelos bilíngues, tanto em BL2 quanto em BL1, foi mais difícil por conta do maior grau de variabilidade observado na produção. É importante destacar que a classificação qualitativa de cada um dos contornos do estudo não foi realizada de forma mecânica, mas apenas após discussão conjunta entre os autores e análise cuidadosa da forma global de tais contornos para que assim fosse tomada a melhor decisão possível para a categorização de um dado contorno. Assim, se por um lado a classificação dos contornos de alguns informantes foi mais simples, por exemplo, a produção da informante *f3*, que produziu três

tipos de contornos diferentes em BL2 (figura 5), um contorno descendente global, um contorno duplo circunflexo e três contornos circunflexos nucleares, por outro a produção de alguns falantes exigiu algum grau de interpretação. Um caso desse tipo foram os contornos do informante m5 em BL2 (figura 5). Seus contornos não podiam ser classificados como bons exemplares do padrão descendente global, típico do PB, mas também não são exemplos típicos nem do padrão circunflexo nuclear nem do padrão duplo circunflexo, ambos possíveis em espanhol. Entre as possibilidades que consideramos mais plausíveis, optamos por classificá-los como exemplares do padrão circunflexo nuclear do espanhol, pois a parte mais alta das curvas estava mais alinhada ao acento nuclear final, que é o traço típico do padrão circunflexo nuclear.

A análise quantitativa da semelhança entre os contornos seguiu o procedimento desenvolvido e apresentado em Silva e Arantes (2021), que usa a técnica *dynamic time warping* (DTW) para calcular a distância entre duas séries temporais de dados. No nosso caso, as duas séries são os valores de f_0 consecutivos em dois contornos. Usamos o pacote *dtw* (GIORGINO, 2009) do ambiente de computação estatístico R (RSTUDIO TEAM, 2021) para obter a distância DTW normalizada entre diferentes pares de contornos de f_0 . A justificativa para o uso de DTW como estimador quantitativo de semelhança/diferença entre pares de contornos segue o seguinte princípio: distâncias DTW próximas a 0 indicam alto grau de semelhança entre os dois contornos comparados e valores maiores indicam dissimilaridade entre eles. As figuras 1 e 2 mostram, cada uma, um par de contornos de f_0 que gera, respectivamente, uma distância DTW de valor baixo e alto. A observação da figura 1 sugere que os dois contornos são similares, impressão que é corroborada pelo baixo valor de 0,21 para a distância DTW. A figura 2, em contraste, mostra dois contornos com pouca correspondência ao longo do tempo, que é refletida em um valor para a distância DTW de 1,25, que é em torno de quatro vezes maior do que a calculada para o par mostrado na outra figura.

Nas análises que reportaremos no estudo, serão comparados de forma sistemática os contornos dos falantes dos quatro grupos experimentais (EM, PM, BL2 e BL1). Para os grupos EM e PM, será calculada a distância média de todas as comparações interfalantes, isto é, cada contorno produzido pelos falantes em um grupo será comparado de forma pareada com todos os contornos dos outros falantes do mesmo grupo (excluindo, aqui, as comparações entre os diferentes contornos produzidos pelo mesmo falante). A média de todos os pares gerados será considerada um estimador da variabilidade intralinguística, isto é, da quantidade de variabilidade na produção dos contornos de f_0 que se pode esperar que os falantes nativos apresentem entre si. No caso do grupo EM, serão comparados entre si apenas contornos do mesmo tipo (subida final com subida final, circunflexo nuclear com circunflexo nuclear etc.), uma vez que a comparação entre os contornos de tipo diferente aumentaria artificialmente a média do grupo. O número de pares comparados para gerar a estimativa de variabilidade interfalante é de 71 no grupo EM e 250 no grupo PM.

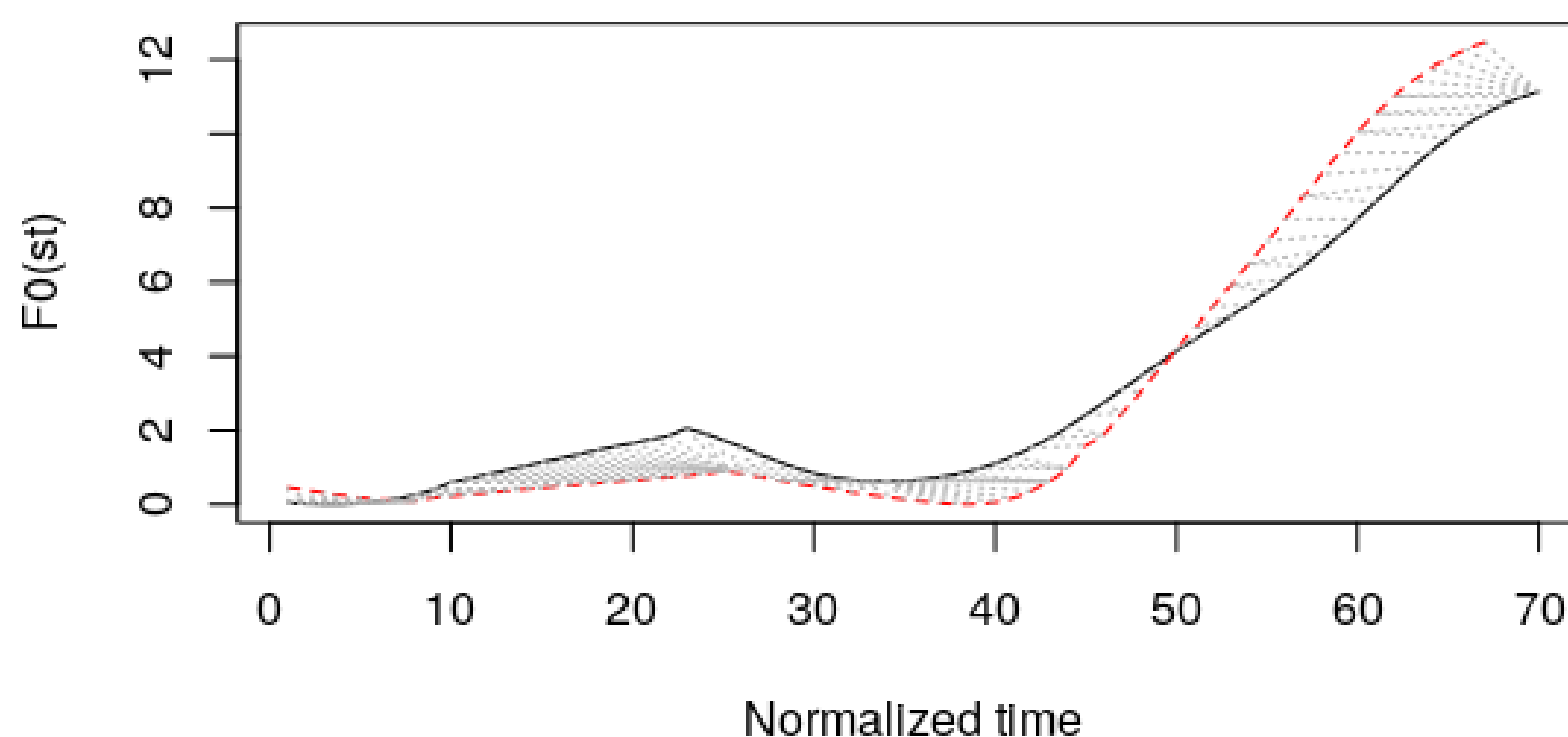


Figura 1: Par de contornos de f_0 de uma interrogativa total em espanhol (*¿Aquellos?*), produzidos por dois falantes nativos. A distância normalizada para o par é 0,21.

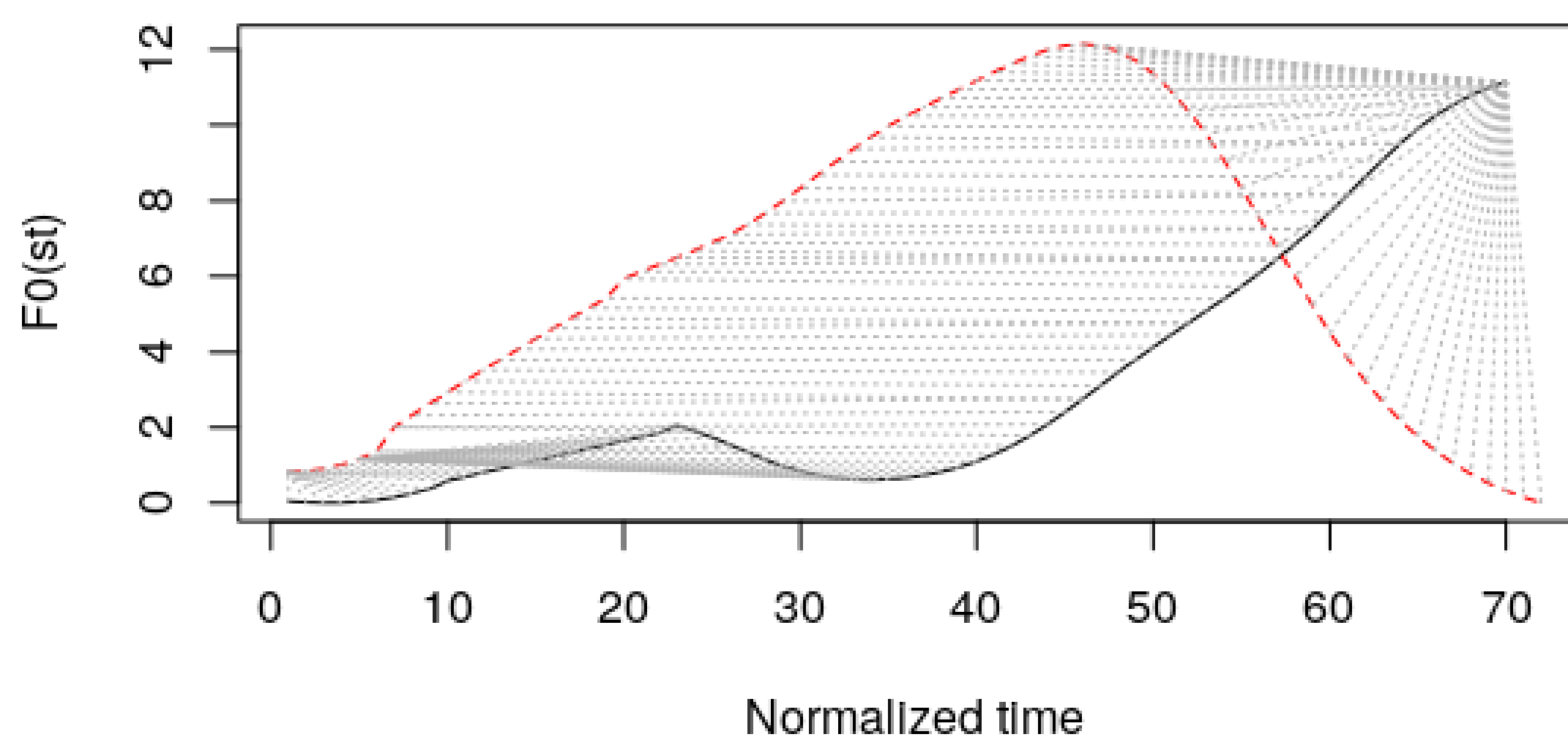


Figura 2: Par de contornos de f_0 de uma interrogativa total em espanhol (*¿Aquellos?*), produzidos por um nativo do espanhol (linha preta contínua) e por um brasileiro bilíngue (linha vermelha tracejada). A distância normalizada para o par é de 1,25.

No grupo BL2, os contornos produzidos por bilíngues que foram classificados como exemplos de aprendizagem, isto é, com configuração próxima a um dos contornos possíveis em espanhol L1, serão comparados com todos os contornos de mesmo tipo do grupo EM e essa média será comparada com a média da variação interfalante do grupo EM – o número de comparações nessa condição é 200. O propósito dessa comparação é verificar quão próximos ao espanhol L1 são esses contornos que podem ser considerados casos de aprendizagem. Os contornos que foram classificados como transferência do PB L1 serão comparados com todos os contornos do grupo PM e a média da distância será comparada com a média da variação interfalante do PM – o número de comparações pareadas nessa condição é 1250. O propósito dessa comparação é estimar se os contornos que podem ser considerados casos de transferência da L1 para a L2 são de fato próximos à produção dos nativos monolíngues da L1 ou se a produção de L1 dos bilíngues foi de alguma forma modificada pela sua experiência com a L2.

No grupo BL1, os contornos produzidos pelos bilíngues que foram classificados como exemplos de atrito, isto é, com configuração próxima a um dos contornos possíveis em espanhol L1, serão comparados com todos os contornos de mesmo tipo do grupo EM e essa média será comparada com a média da variação interfalante do grupo EM – o número de comparações nessa condição é 159. O propósito dessa comparação é verificar o grau de semelhança entre esses contornos e os produzidos pelos nativos do espanhol. Os contornos dos bilíngues que foram classificados como realizações compatíveis com o padrão do PB serão comparados com todos os contornos do grupo PM e a média da distância será comparada com a média da variação interfalante do PM – o número de comparações pareadas nessa condição é 1250. O propósito dessa comparação é estimar quanto os contornos dos bilíngues, que em princípio não foram classificados como casos de atrito, podem ser considerados casos de fato próximos à produção dos nativos monolíngues da L1.

2.4 Análise estatística

Na apresentação da estatística descritiva das diferentes amostras, indicamos o valor da média no corpo do texto seguido, entre parênteses, do valor do desvio-padrão (DP) e do número de observações presentes naquela amostra (N). Para a estatística inferencial, usamos testes t de amostras independentes para as comparações entre dois grupos. O pressuposto de homogeneidade de variância foi verificado usando o teste Fligner-Killeen e a correção dos graus de liberdade de Welch foi aplicada ao teste t nas situações em que o pressuposto não foi atendido. Para estimar o tamanho do efeito nas comparações pareadas, usamos o índice d de Cohen. Para a análise da variabilidade individual, quando a média das comparações de um participante precisou ser comparada a um valor de referência, usamos testes t de amostra única, seguidos do cálculo do d de Cohen. Fixamos em 5% o limiar para estabelecimento de significância em todos os testes de hipótese. A estatística descritiva, as análises estatísticas inferenciais e gráficos foram gerados por meio do ambiente de computação estatística R (RSTUDIO TEAM, 2021).

3. Resultados

Na seção 3.1, apresentamos os contornos de f_0 normalizados dos monolíngues espanhóis e brasileiros (Figura 3 e Figura 4). Na seção 3.2, comparamos tais contornos com as produções dos bilíngues em espanhol e PB (Figuras 5 e 6). Finalmente, na seção 3.3, apresentamos os resultados da análise com DTW e comparamos tais resultados entre bilíngues e monolíngues.

3.1 Análise dos contornos em espanhol e PB – monolíngues

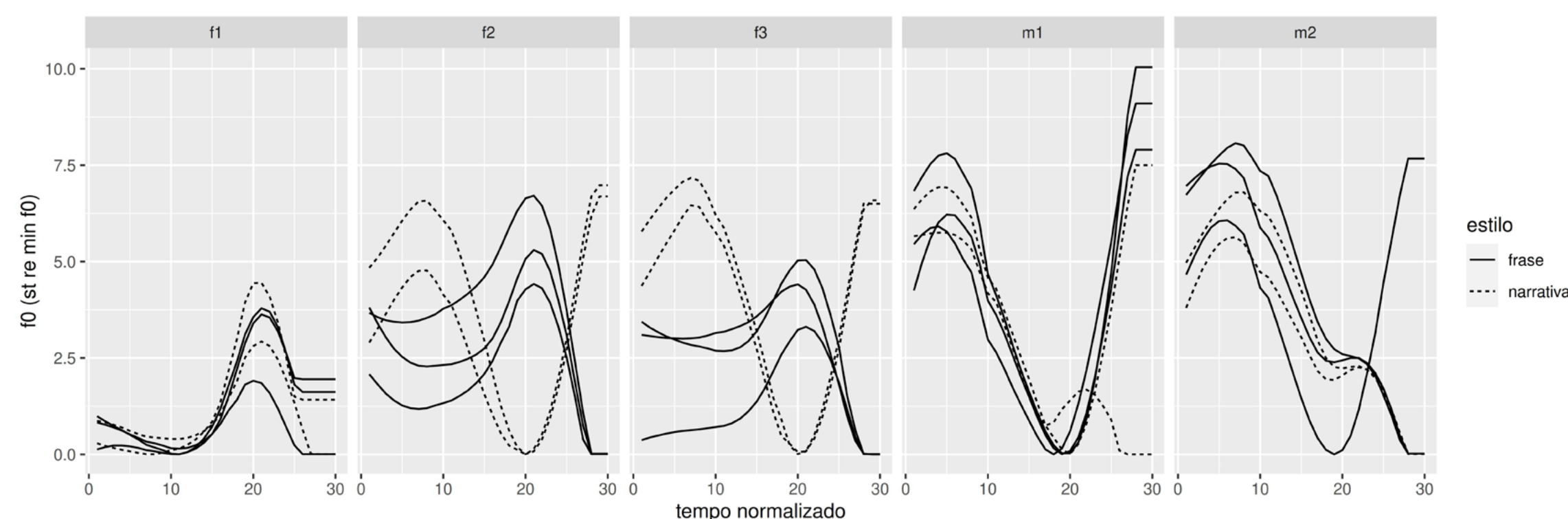


Figura 3: Contornos de f_0 normalizados temporalmente das interrogativas pronominais produzidas pelos falantes monolíngues em espanhol.

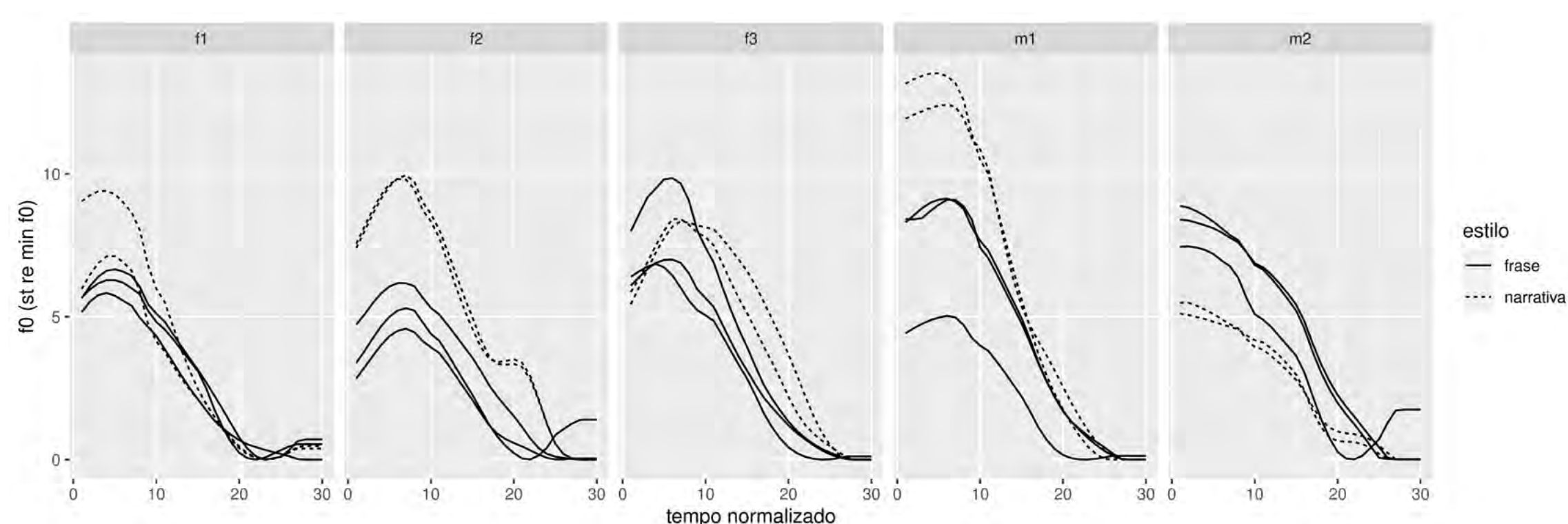


Figura 4: Contornos de f_0 normalizados temporalmente das interrogativas pronominais produzidas pelos falantes monolíngues em português.

Na figura 3, em espanhol, observamos a produção de três tipos de contornos de f_0 distintos: (1) acento pré nuclear alto seguido de um contorno final ascendente (ascendente final); (2) platô inicial seguido por um movimento circunflexo nuclear (circunflexo nuclear); (3) acento pré nuclear alto seguido por um contorno nuclear em um patamar mais baixo (downstepped) (duplo circunflexo). As informantes $f2$ e $f3$ utilizaram o contorno (1) em leitura das frases inseridas na narrativa e o contorno (2) na leitura das frases isoladas, enquanto a informante $f1$ usou apenas o contorno de tipo (2). Já o informante $m1$ utilizou quatro contornos do tipo (1) e apenas um contorno do tipo (3) e o informante $m2$ teve um comportamento inverso, utilizou quatro contornos do tipo (3) e apenas um contorno do tipo (1). Considerando todos os participantes, as porcentagens de uso dos três tipos de contorno são: 44% de circunflexo nuclear; 36% de ascendente final e 20% de circunflexo nuclear.

Na figura 4, em PB, observamos que todos os informantes, de forma consistente, utilizam o tipo de contorno que chamamos de descendente global, caracterizado por um movimento pré nuclear extra-alto (a gama tonal do pico de todos os informantes está aproximadamente entre 5 e 12 semitons) seguido por um movimento descendente suave em direção ao acento nuclear em posição baixa.

3.2 Análise dos contornos em espanhol e PB – bilíngues

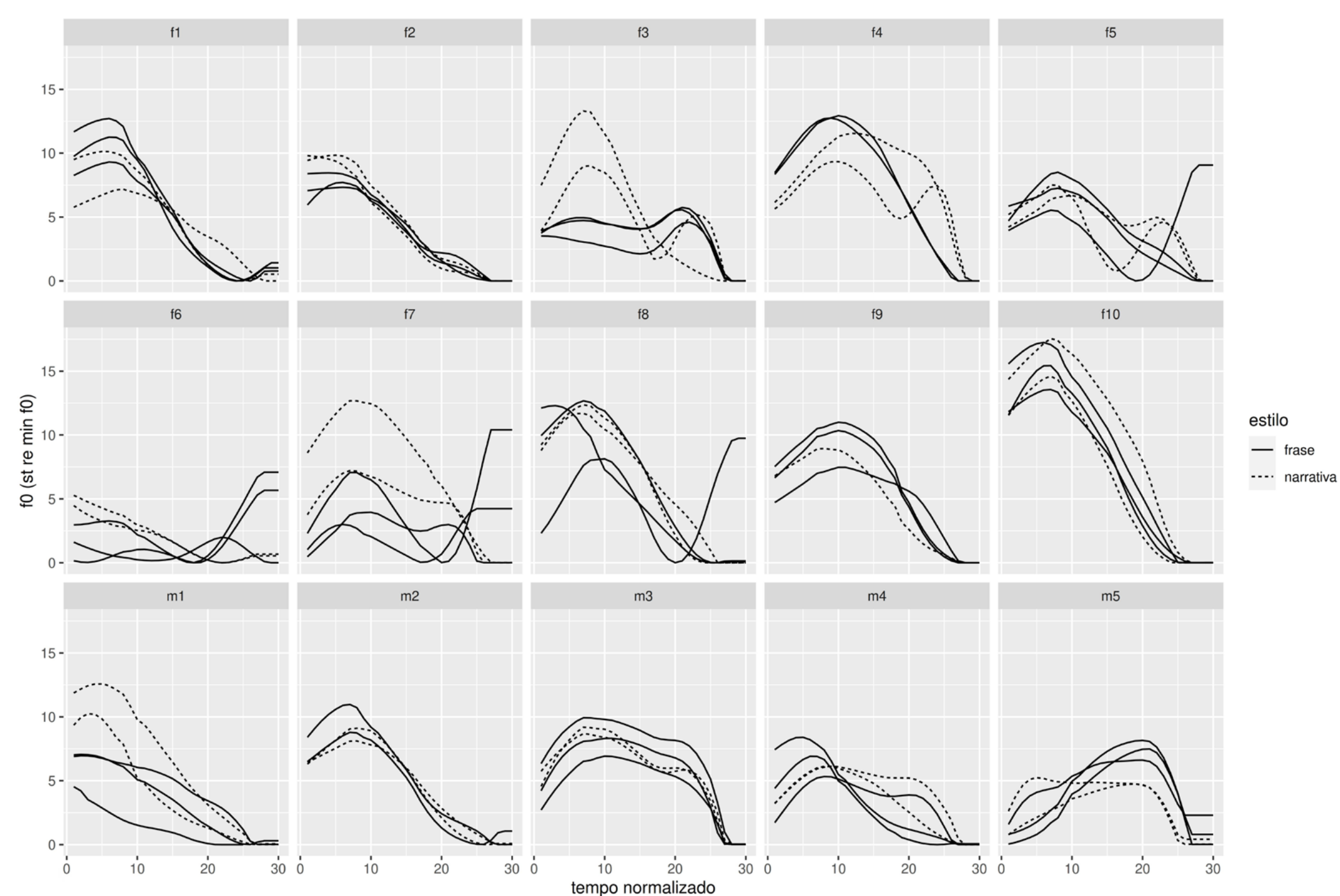


Figura 5: Contornos de f_0 normalizados temporalmente das interrogativas pronominais produzidas pelos falantes bilíngues em espanhol (BL2).

Na figura 5, que mostra os dados da produção de espanhol L2, observamos que, na amostra de 15 falantes bilíngues, sete produziram todos os contornos em L2 usando o padrão descendente global, típico do PB ($f1$, $f2$, $f9$, $f10$, $m1$, $m2$ e $m4$) e um falante produziu todos os seus contornos usando um dos padrões do espanhol ($m5$). Os sete informantes restantes produziram contornos típicos do PB e do espanhol em diferentes proporções ($f3$, $f4$, $f5$, $f6$, $f7$, $f8$ e $m3$).

É importante destacar que embora o contorno descendente global para interrogativas pronominais também faça parte do inventário de contornos possíveis em espanhol, conforme descrição apresentada na seção 1.2, em nosso estudo, assim como no trabalho de Henriksen (2013), esse contorno não foi produzido pelos informantes espanhóis. Além disso, mesmo no estudo anterior de Henriksen (2009), que atestou a produção de tal contorno, sua ocorrência foi de apenas 13% e somente na tarefa de leitura. Sendo assim, nos pareceu mais razoável considerar a produção dos contornos descendentes globais como contornos típicos do PB e não do espanhol.

Na tabela 1, apresentamos a distribuição dos tipos de contornos empregados pelos brasileiros bilíngues em sua produção de espanhol L2:

Tabela 1: Distribuição dos tipos de contornos em espanhol L2 – Bilíngues

Contorno	Valor absoluto	Valor relativo
Descendente global (PB)	50	69%
Circunflexo nuclear (Espanhol)	11	15%
Ascendente final (Espanhol)	6	8%
Duplo circunflexo (Espanhol)	5	7%

Como podemos observar, 31% de contornos de frases em espanhol L2 produzidos por falantes bilíngues podem ser considerados casos de aprendizagem por se aproximarem de padrões típicos da L2, enquanto 69% dos contornos podem ser considerados casos de transferência da L1, por serem ocorrências do padrão típico do PB. Considerando os 22 contornos que foram classificados como típicos do espanhol como um todo, 50% são circunflexo nuclear, 27% ascendente final e 23% circunflexo duplo. A distribuição dos tipos de contorno entre os bilíngues segue, em linhas gerais, aquela que se observa para os nativos do espanhol.

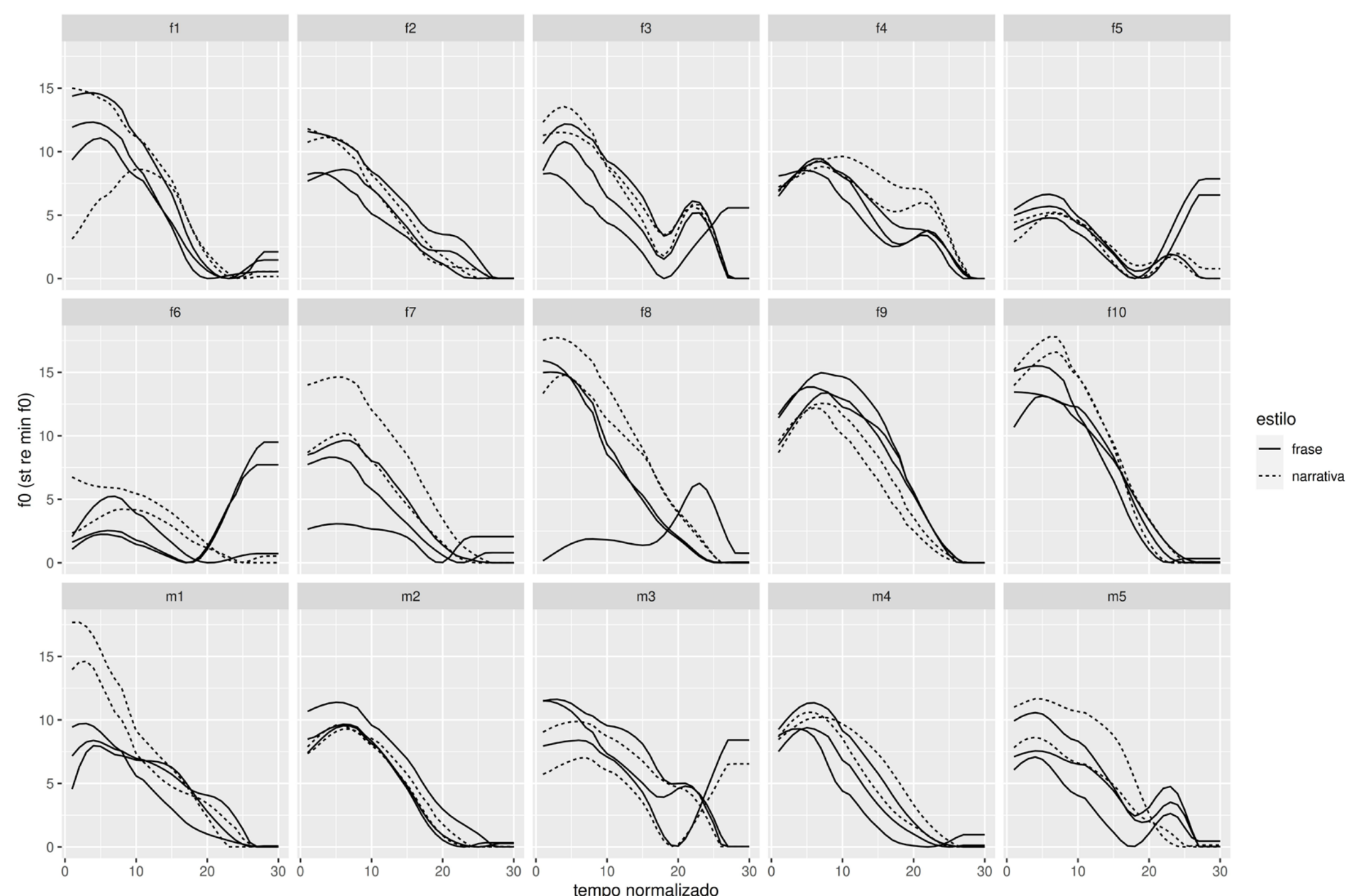


Figura 6: Contornos de f_0 normalizados temporalmente das interrogativas pronominais produzidas pelos falantes bilíngues em português (BL1).

Já na figura 6, que mostra a produção dos bilíngues em PB L1, observamos, na mesma amostra de 15 informantes bilíngues, que oito falantes (f1, f2, f7, f9, f10, m1, m2 e m4) produziram exclusivamente o padrão descendente global típico do PB e os seis informantes restantes (f3, f5, f6, f8, m3 e m5) produziram tanto contornos típicos do PB como também contornos típicos do espanhol (ascendente final, duplo circunflexo e circunflexo nuclear).

Na tabela 2, apresentamos a distribuição dos tipos de contornos empregados pelos brasileiros bilíngues em sua produção de português L1:

Tabela 2: Distribuição dos tipos de contornos em português L1 – Bilíngues

Contorno	Valor absoluto	Valor relativo
Descendente global (PB)	50	67%
Duplo circunflexo (Espanhol)	17	23%
Ascendente final (Espanhol)	7	9%
Circunflexo nuclear (Espanhol)	1	1%

Como observamos na tabela 2, 33% de contornos de frases do PB dos falantes bilíngues podem ser considerados casos de atrito, já que apresentam padrões típicos do espanhol (duplo circunflexo, ascendente final e circunflexo nuclear).

3.3 Análise com DTW

Para o grupo dos falantes monolíngues, primeiramente, estimamos a variação interfalante. Para chegar a essa estimativa, o valor da distância DTW correspondente a cada contorno de cada falante foi comparado com todos os contornos dos demais falantes e, finalmente, foi calculada a média geral. Além da variação interfalante, também estimamos a variação intrafalante, ou seja, a comparação de cada contorno de cada falante com os demais contornos do próprio falante. O esperado é que a variação intrafalante seja menor do que a interfalante. Na estimativa dos tipos de variabilidade, tomamos o cuidado, no caso do espanhol, de comparar apenas contornos de mesmo tipo. Por exemplo, contornos classificados como final ascendente foram comparados apenas com contornos com final ascendente e assim por diante.

Além disso, a variação interfalante na condição EM e PM foi usada como valor de referência com o qual comparamos a produção dos bilíngues (BL2 e BL1). Isso foi feito por meio da comparação de cada tipo de contorno de cada bilíngue com os contornos de cada um dos grupos monolíngues correspondentes (EM e PM), respectivamente. Um contorno circunflexo nuclear, por exemplo, foi comparado somente com os demais contornos do mesmo tipo e da mesma maneira para os demais tipos de contorno.

No grupo BL2, os contornos dos bilíngues que foram classificados como realizações qualitativamente próximas do espanhol foram comparados com o valor de referência do grupo MS. Primeiramente, a média de todos bilíngues foi comparada com o valor de referência e depois cada bilíngue individualmente teve sua média comparada com o valor de referência. Finalmente, os contornos considerados como casos de transferência (classificados como próximos ao PB) foram comparados com os contornos dos monolíngues do grupo PM. A média de todos os bilíngues foi comparada com a referência e a média de cada bilíngue com o valor de referência foi calculada.

Da mesma maneira, no grupo BL1, comparamos os contornos que foram classificados como típicos do PB e cada contorno de cada falante foi comparado com os contornos do grupo MP. Tanto a média global dos bilíngues foi comparada com o valor de referência MP quanto a média das distâncias por falante. No grupo BL1, aqueles contornos considerados atritados foram comparados com o valor de referência do grupo EM, sempre respeitando o tipo de contorno empregado pelos informantes .

3.3.1 Grupo dos monolíngues (EM e PM)

Para o grupo monolíngue espanhol (EM), a variação interfalante tem a média igual a 0,64 (DP 0,57; N=71) e a média da variação intrafalante 0,34 (DP 0,22; N=30). Já para o grupo monolíngue PB (PM), a variação interfalante tem a média igual a 0,54 (DP 0,39; N=250) e a média da variação intrafalante é 0,49 (DP 0,4; N=50). Como era possível prever, a variação intrafalante é menor do que a variação interfalante em ambas as línguas produzidas por monolíngues.

3.3.2 Grupo dos bilíngues

BL2: Contornos típicos do PB (transferência)

Todos os contornos produzidos pelos informantes bilíngues em espanhol L2 que foram identificados como sendo descendentes globais (contorno típico do PB) foram agrupados e foi feito o cálculo da média geral das comparações entre esses contornos e os contornos de mesmo tipo produzidos pelos falantes monolíngues do PB. A média geral dessas comparações é 0,78 (DP 0,63; N = 1250) e a média geral das comparações interfalantes dos monolíngues do PB é 0,54 (DP 0,39; N = 250). Um teste t para amostras heterocedásticas [$\chi^2(1) = 30$ $p < 0,001$] mostra que a diferença entre as médias é estatisticamente significativa e o tamanho do efeito pode ser considerado pequeno [$t(544) = 7,76$; $p < 0,001$; $d = 0,45$]. O resultado sugere que os contornos que podem ser considerados casos de transferência do PB para o espanhol L2 são mais variáveis na sua execução do que o que se observa na amostra de referência, isto é, os falantes monolíngues do PB. Interpretamos essa variabilidade maior na execução dos contornos transferidos como evidência de uma tentativa de modificação do padrão da L1, que, no entanto, não chega a mudar qualitativamente o contorno em direção a um dos alvos típicos da L2. A figura 7 apresenta, à esquerda, gráficos de caixa mostrando a distribuição dos valores da distância DTW entre os contornos de cada falante bilíngue individual com os contornos do grupo monolíngue PM. À direita, os gráficos de caixa mostram a distribuição dos valores de DTW das comparações interfalantes entre os monolíngues do grupo PM. A linha vermelha tracejada indica o valor da média de DTW das comparações interfalantes do grupo PM, usado como referência.

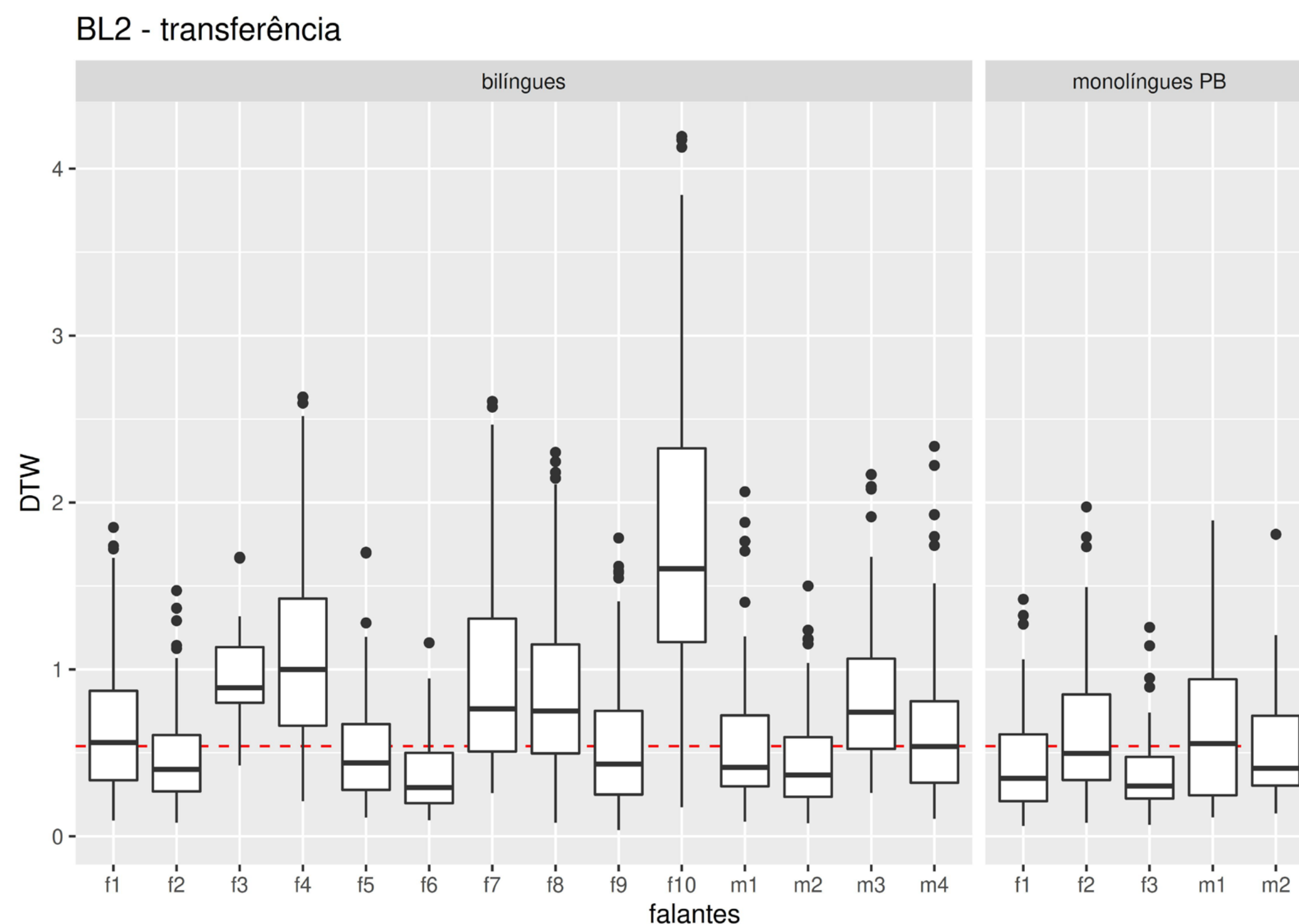


Figura 7: Gráficos de caixa com a distribuição dos valores de distância DTW em espanhol L2 (bilíngues) e português L1 (monolíngues) dos contornos classificados como casos de transferência da L1 (PB) para a L2 (espanhol).

Considerando a variabilidade individual, observamos que dos 15 falantes bilíngues da amostra, 14 produziram em L2 um ou mais contornos considerados casos de transferência de L1. Desses, seis apresentam, do ponto de vista da significância estatística, médias de DTW estatisticamente iguais (*f5*, *f9* e *m1*) ou menores (*f2*, *f6* e *m2*) do que a referência dos monolíngues. A média dos outros oito falantes é superior à referência e essa diferença é estatisticamente significativa. Esse resultado aponta para a ocorrência de variabilidade individual na questão do aprendizado. De modo geral, podemos dizer que há falantes que transferem contornos de L1 quase sem modificação e há aqueles que exibem variações quantitativas em relação ao padrão de L1. Entre essas modificações mensuradas pela distância DTW, há casos de natureza distinta: um exemplo é a falante *f10*, com média de distância de 1,79, que mantém as características gerais do contorno descendente global, embora com alteração na gama de variação; seus contornos apresentam variação entre 15 e 17 semitons relativamente ao valor mínimo, enquanto os falantes monolíngues de PB na amostra variam tipicamente entre 5 e 10 semitons. Há, também, casos como o dos falantes *m3* e *m4*, com médias de distância de 0,88 e 0,63, respectivamente, cujos contornos apresentam modificações no formato geral dos seus contornos que os afastam do padrão típico do PB, sem que com isso tal modificação seja grande o suficiente a ponto de tornar tais contornos semelhantes aos tipos presentes em espanhol monolíngue.

BL2: Contornos típicos do espanhol (aprendizagem)

Todos os contornos produzidos pelos bilíngues em espanhol L2 que foram identificados como bons exemplares da L2 foram agrupados, respectivamente, de acordo com o seu tipo (ascendente

final, circunflexo nuclear e duplo circunflexo) e comparados com os mesmos tipos de contornos produzidos pelos falantes monolíngues de espanhol. A média geral das comparações entre os contornos dos bilíngues em L2 e os contornos dos falantes monolíngues espanhóis é 0,71 (DP 0,53; N = 200) e a média geral das comparações interfalantes dos monolíngues espanhóis é 0,64 (DP 0,57; N = 71). Um teste t para amostras homocedásticas [$\chi^2(1) = 0,059$ $p = 0,81$] mostra que a diferença entre as médias não é estatisticamente significativa e o tamanho do efeito pode ser considerado muito pequeno [$t(269) = 0,87$; $p = 0,38$; $d = 0,12$]. Interpretamos esse resultado como evidência de que esse grupo de falantes implementa, de forma geral, contornos que não apresentam um grau de variabilidade maior do que aquele mostrado pelo grupo de referência, isto é, os monolíngues espanhóis. De maneira geral, podemos considerar que são contornos que reproduzem de forma bem-sucedida as características da L2. A figura 8 apresenta, à esquerda, gráficos de caixa mostrando a distribuição dos valores da distância DTW entre os contornos de cada falante bilíngue individual com os contornos do grupo monolíngue EM. À direita, os gráficos de caixa mostram a distribuição dos valores de DTW das comparações interfalantes entre os monolíngues do grupo EM. A linha vermelha tracejada indica o valor da média de DTW das comparações interfalantes do grupo EM, usado como referência.

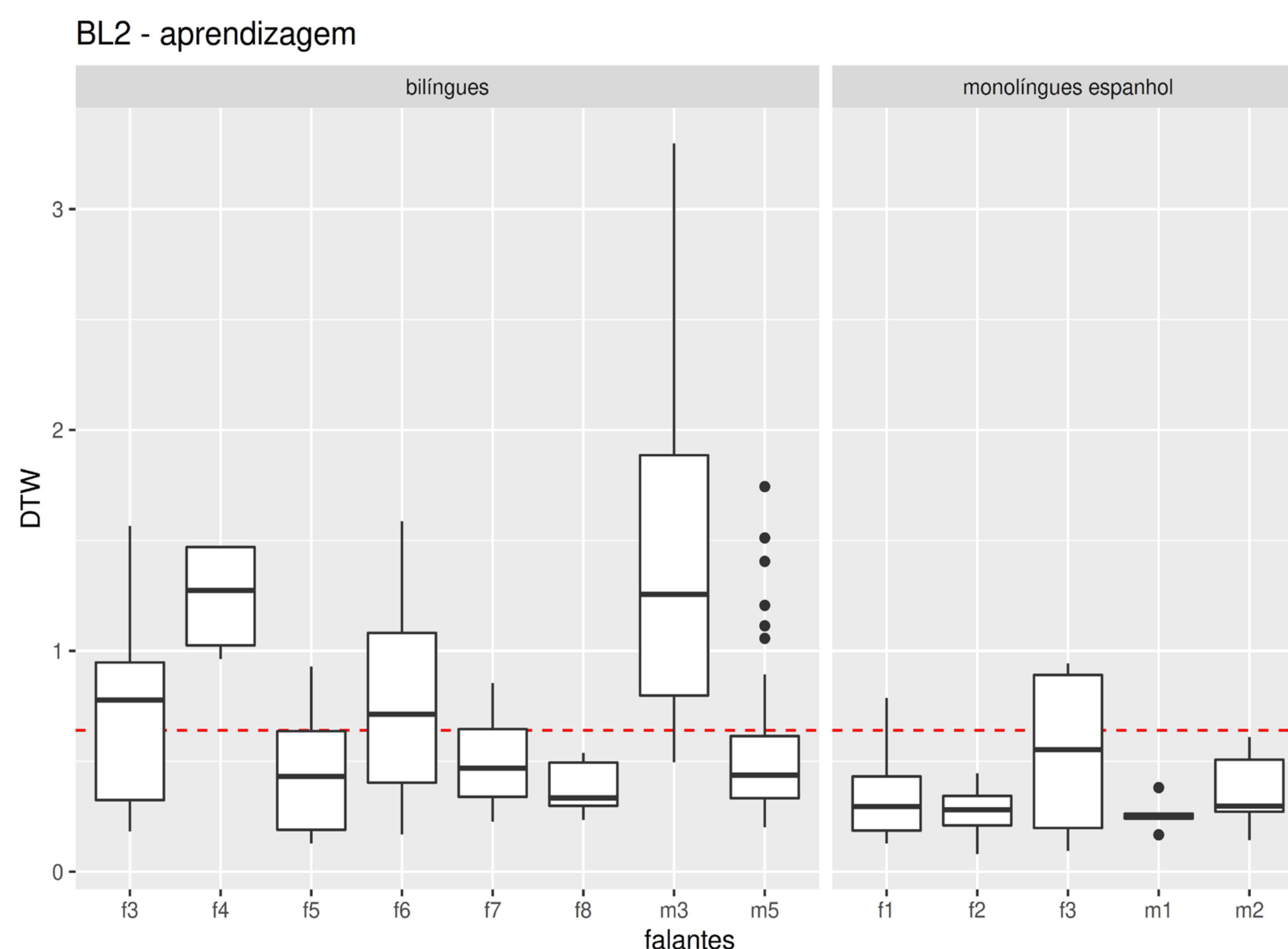


Figura 8: Gráficos de caixa com a distribuição dos valores de distância DTW em espanhol L2 (bilíngues) e espanhol L1 (monolíngues) dos contornos classificados como casos de aprendizagem da L2.

Considerando a variação individual, dos oito falantes que produziram pelo menos um contorno em L2 que foi classificado como um dos contornos típicos do espanhol nativo, dois apresentaram distâncias médias de DTW maiores do ponto de vista estatístico do que o valor de referência: f4 (1,24) e m3 (1,54). O restante apresentou médias estatisticamente iguais ou menores ao valor de referência.

BL1: Contornos típicos do PB

A média geral (todos os falantes incluídos) de DTW nas comparações interfalantes é de 0,94 (DP 0,7; N = 1250) para os contornos do PB produzidos pelos bilíngues brasileiros que foram classificados como realizações do padrão entoacional típico do PB (padrão descendente global). A média de DTW das comparações interfalantes dos monolíngues brasileiros é 0,54 (DP 0,39; N = 250). O teste t de amostras independentes heterocedásticas [$\chi^2(1) = 85$ $p < 0,001$] indica uma diferença significativa nas médias dos dois grupos com um tamanho de efeito de magnitude moderada [$t(615) = 12,5$; $p < 0,001$; $d = 0.7$]. Esse resultado indica que a produção dos bilíngues é mais variável do que a dos monolíngues, mesmo quando os contornos dos bilíngues podem ser classificados como representantes próximos do padrão produzido pelos monolíngues. Consideramos esse resultado como evidência de que a condição bilíngue introduz uma desestabilização dos padrões de produção da L1. No caso desse grupo, a desestabilização não gera uma reestruturação que se poderia chamar de qualitativa dos contornos, mas introduz algum grau de perturbação gradiente na realização do padrão típico da L1, situação que também poderia ser atribuída ao efeito do atrito, embora em grau diferente daquele observado em situações nas quais o bilíngue produz na L1 um padrão inequivocamente típico da L2.

A figura 9 mostra, à esquerda, a distribuição dos valores de DTW para os falantes do grupo BL1 com contornos típicos do PB e à direita a distribuição dos valores no grupo de monolíngues PM. A linha vermelha tracejada indica o valor da média de DTW das comparações interfalantes do grupo PM, usado como referência. O exame visual mostra que a variação entre os falantes monolíngues é bem menor do que entre os bilíngues. Para explorar a variabilidade individual entre os bilíngues, foram feitos testes t de amostra única, comparando a média das distâncias DTW de cada falante bilíngue com o valor de referência 0,54, que é a média geral da variação interfalante entre os monolíngues. Os resultados mostram que apenas três falantes têm médias que são estatisticamente iguais ao valor de referência: f6 (0,59), m2 (0,49) e m4 (0,5). As médias dos demais falantes são estatisticamente diferentes do valor dos monolíngues, com valores entre 0,67 e 1,72 (valores do d de Cohen variam entre 0,34 e 1,47).

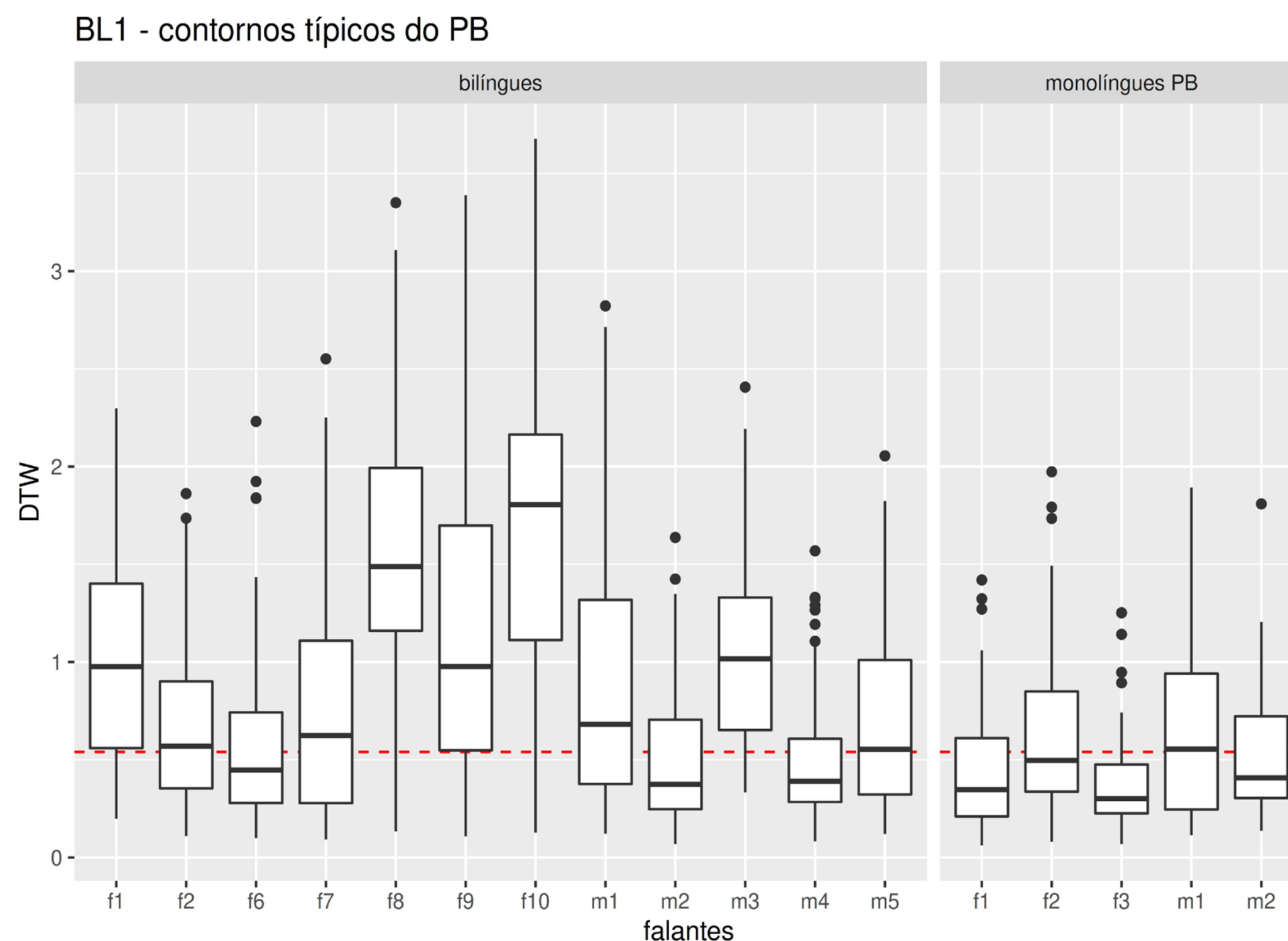


Figura 9: Gráficos de caixa com a distribuição dos valores de distância DTW em português L1 (bilíngues) e português L1 (monolíngues) dos contornos classificados como típicos da L1.

Comparando os contornos normalizados temporalmente dos monolíngues brasileiros (figura 4) com os contornos dos bilíngues na condição BL1 (figura 6), é possível ver que entre os contornos dos falantes com médias estatisticamente diferentes da referência dos monolíngues, a perturbação introduzida pela influência da L2 se manifesta de formas diferentes entre os falantes. Os três falantes com maiores valores médios de DTW, *f8* (1,58), *f9* (1,2) e *f10* (1,72) apresentam contornos que, a princípio, não são qualitativamente tão diferentes do padrão descendente global, embora a observação cuidadosa mostre que a porção inicial da curva, alinhada ao pronome interrogativo, atinge valores em torno de 15 semitons acima do valor mínimo de cada curva, enquanto o padrão típico dos monolíngues do PB gira em torno de 5 a 10 semitons acima do valor mínimo de cada contorno. Falantes como *m1* (0,91), *m3* (1,14) e *m5* (0,72), por outro lado, apresentam contornos que são versões mais alteradas em relação ao padrão descendente global mais típico dos monolíngues do grupo PM em comparação aos outros três falantes mencionados, muito embora a média de DTW não seja necessariamente mais alta.

BL1: Contornos típicos do espanhol (atrído)

A comparação dos contornos do PB atritados produzidos pelos bilíngues (isto é, que foram classificados como realização de um dos tipos de contorno típicos do espanhol dos nativos) com contornos do mesmo tipo dos nativos do espanhol gera uma distância média de DTW de 0,74 (DP 0,64; N = 159). A comparação dessa média com a média da distância DTW entre contornos do mesmo tipo produzidos pelos nativos do espanhol, que é 0,64 (DP 0,57; N = 70), feita por meio de um teste *t* para amostras homocedásticas [$\chi^2(1) = 1,99$ $p = 0,16$], não resulta em diferença estatisticamente

significativa e o tamanho do efeito tem baixa magnitude [$t(228) = 1,13$; $p = 0,26$; $d = 0,16$]. O resultado indica que a variabilidade interfalante dos contornos identificados como atritados é comparável com a variabilidade encontrada entre os nativos do espanhol quando produzem contornos do mesmo tipo. É uma evidência de que, de forma geral, a variabilidade na implementação dos contornos considerados atritados é muito próxima da variabilidade que se observa entre os falantes nativos do espanhol. A figura 10 apresenta, à esquerda, gráficos de caixa mostrando a distribuição dos valores da distância DTW entre os contornos de cada falante bilíngue individual com os contornos do grupo monolíngue EM. À direita, os gráficos de caixa mostram a distribuição dos valores de DTW das comparações interfalantes entre os monolíngues do grupo EM. A linha vermelha tracejada indica o valor da média de DTW das comparações interfalantes do grupo EM, usado como referência.

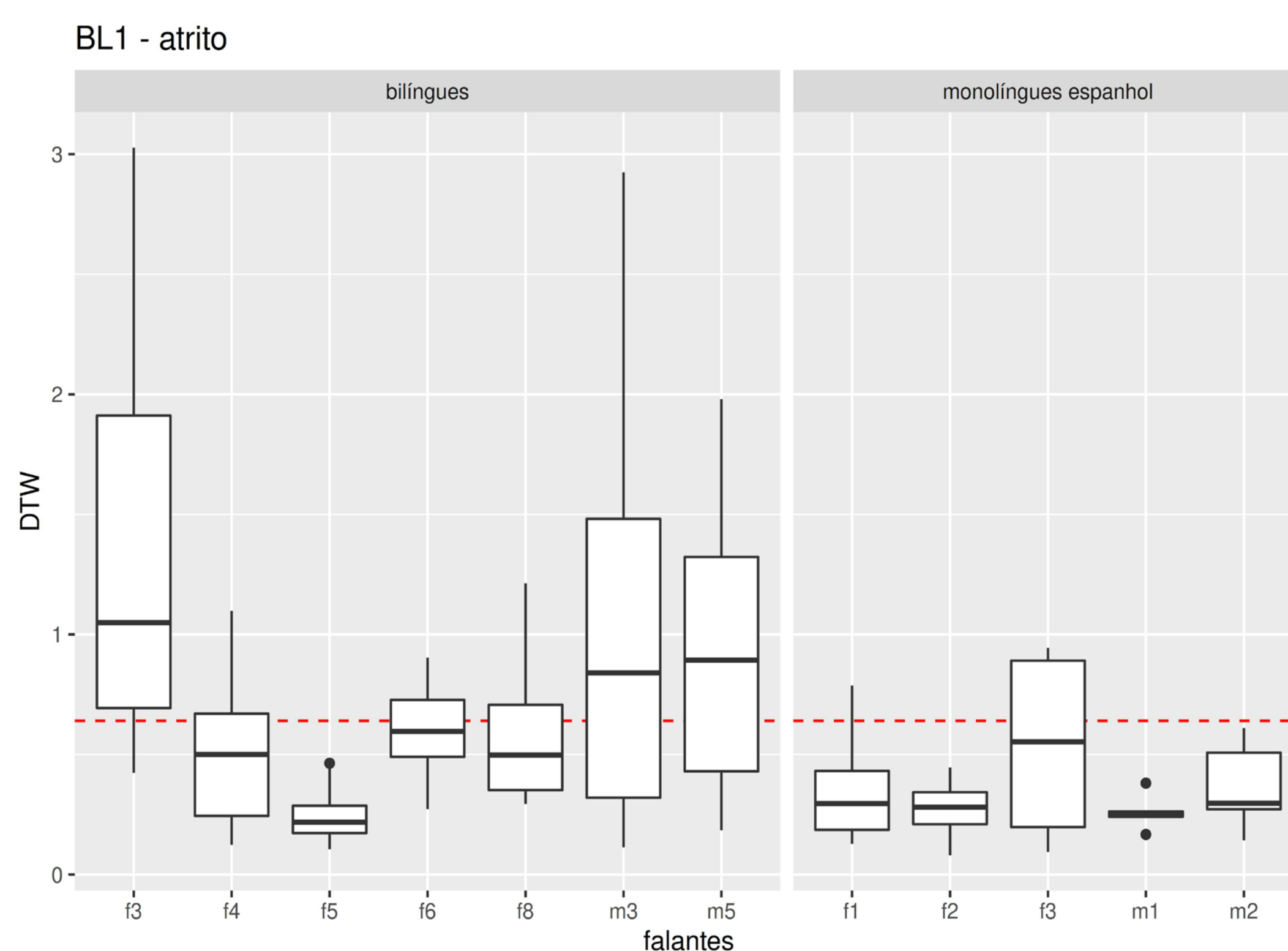


Figura 10: Gráficos de caixa com a distribuição dos valores de distância DTW em português L1 (bilíngues) e espanhol L1 (monolíngues) dos contornos dos bilíngues classificados como típicos do espanhol L1.

Com relação à variação individual, testes t de amostra única comparando individualmente a média dos sete falantes com a média geral dos monolíngues espanhóis mostra que, dos sete falantes, apenas dois têm médias cujo valor é maior do que a referência do ponto de vista estatístico: $f3$ (1,32) e $m3$ (1,04). Os demais têm valores médios que não diferem significativamente do valor de referência – $f6$ (0,6), $f8$ (0,59) e $m5$ (0,88) – ou são significativamente menores do que a referência $f4$ (0,5) e $f5$ (0,24).

3.3.3 Efeito da experiência com L2 e uso de L1

Nesta seção, tentamos correlacionar a ocorrência de atrito e de aprendizagem com a informação a respeito da experiência dos participantes com a L2 e a quantidade de uso de L1 no ambiente bilíngue. Começamos a discussão com os casos de atrito e aprendizagem que resultaram em mudança

qualitativa na produção dos contornos. No caso de atrito, aqueles contornos produzidos na condição BL1 que foram classificados como a realização de um dos tipos de contorno típicos do espanhol. No caso da aprendizagem, incluímos os casos em que contornos produzidos pelos bilíngues na condição BL2 foram considerados qualitativamente equivalentes aos dos falantes monolíngues do espanhol (grupo EM). Reunimos essas informações na tabela 3. Na tabela, a coluna “Atrito” indica a porcentagem de contornos (entre as cinco repetições produzidas por cada participante) produzidos pelos bilíngues em L1 que foram considerados atritados. A coluna “Aprendizagem” indica a porcentagem de contornos produzidos pelos bilíngues em L2 que foram considerados qualitativamente equivalentes a um dos tipos de contornos possíveis em L2. A coluna “Experiência com L2” indica a soma da quantidade de meses de residência do participante na Espanha e de instrução formal que o participante recebeu antes da mudança para a Espanha. A coluna “Uso de L1” mostra a estimativa de uso da L1 na Espanha, em horas por dia, informada pelo participante.

Tabela 3: Comparação da ocorrência de atrito e aprendizagem e experiência com L2 e uso de L1.

Falante	Atrito	Aprendizagem	Experiência L2	Uso de L1
f8	20%	20%	192	8
f3	100%	80%	190	0,1
m3	80%	40%	156	6
f5	100%	60%	78	2
f6	40%	60%	66	2
f7	0%	60%	45	7
m5	60%	100%	24	2
f4	100%	0%	9	0,1

O exame da tabela parece indicar algumas tendências relevantes. A primeira delas é que há uma grande sobreposição entre os participantes que produzem uma porcentagem não nula nas colunas Atrito e Aprendizagem (as exceções são *f7*, que mostra aprendizagem, mas não atrito e *f4*, que exibe atrito, mas não aprendizagem). Tomamos isso como uma indicação da conexão entre os dois fenômenos, isto é, o desenvolvimento da aprendizagem parece estar ligado à ocorrência de atrito. Outra tendência que parece ser relevante é o predomínio de participantes nessa tabela com um nível mais elevado de experiência com a L2 (seis entre oito com experiência de aproximadamente quatro anos ou mais). Sobre o uso de L1, um caso interessante é o da participante *f8*, que tem ao mesmo tempo a maior experiência com L2 e o maior nível de uso de L1. A maior experiência com L2 poderia levar a imaginar que ela apresentaria níveis maiores de aprendizagem e atrito, o que não é o caso. Esse resultado, no entanto, talvez possa ser explicado pelo elevado uso de L1, que contrabalançaria o efeito inibidor da L1 exercido pela elevada com a L2. Contrastamos esse caso com o da participante

f3, que tem um nível de experiência similar com a L2, embora com um uso consideravelmente menor de L1, fator que pode explicar seus índices mais elevados de aprendizagem e atrito.

Cabe, finalmente, dizer que, apesar de uma tendência geral favorável à ideia de uma conexão entre aprendizagem e atrito, ela parece ser bastante afetada por variabilidade individual, além do fato de que seria preciso levar em conta não só o aspecto quantitativo da experiência com L2 e uso de L1, mas também a qualidade do *input* de L2 e as circunstâncias em que o participante tem contato com a L1.

Consideramos, agora, casos na condição BL2 que podem ser considerados intermediários: participantes cuja produção foi considerada qualitativamente como exemplos de contornos típicos do PB, portanto, em princípio, casos de transferência L1-L2, mas que, quantitativamente, apresentam média de distância DTW que é estatisticamente superior em relação ao valor de referência do grupo monolíngue MP. Portanto, poderiam ser caracterizados como contornos que se afastam da norma do PB, mas que não se aproximam qualitativamente dos tipos de contorno do espanhol. Os participantes que se encaixam nesse perfil são: *f1*, *f3*, *f4*, *f7*, *f8*, *f10*, *m3* e *m4*. Destes oito, quatro fazem parte do grupo listado na coluna “Aprendizagem” da tabela 3, isto é, apresentaram casos de contorno classificados como qualitativamente equivalentes a um dos tipos do espanhol (*f3*, *f7*, *f8* e *m3*). Tomamos essa última observação como evidência de que a aprendizagem não implica um comportamento completamente estável, uma vez que um mesmo falante que produz parte dos contornos qualitativamente próximos à norma de L2 também pode ter parte de sua produção em um estado intermediário, isto é, nem coincidente com a norma de L1 nem com a de L2. Existem, ainda, casos como o da participante *f5*, por exemplo, no qual parte da produção em L2 é uma transferência sem perturbação quantitativa e parte pode ser classificada como aprendizagem bem-sucedida, no sentido de se aproximar qualitativamente da norma de L2.

4. Discussão

O estudo apresentou evidência quantitativa tanto em favor da ocorrência de aprendizagem de espanhol L2 quanto do atrito em português L1 na produção dos bilíngues, complementando os resultados de ordem qualitativa apresentados primeiramente em Silva e Arantes (submetido).

Com relação à aprendizagem, os resultados mostraram que em torno de 30% dos contornos produzidos pelos bilíngues na L2 podem ser considerados qualitativamente semelhantes aos padrões dos monolíngues espanhóis. Nesses casos, a variabilidade nos contornos dos bilíngues é muito próxima ou indistinguível da variabilidade interfalante dos monolíngues espanhóis. A respeito dos aproximadamente 70% restantes da produção dos bilíngues, observamos que parte dos contornos pode ser considerada uma transferência direta do padrão descendente global do PB e, nesses casos, a análise quantitativa mostrou que a variabilidade dos bilíngues é a mesma observada entre os

monolíngues do PB. Finalmente, uma parte da produção fica em situação intermediária, isto é, não corresponde qualitativamente a nenhum dos padrões possíveis na L2 e ao mesmo tempo a análise quantitativa da variabilidade desses contornos indica que estão quantitativamente distantes do padrão descendente global típico do PB. Além disso, foi possível observar que um mesmo participante pode apresentar um ou mais desses três padrões de comportamento em sua produção, o que entendemos como evidência para o caráter não estável do aprendizado até quando se trata de um mesmo indivíduo.

Olhando para os dados de atrito, vemos que a proporção de contornos produzidos pelos bilíngues em L1 que podem ser considerados qualitativamente semelhantes aos padrões observados na L2 é de 33%, número muito próximo à proporção de contornos que indicam aprendizagem por parte dos bilíngues. Entre os 33% de contornos em L1 que foram classificados como qualitativamente semelhantes a um dos padrões do espanhol, a variabilidade quantitativa observada foi semelhante estatisticamente àquela mostrada pelos monolíngues espanhóis. Dos 67% restantes dos contornos, a menor parte deles, distribuídos entre apenas 3 participantes, pode ser considerada semelhante ao padrão descendente global do PB e, nesses casos, a análise quantitativa mostrou que a variabilidade entre esses contornos é a mesma observada entre os monolíngues do PB. Finalmente, a parte restante e majoritária dos contornos, distribuída entre 12 participantes, fica em situação intermediária, isto é, não corresponde qualitativamente a nenhum dos padrões possíveis na L2, situação que mais claramente poderia ser considerada ação do atrito, e ao mesmo tempo quantitativamente seu grau de variabilidade indica que estão quantitativamente distantes do padrão descendente global típico do PB e, portanto, não poderiam também ser considerados exemplares típicos da L1.

Outro aspecto importante que o estudo mostrou foi a quantidade significativa de contornos que estão na condição que chamamos de intermediária, tanto no grupo BL1 quanto BL2, o que coloca uma questão a respeito de sua classificação. Em BL2, os contornos que se afastam quantitativamente de forma significativa da norma de L1 sem ao mesmo tempo assumir a forma qualitativa esperada para L2 podem ser considerados como produções em um estágio prévio ao “aprendizado” ou seriam apenas uma forma alterada do contorno de L1? O mesmo questionamento valeria para os casos intermediários em BL1. Seriam contornos já afetados pelo atrito, embora ainda não tenham alcançado qualitativamente a forma necessária para sua classificação como um contorno plenamente compatível com a norma de L2, estes sim, considerados verdadeiramente atritados? Para as duas perguntas anteriores ainda não temos uma resposta definitiva, dado que a literatura tanto sobre aprendizagem de entoação em L2 quanto sobre atrito em L1 parece não levantar essa questão sobre como analisar a gradência na manifestação dos efeitos do bilinguismo sobre a produção dos contornos tanto em L1 quanto em L2.

Finalmente, as observações apresentadas a respeito do possível efeito da experiência com L2 e do uso de L1 apontam para uma possível relação entre mais experiência com L2 e menos uso de L1 com, simultaneamente, maiores níveis de aprendizagem e também de atrito, embora sujeitos à variabilidade individual. Os resultados que apresentamos, embora simples e basicamente tentando correlacionar quantidades, indicam que não é possível apontar uma relação simples e direta entre essas duas variáveis e a produção, como aponta a literatura mais recente sobre o tema, por exemplo, a de Flege e Bohn (2021).

Referências

ARANTES, P. *better_f0*: A Praat script for better f0 extraction. [S.l.]: Zenodo, 2019. Disponível em: <https://zenodo.org/record/3470108>. Acesso em: 4 mar. 2021.

ARANTES, P. *parantes/time-normalized-f0*: Version 2.0 of time-normalized-f0 Praat script. [S.l.]: Zenodo, 2021. Disponível em: <https://zenodo.org/record/4489748>. Acesso em: 4 mar. 2021.

ARANTES, P. Time-normalization of fundamental frequency contours: a hands-on tutorial. In: MEIRELES, A. R. (org.). *Courses on Speech Prosody*. Newcastle upon Tyne: Cambridge Scholars Publishing, 2015. p. 98-123.

DE LEEUW, E.; MENNEN, I.; SCOBIE, J. M. Singing a different tune in your native language: first language attrition of prosody. *International Journal of Bilingualism*, v. 16, n. 1, p. 101-116, mar. 2012.

ESTEBAS-VILAPLANAS, E.; PRIETO, P. Castilian Spanish Intonation. In: ROSEANO, P. (org.). *Transcription of Intonation of the Spanish Language*. München: Lincom, 2010. p. 17-48.

FACE, T. F0 Peak Height and the Perception of Sentence Type in Castilian Spanish. *Revista Internacional de Lingüística Iberoamericana*, v. 3, n. 2(6), p. 49-65, 2005.

FLEGE, J. Second Language Speech Learning Theory, Findings, and Problems. *Speech perception and linguistic experience: Issues in cross-language research*. Timonium, MD: York Press, 1995. p. 233-277.

FLEGE, J. E. The production of “new” and “similar” phones in a foreign language: evidence for the effect of equivalence classification. *Journal of Phonetics*, v. 15, p. 47-65, 1987.

FLEGE, J. E.; BOHN, O.-S. The Revised Speech Learning Model (SLM-r). In: WAYLAND, R. (org.). *Second Language Speech Learning*. [S.l.]: Cambridge University Press, 2021. p. 3-83. Disponível em: https://www.cambridge.org/core/product/identifier/9781108886901%23CN-bp-1/type/book_part. Acesso em: 8 abr. 2021.

FROTA, S.; MORAES, J. A. Intonation in European and Brazilian Portuguese. In: WETZELS, L.; COSTA, J.; MENUZZI, S. (org.). *The Handbook of Portuguese Linguistics*. Chichester: Wiley Blackwell, 2016. p. 141-166.

GIORGINO, T. Computing and Visualizing Dynamic Time Warping Alignment in R The dtw Package. *Journal of Statistical Software*, v. 31, n. 7, p. 1-24, 2009.

HENRIKSEN, N. Style, prosodic variation, and the social meaning of intonation. *Journal of the International Phonetic Association*, v. 43, n. 2, p. 153-193, ago. 2013.

HENRIKSEN, N. Wh-question intonation in Peninsular Spanish: Multiple contours and the effect of task type. *Journal of Portuguese Linguistics*, v. 8, n. 1, p. 45-73, 2009.

KUPSKE, F. *Imigração, atrito e complexidade: a produção das oclusivas surdas iniciais do inglês e do português por sul-brasileiros residentes em Londres*. 2016. Tese (Doutorado em Estudos da Linguagem) – Instituto de Letras, Universidade Federal do Rio de Janeiro, Porto Alegre, 2016.

LUCENTE, L. *Aspectos dinâmicos da fala e da entoação do português brasileiro*. 2008. Dissertação (Mestrado em Linguística) – Instituto de Estudos da Linguagem, Universidade Estadual de Campinas, Campinas, 2008.

LUCENTE, L. *DaTo: um sistema de notação entoacional para o português brasileiro baseado em princípios dinâmicos. Ênfase no foco e na fala espontânea*. 2012. Tese (Doutorado em Linguística) – Instituto de Estudos da Linguagem, Universidade Estadual de Campinas, Campinas, 2012.

MAJOR, R. C. Losing English as a First Language. *The Modern Language Journal*, v. 76, n. ii, p. 190-208, 1992.

MENNEN, I. Beyond Segments: Towards a L2 Intonation Learning Theory. In: DELAIS-ROUSSARIE, E.; AVANZI, M.; HERMENT, S. (org.). *Prosody and Language in Contact: L2 Acquisition, Attrition and Languages in Multilingual Situations*. Berlin: Springer, 2015. p. 171-188.

MENNEN, I. Bi-directional interference in the intonation of Dutch speakers of Greek. *Journal of Phonetics*, v. 32, p. 543-563, 2004.

MIRANDA, L. *Análise da entoação do português do Brasil segundo o modelo IPO*. 2015. Dissertação (Mestrado em Letras Vernáculas) – Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2015.

MIRANDA, L.; RILLIARD, A.; MORAES, J. Audiovisual perception of wh-questions and wh-exclamations in Brazilian Portuguese. In: 19TH INTERNATIONAL CONGRESS OF PHONETIC SCIENCES, 2019, Melbourne, Australia. *Anais...* Melbourne, Australia: [s.n.], 2019.

- MORAES, J. A. Intonation in Brazilian Portuguese. In: HIRST, D.; DI CRISTO, A. (org.). *Intonation Systems: a Survey of Twenty Languages*. Cambridge: Cambridge University Press, 1998. p. 179-194.
- MORAES, J. A. The Pitch Accents in Brazilian Portuguese: analysis by synthesis. In: *FOURTH CONFERENCE ON SPEECH PROSODY*, 2008, Campinas. Anais... Campinas: [s.n.], 2008. p. 389-397.
- NAVARRO, T. *Manual de pronunciación española*. 27. ed. Madrid: Editorial Consejo Superior de Investigaciones Científicas, 1999.
- OLIVEIRA, A. F. *Caracterización de la entonación del español hablado por brasileños*. 2013. Thesis – Universitat de Barcelona, Barcelona, 2013.
- QUILIS, A. *Tratado de fonología z fonética españolas*. Madrid: Editorial Gredos, 1993.
- QUILIS, A.; FERNÁNDEZ, J. *Curso de Fonética y Fonología Españolas*. 18. ed. Madrid: Consejo Superior de Investigaciones Científicas, 2003.
- RSTUDIO TEAM. *RStudio: Integrated Development for R*. Boston: [s.n.], 2021. Disponível em: <http://www.rstudio.com/>. Acesso em: 8 abr. 2021.
- SÁ, P. C. F. *Análise entonacional de enunciados assertivos, continuativos e interrogativos lidos em piadas espanhol/LE e espanhol/LM*. 2008. Dissertação (Mestrado em Letras) – Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2008.
- SÁNCHEZ AGUILAR, A. *Don Quijote Adaptación, notas y actividades*. Barcelona: Vicens Vives, 2004.
- SANCIER, M.; FOWLER, C. Gestural drift in a bilingual speaker of Brazilian Portuguese and English. *Journal of Phonetics*, v. 25, p. 421-436, 1997.
- SAVINO, M. The intonation of polar questions in Italian: Where is the rise? *Journal of the International Phonetic Association*, v. 42, n. 1, p. 23-48, abr. 2012.
- SCHMID, M. *Language attrition*. New York: Cambridge University Press, 2011.
- SCHMID, M. S.; KÖPKE, B. The relevance of first language attrition to theories of bilingual development. *Linguistic Approaches to Bilingualism*, v. 7, n. 6, p. 637-667, 31 dez. 2017.
- SILVA, C. C.; ARANTES, P. *A qualitative study on the variability in intonation learning and attrition in Brazilian Portuguese bilingual speakers of Spanish L2*. (submetido)

SILVA, C. C. *Análise fonético-experimental da entoação de declarativas e interrogativas em espanhol/LE*. 2016. Tese (Doutorado em Linguística) – Instituto de Estudos da Linguagem, Universidade Estadual de Campinas, Campinas, 2016.

SILVA, C. C.; ARANTES, P. Quantitative analysis of fundamental frequency in Spanish (L2) and Brazilian Portuguese (L1): evidence of learning and language attrition. *Journal of Speech Sciences*, v. 10, p. e021003, 8 dez. 2021.

SOSA, J. Wh-questions in Spanish: Meanings and Configuration Variability. *Catalan Journal of Linguistics*, v. 2, p. 229-247, 2003.

SOSA, J. M. *La entonación del español*. Madrid: Cátedra, 1999.

TENANI, L. *Domínios Prosódicos no Português do Brasil: implicações para a prosódia e para aplicação de processos fonológicos*. 2002. Universidade Estadual de Campinas, Campinas, 2002.

TRUCKENBRODT, H.; SANDALO, F.; ABAURRE, M. B. Elements of Brazilian Portuguese intonation. *Journal of Portuguese Linguistics*, v. 8, n. 1, p. 77-115, 2008.

Publique com a gente e compartilhe
o conhecimento



www.lettraria.net

**Prosódia e
Bilinguismo**

**Prosodia y
Bilingüismo**

**Prosody and
Bilingualism**

 **Letraria**[®]