THE LISTENER IN LANGUAGE CHANGE

A DISSERTATION
SUBMITTED TO THE DEPARTMENT OF LINGUISTICS
AND THE COMMITTEE ON GRADUATE STUDIES
OF STANFORD UNIVERSITY
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Simon John Todd
June 2019

Abstract

Language changes constantly, in ways that can be influenced by factors both language-internal, such as word frequency, and language-external, such as social organization and attitudes. A major challenge for linguistic theory is to give a unified explanation of these constraints on language change. In this dissertation, I argue that this challenge can be addressed by looking to spoken language perception, where passive but powerful perceptual biases give rise to many similar constraints on how listeners update the cognitive representations they draw upon for language use.

I present an approach to language change in which perceptual biases in the listener play a central role. I ground this approach in an exemplar-based computational model, which is able to recreate empirically-observed general properties of sound change. I then test the approach by integrating experimentally-supported perceptual biases with computational modeling and novel corpus methods across two studies. In the first study, I apply the computational model to simulate word-frequency effects in sound change. I show that different word-frequency effects in different kinds of sound change follow from a single perceptual bias, whereby high-frequency words are recognized more easily than low-frequency words when acoustically ambiguous. In the second study, I extend the listener-based approach to the effect of improving interethnic social attitudes on the spread of lexical items across ethnic groups in New Zealand. Drawing on biases in the perception of 'other-accented' words, I make specific predictions for the spread of the tag *eh* from indigenous Māori to White Pākehā, which I test with novel corpus methods. Taken together, these two studies highlight how passive but powerful perceptual biases in the listener can give a unified explanation of different constraints on language change.

Acknowledgments

The path to a PhD takes a lot of time, a lot of energy, and a lot of support. I am deeply indebted to everyone who has walked alongside me on parts of this path and helped me get to where I am today, and I am grateful to have this opportunity to thank them all for being there. Reader, beware: I have a lot of people to thank.

Above all, I would like to thank my advisors: Dan Jurafsky, Meghan Sumner, and Jen Hay. Dan, thank you for your inexhaustible energy and enthusiasm, for knowing exactly how to distill extremely broad literatures, and for always telling me to take more time off. Meghan, thank you for your keen focus on the theoretical picture, for giving me sage advice even in difficult situations, and for always making me feel like an equal. Jen, thank you for introducing me to the exciting world of linguistic research, for always staying connected and welcoming me back home, and – together with Janet Pierrehumbert – for your long-lasting collaboration, which led directly to Chapters 2 and 3 of this dissertation. I consider myself lucky to have been able to learn so much from each of you over so many years, about how to be an effective linguist, an innovative researcher, and an all-round good and happy person.

I am indebted to all those who contributed to this dissertation by attending various talks about it and asking interesting questions. My thanks to the attendees of my dissertation defense and practice job talk, to my labmates in the Stanford Phonetics Lab, and to audiences of: the Phonetics and Phonology, Sociolinguistics, and Cognitive Science Workshops at Stanford University; the Fellows' talk series at the Stanford Humanities Center; the CogNetwork and Phonetics and Phonology Workshops at the University of California, Berkeley; the seminar series at the New Zealand Institute of Language, Brain, and Behaviour; and the linguistics colloquia at the University of Illinois at Urbana-Champaign and the University of California, Santa Barbara.

One of the things that I have found particularly enjoyable about my time at both the University of Canterbury and Stanford is a freedom of movement around an intellectual community. They say it takes a village to raise a child; in my case, it took a department (or two!) to produce a scholar. I'd like to thank everyone that I had the pleasure of working with and learning from during my time in higher education. At the University of Canterbury, I am particularly grateful to the faculty and staff who were always happy to welcome me back as if I hadn't been away at all: Lynn Clark, Donald Derrick, Beth Hume, Jeanette King, Vica Papp, Emma Parnell, Heidi Quinn, and Kevin Watson. Special thanks go to Alex James, for introducing me to the world of mathematical and computational modeling and fighting for my right to exist in that world as a linguist. At Stanford, I am particularly grateful to those who met with me to discuss key projects and teaching opportunities, usually over extended periods of time: Arto Anttila, Joan Bresnan, Boris Harizanov, Paul Kiparsky, Rob Podesva, Chris Potts, and John R. Rickford. I'd also like to thank Dan Lassiter for involving me in the organization of the Cognition and Language Workshop, and I must give a special thanks to Beth Levin for going above and beyond the call of duty to manage so much of department life. Finally, I am also grateful to Judith Degen, for being an external member of my dissertation proposal and defense committees, and to Jay McClelland, for chairing my dissertation defense committee.

My time at Stanford was made all the better by the amazing people I got to share it with, and all the ways they helped to balance out my life. I am deeply appreciative of my cohort, Daniel Galbraith, Olek Główka, Sunwoo Jeong, Sara Kessler, Kate Lindsey, and Prerna Nadathur, without whom these past six years would have been depressingly empty. I will never forget our wine-fueled game nights, escape rooms, classic film screenings, adventures to San Francisco for kosher vegan Chinese food, white elephant Christmas parties, combined birthday celebrations, annual brownies, and, above all, daily lunches. Special thanks to Sunwoo, for constantly expanding my vocabulary; to Sara, for baking and saving the world with me at least every other week; to Kate, for being a great support on the job market, even when we were competing for the same jobs; to Prerna, for sharing an office with me for six years, complete with videos of dogs, elephant vocalizations, and countless D&Ms; and to all four of these wonderful friends, for being there for me when I was at my lowest.

Life would also have not been the same without all of the other students in the Linguistics Department, especially those who shared in lunches, volleyball games, the *Quiztel Meth* trivia team, game nights, excursions to the local pub or further afield in San Francisco, and countless other social activities, and who came to visit me after surgery: Sam Bowman, CJ Brickhouse, Ignacio Cases, Branden Chan, Reuben

Cohn-Gordon, Dora Demszky, Tim Dozat, Roey Gafter, Lelia Glass, Chantal Gratton, Sabrina Grimberg, Katherine Hilton, Masoud Jasbi, SK Kim, Ed King, Bonnie Krejci, Judy Kroo, Emily Lake, Daisy Leigh, Yiwei Luo, Zion Mengesha, Eva Portelance, Ciyang Qing, Nay San, Sebastian Schuster, Rob Voigt, Brandon Waldon, and Robert Xu. Nor would life have been the same without Manny (Marcel) Rincon-Cruz, who went beyond the expectations of a simple roommate to become a dear friend and vital lifeline when things got tough. I am 'grateful as' to my fellow Stanford Kiwis (and co.), for sharing in regular celebrations of New Zealand culture and for continuing to expose me to refreshing renditions of the New Zealand accent long after I lost my own: Thil Arachchi, Sam Corbett-Davies, Max Ferguson, Oliver Hinder, CZ Lee, Sarah Poole, Priya Shekar, Robin Wilson, and Liz Wiltshire. Special thanks here to David Wright & Sarah Pratt, for formally registering us as a student group and thus entitling us to free barbecues and beer, and for liking me enough to follow me to the other side of the world after 10 years of shared high school and undergrad and 2 years of living together. Finally, I would like to thank all my fellow fellows at the Stanford Humanities Center in 2018-19, who have enriched my final year both intellectually and socially through countless presentations, shared lunches, and happy hour celebrations.

Going to grad school meant leaving the nest – leaving my hometown for the first time. I am deeply appreciative of my family and friends and all they have done to make this separation easier over the years. I am lucky to have had almost all of my immediate family members travel halfway across the world to visit me during my time here, especially for graduation, and to have had them welcome me back home with open arms on every annual visit. My deepest appreciations to my parents, Derek Todd & Kerry Luke and Leonie Young & Nigel Baker, and to my brothers and their partners, Willie Todd & Chanel Johnson, Kahu Luke & Líanna Fitzsimons, and Ranui Luke & Kris Wong, for always being there. My greatest thanks also to my extended family in New Zealand, especially my grandparents, Stuart & Jenny Young, for always being excited to host me and get updates on my life when I visited. I am also lucky to have had a large group of friends from high school and undergrad come together from far-flung places every year when I visit, and I am grateful to them all for never missing these chances to have fun as though nothing had changed in over 10 years: Steve Bates, Kerianne Butler, Michelle Chim, Emily Craig, Kate Dawkings, Will Douglas, Mason (Garry) Drew, Tynam McCulloch, Adrian McMurray, Shannon O'Brien, Chris Scott, David (Shippy) Shipton, Alex (Frank) Smith, Andrew (Fish)

Wright, and Tomo Yee. Finally, I am thankful to the Crosbie family, Peter, Teresa, Asnakech and Anteneh, for giving me a home away from home and annually inviting me to share their family holidays for Thanksgiving and Christmas.

In the past few years especially, I have come to consider the Bay Area a new home, largely due to the influence of my wonderful partner, Will Hsiao. Will, I cannot thank you enough for sharing your life with me and letting me share mine with you, and I can't wait to see what new adventures we have ahead of us. My greatest thanks also to my āyí, Ruby Poh, and my adopted canine daughter, Marley, for making me feel at home and loved. Finally, my thanks to my new friends, for welcoming me so readily into your circle and sharing countless game nights, escape rooms, meals, and cocktails: Kevin Fok, Aaron Ho, Jeff Hsieh, Helen Huynh, Brian Lan, Naomi Lee, Jennifer Luh, Michelle Tsai, Alfred Wang, Daniel Wong, Anne Wu, and June Yu.

In my time at Stanford, I have taken advantage of a lot of fantastic services, and I am immensely grateful to the staff members who made those services happen. Above all else, I am grateful to the Student Services Managers of the Department of Linguistics, Ann Marie Pettigrew and Jeanne Su, for untangling all sorts of administrative knots and ensuring that my path to degree completion always remained clear. Not a year went by where I didn't injure myself in some way, and I am indebted to Zaldy De La Cuesta and the team at Vaden's More Physical Therapy office for fixing me up every time. I am also indebted to Christine Gibo-Crapps for helping me navigate when I felt lost, for getting me back into on-campus housing when I was homeless, and for being a fantastic GLO Dean when I was a Community Associate. I am grateful to Aneel Chima, for giving me the mindfulness and meditation tools to bounce back from negative mental and emotional experiences and create a balanced environment where I could thrive. Finally, I am grateful to Karen Biestman, Greg Graves, and Denni Woodward, for welcoming me into the Native American Cultural Center as a student of Māori descent and always inviting me to connect with my Ngāi Tahu roots during the annual First Nations' Futures Program.

Doing a PhD is expensive, and I would never have gotten near completing mine without the financial support of numerous institutions. I am immensely grateful to the Department of Linguistics and the School of Humanities and Sciences at Stanford for their generous funding over my six years of graduate study, and to the programs that have contributed additional funding to my graduate career: Fulbright New Zealand; the Stanford University Pigott Scholars Program; and the Stanford Humanities Center Geballe Dissertation Prize Fellowship. I am also grateful for grants that have supported my research with collaborators, from the John Templeton Foundation (Award ID 36617; PI Janet Pierrehumbert) and the Royal Society of New Zealand Te Apārangi (Rutherford Discovery Fellowship to Jen Hay; Grant Number E5909).

Since going to grad school is but one step on the path of education, I am grateful to all of the sources of financial support during my earlier experiences with research, teaching, and learning: the University of Canterbury (including the College of Arts, the Department of Mathematics and Statistics, and the Graduates Association); the New Zealand Institute of Language, Brain, and Behaviour; the Linguistic Society of New Zealand; the New Zealand Mathematical Society; the New Zealand Institute of Mathematics and its Applications; the Māori Education Trust; Freemasons New Zealand; the New Zealand Qualifications Authority; the HMNZS Canterbury; Transpower; St. Andrew's College; and te Rūnanga o Ngāi Tahu.

Finally, since writing a dissertation rooted in empirical observations is impossible without data, I am indebted to all those whose earlier work has made available the data that I use here. I would particularly like to acknowledge the work done in creating the ONZE corpus, which provides an empirical foundation to most of my dissertation. I am grateful to all those who conducted interviews for the corpus: the Mobile Disc recording Unit of the NZ Broadcasting Service, Rosemary Goodyear, Lesley Evans, members of the NZ English class of the University of Canterbury Linguistics Department, and members of the ONZE team. I am also grateful to members of the ONZE team, for all of the additional work in preparing the data, making transcripts, and obtaining background information, and to Robert Fromont, for creating and maintaining the LaBB-CAT software used to access the data. Finally, I am grateful to all of the funding sources that have helped to create and support the corpus: the University of Canterbury; the Foundation for Research, Science and Technology (the New Zealand Public Good Science Fund); the Royal Society of New Zealand; the New Zealand Lotteries Board Fund; and the Canterbury History Foundation.

This dissertation is dedicated to those who never had the chance to see it completed:
my grandparents, Winnie & George;
my great aunt, Lorna;
and my brother, Tamati.

Me te aroha nui ki a koutou, e te whānau.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

This dissertation is about language change. Understanding language change is important for two main reasons. Firstly, change is inherent to the nature of language, as languages are constantly changing. Secondly, change is 'nature's laboratory'; just as Darwinian evolution does for biological entities, language change reveals the factors that are of fundamental importance in giving language its unique shape. The branches of linguistics that are most deeply concerned with language change, historical linguistics and sociolinguistics, have revealed that it is influenced by factors both language-internal, such as how often a word is used, and language-external, such as social organization and attitudes. A major challenge for linguistic theory is to explain both how these different influences work independently, and how they fit together in the complex system of language change.

In meeting this challenge, it is important to ask not just *what* happens when language changes, but also *how* and *why* language changes. By asking how and why, we can address longstanding issues concerning the asymmetry and timing of change. For example, how and why does change spread through some words or some groups of speakers faster than others (e.g. for words: Hay & Foulkes, 2016; Hay, Pierrehumbert, Walker, & LaShell, 2015; for speakers: Labov, 1963, 1972)? How and why is a given instance of language change triggered when it is, and not earlier or later (Weinreich, Labov, & Herzog, 1968)? To answer these questions – and to ultimately arrive at a satisfying unified explanation of language-internal and -external influences on language change – it is necessary to focus not on the *outcomes* of language change, but rather on the *process* by which language changes.

The traditional approach to the process of language change dates back to Hermann Paul in the late 19th century and attributes a primary role to child language *learners* (see e.g. Lightfoot & Westergaard, 2007; Weinreich et al., 1968). According to this approach, language changes cross-generationally through the iterated application of learning biases that structure 'errors' in language acquisition. In effect, the child language learner filters variation in the ambient language environment, magnifying components of this variation that resonate with learning biases and reducing components that do not. This approach has allowed enormous empirical advances, such as the synchronic study of language change-in-progress through the apparent time construct (e.g. Cukor-Avila & Bailey, 2013), and the modeling of the development of linguistic structure in longer-term language evolution both computationally (e.g. Nowak & Krakauer, 1999) and experimentally (e.g. Kirby, Cornish, & Smith, 2008).

However, recent work has made it clear that a focus on the learner does not offer a complete explanation for the process of language change, as an individual's language continues to change throughout their lifetime.[1] These changes are most easily demonstrated at the phonetic level. For example, adults who relocate or otherwise encounter large numbers of speakers of a different dialect throughout a sustained period post-adolescence may exhibit robust changes in accent (Evans & Iverson, 2007; Munro, Derwing, & Flege, 1999). Similarly, adults who remain in one community

[1]Of course, change in adults arises because they continue to learn throughout the lifetime. Nevertheless, I choose not to label adults as 'learners' so as to avoid potential confusion with child language acquisition. My alternative terminology is intended to highlight the ways in which the present proposal deviates from the longstanding tradition in the theoretical literature.

may continue to participate in sound change that community exhibits (Harrington, 2006; Harrington, Palethorpe, & Watson, 2000; Sankoff & Blondeau, 2007). By way of computational modeling, Stanford and Kenny (2013) even suggest that the signature of language change-in-progress – age-graded language variation – may be attributable to a single mechanism of linguistic diffusion that applies equally to adults and children, without the need for additional special mechanisms related to child language acquisition.

To account for language change throughout the lifetime, it is necessary to adopt a perspective that focuses on adult language users, i.e. on the *speaker* and/or the *listener*.[2] In keeping with the highlights of the traditional learner-based approach, the speaker-/listener-based perspective explains the process of language change through biases that structure the production and/or uptake of language variation. However, these biases are activated by in-the-moment communication between adults, rather than by language learning in children.

Under an approach to language change that focuses on in-the-moment variable language usage by adults, it is likely that biases of both the speaker and the listener are relevant. However, much work in this area focuses exclusively on the speaker. For example, H&H theory (Lindblom, 1990) and its modern descendants (Buz, Tanenhaus, & Jaeger, 2016) posit that speakers modulate their speech signal in order to ensure that they are understood by listeners, which means producing highly predictable words differently than unpredictable words (in context). Similarly, theories related to audience design (Bell, 2001) posit that speakers modulate their manner of linguistic presentation in order to evoke certain social evaluations from, and build solidarity with, their listeners.

[2] The focus on the speaker and/or listener need not supplant the focus on the learner – as it remains likely that the child language learner plays an important part in structuring cross-generational language change – but should at least find a place alongside it in theories about the process by which language changes.

It is clear that speaker-based biases exist: speakers *can* and *do* modulate their speech for reasons such as comprehensibility or social appearance. However, it also seems clear that intention plays a large role in this modulation, and that speakers often *do not* modulate their speech in this way, even in experimental situations where the value of modulation is both heightened and made salient (e.g. Arnold, Wasow, Asudeh, & Alrenga, 2004; Bard et al., 2000). On this basis, it seems questionable to view speaker-based biases as exclusively central to the process of language change. The driving of change by biases that arguably rely on intention is at odds with the fact that many cases of language change proceed below the level of consciousness (Labov, 2001). Furthermore, the effect of any such biases is typically small relative to natural variation in speech, and there is presumably a minimum degree of consistency with which small biases need to apply in order to drive large-scale change. If active and sporadic speaker-based biases play a role in language change, then it is reasonable to assume that they do so primarily against a background provided by other, passive and constant biases.

I argue that listener-based biases provide such a background, and therefore that the listener deserves consideration as a central component of the process of language change. Since many listener-based biases are developed by the cognitive system as a means of efficiently extracting linguistic and social information from a spoken utterance (Sumner, Kim, King, & McGowan, 2014), they can apply without intention or even awareness. They are also difficult to consciously disable, and apply more often that speaker-based biases simply because people spend more time listening than speaking (see e.g. Emanuel et al., 2008, and references therein). Listener-based biases have a direct link to linguistic behavior, because speech that is listened to is stored in memory (Goldinger, 1996) and can affect future speech production and perception (see e.g. for production: Fowler, Brown, Sabadini, & Weihing, 2003; Goldinger, 1998; Nielsen, 2011; Pardo, 2006; and for perception: Bradlow & Bent, 2008; Clarke-Davidson, Luce, & Sawusch, 2008; Dahan, Drucker, & Scarborough, 2008; Kraljic & Samuel, 2006; Norris, McQueen, & Cutler, 2003). This link provides a clear process through which listener-based biases can shape language change in both the individual and the community, as the process of perceiving (spoken) language updates the cognitive representations that may be drawn upon for producing language at later times.[3] Through this process, listener-based biases have the potential to situate language-internal and language-external influences on language change within a single unified system, as many perceptual biases concerning both kinds of influence have been experimentally observed (see e.g. Connine, Titone, & Wang, 1993; Ganong, 1980; Hay & Drager, 2010; Hay, Warren, & Drager, 2006; Kleinschmidt & Jaeger, 2015; Niedzielski, 1999; Pitt & Samuel, 1993; Strand & Johnson, 1996).

In this dissertation, I advance a listener-based approach to language change. I ground this approach in the results of speech perception experiments, which have revealed passive but powerful perceptual biases in the listener, and in the results of corpus studies of language change, which have revealed structured patterns in language change. I demonstrate plausible causal links between these perceptual biases and empirical patterns of change through formal computational modeling, which I then draw upon to support new empirical research of language change-in-progress. To underline the generality of the listener-based approach and the potential it holds for explaining the influences of both language-internal and language-external factors on language change, I develop analyses for change at two different levels of linguistic

[3] In stating that production may draw upon the representations formed and updated through perception, I emphasize the *may*. One interpretation of the link between production and perception is that they access the same representational space, but this does not mean that they do so in the same way. At a minimum, production accesses a smaller portion of the space than perception, because the range of variation produced as a speaker is much smaller than the range of variation perceived as a listener. While all perceptual experiences update the representational space, only some of them do so in the portions accessed by production. Thus, only *some* perceptual experiences have direct implications for later production.

structure – the phonetics-phonology interface and the lexicon – under two different kinds of influence – word frequency and social attitudes, respectively.

In Chapter 2, I implement the listener-based approach in a computational model of regular sound change. In doing so, I formalize the representations and processes that are involved in the approach, and I show how they work together to generate language change. In particular, I show that the model is capable of generating changes with properties seen empirically in corpus data from New Zealand English, which no previous model has been able to do, and that this capability stems primarily from the adoption of a heavy focus on the listener. I also discuss how the model clarifies key distinctions between the listener-based approach developed in this dissertation, which relies on perceptual biases, and the well-known previous listener-based approach developed by Ohala (1981), which relies on misperception. The model developed in this chapter provides a formal foundation for predictions that specific perceptual biases make about patterns of language change, which I explore in the later chapters.

In Chapter 3, I explore the potential of the listener-based approach for explaining a particular language-internal influence on a particular kind of language change, the effect of word frequency on rates of regular sound change. I consider a puzzle presented by the three existing corpus studies of word frequency effects on rates of sound change, which have each found different effects. I argue that a listener-based approach can solve this puzzle, as the different effects correspond to different kinds of sound change, with different implications for the listener. I extend the model developed in Chapter 2 to incorporate an experimentally-established perceptual bias relating word frequency to word recognizability, and I show that this single bias generates all of the different word frequency effects seen empirically. Crucially, by comparing the model results to those obtained when certain processes in the speaker or listener are switched off, I show that the listener plays a causal role in generating the word frequency effects, and that the speaker plays only a minor facilitatory role. I conclude by discussing the implications of these results and the predictions they make for other kinds of regular sound change.

Finally, in Chapter 4, I explore the potential of the listener-based approach for explaining a particular language-*external* influence on a *different* kind of language change, the effect of social attitudes on interethnic lexical adoption. I first show how a listener-based approach can be extended to socially-distributed discrete linguistic elements, and how it can draw upon experimentally-established perceptual biases associated with social information and attitudes. I then consider a puzzle presented by previous corpus studies of the discourse tag *eh* in New Zealand, which is heavily used by – and strongly associated with – indigenous Māori, but also used by young White Pākehā (i.e. non-indigenous New Zealanders, typically of European descent). Previous studies have disagreed whether the use of *eh* by young Pākehā is an isolated age-graded phenomenon or whether it represents a larger language change-in-progress whereby Pākehā in general are beginning to adopt *eh* from Māori. I argue that a listener-based approach provides a solution to this puzzle, both by making salient a way to disentangle the predictions of the age-grading and change-in-progress hypotheses, and by giving a basis for actuation of change-in-progress in the perceptual effects of recent improvements in Pākehā attitudes toward Māori. Through using a new methodological tool in the study of a much larger corpus than previously analyzed, I find support for the change-in-progress hypothesis, as well as evidence that this change is associated with improvements in social attitudes, as predicted by the listener-based approach. I conclude by discussing the general advantages of a listener-based approach to social factors in language change, together with the general predictions the approach makes for such cases.

I draw the three core chapters together in Chapter 5 by highlighting their major theoretical and practical contributions and their implications for future work, both narrowly within the study of language change, and more broadly across other subfields of linguistics. In doing so, I argue that a listener-based approach to language change lays the foundation for a theory that can unify the influences of language-internal and -external factors on language change.

Chapter 2

A listener-based model*

As described in Chapter 1, a listener-based approach to language change relies on the notion that the process of perceiving language updates the cognitive representations that may be drawn upon for producing language. In order to draw predictions for change from this approach, it is necessary to formulate a model for *how* this process of updating works. For the sake of understanding, this model need not capture the exact, intricate details of the accessing and updating of linguistic representations; rather, it should capture the broad strokes in as simple a way as possible. Maximal clarity will be offered by a model that is computationally implemented, as such implementation forces assumptions and hypotheses about the representations and processes involved in production and perception to be made explicit and precise, and allows simulations to reveal concrete patterns of change that follow from these assumptions and hypotheses.

In this chapter, I implement the listener-based approach in a computational model, providing a foundation for predictions that specific perceptual biases make about patterns of language change in Chapters 3 and 4. To do this, I first lay out the

*This chapter is based on work published as Todd, Pierrehumbert, and Hay (2019). The idea of developing a model was initially proposed by my coauthors. The design, implementation, and analysis of the model is primarily my own work, but it has benefited from the input of my coauthors.

concrete focus of the model, on a particular kind of language change, and develop desiderata for the simulations based on corpus evidence from New Zealand English (Section 2.1). I then establish the theoretical and implementational framework of the model at a high level (Section 2.2), before zooming in to lay out the details of the representations and processes in the model (Section 2.3). Next, I show how the processes work together to simulate regular sound change, and how the model can be tuned to meet the desiderata (Section 2.4). Finally, I discuss the general advantages offered by the computational implementation of a listener-based approach to language change (Section 2.5).

2.1 Focus

For concreteness, the model in this chapter focuses on *regular sound change*, defined as the gradual transformation of the phonetic realization of a phoneme over time (Labov, 2010). Within this realm, it focuses on simplified versions of two kinds of regular sound change: *phonetic drift*, which involves isolated changes in the realizations of a single phoneme; and *push chains*, which involve joint changes in the realizations of two phonemes.

The focus on regular sound change makes the model directly applicable to the effects of word frequency on rates of sound change (Chapter 3). Though the model itself cannot be directly applied (as stated) to other kinds of language change, the insight that it offers can be applied generally. It is this insight that provides a foundation for understanding the effects of social attitudes on the adoption of a lexical item across ethnic groups (Chapter 4).

2.1.1 Why regular sound change?

Regular sound change is a good candidate for a listener-based model for three primary reasons: (i) it can occur within an individual, across the lifespan; (ii) it has parallels with synchronic phonetic accommodation; and (iii) it often proceeds below the level of awareness, in a way that is not subject to individual control.

Firstly, regular sound change can occur within an individual, across the lifespan. For example, Harrington et al. (2000) document changes in 10 of 11 monophthongal vowels of Queen Elizabeth II over 50 years, and Harrington (2006) show that (at least one of) these changes are above and beyond those expected based on age-related factors. The fact that regular sound change can occur within the individual implies that it is not solely driven by the child *learner*, through biases in language acquisition. Instead, there is a clear role in regular sound change for the adult *speaker* or *listener*, consistent with a listener-based approach.

Secondly, the diachronic phenomenon of regular sound change is paralleled by the synchronic phenomenon of phonetic accommodation, whereby a speaker's phonetic realizations change on the basis of those they hear from others (e.g. Fowler et al., 2003; Goldinger, 1998; Nielsen, 2011; Pardo, 2006). This parallelism is heightened by the finding from the aforementioned studies (Harrington, 2006; Harrington et al., 2000) that the Queen's phonetic realizations changed over time to become more similar to those of Standard Southern British, which is a prevalent variety. As in diachronic sound change, there is evidence that synchronic accommodation is mediated by social factors, but it appears to be mostly automatic (Dijksterhuis & Bargh, 2001; Pickering & Garrod, 2004). For example, Babel (2010) found convergence across a range of vowels for New Zealanders exposed to Australian speech, with more convergence among participants that held positive attitudes toward Australia and less (but convergence nevertheless) among those that held negative attitudes. The automaticity of phonetic accommodation implies a strong connection between the linguistic representations accessed during speech perception and speech production, and the parallelism with regular sound change implies that this connection can have lasting diachronic effects, as is expected from a listener-based approach.

Finally, and relatedly, regular sound change has been classified as beginning – and often proceeding to completion – below the level of awareness (Labov, 1994). This observation accords with an intuition that, in general, phonetic realization is less subject to active speaker planning or strategizing than things like the choice or arrangement of words. That is not to say that speakers cannot or do not manipulate their phonetic realizations actively, for strategic reasons – many aspects of phonetic variation have been analyzed to reflect speakers' stylistic practice (e.g. Eckert, 2008; Podesva, 2007) or communicative intent (e.g. Buz et al., 2016; Lindblom, 1990). However, it is not clear that speakers consistently exhibit such active design processes in speech production; for example, Bard et al. (2000) observe that speakers reduce intelligibility when repeating referring expressions, even when they know that the listener has no knowledge of the referent. Conversely, the passive processes involved in speech perception are automatic and inescapable; for example, it is impossible to listen to someone speak without forming inferences and evaluations about who they are and/or what the situation is (e.g. Sumner et al., 2014). Since a large component of regular sound change occurs below the level of awareness, it is plausibly not a consequence of active processes in the speaker, but rather a consequence of passive, automatic processes in the listener, consistent with a listener-based approach.

2.1.2 Kinds of sound change modeled

The model in this chapter is designed to capture two kinds of regular sound change: phonetic drift and push chains.

Phonetic drift involves the movement of a single phoneme representation in the acoustic space over time. Crucially, this movement is isolated with respect to other phoneme representations in the acoustic space: it is not motivated by proximity to another phoneme representation, nor does it result in encroachment on the acoustic territory of another phoneme representation. In this way, the movement may be considered to be caused by a bias that is external to the phoneme representation itself, such as a production bias based in the reduction of articulatory effort or in the identification with a particular social group. An example of phonetic drift is /t/-glottaling in Manchester English (Bermúdez-Otero, Baranowski, Bailey, & Turton, 2015), where the phonetic realization of intervocalic /t/ has drifted to [ʔ] over time.[4] /t/-glottaling constitutes phonetic drift because it only affects /t/ and because it produces realizations that remain unlike any other phoneme of English (as there is no phoneme /ʔ/).

Push chains involve the interaction of two phoneme representations, where one moves in the acoustic space toward the other, which in turn moves away. The movement of the first phoneme representation may be considered to be caused by an external bias, like in phonetic drift. Unlike in phonetic drift, however, this movement causes the first phoneme representation to encroach on the acoustic territory of the second. The movement of the second phoneme representation, away from the first, can therefore be considered to be motivated by this encroachment, rather than by any external bias. Push chains – and their complements, pull chains – were first postulated by Martinet (1952), and have since been empirically observed in typologically diverse languages (Łubowicz, 2011). One example is the New Zealand short front vowel shift (Gordon et al., 2004), in which the raising (and fronting) of /æ/ over time

[4] Realization of /t/ as [ʔ] entails the removal of oral gestures. This phenomenon, also known as *glottal replacement*, is common across British varieties of English and almost exclusively affects /t/ (see e.g. Milroy, Milroy, Hartley, & Walshaw, 1994, for evidence from Tyneside English). It is articulatorily and acoustically distinct from the application of glottal constriction to oral gestures (*glottal reinforcement*), which more commonly affects /p/ and /k/.

triggered the raising (and fronting) of /ɛ/, and, in turn, the centralization of /ɪ/.[5]

2.1.3 Model desiderata

To be of primary use in the study of regular sound change, the model must generate phonetic drift and push chains that resemble empirical sound changes. That is, the movement of phoneme representations generated by the model must display certain key properties that are observed in corpus data. For present purposes, I assume that these properties do not differ between phonetic drift and push chains (or other kinds of regular sound change), and I infer them on the basis of data from the New Zealand short front vowel shift.

The key properties relate to the maintenance of structure over the course of change. As vowel distributions moved in the New Zealand short front vowel shift, they maintained their distance from one another, their shapes (width and skewness), and their degree of overlap with one another. At all times, they exhibited little skewness and substantial overlap relative to their widths; such properties are also seen in vowels in American English (Hillenbrand, Getty, Clark, & Wheeler, 1995). I illustrate these key properties for /æ/ and /ɛ/ over a 60-year period of the data in Figure 2.1.

The basic desiderata for the model are therefore that it: (i) generates movement of each phoneme representation; (ii) maintains the shape (width and skewness) of each phoneme representation; and, in push chains, maintains the (iii) distance between and (iv) overlap of the phoneme representations. To my knowledge, no other exemplar-based model has met all these desiderata; see Section 2.2.2 for further discussion.

[5]Gordon et al. (2004) suggest that the New Zealand short front vowel shift may have begun with the fronting of /a/. An acoustic study of the diachronic trajectory of /a/ has not yet been conducted, but Hay et al. (2015) find indirect acoustic evidence to corroborate this suggestion: the change in /æ/ is most advanced in words where the vowel precedes a voiced sound and is therefore most similar in length to /a/ (Chen, 1970; Mack, 1982).

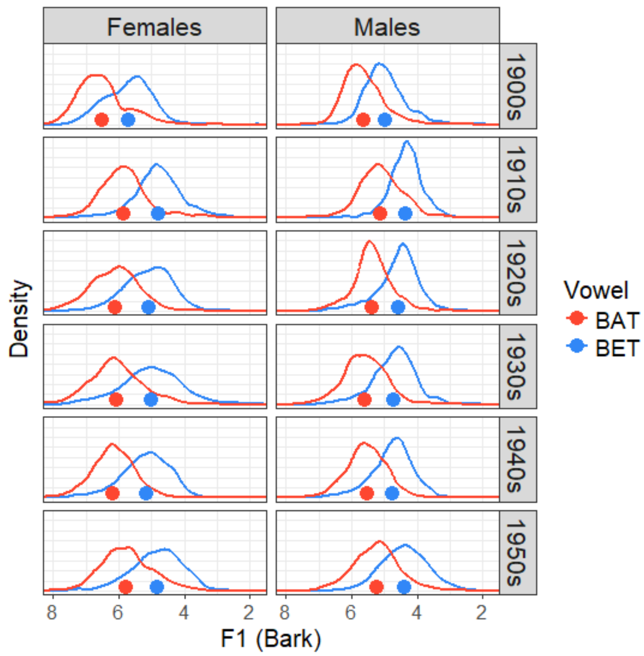

Figure 2.1: Vowel distributions in New Zealand English over time. Centroids (points) and distributions (densities) for the F1 values of the BAT (/æ/; red) and BET (/ɛ/; blue) vowel categories for speakers of New Zealand English born each decade from 1900 to 1959 (top to bottom), based on the raw data from Hay et al. (2015). While the category centroids move over time, their distance from one another stays approximately constant. The shapes (width and skewness) of the category distributions also stay approximately constant over time, as does the substantial degree of overlap between the two distributions.

2.2 Framework

The model of regular sound change presented in this chapter is couched in the framework of Exemplar Theory, according to which the cognitive representation of a phoneme is made up of memories of experienced instances of that phoneme in words. I do not intend the model to be interpreted as a claim about the nature of linguistic representations, but rather as a convenient way to capture two ideas about linguistic representations that are key to a listener-based approach: that they are shared across production and perception; and that they are updated through perception.[6] In doing so, it allows perception to influence production, which is key to the central claim of a listener-based approach.

In this section, I lay out Exemplar Theory, showing how it captures keys ideas for the listener-based approach, and situate its motivation in historical developments in the literature (Section 2.2.1). I then describe applications of exemplar-based models to regular sound change in previous work, highlighting key results and challenges (Section 2.2.2). Finally, I zoom out to formalize key concepts that describe how exemplar-based models generate forces on the shape and position of linguistic representations, through processes acting on individual exemplars (Section 2.2.3).

2.2.1 Exemplar Theory

Exemplar Theory claims that, in the individual's cognitive system, linguistic representations – such as phonemes – are composed of memories of episodes where those representations were externally instantiated – such as phonetic realizations of phonemes

[6]In the framework of Exemplar Theory, representations may be *partially* shared across production and perception, meaning that only *some* representational updates through perception have an impact on the representations accessed in production (see also Footnote 3 of Chapter 1). Nevertheless, when it comes to implementing a model in this framework, it is typically convenient to make the simplifying assumption that representations are *fully* shared across production and perception, meaning that *every* update through perception has an impact on production.

(in uttered words).[7] For example, the representation of the vowel /æ/ is made up of memories of people saying [æ], in words like *map*, *cat*, and *rag*. These memories, known as episodic traces or *exemplars*, are constantly being replaced as the individual accumulates new experiences and forgets old ones. In this way, Exemplar Theory captures the first key idea of a listener-based approach to language change, which is that linguistic representations are updated through perception. Furthermore, exemplars are assumed to be recruited for both language production and language perception, capturing the second key idea of a listener-based approach: that linguistic representations are shared across both production and perception. In this way, Exemplar Theory allows perception to exhibit influences on production over time; in particular, it supports the central claim of a listener-based approach, which is that perceptual biases that treat linguistic experiences differentially in perception can give rise to differential patterns of language change.

Exemplar Theory has its roots in a formal model of general perception from the psychological literature, the Generalized Context Model (GCM; Nosofsky, 1986). The GCM takes key ideas from the literatures on identification (discriminating a stimulus from others that are similar) and categorization (grouping a stimulus with others that are similar). From the identification literature, it takes the idea that perception involves pinpointing a stimulus in a multidimensional perceptual space (Shepard, 1957). From the categorization literature, it takes the idea that perception involves comparing a stimulus to exemplars of previously-experienced stimuli, stored in memory (Medin & Schaffer, 1978). It unifies these ideas by proposing that perceivers store exemplars of stimuli in memory as they experience them, in a multidimensional perceptual space, and recruit these exemplars directly in the perception (identification and categorization) of future stimuli.

[7]Though I discuss only applications of Exemplar Theory to representations of phonemes, in keeping with the present focus on regular sound change, it has been applied to representations of other levels of linguistic structure as well (see e.g. Walsh, Möbius, Wade, & Schütze, 2010).

The exemplar-based framework of the GCM was not developed with linguistic perception in mind. However, it was soon shown to be valuable for speech perception by Goldinger (1996), who found experimental evidence that listeners store richly-detailed phonetic memories and draw upon them in lexical access. In Goldinger's (1996) experiments, participants first performed a study session, where they listened to a set of words in multiple voices, and later – after a delay of up to a week – performed a test session, in which some of the words were repeated (in the same voice or in a different voice). Each participant performed one of two tasks, either *recognition memory* or *perceptual identification*. In the recognition memory task, the study session consisted of transcribing words in the clear, and the test session consisted of a surprise recognition test where participants judged whether words had been presented during study. In the perceptual identification task, both the study and test sessions consisted of transcribing words in noise. Across both tasks, test accuracy was higher for words repeated in the same voice than for words repeated in a different voice, at delays of at least a day for recognition memory and at least a week for perceptual identification. Furthermore, among words that were repeated in a different voice, test accuracy increased with the similarity between test and study voices. These results are predicted by the exemplar-based framework of the GCM: listeners stored exemplars during the study session and (implicitly or explicitly) drew upon them during the test session, gaining advantage the closer a relevant study exemplar was to the test stimulus in a multidimensional perceptual-acoustic space.

Following Goldinger's (1996) empirical demonstration of the value of an exemplar-based framework for speech perception, Johnson (1997) adapted the GCM to a formal exemplar model of speech perception. As a direct adaptation of the GCM, Johnson's (1997) model implies that linguistic representations are updated through perception, capturing the first key idea of a listener-based approach. However, it says nothing of the role of exemplars in speech production.

The potential of an exemplar-based framework for speech production, capturing connections between perception and production, was made explicit through shadowing experiments by Goldinger (1998). These experiments consisted of interleaved listening blocks, in which participants listened to stimulus words or nonwords[8] repeated up to 12 times each, and shadowing blocks, in which participants listened to each stimulus again and repeated it either immediately or after a 3–4 second delay. The results concern *reaction times* – how quickly participants began responding – and *imitation* – the degree to which responses were perceived to sound like the shadowed stimulus (by separate participants, in AXB experiments). In immediate shadowing, stimuli that had more repetitions in the listening block were associated with faster reaction times and more imitation.[9] The same effects were observed in delayed shadowing, but were attenuated. Through simulation in an exemplar-based model[10] (Hintzman, 1986), Goldinger (1998) shows that these results are predicted by an exemplar-based framework that links production and perception. In the shadowing task, exemplars are recruited according to how similar they are to the stimulus. Repetition in the listening block creates additional exemplars that are (almost) identical to the stimulus, and thus easily recruited. With more exemplars of this sort, the stimulus can be recognized faster, which permits faster reaction time, and it exerts more influence on the production target, which encourages imitation. Delay allows for the recruitment of additional exemplars, which correspond to the stimulus in terms of lexical identity but are acoustically dissimilar to it. These additional exemplars also exert influence on the production target, counterbalancing the influence of the stimulus, and thus attenuating the imitation effect. In further experiments, Goldinger (1998) also showed

[8]In experiments involving nonword stimuli, participants were trained on the nonwords the day before, to build familiarity and establish an artificial lexicon.

[9]Goldinger (1998) also observed effects of lexical frequency, which are consistent with an exemplar-based framework. For brevity and simplicity, I do not discuss these effects here.

[10]The framework of Hintzman (1986) is not the same as the GCM in its implementational details, but it shares the high-level conceptual framework of exemplar-based models.

that the reaction time and imitation effects are attenuated if stimuli are presented in different voices in the listening block and shadowing block, drawing connections with the exemplar-based framework of speech perception motivated in earlier work (Goldinger, 1996).

To formally capture the connections between speech perception and speech production in an exemplar-based framework, Pierrehumbert (2001) extended Johnson's (1997) model to claim that exemplars may be recruited in production as well as in perception. This claim implies that linguistic representations are shared across production and perception, capturing the second key idea of a listener-based approach. In particular, production in Pierrehumbert's (2001) model is based on sampling from the exemplar distribution. If this sampling is not strategic in any way, production will reflect perceptual influences, implying that biases in perception can give rise to patterns of change, which is the central claim of a listener-based approach.

2.2.2 Insight from past models

Most formal exemplar-based models of regular sound change in the literature have been developed to explore questions broadly related to contrast maintenance. For example, some models have explored the conditions under which two phonemes can be expected to merge or remain acoustically distinct (Pierrehumbert, 2001; Tupper, 2015; Wedel, 2004, 2006; Wedel & Fatkullin, 2017). Others have explored the role of contrast maintenance in creating stable phonological inventories (Sóskuthy, 2013; Wedel, 2012), and how such stability promotes interlinked changes in different phonemes, as in push chains (Wedel, 2012) and pull chains (Ettlinger, 2007). Since phonemic contrast is central to sound systems, both synchronically and diachronically, I focus here on the highlights and challenges presented by models engaging with such questions.

One of the major highlights of past exemplar-based models is that the maintenance of acoustic contrast can emerge from local competition between phonemes in the categorization of acoustically ambiguous signals (Sóskuthy, 2013; Tupper, 2015; Wedel, 2004, 2006, 2012; Wedel & Fatkullin, 2017). For example, consider the situation illustrated in Figure 2.2, where there are just two phonemes of relevance, /s/ and /ʃ/, the representations of which overlap in acoustic space. Suppose a speaker intends to say "sip" but utters [s̠ɪp], with a retracted sibilant that falls in the region of acoustic overlap (arrow 1 in the figure). Absent any influences of context, a listener will perceive this utterance as ambiguous, and may categorize it either as *sip* or as *ship*. If the listener correctly categorizes the token as *sip*, then storing it as an exemplar will strengthen the representation of /s/ in the region of acoustic overlap. However, if the listener incorrectly categorizes the token as *ship*, then storing it as an exemplar will instead strengthen the representation of /ʃ/ in the region of acoustic overlap, thus weakening the representation of /s/ in comparison. Now suppose a speaker utters [s̟ɪp], with a fronted sibilant that falls far from the region of acoustic overlap (arrow 2 in the figure). A listener will perceive this utterance as an unambiguous instance of *sip*; thus, storing it as an exemplar will only ever strengthen the representation of /s/ in the corresponding part of acoustic space. Since phoneme representations can both strengthen and weaken in the region of acoustic overlap, but can only strengthen outside of the region of overlap, representations tend to become stronger outside of the region of overlap than within it. By consequence, phonemes tend to evacuate the region of overlap over time, effectively repelling each other.

The mutual repulsion of phonemes is known as *dispersion*, and is a central component of functionalist approaches to phonology (e.g. Flemming, 2004; Liljencrants & Lindblom, 1972; Martinet, 1952). Exemplar-based models have contributed to the theory of dispersion by demonstrating that it is a natural consequence of perceptual competition between phonemes. Additionally, exemplar-based models have highlighted that dispersion is extremely robust to modeling assumptions. Dispersion

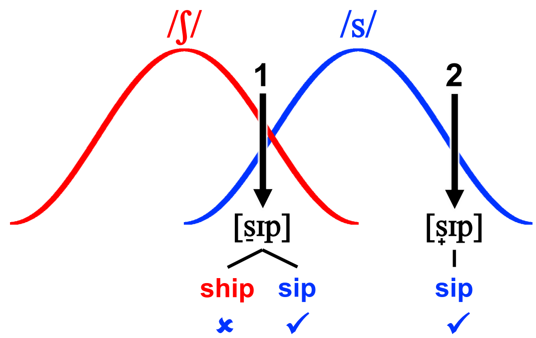

Figure 2.2: Ambiguity and acoustic overlap. Suppose the speaker intends to produce an instance of /s/, in the word *sip*. (1) If the speaker produces a token in the region of acoustic overlap between /ʃ/ and /s/, it may be miscategorized as *ship*. Storing the token as an exemplar may weaken the representation of /s/ in the region of acoustic overlap. (2) If the speaker produces a token outside of the region of acoustic overlap, it will always be correctly categorized as *sip*. Storing the token as an exemplar will strengthen the representation of /s/ outside of the region of acoustic overlap.

has been observed in a wide range of exemplar-based models, regardless of assumptions about perception (e.g. whether categorization is probabilistic or deterministic), memory (e.g. whether exemplars decay or are overwritten over time), and number of interacting agents (see especially Sóskuthy, 2013).

Wedel and Fatkullin (2017) provide a useful analogy for understanding dispersion in exemplar-based models. As the exemplars constituting phoneme representations drift quasi-randomly about an acoustic space, they may encounter a wall through which they cannot pass. Such a wall may be constituted by a categorization boundary between two phonemes, or by an articulatory boundary representing limits on the physical production of sounds (Sóskuthy, 2013). By virtue of random drift, phonemes may move in any direction at any time. But since movements through a wall are blocked, only movements away from it will be successful, making it increasingly likely over time that the phoneme will move away from the wall.

Just as categorization of acoustically ambiguous stimuli leads to the movement of phoneme representations under dispersion, exemplar-based models have shown that any recurring process that distorts production or diminishes perception can lead to the movement of phoneme representations. Sóskuthy (2013) provides a useful characterization of how different processes interact, through forming an *adaptive landscape* that defines the long-term states that are likely to arise from the corresponding movements over time. In this characterization, a system of phoneme representations drifts quasi-randomly about a state space, which corresponds to the joint position of the phoneme representations in the acoustic space. The state space is overlaid with a series of peaks and valleys, constructed by processes in production and perception. A system of phoneme representations is more likely to drift downhill than uphill, and, as time passes, is increasingly likely to settle in one of the valleys.

Sóskuthy's (2013) characterization of adaptive landscapes shows how exemplar-based models can be applied to both stability and dynamism in sound systems. If the adaptive landscape is fixed, then – given enough time – phoneme representations will settle into stable states in the valleys. However, if an external force causes the adaptive landscape to change – for example, due to the introduction of a production bias in one phoneme – then the valleys will move, and the phoneme representations will move with them. In this vein, exemplar-based models have been successful in demonstrating that the forced movement of one phoneme representation will trigger movement of another, as in push chains (Wedel, 2012) and pull chains (Ettlinger, 2007).

One of the biggest challenges facing exemplar-based models is the maintenance of acoustic overlap of phoneme representations in the face of dispersion and chain-shifting. As explained above, the process of categorization shapes the adaptive landscape such that phoneme representations repel each other and evacuate any regions of acoustic overlap over time. This is a problem because empirical studies have shown that phoneme representations can and do overlap, at least at the population level; for example, Hillenbrand et al. (1995) document extensive overlap of American English vowels in the F1-F2 space, which causes a discriminative statistical model to obtain just 68.2% classification accuracy.[11] Because dispersion is extremely robust to modeling assumptions, this problem is pervasive; for example, Pierrehumbert (2002, p. 133) states that she "[has] not actually been able to find a parameter range for this model which shows stable overlapping distributions". It creates a major challenge for meeting the present model desideratum of maintaining acoustic overlap of phoneme representations.

Tupper (2015) shows that a promising strategy to meeting the desideratum of overlap maintenance involves biasing tokens of each phoneme toward the other in production, categorizing them probabilistically, and then not storing miscategorized tokens as exemplars. This strategy can be illustrated by returning to the example illustrated in Figure 2.2. Under this strategy, both *ship* and *sip* would be biased to be produced somewhat like [s̠ɪp] (arrow 1 in the figure). The resultant token would be stored as an exemplar of /ʃ/ only if the word *ship* fits the context, and as an exemplar of /s/ only if the word *sip* fits the context. The production component of this strategy ensures that phoneme representations do not drift away from each other, and the perception (categorization/storage) component counteracts the weakening of each representation in the region of acoustic overlap. However, the adaptive landscape associated with this strategy is one in which phoneme representations have extremely skewed acoustic distributions, because movement toward the categorization boundary is easy, but movement through it is very difficult. Thus, the maintenance of acoustic overlap of phoneme representations may come at the cost of changes to distributional shape, counter to another desideratum of the present model. The joint maintenance

[11] Even when considering a higher-dimensional acoustic space – including dynamic measurements of F0–F3, as well as duration – Hillenbrand et al. (1995) still observe overlap in vowel distributions that limits their classification accuracy to 94.8%. Though this number may appear high, it is important to point out that the tokens in question are taken from high-quality recordings of carefully-articulated speech, which are likely much more precise than those that might be encountered in everyday situations.

of acoustic overlap and distributional shape remains a major challenge for exemplar-based models.

The work outlined above is all concerned with the maintenance of acoustic contrast between phonemes, but exemplar models have also highlighted the assumptions that lead to the *loss* of contrast, as in phoneme merger. Pierrehumbert (2001) developed an early model of merger, in which the loss of acoustic contrast was identified with the extinction of a phoneme representation altogether. She argued that merger followed from one phoneme representation leeching the other to extinction, due to the combination of strong priming in production – where the probability of uttering a given word is determined solely by the number of times it has been said recently – and competition in perception – where an acoustically ambiguous token may be miscategorized and not stored as an exemplar of its source phoneme. However, Wedel and Fatkullin (2017) show that extinction becomes vanishingly unlikely if production is based not on strong priming but on externally-grounded lexical frequencies. Thus, the extinction of a phoneme representation is an artifact of (overly) strong priming mechanisms in production; without such mechanisms, the maintenance of acoustic contrast between phonemes is the norm.[12]

A final highlight offered by past exemplar-based models concerns correspondences between representations of words and phonemes. Intuitively, word representations and phoneme representations have the same two basic needs: the representations of two different words, or phonemes, should be acoustically contrastive; and every instance of a given word, or phoneme, should be acoustically similar. Exemplar-based models meet these needs by assuming that exemplars can be labeled for membership in both word representations and phoneme representations, at the same time

[12] Merger remains a challenge for exemplar-based models, but it is not one that I will take up here. Following other authors, I assume that merger results from extreme overlap between phoneme representations (Sóskuthy, 2013; Wedel, 2004, 2012), which triggers reanalysis of the phoneme inventory, or of the phonological form of lexical entries, between speakers or generations (Blevins, 2006; Harrington, Kleber, Reubold, Schiel, & Stevens, 2018).

(Pierrehumbert, 2002). Given this assumption, acoustic contrast is maintained by competition in categorization, as described previously. Acoustic similarity is maintained by incorporating references to other exemplars of the same word, or phoneme, in production or perception; for example, in producing or perceiving the word *sip*, multiple exemplars of *sip* may be referenced. Exemplar-based models have highlighted the necessity of both word and phoneme representations: word representations are necessary to allow for consistent differences according to properties such as lexical frequency (see Chapter 3), while phoneme representations are necessary to ensure acoustic similarity between instances of the same phoneme in different words (Wedel, 2012). In models developed to date, however, only a few word representations have typically been included, all with similar lexical frequency; the modeling of a large number of words, spanning a range of frequencies, remains an open challenge.

2.2.3 Formalizing key concepts

The previous two sections have informally introduced the notions of *word/phoneme representations* and the *adaptive landscape*. In this section, I show how these notions are formalized within exemplar-based models, and how they relate to processes in production and perception, through the conceptualization of forces.

In exemplar-based models, word and phoneme representations are made up of exemplars, which have acoustic values. Formally, then, word and phoneme representations can be understood as distributions over an acoustic space. Word representations are acoustic distributions formed by collecting different exemplars of that word. Similarly, phoneme representations are acoustic distributions formed by collecting different exemplars of that phoneme as it is instantiated in different words. The storage of a new exemplar updates the distribution of acoustic values, and thus the corresponding representations. In this way, processes acting on individual exemplars aggregate to shape entire representations, in an emergent fashion.

Given enough time, the shaping of representations by processes acting on individual exemplars will lead to a stable state, where representations hardly change any more. Reaching this stable state represents settling into a valley in the adaptive landscape. From this insight, it follows that the adaptive landscape can be formed by collecting all of the stable states that result from many different runs of the model over a long period of time. Formally, it is obtained from the joint probability distribution over the acoustic space for the long-term location of the centroid of each representation. While the adaptive landscape is a useful concept for thinking about the long-term state of an exemplar-based model, it does not give much insight into the path by which that long-term state is reached, as a direct consequence of short-term processes acting on individual exemplars. To make the connection between short-term processes and the long-term adaptive landscape, it is useful to think about *forces*.

The shaping of representations by processes acting on individual exemplars can be abstracted, to be considered the action of forces on acoustic distributions. That is, each process in production or perception may be considered to exert a force on the acoustic distributions that constitute representations. Formally, such a force can be understood as the expected effect of applying the process to an exemplar sampled at random from the distribution. For every state of the system, there will exist forces on every representation. The net force on each representation, formed by composing all of the independent forces, indicates the way in which it is expected to change, averaged over many different runs of the model. The representations will continue to change until they reach a combined state in which all forces are balanced – a valley in the adaptive landscape. In this way, forces on distributions ultimately produce the adaptive landscape, and can also be used to describe paths taken around it.

There are many possible forces on acoustic distributions, because every process in production or perception may give rise to one. However, these forces fall into just a few basic kinds, based on how they affect the distribution. Figure 2.3 illustrates the

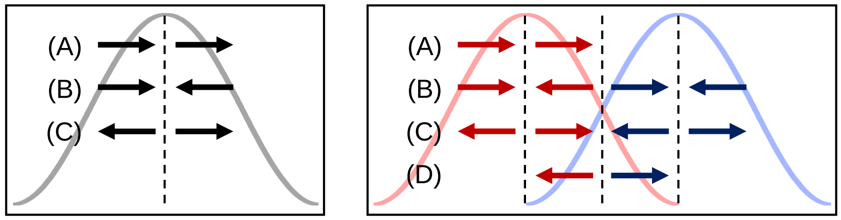

Figure 2.3: Forces on acoustic distributions. Left: a single phoneme representation undergoing phonetic drift is subject to a pushing force (A), a spreading force (B), and a squeezing force (C). Right: two phoneme representations in a push chain are also subject to a repulsive force (D).

forces that apply in the model presented in this chapter, in situations with one or two interacting phoneme representations.[13] A single, isolated phoneme representation undergoing phonetic drift is subject to: a pushing force, which pushes it *along* in a particular direction; a spreading force, which spreads it *outward* from its center; and a squeezing force, which squeezes it *inward* toward its center. For two phoneme representations interacting in a push chain, only one is subject to the pushing force, but both are subject to the spreading and squeezing forces, and also to an additional repulsive force, which repels each phoneme *away* from the other.

2.3 Model description

Following Exemplar-Theoretic principles, the model constructed in this chapter forms a production-perception loop (Pierrehumbert, 2001) and consists of a cycle of processes applying to exemplars (one per iteration). It is a *listener-based* model because the processes in production are minimal and do not correspond to active strategizing by the speaker in any way,[14] while the processes in perception make room for passive

[13]Since articulatory boundaries give rise to dispersion, they also generate forces that are not pictured in Figure 2.3, which push inward from the boundaries of the acoustic space. I have not included these forces in the present model because they do not bear directly on the kinds of change I model, where movement of phonemes is not directed toward or away from articulatory boundaries.

[14]Of course, speakers *can* and *do* strategize in production, at least some of the time, and the model can be extended to incorporate such speaker-based production biases. I have not made such an extension at present in order to make clear the potential of listener-based perceptual biases for

and powerful biases that are known to affect listeners in the short-term and that shape the behavior of the model in the long-term.

In this section, I describe and formalize the representations in the model, the processes constituting production and perception, and the way in which these processes yield the forces that drive the evolution of the system. Throughout, I compare the low-level details of the model with those of previous models, and I conclude by highlighting the major high-level areas in which the model makes important contributions through new decisions.

2.3.1 Representations in the model

The model describes how the realization of a phoneme occurring within words changes over time as those words are used in successful communication. For concrete illustration, it can be assumed that the phoneme in question is a vowel, and the words in question are monosyllabic.

The model contains three levels of representation in memory. At the lowest level are *exemplars*, arranged within a perceptual-acoustic *exemplar space*. Exemplars are collected into the intermediate-level representation of *types*, which correspond to words. The *frequency* of a type is represented by the number of exemplars it contains. The highest-level representations are *categories*, which correspond to phonemes. I include a glossary of representation terms in Table 2.1, along with measures of the representations in simulations, for ease of reference.

---

the following chapters. However, the incorporation of speaker-based biases alongside listener-based biases constitutes a valuable avenue for future research, as it enables a comparison between the two kinds of biases in terms of their capacity to account for established patterns of language change.

Table 2.1: Glossary of representation terms

Term	Meaning	Example	Simulations
Exemplar	A memory trace of an experienced instance of a particular type, i.e. a spoken word.	"map"	492/category
Exemplar space	The distribution of exemplars across a granularized perceptual-acoustic dimension (e.g. vowel F1). Assumed to be shared across perception and production.		Grain: 0.1
Type	An abstract template for a word containing a particular phoneme, collecting together experienced instances of that word in memory. Contains information about the *frame* (e.g. onset and coda consonants of a monosyllabic word) and the category (e.g. nucleus vowel).	*map*	92/category
Type frequency	The number of exemplars of a given type. Based on word log-frequency in a large corpus (see Appendix A).		Range: 1-12
Category	A generalization over experienced instances of a phoneme (e.g. a vowel), stored in memory.	/æ/	1–2 categories

2.3.1.1 Exemplars

As in all exemplar-based models (Section 2.2), the building blocks of the present model are *exemplars*, which are detailed memory traces of linguistic events. For present purposes, exemplars in the model can be understood as capturing the experienced perceptual-acoustic value of the nucleus vowel in a word. For example, exemplars of the word *map* are remembered instances of spoken “map”, each one containing a slightly different realization of the vowel /æ/.

Exemplars are distributed in a perceptual-acoustic *exemplar space*. Following Kruschke (1992), the exemplar space is granularized, such that acoustic values that are different but nevertheless perceived identically are all represented by a single, shared value (see also Pierrehumbert, 2001). For this early work, I make the simplifying assumption that the exemplar space is one-dimensional, e.g. corresponding to F1 at the vowel midpoint.

The model contains a single exemplar space, shared across production and perception. This can be interpreted as a single agent talking to themself, or as an aggregate over a homogeneous community talking amongst itself. The modeling of multiple agents with distinct spaces is left for future work.

2.3.1.2 Types

Exemplars are labeled as instances of words, which in the model are represented by *types*. A type has a dual nature: it is both a static entity with abstract properties, defining a template for relating to exemplars, and a dynamic entity with acoustic properties, constructed by exemplars. Types mediate the processing of exemplars in production and perception.

Construed as a static entity, a type is a template with two abstract properties of current relevance: the phonological frame, which specifies the parts of the word that are *not* at issue, and the category to which the type belongs, which specifies the phoneme that *is* at issue (see next section). For example, the type corresponding to the word *map* has the phonological frame /m_p/ and belongs to the category /æ/. For simplicity, I assume that there are no minimal pairs, so that the category membership of a given type may be uniquely determined from its phonological frame. Such an assumption is warranted by the fact that only a minority of words in real lexicons are part of a minimal pair, and it is unlikely that a minority of words can drive change across an entire phoneme representation; see Appendix C for further discussion and supporting simulations.

Construed as a dynamic entity, a type is a collection of exemplars that fit its abstract template (i.e. that all correspond to the same word). Consequently, it can be understood as a distribution in the perceptual-acoustic exemplar space, formed by collecting the acoustic values of these exemplars.

Different types have different numbers of exemplars. I follow the multiple-trace hypothesis (Hintzman & Block, 1971) in assuming that the number of exemplars for a given type represents that type's frequency. For the simulations presented here, I model type frequencies on word log-frequencies in a large corpus. Consequently, types in the model have between 1 and 12 exemplars (see Appendix A for details).

Type frequency can be understood as an individual's subjective frequency of a word, as compared with the objective frequency of that word in actual experience. The decision to represent subjective frequencies by the log-transformation of objective frequencies reflects the fact that participants underestimate the frequency of common words (Begg, 1974). It is also consistent with the "negatively accelerated, increasing relation between represented and actual frequency" observed by Nosofsky (1991, p. 15). Such log-transformation is widely used in processing models and empirical studies assessing a relationship between word frequency and behavior (in terms of both behavioral response properties – e.g. reaction time and categorization probability –

and word realization properties – e.g. duration and acoustic quality), both for words in isolation (e.g. Murray & Forster, 2004, and studies cited therein) and for words in context (e.g. Smith & Levy, 2013, and studies cited therein).

The model uses the same (subjective) type frequencies for production and perception. The use of subjective rather than objective frequencies in production might seem inappropriate; after all, in the real world, words are produced according to their frequency rather than their log-frequency. However, it is not, because the model is fundamentally concerned only with productions *that are assessed for storage in memory*, which need not include all words in a stream of speech (see also Landauer, 1986, for a model of memory in which not every input is stored). Similarly to recent subsampling approaches in Natural Language Processing (Mikolov et al., 2013), I assume that listeners may filter out (or otherwise downweight) some instances of high-frequency words due to their high predictability. I take the liberty not to model such filtered instances for the sake of computational efficiency.

2.3.1.3 Categories

Each type in the model belongs to a particular *category*, which corresponds to a phoneme and represents an abstract generalization over experienced instances of words containing that phoneme. For example, the category representation of the phoneme /æ/ is a generalization over experiences with words like *map*, *lab*, *cat*, etc., which picks up on the qualities of the common nucleus vowel.

A category can be construed as the set of exemplars of all types that it contains. In the same way as for the constituent types, collecting the acoustic values of these exemplars yields a representation of the category as a distribution in the perceptual-acoustic exemplar space.

In the present model, I include 92 types per category, for a total of 492 exemplars per category. The initial distribution of exemplars for each category is constructed in such a way as to avoid asymmetries across categories, across classes of type frequency, or across exemplars within a frequency class. See Appendix A for further details.

For modeling phonetic drift, I include a single category. For modeling push chains, I include two categories: a Pusher category that receives an external bias, and a Pushee category that does not. The initial width of the categories, σ, and the initial distance between them, μ, are parameters of the model.

2.3.2 Processes in the model

Each iteration of the model begins with the production of a *token* – an instance of a type with a particular *target* acoustic value – based on the existing exemplar space. The produced token is then transmitted to the listener, as the acoustic value of an unknown phoneme residing in a known phonological frame.[15] In perception, the listener uses the value and frame to recover the type (and thus the category) intended by the speaker, and then decides whether the token should be stored as an exemplar of that type and thus update the corresponding category distribution.

Together, production and perception form a closed loop, gradually updating the distribution of exemplars in the space through the generation and storage of tokens. This loop is composed of multiple processes, both on the production side and the perception side, as illustrated in Figure 2.4. It is these processes that yield the category-level forces illustrated previously in Figure 2.3.

[15]For simplicity, I assume that both the acoustic value and the phonological frame of the token are perceived exactly as produced.

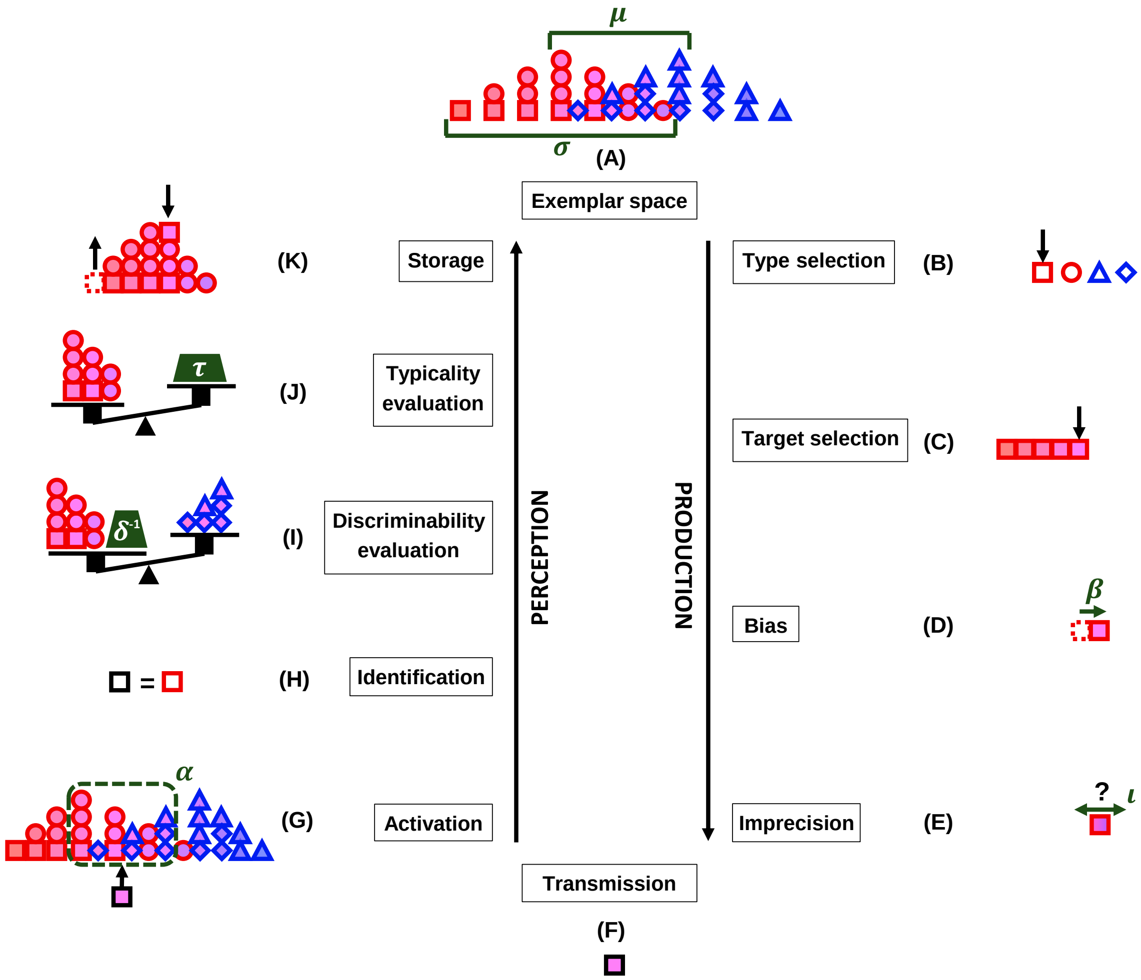

Figure 2.4: Schematic illustration of processes in the model. Outline colors represent phoneme category membership (e.g. /æ/), shapes represent phonological frame (e.g. /m_p/), and fill colors and horizontal positions represent perceptual-acoustic value (e.g. vowel F1). Dark green components with Greek letters indicate parameters of the model. (A) Two partially-overlapping categories exist in an exemplar space. (B) The speaker randomly selects a type for production, according to its frequency. (C) An exemplar of that type is randomly selected to provide an acoustic target for the production. (D) For the Pusher, the target is shifted by a constant bias toward the Pushee. (E) The actual realization of the target is imprecise, causing it to shift by a random amount in either direction. (F) The realized token is transmitted to the listener, with its acoustic value and phonological frame but without its category membership. (G) The listener locates the token in their exemplar space, activating surrounding exemplars of both categories within a fixed activation window. (H) The candidate type of the token is identified based on context (represented here by phonological frame), yielding identification of the intended category. (I) The activation of the intended category (red) is compared to the activation of the other category (blue); if the ratio of activations is below a fixed discriminability threshold, the token is unlikely to be stored. (J) The activation of the intended category is compared to a typicality threshold; if the activation is below the threshold, the token is unlikely to be stored. (K) If the token is sufficiently discriminable and typical, it is stored in the listener's exemplar space, replacing a random exemplar of the same type. Storage updates the exemplar space for future production and perception.

2.3.2.1 Type selection

Production begins with a speaker selecting a type to produce, T_k. The selection of a type is random and conducted across types from all categories, weighted by type frequency[16] (f_k), as shown in Equation (2.1). Consequently, production of a type is akin to drawing from a unigram 'bag-of-types' language model. On each iteration, this draw is independent; production is not influenced by history or context.

$$P(T_k) = \frac{f_k}{\sum_j f_j} \tag{2.1}$$

The weighting of type selection by frequency yields production rates that are anchored in the lexicon, external to the exemplar system (Wedel & Fatkullin, 2017). This fact allows the model to avoid the problem of categories leeching each other to extinction, as seen in other models where production rates are anchored entirely in system-internal exemplar priming (e.g. Pierrehumbert, 2001).

2.3.2.2 Target selection

Having decided *what* to say, a speaker then decides *how* to say it. For this, the speaker begins by selecting an initial acoustic target of production. Target selection is a random (uniform) draw from the set of exemplars of the to-be-produced type. The probability of choosing exemplar j of type T_k, with acoustic value $x_{j,k}$, is given by Equation (2.2).

$$P(v = x_{j,k}|T_k) = \frac{1}{f_k} \tag{2.2}$$

Unlike in other models (Ettlinger, 2007; Pierrehumbert, 2001; Sóskuthy, 2013; Tupper, 2015; Wedel, 2006, 2012; Wedel & Fatkullin, 2017), target selection is not influenced by exemplar age. It is also not influenced by properties of the type, such

[16] The type frequencies underlying production are the same as those underlying perception, and thus reflect real-world log-frequency. See Section 2.3.1.2 for discussion.

as frequency, or of the occasion, such as social situation. Consequently, type and target selection together form a random (uniform) draw from the set of all exemplars, meaning that they do not reflect any active strategizing on the part of the speaker.

Having chosen an initial production target, the speaker does not just produce it outright. Prior to realization, the target value is adjusted under two influences, bias and imprecision.

2.3.2.3 Bias

The first adjustment of the target value is due to *bias*, which shifts the target by a small and consistent amount. Bias represents external influences such as reduction of articulatory effort. At the level of the category distribution, it yields the intrusive force (henceforth, the *bias force*). The degree of bias, β, is a parameter of the model; as β grows, so does the bias force.

In the simulations presented here, I apply bias to all productions in the single-category case, and to all productions of the Pusher category in the two-category case. Bias consists of the addition of β to the target v, yielding a new target v'. The bias process is illustrated for the two-category case in Equation (2.3).

$$v' = \begin{cases} v + \beta & \text{if target is Pusher} \\ v & \text{if target is Pushee} \end{cases} \tag{2.3}$$

The function of bias is to enforce sustained category interaction and promote long-term movement in one direction. Thus, bias itself does not *cause* categories to interact, but rather gives categories sustained opportunities to interact. Simulations of two categories without bias exhibit decreasing category interaction over time (see Section 3.5.2 of Chapter 3).

The treatment of bias as systematic, i.e. applied to all tokens (of the Pusher) equally, follows that presented by Pierrehumbert (2001). The major downside to this treatment is that the bias is unconstrained and continues acting in the same way throughout the simulation, generating perpetual category movement. Other authors (Sóskuthy, 2013; Tupper, 2015; Wedel, 2006; Wedel & Fatkullin, 2017) use instead a bias that applies to tokens differentially, based on their distance from some fixed attractor point. This alternative treatment places constraints on the movement induced by the bias, causing movement to cease when the Pusher reaches the attractor. While it is easy to understand how such an attractor may arise in the case of leniting biases (i.e. through the minimization of articulatory effort), it is harder to understand how an attractor may arise in sound change more generally, assuming that it is not something the speaker can agentively establish. Instead, the *illusion* of an attractor may result from the interaction of counteracting forces (Sóskuthy, 2013). Consequently, the generation of perpetual movement under the present treatment is a reflex of the simplicity of the modeling environment: with the inclusion of additional repellers in the system (provided by other categories and/or articulatory limits), movement would no longer be unconstrained.

2.3.2.4 Imprecision

The second adjustment of the target value is due to *imprecision*, which shifts the target by a small (random) amount in either direction. Imprecision represents natural variability in the application of motor routines in realization.[17] At the level of the category distribution, it yields the spreading force (henceforth, the *imprecision force*). The degree of imprecision, ι, is a parameter of the model; as ι grows, so does the imprecision force.

[17] In principle, imprecision may be experienced by the speaker or by the listener. The number of imprecision processes does not affect the high-level model behavior, so I choose a single process for simplicity. I choose to locate this process in the speaker, to make it clear that the core claim of the listener's centrality to sound change is based on processes that have strong justification for being listener-based.

I implement imprecision through the addition of random noise n to the target v', for all productions. This process yields a final target v'' for the transmitted token, as shown in Equation (2.4). n is a single sample from a normal distribution with standard deviation ι; the larger ι, the more the target may deviate. In keeping with the granularization of the exemplar space (Section 2.3.1.1), the final target v'' is rounded to the nearest 0.1 before the token is transmitted.

$$v'' = v' + n \qquad n \sim \mathcal{N}(0, \iota^2) \tag{2.4}$$

The function of imprecision is to allow a discrete set of exemplars to generate a continuous distribution over the acoustic space, from which targets can be sampled in production. In this way, imprecision allows for novelty in production targets.

The use of token-wise imprecision generates a non-parametric sampling distribution. This approach is standard in exemplar-based models, but other approaches are also possible. For example, Harrington et al. (2018) generate a parametric sampling distribution by inferring a Gaussian distribution over all exemplars of a category. A parametric approach forces all exemplar distributions to have a common shape, with fixed kurtosis and zero skewness. While this enforcement makes the maintenance of category shape almost trivial, it doesn't allow for the modeling of distributions that differ substantially from the parametric (Gaussian) shape.

2.3.2.5 Activation

For the listener, the incoming token *activates* exemplars of both categories within a window around the target. These activations are aggregated within each category to yield overall category activation, which underlies key processes in perception. The size of the activation window, α, is a parameter of the model, and modulates the size of perceptual forces.

The activation A_i of category C_i is given by the sum of the activations of all exemplars belonging to that category, as shown in Equation (2.5). The activation of each exemplar x is determined by its distance from the token v'', through the use of a Gaussian activation window w_a with width α, as shown in Equation (2.6). Exemplars that are very near the token are given activations close to 1, while exemplars that are very far away are given activations close to 0. Increasing α causes exemplars within a wider radius to be given non-negligible activations.

$$A_i = \sum_{x \in C_i} w_a(v'' - x) \tag{2.5}$$

$$w_a(d) = \exp\left(\frac{-d^2}{2\alpha^2}\right) \tag{2.6}$$

By mapping the category activation that would arise from tokens at all points in the exemplar space, we obtain the *activation field* of the category. The activation field measures how the strength of the category changes across the space, and is of greater representational utility than the raw exemplar distribution. Accordingly, the plots of exemplar systems that I will display later all show activation fields, not raw exemplar distributions. Mathematically, mapping the activation field is akin to kernel density estimation, a statistical technique for estimating a continuous probability distribution from a discrete set of measurements (Ashby & Alfonso-Reese, 1995).

Most previous exemplar-based models of regular sound change have used a rectangular (Ettlinger, 2007; Pierrehumbert, 2001, 2002) or exponential (Wedel, 2006, 2012; Wedel & Fatkullin, 2017) activation window. The use of a Gaussian window here is motivated by discussion in the psychological literature of an equivalent parameter (p) in exemplar-based models of categorization using Multi-Dimensional Scaling. For example, Nosofsky (1985) found that asymptotic human categorization data (i.e. highly successful categorization which accesses pre-learnt structures) is better modeled with a Gaussian activation window than an exponential one. Similarly, Shepard (1958) developed an underlying process model predicting that a Gaussian activation window should arise under cases of infrequent feedback of categorization correctness, while an exponential window should arise under continuous feedback (which arguably does not occur in language, at least directly). The Gaussian window also has the practical advantage that it is *smooth*, whereas the rectangular and exponential windows are not (they contain jumps and a sharp peak, respectively); this ensures that the activation fields obtained in the modeling process are also smooth, even when exemplar distributions are sparse.

2.3.2.6 Identification

Based on the phonological frame, the listener *identifies* the type – and thus the category – corresponding to the token. Since we assume no minimal pairs and perfect transmission of the phonological frame, the type intended by the speaker is the only candidate for identification of the token. In principle, however, a set of types may be plausible candidates, and each may be assessed for the extent to which it is compatible with the transmitted token (Norris & McQueen, 2008). I discuss the introduction of architecture to handle multiple candidate types in Appendix C.

Not every identified token is stored as an exemplar, updating the category representation. The token will only be stored if it is assessed to be 'good' enough, meaning that it must not strongly resemble a competing category, and must strongly resemble the intended category. These assessments are performed by two probabilistic evaluations: the discriminability evaluation and the typicality evaluation.

2.3.2.7 Discriminability evaluation

The first evaluation, the *discriminability evaluation*, poses the question: how likely is the token to be a realization of its identified category, as opposed to any other relevant category, based on its acoustic value? It follows results in speech perception that tokens that are acoustically ambiguous between categories incur processing costs, causing errors and delays in recognition (e.g. Connine, Blasko, & Hall, 1991). Tokens that do not pass the evaluation are not stored, and hence do not update the category distribution. In the case of phonetic drift, the discriminability evaluation cannot be failed, since there is no relevant 'other' category. Thus, the following description is based on the case of push chains, where the identified category competes with one other category.

The discriminability evaluation is probabilistic, based on the ratio of category activations (identified category activation, A_i, divided by other category activation, A_o). Hence, tokens outside of the region of category overlap (where the ratio is large) are more likely to pass than tokens inside it (where the ratio is small). At the level of the category distribution, this asymmetry yields the repulsive force (henceforth, the *discriminability force*).

The evaluation proceeds by comparing the category activation ratio to a discriminability threshold, δ, as shown in Equation (2.7). δ is a parameter whose size determines the size of the discriminability force: as δ grows higher, passing the evaluation becomes harder, and the force grows stronger.

$$P(\text{pass discriminability evaluation}|A_i, A_o) = \frac{\dfrac{A_i}{A_o}}{\dfrac{A_i}{A_o} + \delta} \tag{2.7}$$

The formulation of the discriminability evaluation in Equation (2.7) is equivalent to an application of the Generalized Context Model (Nosofsky, 1986), which extends the application of Luce's Choice Rule (Luce, 1959) over category activations in the Context Model (Medin & Schaffer, 1978) by incorporating category response biases. Here, the bias towards the identified category C_i is $1/\delta$ and the bias towards the other category C_o is 1, as shown by Equation (2.8).

$$P(\text{pass discriminability evaluation}|A_i, A_o) = \frac{\frac{1}{\delta} \cdot A_i}{\frac{1}{\delta} \cdot A_i + 1 \cdot A_o} \tag{2.8}$$

This equivalence allows for an alternative interpretation of discriminability evaluation, as the act of categorizing the input. In particular, a probabilistic discriminability evaluation translates to a probabilistic model of categorization, in which the boundary between categories is permeable (cf. Wedel & Fatkullin, 2017). Similarly, not storing tokens that fail the discriminability evaluation translates to not storing tokens that are miscategorized, which is a source of contrast maintenance in many exemplar-based models (Harrington et al., 2018; Sóskuthy, 2013; Tupper, 2015; Wedel, 2006, 2012). Consequently, the discriminability evaluation is expected to play a major role in meeting the model desiderata of maintaining category overlap while generating interlinked movement.

Following the correspondence of $1/\delta$ to categorization bias, I assume $\delta \leq 1$, to obtain a positive bias toward the identified category. Given the lack of minimal pairs in the model, this assumption means that an acoustically ambiguous signal is likely to be recognized as a real word rather than a nonword, consistent with experimental results (e.g. Ganong, 1980). It also means that the discriminability force grows with the size of the activation window, α, in most cases.[18] A wider activation window encapsulates more exemplars, yielding a category activation ratio closer to 1, and hence to $\delta \leq 1$.

[18]Increasing α does not increase the discriminability force for the few tokens produced on the 'wrong' side of the category boundary, for which the category activation ratio is less than 1. In this case, taking the ratio closer to 1 actually takes it further from $\delta \leq 1$.

2.3.2.8 Typicality evaluation

If the token passes the discriminability evaluation, then it has been confidently assessed as a better fit for the identified category than for any other category. However, this assessment does not necessarily mean that the token is a *good* fit for the identified category. The question of the absolute value of token's fit for the identified category is addressed by the second evaluation, the *typicality evaluation*.

The typicality evaluation poses the question: how good is the token as a realization of its identified category, in absolute terms? It follows results in speech perception that tokens that are 'good' instances of their category – i.e. that are similar to many other instances of the category – are encoded strongly in memory (e.g. Clopper, Tamati, & Pierrehumbert, 2016), giving them advantages in immediate processing (e.g. Johnson, 2006) and long-term recall (e.g. Sumner et al., 2014). Tokens that are poor instances of their category, i.e. that do not pass the evaluation, are not stored.

The typicality evaluation is probabilistic, based on the activation of the identified category. Hence, tokens that are near the mode of the category (where activation is high) are more likely to pass than tokens that are far from it (where activation is low). At the level of the category distribution, this asymmetry yields the squeezing force (henceforth, the *typicality force*).

The evaluation proceeds by comparing the activation of the identified category, A_i (normalized for the number of exemplars of the category, N_i), to a typicality threshold, τ, as shown in Equation (2.9). τ is a parameter whose size determines the size of the typicality force: as τ grows higher, passing the evaluation becomes harder, and the force grows stronger. The typicality force also grows as the size of the activation window, α, shrinks. A narrower window encapsulates fewer exemplars and yields lower category activation.

$$P(\text{pass typicality evaluation}|A_i) = 1 - \exp\left(-\ln 2 \cdot \frac{A_i}{N_i\tau}\right) \tag{2.9}$$

The formulation of typicality evaluation in Equation (2.9) is inspired by the Complete Set Model of Busemeyer, Dewey, and Medin (1984). In this model, a “junk” category competes with established categories in the classification of a token; when the token does not yield sufficient activation, it is discarded as junk. The junk category has no exemplar basis of representation and thus is not included in the application of Luce’s Choice Rule (i.e. in the equivalent of Equation (2.8)); instead, it discounts the probability mass of each category (as derived from Luce’s Choice Rule) by a scale factor. As shown in Equation (2.10), this is equivalent to a two-stage process where the scale factor represents the probability associated with a junking decision that is contingent on categorization.

$$P(\text{member of } C_i \text{ and not junk}) = P(\text{member of } C_i) \cdot P(\text{not junk}|\text{member of } C_i) \tag{2.10}$$

I equate this post-categorization junking decision with typicality evaluation. Busemeyer et al. (1984) assume that junking is independent of (and thus potentially precedes) categorization, with the probability of junking decreasing exponentially with total activation across all categories. I keep the same form for the present treatment of typicality evaluation (Equation (2.9)), but assume that only the activation of the identified category contributes. This assumption follows from the treatment of typicality evaluation as occurring after identification and discriminability evaluation, and hence as assessing the extent to which the token is typical for the category to which it has been confidently assigned.

Typicality is a novel process for an exemplar-based model. In previous models, the squeezing force that is generated here by typicality is instead generated by a different process. Following Pierrehumbert (2001), previous models generate the squeezing force by *entrenchment*, a process of averaging targets in production (but see also Hintzman, 1986, and Goldinger, 1998, for a proposal of entrenchment in perception, through *echoes*). Though typicality and entrenchment both generate squeezing forces, they generate different kinds of squeezing forces, and the kind offered by typicality is superior for the model desiderata of maintaining category overlap without sacrificing category shape. For further discussion, see Section 2.3.3.

2.3.2.9 Storage

If the token passes both the discriminability and typicality evaluations, then it is *stored* as an exemplar of the identified category, updating the category representation. The conditioning of storage by the two evaluations implies that perceptually 'poor' (indiscriminable and/or atypical) productions are much less likely to influence representations – and thus are much less likely to be repeated – than perceptually 'good' productions.

When a token is stored, it overwrites a random exemplar of the same type (see e.g. Landauer, 1986, for discussion of overwriting as a principle of memory). All exemplars are stored with the same strength, which does not decay over time. This approach is nonstandard; following Pierrehumbert (2001), most exemplar-based models assume that exemplars decay over time and are never overwritten (Ettlinger, 2007; Sóskuthy, 2013; Tupper, 2015; Wedel, 2006, 2012; Wedel & Fatkullin, 2017). However, it is not without precedent: a similar approach is taken in the models presented by Wedel (2004) and Harrington et al. (2018). Moreover, it has no negative impact on the high-level model results, apart from preventing the problematic extinction of categories through leeching (cf. Wedel & Fatkullin, 2017). In fact, averaged over many runs, the expected behavior of a model with random overwriting of exemplars is equivalent to a special case of a model with decaying exemplars; see Appendix D for mathematical discussion.

Because tokens that fail a perceptual evaluation are not stored, they have no influence on category representations. However, since perceptual evaluations are probabilistic, tokens that fail an evaluation on one run of the model may pass it on another. Consequently, averaged over many different runs, the impact of perceptual evaluations is not to prevent perceptually 'poor' tokens from influencing category representations, but rather to *decrease* their influence on category representations (relative to perceptually 'good' tokens). In this way, the approach taken to storage here is equivalent to one in which all exemplars are stored, with a strength determined by the discriminability and typicality evaluations.

2.3.3 Comparison to existing models

At a high level, most of the representations and processes in the present model are common to previously-proposed exemplar-based models (Ettlinger, 2007; Pierrehumbert, 2001; Sóskuthy, 2013; Tupper, 2015; Wedel, 2004, 2006, 2012; Wedel & Fatkullin, 2017), though some low-level details of implementation differ. There are two components of the model that stand out from this commonality.

Firstly, the model detailed here includes both category- and type-level representations, and includes many types per category, of many different frequencies. Most previous exemplar-based models (Ettlinger, 2007; Pierrehumbert, 2001; Sóskuthy, 2013; Tupper, 2015; Wedel & Fatkullin, 2017) include just one of these two levels of representation, and those that include both (Wedel, 2004, 2006, 2012) include just a few types, all of the same frequency. Using a prototype-based model (which is closely related to exemplar-based models), Sóskuthy (2014) shows that the inclusion of many types per category, of many different frequencies, has implications both for changes in category shape and for frequency effects on rates of change. Given the focus on maintaining category shape here, and on frequency effects in Chapter 3, the novel treatment of type representations in the present model ensures that the results are applicable to real-world sound change, which of course includes many different words of many different frequencies.

Secondly, as described in Section 2.3.2.8, the model detailed here generates a squeezing force with the novel process of typicality evaluation in perception, rather than the usual process of entrenchment in production (cf. Pierrehumbert, 2001; Tupper, 2015; Wedel, 2006, 2012; Wedel & Fatkullin, 2017). This difference is not merely cosmetic, in shifting the origin of the squeezing force to perception from production; rather, it results in an important difference in the nature of the force. The force generated by typicality evaluation squeezes a category toward its mode, while the force generated by entrenchment squeezes it toward its mean. Squeezing toward the mode is superior to squeezing toward the mean for the present model desiderata, as it better maintains category overlap while better resisting changes in category shape.

To see how the differences between squeezing toward the mode and mean work out, consider the case of partially overlapping categories with short tails in the overlapping region (as created by the discriminability force, if left unchecked). Since the mode of each category is located closer to the overlapping region than the mean, squeezing toward the mode will push categories away from each other less than squeezing toward the mean. With less compounding of category repulsion, squeezing toward the mode will maintain overlap better than squeezing toward the mean. Furthermore, if the squeezing force grows superlinearly with the distance from the center (mode or mean), then squeezing toward the mode will shorten the short tail less than squeezing toward the mean, since the short tail is closer to the mode than it is to the mean (and vice-versa for the long tail). With less shortening of the short tail and more shortening of the long tail, squeezing toward the mode will resist increases in category skewness better than squeezing toward the mean.

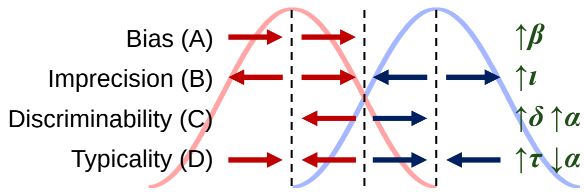

Figure 2.5: Forces from processes in the model. Iterated over time, processes in the model exert forces on the distributions of exemplars within each category. (A) Bias pushes one category along, toward the other. This force increases with β. (B) Imprecision spreads each category outward. This force increases with ι. (C) Discriminability pushes exemplars out of the region of overlap between categories, repelling categories away from one another. This force increases with δ and α. (D) Typicality lightens the tails of distributions, squeezing each category inward toward its mode and countering skewness. This force increases with τ and decreases with α.

2.4 Modeling regular sound change

The processes described in Section 2.3.2 work together to model regular sound change by generating different (complementary and counteracting) forces on category distributions, as illustrated in Figure 2.5. The size of each force is determined by parameters of the relevant processes. Since the forces act to construct the adaptive landscape, which shapes the long-term behavior of the model, the output of the model changes as the parameters change. For each choice of parameters, the model generates behavior that may or may not be appropriate for regular sound change.

How can it be determined which choices of parameters are appropriate for modeling regular sound change? Recall from Section 2.1.3 the high-level desiderata for a model of regular sound change: for both phonetic drift and push chains, it should (i) generate movement of each category; (ii) maintain the shape (width and skewness) of each category; and, for push chains, it should further maintain the (iii) distance between and (iv) overlap of the categories. The task for appropriate modeling of regular sound change is to identify parameter values that allow the model to meet these desiderata.

2.4.1 Meeting desiderata

As the model is run, the system of categories will move about the adaptive landscape defined by the category-level forces. The system will stop changing when it reaches a valley in the adaptive landscape, where the forces are balanced.[19] To meet the model desiderata, parameter values have to be chosen in such a way that the forces are close to balanced right from the beginning. Then, the structural properties with which the categories are initialized will be maintained perpetually.

For phonetic drift, three forces are active: bias, imprecision, and typicality. From the theoretical perspective of forces, the model desideratum of movement should be guaranteed by the bias force, for any nonzero value of β. However, since the category updates one exemplar at a time rather than all at once, the bias force will cease to act on occasions where it creates atypical tokens, and it may introduce 'motion blur' that increases category skewness (see also Pierrehumbert, 2001). For this reason, in practice, large values of β will be ineffective or counterproductive for meeting the model desiderata. The other forces in phonetic drift are more straightforward. To meet the model desideratum of maintaining category width, values of ι, τ, and α must be chosen that balance the imprecision force against the typicality force. A larger value of ι will cause a larger imprecision force that tends to widen the category, and must be countered by a larger value of τ or smaller value of α, which will cause a larger typicality force that tends to narrow the category.

For push chains, all four forces are active in the Pusher, and the three forces except bias are active in the Pushee. Movement of the Pusher is created by the bias force, but is countered by the discriminability force, and movement of the Pushee

[19] Since the present model has an unbounded exemplar space and no external categories, movement due to the bias force will persist perpetually. This means that the bias force need not be balanced out, but rather used to ensure that the net force in one direction is the same for both categories. In terms of the adaptive landscape, perpetual movement under the bias force corresponds to a long valley that has a constant downward slope, forever.

is created by the discriminability force. To meet the desideratum of maintaining category distance, both categories must move at the same rate. Movement of both categories at the same rate can be obtained only if the difference between the bias and discriminability forces in the Pusher is equal to the discriminability force in the Pushee. Thus, given a set of parameters that balance forces in phonetic drift, a starting point for push chains requires an appropriate choice of δ. This is only a starting point because the discriminability force is asymmetric, acting only on the side of each category that overlaps with the other category. Consequently, any choice of $\delta > 0$ will introduce an additional tendency for categories to narrow. To meet the model desideratum of maintaining category shape, this tendency must be countered by increasing ι relative to τ, thus increasing the imprecision force relative to the typicality force. The fact that the typicality force squeezes toward the mode rather than the mean will help to prevent increases of category skewness, in service of the desideratum of maintaining category shape (see Section 2.3.3).

2.4.2 Model tuning

As suggested by the descriptions in Section 2.4.1, the choices of parameter values that allow the model to meet the desiderata are highly interdependent and sensitive. Making effective choices therefore requires a systematic approach that explores the interactions between different parameters in a thorough manner, *tuning* the model to the desiderata. For both phonetic drift and push chains, I used a stepwise approach, tuning sets of parameters in order to meet different desiderata in turn. In this section, I describe the tuning processes at a high level; for the details of parameter values explored, see Appendix B.

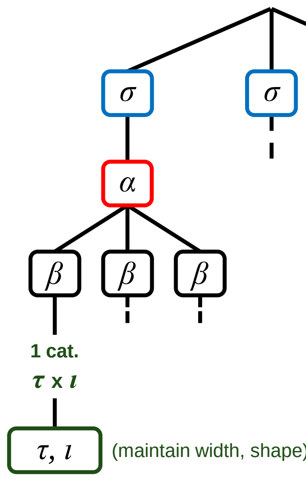

Figure 2.6: Illustration of the model tuning process for phonetic drift. The initialization parameter σ (blue box) was set to several values to define the objectives of the modeling process. The activation window size parameter α (red box) was arbitrarily fixed to $\sigma/2$ to provide a scale (without loss of generality). The bias size β (black boxes) was set to pre-defined (controlled) values. The other relevant parameters, typicality threshold τ and imprecision degree ι (green boxes) were tuned in order to meet the desiderata of maintaining category width and shape.

2.4.2.1 Tuning for phonetic drift

I illustrate the approach to model tuning for phonetic drift in Figure 2.6. The general strategy was to pre-determine values for the initial category width, σ, and then choose values for the other parameters so as to obtain category movement with maintenance of category shape and width.

I pre-determined three values for σ, representing narrow, medium-width, and wide category distributions. I fixed the activation window size, α, to be half the width of the category, σ, reflecting the observation that perception should draw on neither too many nor too few exemplars within a category. Fixing α in this way is not problematic because its role is to provide a perceptual scale, moderating the effect of the typicality threshold, τ, on the typicality force.

The goal of parameter tuning is to balance three forces: bias, imprecision, and typicality. This can be accomplished by adjusting two forces while keeping the other one fixed (providing the scale), because what matters for the qualitative dynamics of the system is the size of each force relative to the others. I therefore pre-determined three values for the bias, β, yielding weak, medium-strength, and strong bias forces.

These decisions gave me 9 systems (one for each pair of σ and β), each with two parameters to tune: τ, which determines the strength of the typicality force; and ι, which determines the strength of the imprecision force. I tuned these parameters by varying them independently among 10 values each. For each of the 9 systems and each of the 100 pairs of values of τ and ι, I ran the model 100 times for 5000 iterations (enough to indicate the stable state). For each system, I chose the value of τ that best maintained category kurtosis (tail length), and the value of ι that best maintained category width and skewness.

2.4.2.2 Tuning for push chains

The approach to model tuning for push chains strongly resembled the approach for phonetic drift. The general strategy was to pre-determine values for the initial category width, σ, and initial category distance, μ, and then choose values for the other parameters so as to maintain category shape, width, distance, and overlap, building on the existing results for phonetic drift. I illustrate the approach in Figure 2.7.

I used the same three pre-determined values of σ as for phonetic drift, and also fixed α in the same way. I pre-determined two values for μ for each value of σ, representing two different category distances and degrees of category overlap. These initialization parameters both define the structural properties to be maintained and contribute to the initial behavior of the model – causing, for example, greater initial discriminability force when the overlapping region is initially dense.

The goal of parameter tuning for push chains is to balance four forces: bias, imprecision, and typicality (as for phonetic drift), as well as discriminability. To

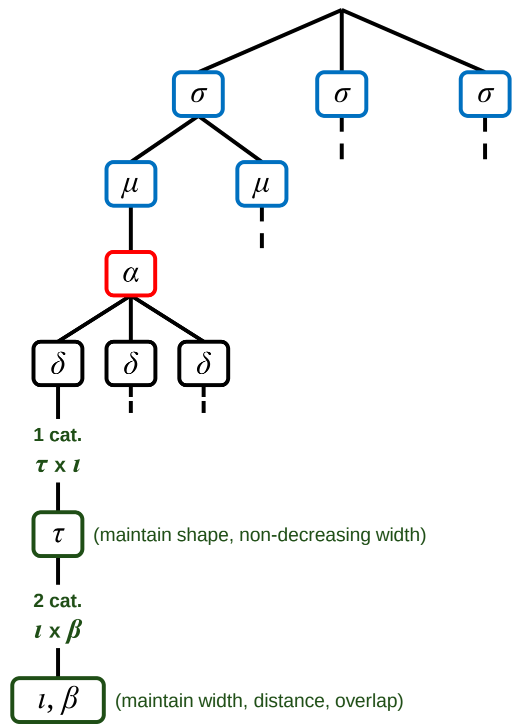

Figure 2.7: Illustration of the parameter-tuning process for push chains. The initialization parameters σ and μ (blue boxes) were set to several values to define the objectives of the modeling process. The activation window size parameter α (red box) was arbitrarily fixed to $\sigma/2$ to provide a scale (without loss of generality). The discriminability threshold δ (black boxes) was set to pre-defined (controlled) values. The other parameters (τ, ι, β; green boxes) were tuned in order to meet the objectives of the modeling process, in two steps. In the first step, I drew on the results of tuning for phonetic drift to choose a value of τ yielding maintenance of category shape alongside a range of non-decreasing category widths (for different values of ι). In the second step, I simulated push chains with a range of values of ι and β and chose the value of ι that yielded best maintenance of category width and distance and the value of β that additionally yielded best maintenance of category overlap.

balance four forces, one can be fixed while the others are adjusted. I therefore predetermined three values for the discriminability threshold, δ, yielding three different (fixed) strengths of the discriminability force: weak, medium-strength, and strong.

These decisions gave me 18 systems (one for each combination of σ, μ, and δ), each with three parameters to tune: β, which determines the strength of the bias force; ι, which determines the strength of the imprecision force; and τ, which determines the strength of the typicality force. I tuned these parameters in two steps. In the first step, for each system, I drew upon the results from tuning for phonetic drift to choose a suitable value for τ and 4 potential values for ι, which would best maintain category shape.[20] In the second step, I independently varied ι among the 4 potential values per system and β among 25 values. For each of the 18 systems and each of the 100 pairs of values of ι and β, I ran the model 100 times for 5000 iterations. For each system, I chose the value of ι that best maintained category width and distance, and the value of β that best maintained category overlap.

2.4.3 Results

The tuning process successfully identified sets of parameter values that allowed the model to meet the desiderata under a variety of circumstances, for both phonetic drift and push chains. To my knowledge, the generation of a push chain (i.e. interlinked category movement) with maintenance of category width, shape, distance, and overlap is a first in the literature on exemplar-based models of regular sound change. The results of push chain simulations for one set of parameter values are illustrated in Figure 2.8. For details of the parameter values chosen by the tuning process, and the structural properties they generated, see Appendix B.

[20] The values of ι were chosen to yield a range of increases of category width in phonetic drift. Increases in width were deemed potentially suitable because the addition of the discriminability force favors additional narrowing of categories, which must be countered to meet the model desiderata.

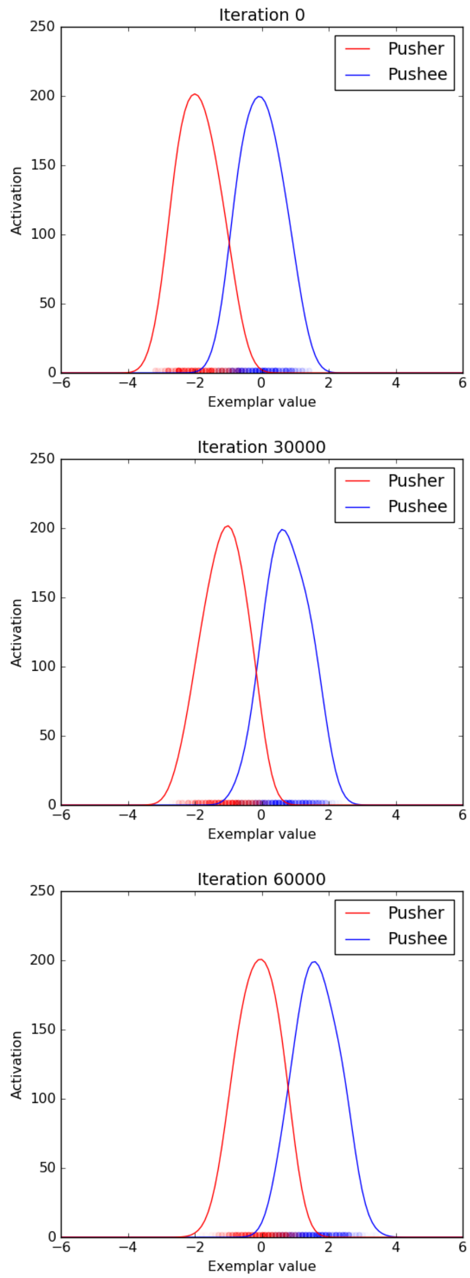

Figure 2.8: Successful modeling of a push chain. The plot traces the evolution of exemplar distributions (rugs on horizontal axis) and corresponding activation fields for one run of the model (parameter set (2) from Table B.2). Over time (from top panel to bottom), the two categories move to the right, maintaining their distance from one another, their degree of overlap, and their widths and skewnesses.

In addition to demonstrating that the model is capable of meeting the desiderata, the tuning process confirmed that parameter choices had the expected implications for category properties. For example, for phonetic drift, increasing bias (β) increased category displacement by a proportionate amount, and also increased category skewness and kurtosis due to the introduction of 'motion blur'. Increasing imprecision degree (ι) increased category width, and increasing the typicality threshold (τ) decreased category width. For push chains, higher discriminability thresholds (δ) created more displacement of the Pushee, whilst also decreasing category width and increasing category skewness. Thus, the parameter values affect the forces acting on category distributions in the way outlined in Figure 2.5.

Relatedly, the tuning process also identified relationships between forces / parameter values that are necessary for meeting the model desiderata. For phonetic drift, both the imprecision force (ι) and the typicality force (τ) are required to be sufficiently large relative to the bias force (β) in order to prevent the category becoming excessively skewed. When ι is small relative to β, little of the variation in production can be attributed to imprecision, meaning that the effect of bias is very clear. When τ is small relative to β, few extreme tokens are discarded for being atypical, meaning that bias is permitted to continue unchecked in the creation of extreme tokens. In addition, an increase in ι requires a concomitant increase in τ in order to maintain category width, and vice-versa.

For push chains, ι is required to be larger than it is for phonetic drift, in order to maintain category width in the face of the additional discriminability force. In fact, ι is required to be very large for push chains (near σ in value, representing variation that spans the width of the category), as it is only through imprecision that Pushee targets can be produced outside of the existing exemplar distribution, facilitating retreat. Additionally, greater bias force (higher values of β) is required to maintain smaller category distances (smaller values of μ). It is bias that causes the Pusher to move toward the Pushee, countering the repulsion due to the discriminability force, and nearer categories have denser overlapping regions and thus greater discriminability force. For the same reason, a great bias force is also required to counteract a great discriminability force, stemming from a high discriminability threshold (δ).

Finally, the tuning process highlighted the generality of the model's ability to meet the desiderata. The tuning process began with arbitrary decisions of values for a number of parameters. That these decisions did not limit the ability to identify sets of parameter values that meet the model desiderata suggests that there are many such suitable sets of parameter values. Furthermore, the model was not highly sensitive to the particular value of some tuned parameters. For example, slightly larger values of τ and ι would have yielded similar category shapes and widths to the values chosen, and thus would also have been appropriate.

2.5 Discussion

The model I have presented in this chapter makes an important contribution to the literature on regular sound change. To my knowledge, it is the first exemplar-based model to generate phonetic drift and push chains while maintaining key structural properties such as category shape, distance, and overlap. In fitting with its status as *listener-based*, the model owes a good deal of its success in these areas to processes in perception.

The model's success in meeting desiderata of structural maintenance can be attributed to three aspects of the perceptual processes undertaken by the listener. Firstly, by using a discriminability threshold $\delta < 1$, the model induces a lexical bias (Ganong, 1980) in cases of low discriminability, effectively shrinking the instability that acoustic overlap between phoneme categories creates for the perception of attested (real-word) types. Secondly, by not storing tokens that fail the discriminability evaluation (i.e. tokens that are likely to be recognized as nonwords), the model avoids skewness-inducing overpopulation of the overlapping region between categories (see e.g. results of "competition with discards" in Tupper, 2015). Thirdly, by including the novel process of typicality evaluation, the model generates a squeezing force that keeps skewness in check while facilitating overlap (see Section 2.3.3). The importance of these perceptual processes supports the claim that the listener plays a central role in sound change.

Further support for the centrality of the listener to the model is provided by the fact that, without the listener, push chains would not occur at all in the current setup: the Pusher would simply float over the Pushee, because the speaker samples production targets without concern for potential ambiguity. The listener prevents this behavior by providing an indirect, non-teleological influence of perceptual filtering on production: whenever the speaker is ambiguous, the listener is unlikely to store the token, and is thus unlikely to use it as a basis for future productions. The listener thus drives category interaction in the model by creating category repulsion via the discriminability force, with the speaker's constant bias serving to ensure that interaction persists in the face of this repulsion. More generally, in the current setup, the listener drives any self-organizational response to the system as a whole, as perceptual processes involve the activation of multiple exemplars (from both categories), whereas production processes involve no more than the single initial target exemplar.

The notion that the listener could be important for sound change is not new, but the present approach to it is. Ohala (1981) claimed that the listener could be a source of sound change by under- or over-applying perceptual compensation for coarticulation and thus misperceiving one sound as another. For example, the listener could incorrectly compensate for coarticulation that wasn't present in the realization of /yt/ as [yt] and thus reconstruct it as /ut/. There are three main differences between Ohala's model and the present model.

Firstly, Ohala's model represents sound change as a change in the phonological representation of words (e.g. /yt/ changing to /ut/), while the present model represents sound change as a change in the phonetic representation of a phoneme (e.g. /y/ changing its realization from [y] to [u]). Consequently, Ohala's model is not designed to capture the gradient changes that are observed in regular sound change, while the present model is.

Secondly, Ohala's model attributes the influence of the listener to misperceptions, while the present model attributes it to memory disadvantages of acoustically ambiguous tokens. While listeners certainly can misperceive one sound as another, especially in the presence of noise (Miller & Nicely, 1955), I do not believe that misperception is as widespread in practice (especially given context) as would be required for it to really drive sound change (see Appendix C for related simulations showing that misperception of minimal pairs is insufficient to generate robust push chain behavior).

Finally, in Ohala's model, a sound change that spreads across the lexicon (outside of the conditioning environment) must do so via analogy. Similarly, a change that affects one phoneme category has little impact on other categories. Consequently, the listener is a *source* of sound change, but does not *drive* sound change across the lexicon. In the present model, the speaker is the source of sound change (via biases in production), but the listener drives it, by forcing categories to interact (via the discriminability evaluation).

The model presented here also makes important contributions to the listener's role in language change more generally. These contributions stem from its formulation as a production-perception loop, where the representations that are drawn upon for production are also updated through perception. Within such a loop, the diachronic trajectory of a language is formed from the way in which the language is used at different points in time, and thus shaped by the forces (social, cognitive, physiological, etc.) that act during any synchronic communicative event (Beckner et al., 2009).

The interdependence between production and perception predicts that *any* sort of synchronic asymmetry in the way language is produced or perceived has the potential to shape patterns of diachronic change, provided it is sufficiently widespread, robust, and persistent. Research in speech perception has revealed many powerful and passive perceptual biases, relating to factors both internal to the language system – such as word frequency (see Chapter 3) – and external to the language system – such as social demographics and attitudes (see Chapter 4). Under the present model, these perceptual biases are expected to play a large role in shaping language change.

2.6 Summary

In this chapter, I have detailed a computationally-implemented listener-based model of regular sound change. I have illuminated how this model fits into the literature, by comparing it to other models based in the same framework. Through such comparison, as well as through simulations, I have established a thorough understanding of how the model generates sound change from interactions between the production and perception of individual words, across numerous communicative events.

The model presented in this chapter has demonstrated the utility of a listener-based approach to language change. Its approach to regular sound change is distinct from previous approaches and models, and it has been successful in generating multiple instances of phonetic drift and push chains that meet empirically-grounded desiderata. It lays important groundwork for the remaining chapters of this dissertation, where I consider how biases in speech perception – which I relate to the discriminability and typicality evaluations – can predict different words and speakers to participate in language change at different rates.

Chapter 3

Word frequency in sound change[*]

In Chapter 2, I introduced a listener-based computational model of regular sound change and demonstrated its ability to capture key properties of certain kinds of sound change involving one or two phonemes. In this chapter, I show how the model can be extended to capture the effects of word frequency in such changes, by incorporating experimentally-established perceptual biases. From one perceptual bias, I generate many different word frequency effects, for different kinds of regular sound change. I show that these effects match all those that have been empirically established through corpus studies of three kinds of regular sound change in the literature. Finally, I derive general predictions for the effects of word frequency in other kinds of regular sound change.

My primary goal in this chapter is to solve the major theoretical puzzle of explaining word frequency effects in sound change, and thereby demonstrate the potential of a listener-based approach. In doing so, I also address a secondary goal, of countering the common criticism that exemplar-based models overpredict advantages of high-frequency words in regular sound change (Abramowicz, 2007; Bermúdez-Otero

[*]This chapter is based on work published as Todd et al. (2019). The analysis and discussion of model results is primarily my own work, but it has benefited from the input of my coauthors.

et al., 2015; Dinkin, 2008; Tamminga, 2014). In showing that the listener-based model generates different effects of word frequency in different kinds of sound change, I dispel the misunderstandings underlying this criticism and point out the value that exemplar-based computational modeling can bring to the study of sound change.

This chapter is structured as follows. First, I lay out the focus on particular kinds of word frequency effects, which I situate against the predictions of different theories of language change (Section 3.1). Next, I describe empirically-observed word frequency effects in diachronic sound change and synchronic speech perception (Section 3.2), and I extend the model to connect these two effects (Section 3.3). Then, I present simulations showing that a single perceptual bias can cause different word frequency effects in different kinds of regular sound change (Section 3.4), together with additional simulations showing that these effects are not driven by biases in the speaker (Section 3.5). Finally, I discuss the implications of these results and the predictions they make for other kinds of regular sound change (Section 3.6).

3.1 Focus

Traditionally, linguists have considered the effects of word frequency on the *actuation* of sound change – i.e. whether high- or low-frequency words change *first*. In this chapter, I focus instead on *rates* of change – i.e. whether high- or low-frequency words change *fastest*. I focus on rates rather than actuation for empirical reasons: they can be identified more easily and robustly in corpora without extensive time depth (as a statistical interaction between word frequency and time), and they allow for easier disentanglement of change from natural phonetic variation in a continuous acoustic system. I focus on word frequency effects (of any sort) because different approaches to sound change make different predictions.

According to the Neogrammarian hypothesis (see e.g. Garrett, 2014), regular sound change affects all eligible words in the same way – and thus at the same rate, regardless of frequency. This lexical independence in sound change follows from the assumption of *strict modularity*, where the representation of the phoneme is independent of its instantiation in words. Under this assumption, regular sound change involves changes to the phoneme representation rather than to words directly.

By contrast, recent usage-based approaches relax the assumption of strict modularity, contending that the instantiation of the phoneme within words is central to the way that the phoneme is represented both cognitively and theoretically (e.g. Beckner et al., 2009; Blevins & Wedel, 2009; Bybee, 2002; Harrington et al., 2018; Hay & Foulkes, 2016; Hay et al., 2015; Johnson, 1997; Pierrehumbert, 2001, 2002; Wedel, 2006, 2012). Such approaches assume that instances of the same phoneme in different words may have different (but related) representational bases. Consequently, while sound change is expected to affect all words containing the changing phoneme over a certain period of time, it is not assumed to affect all words at the same rate.

In particular, the Frequency Actuation Hypothesis (henceforth, FAH; Bybee, 2002; Phillips, 1984) claims that word frequency effects[21] will be different in different kinds of sound change, depending on the motivation of the change. Phillips (1984) presents a two-way distinction between physiologically motivated changes and non-physiologically motivated changes. Physiologically motivated changes result from the iteration of articulatory biases and affect the surface phonetic form of phonological segments. An example is /t/-tapping, where a word like *matter* comes to sound more like *madder*, reducing articulatory effort. In physiologically motivated sound changes, high-frequency words are predicted to change faster than low-frequency words, since

[21] The FAH was originally posited with reference to the question of whether high- or low-frequency words change first. For the reasons previously outlined, I reinterpret it to generate predictions about whether high- or low-frequency words change fastest, under the assumption that change begins from neutral initial conditions.

they are produced and thus subjected to the articulatory bias more often. Non-physiologically motivated changes result from lexical analogy of sound patterns to new environments and yield new constraints on underlying phonological or phonotactic structures. An example is the deletion of glides after coronal stops /t d n/, where a word like *tune* comes to sound more like *toon*, generalizing the constraint banning glides after other coronal consonants (Phillips, 1981). In non-physiologically motivated sound changes, high-frequency words are predicted to change more slowly than low-frequency words, since their frequent use allows them to persist as exceptions in the phonological grammar.

The most intuitive application of the FAH to regular sound change makes the assumption that gradient phonetic change results primarily from iterated biases in the speaker's phonetic implementation, and thus predicts that high frequency words should always change faster than low-frequency words. This prediction found support in claims based on early exemplar-based modeling work by Pierrehumbert (2001), in which the speaker was central. However, it does not hold uniformly in empirical data. I am aware of three corpus studies of word frequency effects on rates of sound change across the lexicon.[22] One of these studies found a result that is inconsistent with predictions of the FAH while not directly opposing them: Bermúdez-Otero et al. (2015) found that /t/-glottaling in Manchester English is affecting words of all frequencies *at the same rate*. Another found a result that is fully consistent with predictions of the FAH: Hay and Foulkes (2016) found that /t/-tapping in New Zealand English is affecting high-frequency words faster than low-frequency words. The final study, however, found a result that opposes predictions of the FAH: Hay et al. (2015) found that /ɛ/-raising (and other processes in the same vowel shift) in New Zealand English

[22] I include only studies that aggregate across the lexicon because the assumption that speech and/or sound change contain stochastic elements implies that sampling a few words is not sufficient to reflect upon the existence of statistical tendencies tied to word frequency. For this reason, I exclude from consideration a study by Tamminga (2014), which explores /ai/-raising in Philadelphia English for various senses of *like*.

affected high-frequency words *slower* than low-frequency words. I discuss these three changes in more detail in Section 3.2.1; for now, I simply note that the FAH does not explain why they should each show different word frequency effects. The existence of different effects of word frequency on rate of change in different kinds of change remains an unsolved puzzle in studies of regular sound change.

I propose that these differences can be understood by making the listener central to regular sound change, as proposed in Chapter 2. I hypothesize that asymmetries in the rates with which regular sound change affects different words follow from experimentally-established asymmetries in the robustness with which those words can be recognized: high-frequency words can be recognized more robustly than low-frequency words in the face of acoustic ambiguity. Under this hypothesis, different asymmetries are observed in different kinds of regular sound change because of the different implications they have for the acoustic ambiguity of the involved phonemes. High-frequency words change at the same rate as low-frequency words when a phoneme moves without encroaching on the acoustic space of another, with no bearing on acoustic ambiguity; faster than low-frequency words when a phoneme moves toward another, potentially increasing acoustic ambiguity; and slower than low-frequency words when a phoneme moves away from another, potentially decreasing acoustic ambiguity.

3.2 Word frequency effects

As described in Section 3.1, recent corpus-based studies have demonstrated a range of different effects of word frequency on rates of regular sound change, but existing theories struggle to account for these different effects satisfactorily. At the same time, a host of experimental studies have identified that spoken word perception is subject to biases related to word frequency, but these biases have not been incorporated into theories of sound change. In this section, I describe both sets of empirical observations, highlighting their parallels, to feed into the argument that they are connected.

3.2.1 Word frequency effects in sound change

To my knowledge, only three studies to date have explored the effect of word frequency on rates of regular sound change across the lexicon. Intriguingly, these studies have all yielded different results.

The first study concerns /t/-glottaling in Manchester English (Bermúdez-Otero et al., 2015). /t/-glottaling refers to a sound change whereby /t/ between vowels becomes increasingly likely to be realized as a glottal stop [ʔ]; an example is *mitten* coming to be pronounced as "mi'en". This sound change is an instance of *phonetic drift*: the phonetic realization of /t/ has drifted to [ʔ] over time, without retreating from or encroaching on the acoustic territory of any other phoneme. Bermúdez-Otero et al. (2015) find that, while high-frequency words exhibit more /t/-glottaling at every point of time, words of all frequencies have exhibited a change toward increased /t/-glottaling at the *same rate*.

The second study concerns /t/-tapping in New Zealand English (Hay & Foulkes, 2016). /t/-tapping refers to a sound change whereby intervocalic /t/ is weakened to a short voiced sound that may be notated [ɾ] or [d]; an example is *matter* coming to be pronounced more like "madder". This sound change resembles the *Pusher* component of a push chain: the phonetic realization of /t/ has increasingly encroached on the acoustic territory of /d/ over time.[23] Hay and Foulkes (2016) find that high-frequency

[23]Hay and Foulkes (2016) do not investigate whether the changes in /t/ have triggered related changes in /d/, as would be expected in a push chain. For present purposes, however, the behavior of /d/ does not matter; what matters is that /t/ advances toward it. The modeling of a situation where /d/ exhibits minimal reaction to the advancement of /t/ is beyond the scope of the current work, but could easily be accomplished by the assumption of a repulsive boundary provided by paradigmatic or articulatory limits on the realization of voiced stops.

words have exhibited a change toward increased /t/-tapping at a *faster rate* than low-frequency words.

The third and final study of word frequency effects on rates of sound change concerns /ɛ/-raising in New Zealand English (Hay et al., 2015). /ɛ/-raising causes the vowel /ɛ/ to be pronounced with the tongue higher in the mouth, as [e], making New Zealand *bet* sound to non-New Zealanders like "bit". This sound change is a *Pushee* in a larger push chain[24]: /æ/ has increasingly moved toward /ɛ/ in acoustic space, pushing it along a related trajectory of change. Hay et al. (2015) find an effect of word frequency which is different to that seen in the previous two studies: high-frequency words have changed *slower* than low-frequency words.

In light of the difficulty of amassing enough good-quality data to reliably span time, speakers, and the lexicon, it is not surprising that there exist only three studies on word frequency effects on rates of sound change. Given the paucity of work in this area, it is not currently possible to definitively solve the puzzle of word frequency effects – but it is possible to make progress. To this end, I take the strong approach of assuming that the three cases described here are representative, and I use them to construct general predictions that can be fully tested as more data become available.

Under this strong approach, it is notable that word frequency effects on rate of sound change only emerge in cases where two phoneme categories interact, with sound change having implications for – or motivations in – the region of acoustic ambiguity caused by category overlap. In /t/-glottaling, where the change has no connection to category overlap, there are no word frequency effects. Conversely, in /t/-tapping, where the change in question moves a category *toward* an ambiguous region of category overlap, there is a high-frequency advantage, and in /ɛ/-raising, where the change moves a category *away* from an ambiguous region of category

[24]While the larger push chain has implications for multiple phoneme categories, I focus here on establishing a thorough understanding of the /ɛ/-raising component because it is least affected by simplifications in the modeling of the exemplar space (see Section 3.3.3 for discussion).

overlap, there is a low-frequency advantage. As I will show in Section 3.2.2, this connection between acoustic ambiguity and word frequency effects is paralleled in speech perception.

3.2.2 Word frequency effects in speech perception

The literature contains numerous empirical results showing that high-frequency words are privileged over low-frequency words in speech perception, both when there is no salient lexical competitor, and when there is. In situations without a salient lexical competitor, relative to low-frequency words, high-frequency words are intelligible in larger amounts of masking noise (Howes, 1957) and are classified as real words more often (Luce & Pisoni, 1998) and faster (Forster & Chambers, 1973) in lexical decision. In situations where multiple salient words compete for recognition, higher-frequency words attract more attention early in processing (Dahan, Magnuson, & Tanenhaus, 2001) and are favored responses to degraded stimuli (Savin, 1963) or stimuli from a dialect other than one's own (Clopper, Pierrehumbert, & Tamati, 2010).

Furthermore, a series of phonetic categorization studies have shown word frequency effects in the mapping of an acoustically ambiguous stimulus to one of two words in a minimal pair. Fox (1984) observed that, when presented with ambiguous stimuli on a "bad"−"dad" continuum, listeners were more biased toward *bad* responses than expected (based on their responses to a /bæ/−/dæ/ continuum). He suggested this might be because *bad* is more frequent than *dad*. Connine et al. (1993) provided support for this suggestion from a range of continua between high- and low-frequency words that differ in initial stop voicing (e.g. "best"−"pest"). Ambiguous stimuli on these continua were more likely to trigger the high-frequency response (e.g. *best*) than the low-frequency response (e.g. *pest*). Similar results were found by VanDam (2007). Finally, de Marneffe, Tomlinson, Tice, and Sumner (2011) replicated this result using manipulated French-accented English words with final stops (e.g. "tag" and "tack"), showing that the high-frequency response bias is not limited to situations where the stimulus begins with an ambiguous sound. Furthermore, they showed that the bias is not limited to minimal pairs with extreme frequency differences, but is found across minimal pairs, with strength related to the ratio of word frequencies.

The experimental results reviewed above imply that – all else being equal, i.e. absent effects of speech style or semantic context – the perceptual system is biased toward the recognition of high-frequency words, especially in the case of acoustically ambiguous tokens. The fact that word frequency effects on speech perception are observed in the presence of acoustic ambiguity establishes a parallel with regular sound change, suggesting that the two may be connected. In Section 3.3, I formalize this connection within the listener-based model described in Chapter 2, establishing a route through which word frequency effects on speech perception may provide a plausible explanation for word frequency effects on rates of regular sound change.

3.3 Modeling word frequency effects

The effects of word frequency on sound change (Section 3.2.1) can be summarized in the statement that high-frequency words are faster than low-frequency words to move into, and slower to move out of, regions of acoustic ambiguity. Put another way, high-frequency words can advance into an ambiguous region, and stay in it, in spite of the strong repulsive force associated with it; thus, they are less sensitive than low-frequency words to this repulsive force. As established in Chapter 2, in a listener-based model, the repulsive force is generated by failures of word recognition, through the process of the discriminability evaluation. Experimental results in speech perception (Section 3.2.2) have shown that listeners are less likely to fail to recognize high-frequency words than low-frequency words, particularly when acoustically ambiguous. Consequently, passive processes in speech perception provide precisely the right frequency-based asymmetries in the repulsive force that are required to explain word frequency effects on rates of sound change. In this section, I formalize the connection between frequency effects in sound change and in speech perception in an extension of the listener-based model described in Chapter 2.

As in Chapter 2, I do not attempt to model the exact details of the three empirical sound changes described in Section 3.2.1. Instead, I identify them with components of the models of phonetic drift and push chains described in Chapter 2, and I aim to capture the qualitative properties of word frequency effects in each of these components. I identify Manchester English /t/-glottaling (Bermúdez-Otero et al., 2015) with phonetic drift, in which I aim to generate no word frequency effects. I identify New Zealand English /t/-tapping (Hay & Foulkes, 2016) with the Pusher in a push chain and New Zealand English /ɛ/-raising (Hay et al., 2015) with the Pushee. I combine these two cases into a single model for convenience, because each represents the interaction of two phoneme categories in a similar way, and each focuses on a category fulfilling a different role in this interaction. In the modeled interaction, I aim to generate faster change of high-frequency words than low-frequency words in the Pusher, and slower change in the Pushee.

3.3.1 Approach

I assume that acoustically ambiguous tokens of high-frequency words are more robustly recognized and stored than similarly-ambiguous tokens of low-frequency words (see also Hay et al., 2015, for more discussion). I encode this perceptual bias in the model by varying the discriminability threshold, δ, with type frequency. Specifically, I give tokens of high-frequency types lower δ than tokens of low-frequency types, making them more *discriminable*, i.e. more likely to pass the discriminability evaluation and be stored when encountered in the overlapping region between categories. This assumption has no implications for the case of phonetic drift, as the discriminability evaluation cannot fail in a system containing only one category.

To test the hypothesis that frequency-based asymmetries in discriminability could give rise to empirically observed frequency effects on rate of change, I conducted simulations with frequency-sensitive δ, keeping all other parameters fixed at the previously-tuned values (Section 2.4.2). I set the discriminability threshold (δ) to be a linear function of type frequency (f):

$$\delta(f) = \left[\lambda + \left(\frac{2(f-1)}{M-1} - 1\right)\phi\right]_0^1 \tag{3.1}$$

where M is a constant representing the maximum type frequency in the system (here $M = 12$) and where $[x]_0^1$ evaluates to 0 if $x < 0$, 1 if $x > 1$, and x otherwise. I set a ceiling at $\delta = 1$ because $\delta > 1$ would imply a *dis*advantage for real words in the recognition of phonetically ambiguous stimuli (contra Ganong, 1980). I set a floor at $\delta = 0$ because it represents the limit case where tokens pass the discriminability threshold regardless of the activations they incite.

Using Equation (3.1), I constructed 15 frequency-sensitive δ functions, which are illustrated in Figure 3.1. Each δ function has an equivalent average (median-frequency) value to one of the 3 constant δ values from the original simulations in Chapter 2 ($\lambda \in \{0.25, 0.50, 0.75\}$), and has one of 5 different slopes ($\phi \in \{0.00, 0.25, 0.50, 0.75, 1.00\}$), representing 5 different degrees of asymmetry in the discriminability of high-frequency types relative to low-frequency ones.

The implications of a frequency-sensitive discriminability threshold can be illustrated by considering the mathematical formulation of the discriminability evaluation, which is repeated in Equation (3.2). The value of the threshold, δ, has an inverse multiplicative effect on category activations: the lower δ, the higher the category activation for the purposes of the discriminability evaluation. In effect, then, δ provides an activational boost to the identified category (through the identified

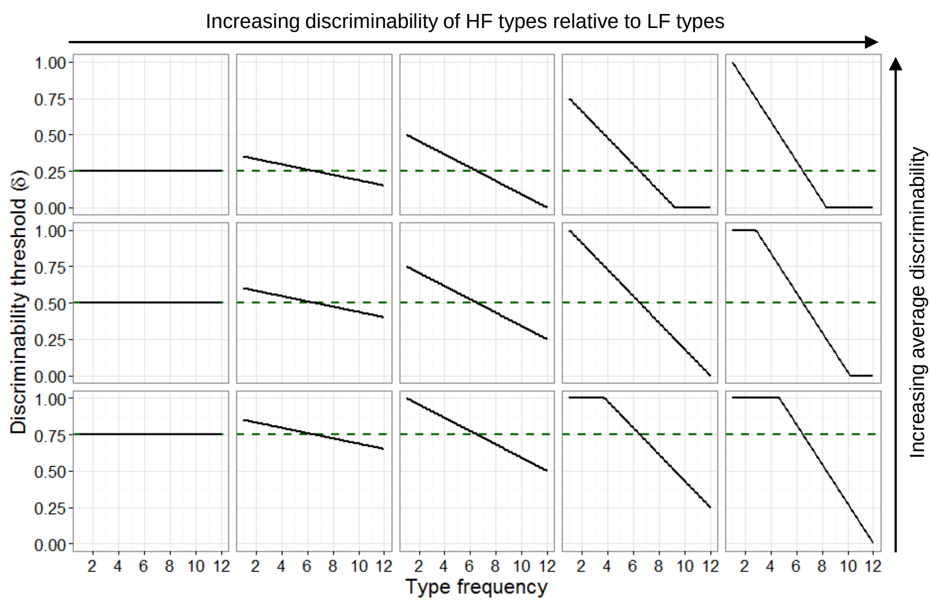

Figure 3.1: Frequency-sensitive discriminability thresholds. δ functions investigated (black lines). Lower δ indicates lower discriminability threshold, and hence greater ability to pass the threshold, i.e. greater *discriminability*. Across all panels in a given row, δ is kept constant for median-frequency types (dashed green lines). This median-frequency δ decreases moving up the rows, making discriminability higher on average. Across all panels in a given column, the difference between δ for high-frequency types and δ for low-frequency types (slope) is kept constant. This difference increases (slope steepens) moving rightward across the columns, making high-frequency types increasingly more discriminable than low-frequency types.

type), which can help it overcome the difficulties of recognition in the face of acoustic ambiguity. In this way, the present formulation resembles the Logogen model of Morton (1969), where high-frequency words have a higher resting activation than low-frequency words. However, there is an important difference between the present model and the Logogen model, concerning the form of the activational boost: in the present model, it is multiplicative, but in the Logogen model, it is additive. The use of a multiplicative boost here means that the activation of an exemplar of a high-frequency type is not on average higher than the activation of an exemplar of a low-frequency type, independent of acoustic value. Rather, a token of a high-frequency type garners more activation than an otherwise identical token of a low-frequency type (i.e. one with the same acoustic value).

$$P(\text{pass discriminability evaluation}|A_i, A_o) = \frac{\frac{1}{\delta} \cdot A_i}{\frac{1}{\delta} \cdot A_i + 1 \cdot A_o} \tag{3.2}$$

3.3.2 Potential mechanisms

While the assumption that high-frequency types pass the discriminability threshold more easily than low-frequency types is justified by results in the literature (Section 3.2.2), its implementation here – directly varying the discriminability threshold, δ, with type frequency (Section 3.3.1) – does not follow from anything else within the exemplar-based framework. On one hand, the implementational details are unimportant; the high-level impact of the corresponding perceptual bias on patterns of sound change can be sufficiently demonstrated through any implementation within a listener-based model. On the other hand, however, the appeal of a listener-based computational model framed in Exemplar Theory is that it makes explicit the mechanistic details of processes that give rise to observed behaviors. In this light, it is worthwhile to consider the mechanisms through which the assumed frequency-based asymmetry could arise, even though implementing them is beyond the scope of the present work.

In what follows, I outline two theoretically-justified mechanisms that would give rise to frequency-based asymmetries in the discriminability evaluation, and I discuss their implications for frequency-based asymmetries in the typicality evaluation.

Under the first mechanism, when an incoming token is perceived, the activation of exemplars is weighted by their structural compatibility with the token (their similarity in phonological frame identity) in addition to their position within the activation window (their similarity in acoustic quality of the target phoneme).[25] Such weighting represents a recognition of the fact that the exemplar space is multidimensional, with dimensions corresponding to the phonological frame as well as the quality of the target category realization (Pierrehumbert, 2002). Thus, the token "map" would activate an exemplar of the type *map* with a given F1 value more than an exemplar of the type *pat* with the same F1 value; exemplars of *map* would receive an activational boost from their high structural compatibility with the token "map" (proportional to their position within the activation window). Because a high-frequency type is represented by more exemplars than a low-frequency type, its category receives more of these activational boosts than it would in an equivalent situation with a low-frequency type, yielding greater expected category activation for high-frequency types than for low-frequency types. The ratio of category activations is thus expected to be greater for a high-frequency type than for a low-frequency type, making it easier to pass the discriminability evaluation. A consequence of this mechanism is that the greater expected category activation for high-frequency types also makes them more likely to pass the typicality evaluation.

Under the second mechanism, high- and low-frequency types project activation windows of different sizes. In defining the exemplar-based Generalized Context Model,

[25] Exemplars may also be weighted by their contextual similarity with the token more generally. Weighting according to the broad context provided by talker or situation may generate perceptual adaptation effects, where the listener rapidly adjusts perceptual expectations and representations while listening (Bradlow & Bent, 2008; Clarke-Davidson et al., 2008; Dahan et al., 2008; Kraljic & Samuel, 2006; Norris et al., 2003).

Nosofsky (1986, p. 41) states that the perceptual sensitivity parameter[26], c, (inversely related to the present activation window size parameter, α) "would be expected to increase... as subjects gained experience with the stimuli". Thus, the perception of a token of a high-frequency type is expected to draw on fewer exemplars that are far from the token in the perceptual-acoustic space than the perception of a token of a low-frequency type. Consequently, the activations of both the intended and the other category are expected to be lower for a token a high-frequency type than for an equivalent token of a low-frequency type; in particular, the activation of the other category is expected to be very small for a token of a high-frequency type relative to a token of a low-frequency type, since most exemplars of the other category are located far from the average token of the intended category. Thus, the ratio of activations (identified/other) would generally be greater for high-frequency types than for low-frequency types, making it easier to pass the discriminability evaluation. A consequence of this approach is that the lower expected category activation for high-frequency types would make them less likely to pass the typicality evaluation.

At present, I have not incorporated either of the mechanisms presented above into the model, in order to focus precisely on a single process, but I believe that doing so could be fruitful for future research. Since both mechanisms have implications for type frequency effects in the typicality evaluation, which is akin to a filter on storage in memory, they have the potential to speak to a large number of results in the literature about frequency-based asymmetries in the way that tokens are stored in memory (Benjamin, 2003; Bowers, 2000; Chee, Goh, Lim, Graham, & Lee, 2004; Diana & Reder, 2006; Forster & Davis, 1984; Goldinger, Luce, & Pisoni, 1989; Kinoshita, 1995;

[26]The sensitivity parameter is assumed by Nosofsky (1986) to be constant across all types experienced by a given subject, but it could plausibly be extended to vary across types, given that exemplar-based models assume that experience is accrued in a type-specific manner (Pierrehumbert, 2002). Indeed, Nosofsky (1991) considers the equivalent of such an extension and finds that it gives superior description of human recognition data (in the visual mode), though not of classification data.

Schulman, 1967; Wagenmakers, Zeelenberg, & Raaijmakers, 2000; Wierda, Taatgen, van Rijn, & Martens, 2013). The linking of implications for different perceptual effects through a single mechanistic assumption highlights the unifying power that such an assumption may have within a listener-based computational model.

3.3.3 Limitations

The formulation of a tractable computational model of sound change requires making many simplifications. While these necessary simplifications enable isolated investigation of the role of a single perceptual process, they also place limitations on the ways that the model results can be understood or applied. In this section, I discuss major limitations resulting from necessary model simplifications, around three questions: what is being predicted, how the predictions can be interpreted, and what empirical sound changes can be used to test them.

The first limitation concerns the kind of behavior that the model predicts. The force-balancing conception of the model ultimately concerns the *eventual* behavior of a system, when all forces are balanced. Thus, the model predicts the relative positions of high- and low-frequency types upon converging to a stable equilibrium state[27]; I illustrate how these predictions work in Figure 3.2. Though the model predicts convergence to an equilibrium state, the path that the system takes to get to that equilibrium state depends upon the initial conditions (e.g. initial difference between high- and low-frequency types). This dependence strongly affects the question of whether high- or low-frequency types will be 'ahead' in a change at any particular point in time, and is responsible for my decision to focus on *rates* of change. While

[27]Because the model assumes that the representations (e.g. lexical and phonological inventories) and forces (e.g. production bias) are constant for all time, it is always driven toward a single equilibrium state, even if this state is not reached within the specified iterations. In real sound systems, the representations and forces may change readily, meaning that 'convergence' never actually occurs, and instead the expected frequency effects change with the system.

not a direct prediction of the model, the effect of frequency on rates of change is a fairly robust indirect prediction that holds across most initial conditions.

The second limitation concerns the way that the model predictions can be interpreted. Because the model is stochastic in nature, its predictions reflect what is expected *on average* for a particular kind of change, not what actually happens in any *instance* of that change. The model's predictions are tendencies that should be observable *across* many different instances of the same kind of change. For example, even under the model settings that generate the strongest frequency effects in the exploration presented here (see Section 3.4.2), only 70% of simulations yield the corresponding qualitative pattern of results, some of which represent extremely small effect sizes. Thus, the model predicts the existence of a certain number of null or even conflicting results from empirical studies. Similarly, the model's predictions for frequency-based differences across types within a particular change are also tendencies. The model does not predict that *every* pair of high- and low-frequency types within a category will exhibit the relationship that is expected for the given change; rather, the relationship is only expected to manifest when aggregating over the entire lexicon (or a sufficiently large representative sample).

The final limitation concerns the empirical sound changes that can be used to test the model predictions. This limitation is caused by simplifying assumptions about the exemplar space: (i) it contains just one or two categories; (ii) it consists of a single perceptual-acoustic dimension; and (iii) it extends without bound, with all areas being equally 'hospitable' for exemplars. Relaxing any of these assumptions would introduce complexities that are beyond the scope of the present work. Relaxing the first assumption – allowing more than two categories into the model – would cause the predicted frequency effects to depend upon the precise initial configuration of categories. For example, in a system that includes three categories along a single dimension, different effects will be predicted based on whether the middle category

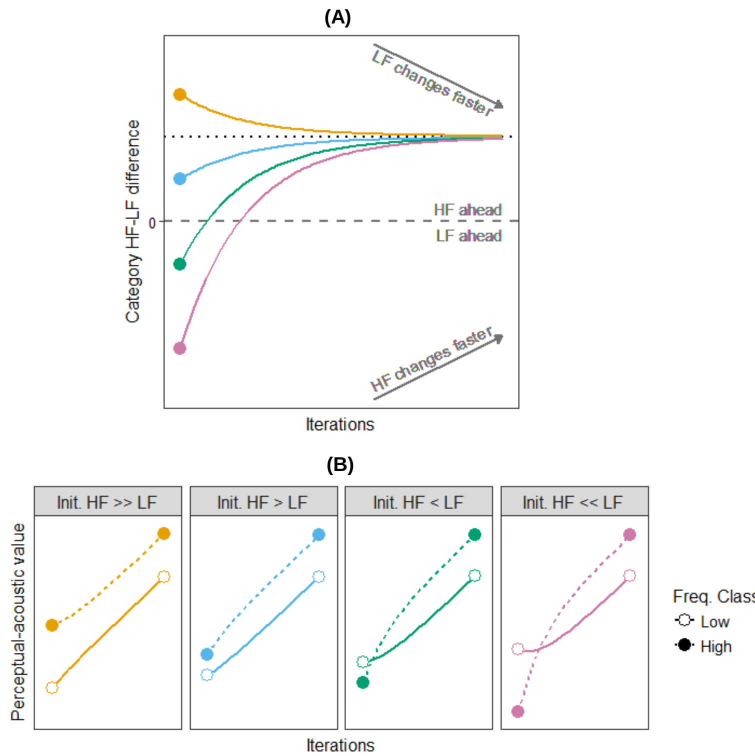

Figure 3.2: Illustration of how the model predictions work. Assume that the sound change is causing increase along some perceptual-acoustic dimension. (A) illustrates change in the difference along the perceptual-acoustic dimension between high- and low-frequency subdistributions of the category over time, while (B) illustrates potential trajectories of sub-distribution movement that could generate the patterns in (A). Both illustrations include possibilities from different initial conditions (matched colors), i.e. different relative positions of the high- and low-frequency subdistributions at the beginning of the change. The model predicts eventual convergence to a fixed difference (dotted black line in (A); fixed separation of trajectories in (B)). This prediction is not robustly related to a prediction of which types are 'ahead' in the change at any given point in time. Half of the initial conditions (orange, blue) have high-frequency types ahead (above the dashed gray line in (A); higher value of the high-frequency trajectory in (B)) at all points in time, while the other half (green, purple) have low-frequency types ahead at first and high-frequency types ahead later. The model's prediction is more robustly related to a prediction of differences in rates of change. Most of the initial conditions (blue, green, purple) have high-frequency types change faster than low-frequency types (positive slope in (A); steeper slope of the high-frequency trajectory in (B)).

begins closer to the first or last category. Relaxing the second assumption – expanding to a multi-dimensional exemplar space – would cause the predicted frequency effects to depend upon the alignment of the category trajectories. For example, if the Pusher is moving horizontally, different frequency effects will be predicted based on whether the Pushee also moves horizontally or instead moves with a vertical component. Finally, relaxing the third assumption – adding bounds to the exemplar space that repel nearby exemplars, e.g. via physical limits of articulation – would cause the predicted frequency effects to depend upon the role of those bounds in the change. For example, in a two-dimensional system[28], different frequency effects will be predicted based on whether Pushee movement along a certain dimension is caused by repulsion from the Pusher or from the bounds of the space. The limitations surrounding the exemplar space are the reason that I focus on just the /ɛ/-raising component of the New Zealand short front vowel shift (Hay et al., 2015). The whole push chain involves up to four categories on different trajectories in a two-dimensional space; /ɛ/-raising is the only component that involves movement of a category (/ɛ/) along the same trajectory as its pusher (/æ/) but not its pushee (/ɪ/).

3.4 Results

I hypothesized that incorporating an experimentally-established perceptual bias into the listener-based computational model of sound change would generate effects of type frequency on rate of change, mirroring the effects seen empirically at a high level. In Section 3.3.1, I laid out the details of this approach, as varying the discriminability threshold with type frequency. In this section, I report the results of simulations

[28]In a one-dimensional system, a bounded exemplar space would cause movement to cease eventually, but would not otherwise interact with frequency effects. This lack of interaction allows the model to be applied to movement that eventually ceases – as in /t/-glottaling and -tapping – even though movement never actually ceases in simulations.

with frequency-sensitive discriminability threshold, both for phonetic drift of a single category and for the interaction of two categories.

3.4.1 Phonetic drift

For phonetic drift – where an isolated category moves about the exemplar space – the goal is to produce change of all types at the same rate, mirroring empirical data from Manchester English /t/-glottaling (Bermúdez-Otero et al., 2015). Under a listener-based approach, this goal seems readily attainable, because phonetic drift involves no appeal to acoustic ambiguity and thus contains no role for the discriminability evaluation.

Figure 3.3 shows how both low- and high-frequency types move over time, averaged across 1000 simulations, for representative parameter settings under different degrees of bias. The centroids of the low- and high-frequency sub-distributions change in parallel (i.e. at the same rate), supporting the hypothesis that sound change without adverse implications for the listener is frequency-independent. I take this result to mean that a listener-based focus can explain the lack of word frequency effect on rates[29] of /t/-glottaling in Manchester English (Bermúdez-Otero et al., 2015).

The demonstration of a lack of frequency effect on rates of phonetic drift is an important contribution not only empirically, but also theoretically. It counters prominent intuitions in previous usage-based approaches (Bybee, 2002; Phillips, 1984; Pierrehumbert, 2001) that the automation of biased production strategies should cause

[29]I emphasize that my focus is on the lack of word frequency effects on *rate of change*, and not on the existence of a stable frequency effect whereby high-frequency words exhibit a fixed amount more /t/-glottaling than low-frequency words at every point in time. To account for this stable effect, the model would have to assume that high-frequency words are more prone to hypoarticulation than low-frequency words (e.g. Bell, Brenier, Gregory, Girand, & Jurafsky, 2009; Gahl, 2008) and that the initial conditions reflect this asymmetry (rather than being neutral, as at present). The same assumptions are required in the explanation put forward by Bermúdez-Otero et al. (2015).

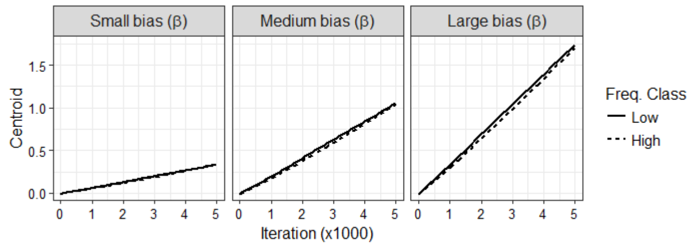

Figure 3.3: Frequency effects in phonetic drift: category movement. Results of simulations with a single category ($\sigma = 0.8$) subject to varying degrees of bias, illustrating differences between low-frequency (solid) and high-frequency (dashed) types. For all degrees of bias, the centroid of the category distribution advances at the same rate for both low- and high-frequency types.

high-frequency types to change faster than low-frequency types, since they are produced (with bias) more often. The primary reason for this difference from previous approaches is that the present model considers both the type (word) level and the category (phoneme) level, allowing types of different frequencies to be represented by different numbers of exemplars. Even though high-frequency types are produced (with bias) more often than low-frequency types, they are also represented by more exemplars (Hintzman & Block, 1971). Thus, an isolated production has less influence on the representation of a high-frequency type than on that of a low-frequency type, counterbalancing the difference in rates of production (see also Sóskuthy, 2014). I illustrate the differences between previous usage-based models and the present model schematically in Figure 3.4; for discussion from a mathematical point of view, see Appendix D.

Although the centroids of the sub-distributions of exemplars of high- and low-frequency types changed at the same rate in these simulations, the sub-distributions themselves did not evolve identically; rather, they exhibited differences in changes in width. While the category as a whole maintained its width throughout the simulations, the sub-distribution corresponding to high-frequency types narrowed and the sub-distribution corresponding to low-frequency types widened, as demonstrated in Figure 3.5. This result reflects a difference in sensitivity to the typicality force:

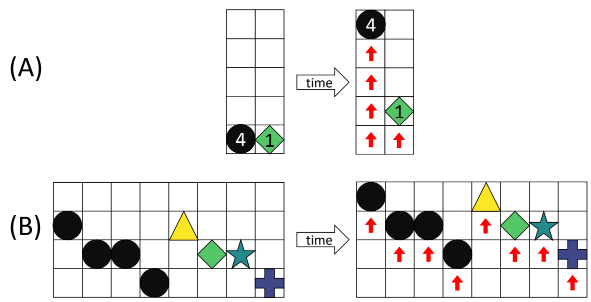

Figure 3.4: The interaction of type frequency and bias. Schematic illustration of underlying intuitions in prominent usage-based models (A; e.g. Pierrehumbert, 2001) and the present model (B). Different types are represented by different shapes and colors, acoustic value is represented by vertical position, and movement due to bias is represented by red arrows (one per production of the type). Each case depicts a high-frequency type (black circles; frequency 4) and some low-frequency types (colored angled shapes; frequency 1). In (A), shapes represent the location of the mean acoustic value for each type, and type frequency is indicated by numbers; in (B), the exemplar distribution of each type is represented by individual exemplars, and type frequency is indicated by number of exemplars. The left panel shows an initial condition, and the right panel shows the expected (average) result after the amount of time it takes to produce the high-frequency type 4 times. (A) Intuition underlying prominent models; comparing a single type of each frequency. In the expected amount of time it takes to produce the high-frequency type 4 times (5 iterations), the low-frequency type is expected to be produced once. Each production of a type adds bias and updates that type's mean. Since the high-frequency type is produced 4 times more often than the low-frequency type, it is subjected to 4 times as much bias. The high-frequency type thus evolves at a faster rate than the low-frequency type. (B) Present model; comparing the aggregate over exemplars of each frequency. In the expected amount of time it takes to produce the high-frequency type 4 times (8 iterations), the four low-frequency types are each expected to be produced once. Each production of a type adds bias to a single exemplar of that type. Since each exemplar of each type is produced once, the bias is distributed over the exemplars. While the high-frequency type is subjected to more bias than any low-frequency type, it distributes this bias over 4 exemplars, which is equivalent to the distribution of bias over 4 exemplars of the 4 different low-frequency types in aggregate. The sub-distributions of high- and low-frequency exemplars thus evolve at the same rate.

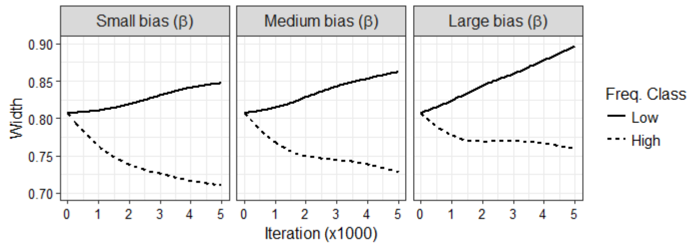

Figure 3.5: Frequency effects in phonetic drift: category width. Results of simulations with a single category ($\sigma = 0.8$) subject to varying degrees of bias, illustrating differences between low-frequency (solid) and high-frequency (dashed) types. For all degrees of bias, the category distribution widens for low-frequency types, but narrows for high-frequency types. Category skewness increases with bias (β), causing apparent category width also to increase.

high-frequency types are more sensitive to it than low-frequency types.

High-frequency types have increased sensitivity to the typicality force – and to perceptual forces in general – as an indirect result of an interaction between assumptions about production and storage. Recall that the selection of an initial target for a token proceeds by copying the acoustic value of an exemplar of the given type (Section 2.3.2.2), and that a token is unlikely to be stored if it falls in a perceptually-disadvantaged (e.g. low-typicality) area of the exemplar space (e.g. Section 2.3.2.8). Consequently, an exemplar in a perceptually-disadvantaged area (e.g. an atypical exemplar) is less likely to generate a token that will overwrite an exemplar in a perceptually-advantaged area (e.g. a typical exemplar) than vice-versa. In this way, the target-copying production mechanism provides escape routes from perceptually-disadvantaged areas of the exemplar space, yielding boosts to perceptual forces. When an exemplar of a high-frequency type falls in a perceptually-disadvantaged area, it is likely to have many escape routes, as there are many other exemplars of the same type that can serve as production targets in perceptually-advantaged areas. The opposite is true for an exemplar of a low-frequency type, as there are few other exemplars of the same type. With more escape routes, a high-frequency type gets more boosts to the perceptual forces.

While this sensitivity difference is interesting, it is a side-effect of model assumptions and should not be interpreted as theoretically meaningful. If anything, it serves to illustrate the importance of being thorough and careful with modeling assumptions, and of not taking results at face value without considering the possibility that they may come from adverse interactions of assumptions. I present the sensitivity difference only in the interest of completeness; none of the key results presented here or elsewhere rely on it in any way.

3.4.2 Two-category interactions

For two-category interactions (modeled as push chains) – where the movement of one category in the exemplar space has implications for another – there are two goals, one for each category. For the Pusher (i.e. the category subject to production bias), the goal is to produce *faster* change in high-frequency types than in low-frequency types, mirroring empirical data from New Zealand English /t/-tapping (Hay & Foulkes, 2016). For the Pushee (i.e. the category retreating from the other due to dispersion), the goal is to produce *slower* change in high-frequency types than in low-frequency types, mirroring empirical data from New Zealand English /ɛ/-raising (Hay et al., 2015). The simulations presented here attempt to meet both goals through the single approach of varying the discriminability threshold, δ, with type frequency.

The results of varying δ with type frequency are shown in Figure 3.6. When average discriminability is sufficiently high and high-frequency types are sufficiently more discriminable than low-frequency types, the goals are met: high-frequency types change faster than low-frequency types in the Pusher and slower than low-frequency types in the Pushee. More generally, as the frequency-based asymmetries in the discriminability evaluation grow, the desired frequency-based asymmetries in rates of change become stronger and more robust, for both categories. I take this result to mean that a listener-based focus containing a single, experimentally-established perceptual bias can explain the different word frequency effects on rates of /t/-tapping (Hay & Foulkes, 2016) and /ɛ/-raising (Hay et al., 2015) in New Zealand English.

The result obtains because a frequency-based discriminability threshold changes the sensitivity of high- and low-frequency types to the category-level discriminability force. As the chosen asymmetry in thresholds grows, the sensitivity shrinks for high-frequency types and grows for low-frequency types. To balance category-level forces, the sub-distribution of high-frequency exemplars is shifted closer to the overlapping region, where the local discriminability force is larger, and vice-versa for the sub-distribution of low-frequency exemplars.[30] The size of the frequency effect on rate of change is thus a function both of the average discriminability, which determines the size of the discriminability force on average, and of the degree of frequency-based asymmetry in discriminability thresholds, which determines the difference in sensitivity to the discriminability force. Since high-frequency types are a priori more sensitive to perceptual forces than low-frequency types, due to the interaction between assumptions about production and storage discussed in Section 3.4.1, reversed frequency effects are observed when the frequency-based asymmetry in discriminability thresholds is small (Figure 3.6). As the threshold asymmetry grows, it first overcomes this prior difference in sensitivity, and then introduces the opposite difference, which generates the desired frequency effects.

[30]The separation of exemplar sub-distributions is insufficient on its own to lead to a category split. All exemplars within a category are subjected to the same typicality force, which squeezes them toward a single mode. Moreover, the model treats frequency as a continuous variable, and a category is made up of sub-distributions spanning the entire frequency range; the discussion of two extremes here is for illustrative purposes only.

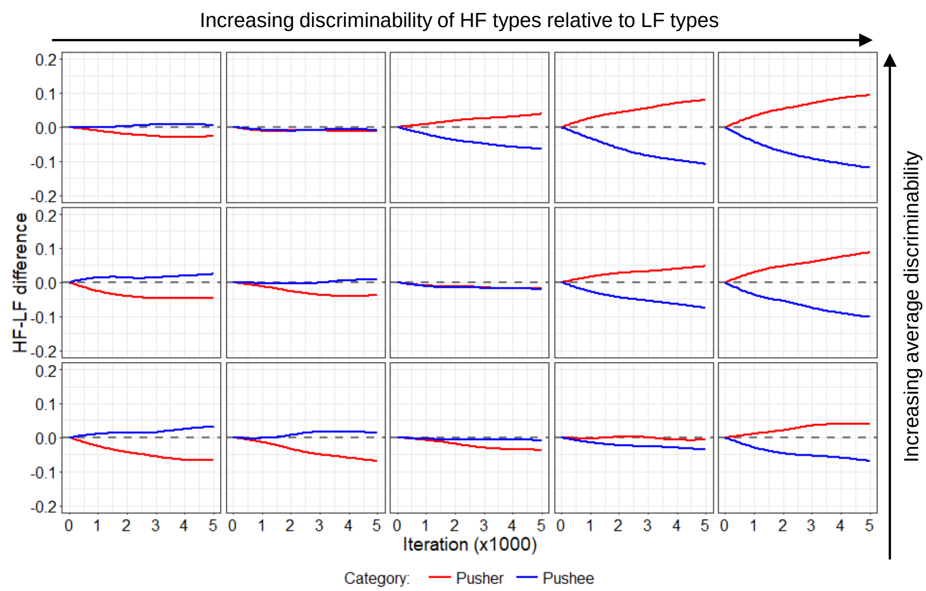

Figure 3.6: Frequency effects in two-category interactions. Results of varying discriminability threshold (δ) with type frequency for representative sets of parameter values (sets (4), (10), and (16) from Table B.2; all other sets give similar results). The vertical axis shows the extent to which high-frequency types are ahead of low-frequency types in the Pusher (red) or Pushee (blue), averaged over 1000 runs for each parameter setting. A positive slope represents a faster rate of change of high-frequency types compared to low-frequency types. All curves end with a horizontal section corresponding to a stable equilibrium. Panels are laid out according to δ function, as in Figure 3.1. Moving rightward across the columns, high-frequency types become increasingly more discriminable than low-frequency types. This shifts the end of the curve upward for the Pusher (red), causing positive-sloping sections where high-frequency types change at a faster rate than low-frequency types, and the reverse for the Pushee (blue). This effect grows more pronounced moving upward across the rows, as discriminability increases on average. When average discriminability is sufficiently high and high-frequency types are sufficiently more discriminable than low-frequency types (i.e. sufficiently close to the upper-right panel), the model generates robust frequency effects resembling those seen empirically. The reverse effects seen close to the lower-left panel result from an interaction between assumptions about production and storage.

3.5 Additional simulations

This dissertation is about the centrality of the listener to language change, and the discussion in this chapter has correspondingly focused on the role of perceptual biases in the generation of word frequency effects on rates of sound change. However, the speaker also plays a role in language change, and has been considered crucial to explanations of word frequency effects in sound change in previous usage-based approaches such as the FAH (Bybee, 2002; Phillips, 1984). In this section, I present two sets of additional simulations designed to show that the speaker is not as important as the listener for generating frequency effects in the present model (i.e. without some form of active strategy).

In these simulations, I consider two interacting categories, and I alter the impact of the speaker by varying the application of production bias. I first present simulations where both categories receive production bias, to demonstrate that the existence of frequency effects is not tied to overall category movement. I then present simulations where neither category receives productions bias, to demonstrate that frequency effects are also not tied to bias and can be obtained even in the absence of speaker influence. In both cases, I continue to refer to the categories as “Pushee” and “Pusher” to allow comparison with the earlier results. These names should only be taken as convenient, not as indicating the kind of movement exhibited by the category or the bias to which it is submitted.

3.5.1 Categories biased together

To demonstrate that the model’s results on frequency effects stem from the internal reorganization of categories due to perceptual asymmetries, rather than from the movement of categories due to the application of production biases, I conducted simulations in which the two categories were biased together and didn’t move. I

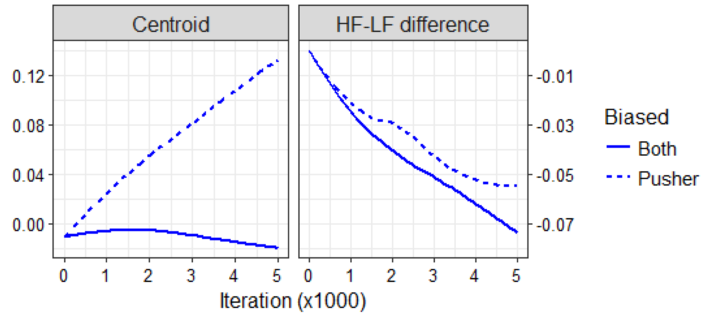

Figure 3.7: Frequency effects without category movement. Results of simulations involving two categories biased together (solid line) or a single Pusher category biased toward the other (dashed line), showing the centroid (left panel) and the frequency effect (right panel) of the Pushee. While the centroid hardly moves in the case with two categories biased together, a robust frequency effect is still observed, which is at least as large as the effect observed when just the Pusher is biased.

subjected the Pusher to a positive bias (half the size of that used in Section 3.4.2) and the Pushee to a negative bias of the same magnitude. I ran 1000 models for 5000 iterations each, using a single set of parameters (parameter set (10) from Table B.2) and a single δ function (in which the discriminability threshold decreased linearly from a value of 1 for the lowest-frequency types to a value of 0 for the highest-frequency types); the results for other parameters and δ functions are similar.

In these simulations, both the Pushee and the Pusher stayed approximately still (with slight movement of the centroids due to reversion to the modes, which were slightly off-centered in the initialization data). The results for the Pushee are shown in Figure 3.7. As can be seen, while biasing the categories together results in almost no overall movement, it still yields a Pushee frequency effect: high-frequency types in the Pushee become peripheral slower than low-frequency types. This frequency effect is at least as large as the effect observed when just the Pusher is biased.

The fact that a frequency effect is observed even without category movement implies that the frequency effects predicted by the model are not dependent on category movement. Rather, frequency effects arise as a result of internal reorganization of categories to balance forces from production and perception, as discussed in Section

3.4.2. The fact that the same kind of frequency effect is obtained under two qualitatively different kinds of production force implies that it is driven by the perceptual forces, i.e. by processes in the listener rather than the speaker.

3.5.2 No bias

Having established that the model's results on frequency effects are independent of category movement due to production bias, I conducted further simulations to demonstrate that they are independent of production bias altogether. To accomplish this, I removed the bias entirely, such that neither category was biased in any way. I ran 1000 models for 5000 iterations each, using the same settings as in Section 3.5.1.

In these simulations, the Pushee and the Pusher were repelled from one another, and gradually drifted apart. This repulsion was caused by the discriminability force, which caused perceptual downweighting of tokens produced near the region of category overlap. Its gradualness was a result of the typicality force, which caused similar downweighting of tokens produced far from the mode of each category. Taken together, these two forces yield the *hyperspace effect* (Johnson, Flemming, & Wright, 1993): from the listener's perspective, the optimal production target for a category is hyperarticulated – i.e. located further away from other categories than the mode – but not so much as to no longer resemble natural speech.

The results for the Pushee are shown in Figure 3.8. As can be seen, the degree of repulsion decreases in the absence of Pusher bias as the two categories separate, yielding less category movement over time. However, the frequency effect appears to be unaffected: high-frequency types change slower than low-frequency types in the Pushee, regardless of whether there is production bias in the system or not. The fact that frequency effects are still observed even when production bias is removed confirms that such results in the model follow from the listener, not the speaker. The role of the speaker is to enable prolonged category interaction, by counteracting

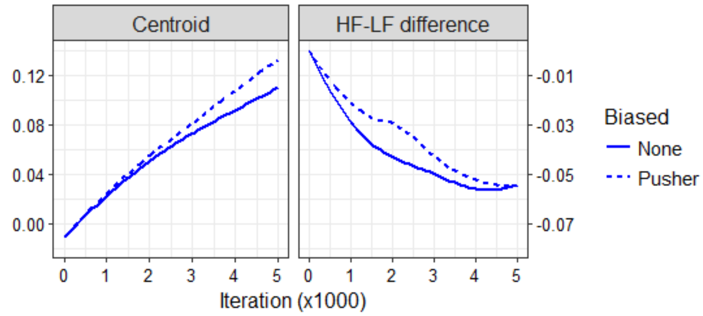

Figure 3.8: Frequency effects without production bias. Results of simulations involving two categories without bias (solid line) or a single Pusher category biased toward the other (dashed line), showing the centroid (left panel) and the frequency effect (right panel) of the Pushee. While the removal of bias causes decreased centroid movement, it does not affect the frequency effect.

the repulsion of the Pusher with production bias – but interaction does not have to be prolonged in this way in order for categories to internally reorganize and display frequency effects.

3.6 Discussion

The simulations in this chapter have demonstrated that perceptual biases likely play a causal role in the generation of word frequency effects in real-life sound changes. I have assumed a single perceptual bias in the form of a frequency-based discriminability threshold, which is experimentally supported (Section 3.2.2). Word frequency effects on sound change that match those seen empirically in two-category interactions are only generated when the bias is included in the model, at sufficient strength; when it is excluded, or set too weakly, the model generates qualitatively different (unattested) effects (Section 3.4.2). Thus, in the model at least, the bias may be said to *cause* the word frequency effects.

The fact that the cause of empirically-established frequency effects in the model is a *perceptual* process lends crucial support to the listener-based approach to language change. The frequency effects arise through the listener, not the speaker; correspondingly, they are not removed when the influence of the speaker is minimized (Section 3.5). In the present model, high-frequency types and low-frequency types all show the same sensitivity to speaker-based forces (bias and imprecision), and there is no impetus for the speaker to produce tokens of high-frequency types in a hypoarticulated manner or tokens of low-frequency types in a hyperarticulated manner (cf. Lindblom, 1990). Rather, tokens of high-frequency types are more robustly recognized than tokens of low-frequency types, leading them to be more likely to be stored when they are in the overlapping region between categories. Consequently, the high-frequency sub-distribution will come to be dominated less by 'clear' exemplars (from outside of the overlapping region) than the low-frequency sub-distribution, and the asymmetry in perception will drive asymmetries in production without the speaker ever needing to make an explicit effort to adjust the clarity of their productions.

The downplaying of the role of the speaker runs counter to the intuition underlying prominent usage-based approaches such as the Frequency Actuation Hypothesis (FAH; Bybee, 2002; Phillips, 1984). In the FAH, the speaker's iteration of a production bias is assumed to cause high-frequency words to change at a faster rate than low-frequency words, as they are produced (with the bias) more often. The simulations in this chapter have demonstrated that this assumption does not hold in a listener-based approach (Section 3.4.1): the fact that high-frequency words are produced very often is counterbalanced by the fact that each stored token of a high-frequency word has very little influence on the word representation, due to its high exemplar density. This demonstration of a *lack* of word frequency effect on rates of sound change – together with the demonstration of a low-frequency advantage in some circumstances (Section 3.4.2) – is a novel contribution to the literature on Exemplar Theory. It undermines criticisms that have been levied against exemplar-based models based on their perceived overprediction of high-frequency advantages in sound change (Abramowicz, 2007; Bermúdez-Otero et al., 2015; Dinkin, 2008; Tamminga, 2014).

In addition to contributing to the theory of language change, the extension of the listener-based model in this chapter also contributes to the empirical study of language change. The extended model was successful in capturing key properties of three known empirical studies of word frequency effects on rates of sound change (Bermúdez-Otero et al., 2015; Hay & Foulkes, 2016; Hay et al., 2015), but its implications extend more broadly. The success of the model follows from the conception of sound change as the result of balancing emergent forces that stem from both the speaker and the listener, where words of different frequencies are crucially assumed to be differentially sensitive to the perceptual forces in the listener. This force-balancing conception of sound change is entirely general, allowing the model to make predictions for word frequency effects in sound changes beyond the three cases examined here. In addition, the formal statement and computational implementation of the model allows it to make such predictions with clarity, avoiding the gaps and uncertainties that can befall informally-stated, unimplemented approaches.

3.6.1 General predictions

Beyond the three sound changes examined in this chapter, the extended listener-based model predicts a (probabilistic) typology of word frequency effects in different kinds of sound change. These predictions all follow from the implications of the change for acoustic ambiguity, given the perceptual bias in the discriminability evaluation that allows high-frequency types to be more robustly recognized than low-frequency types in the face of acoustic ambiguity. In changes that do not affect the acoustic ambiguity of the phoneme undergoing change, the model predicts all words to change at the same rate. In changes that act (locally) to increase the acoustic ambiguity of the phoneme undergoing change (via movement toward another phoneme in the acoustic space), the model predicts high-frequency words to change at a faster rate than low-frequency words. The reverse result is predicted for changes that act to decrease the acoustic ambiguity of the phoneme undergoing change (via movement away from another phoneme in the acoustic space).

By considering the implications for acoustic ambiguity in different kinds of regular sound change, the listener-based model generates the typology of predictions laid out in the first column of Figure 3.9. These predictions are particularly interesting because they fit the existing empirical results more closely than those of other approaches. As previously stated, empirical results in the literature align better with the listener-based model's predictions than with those of the FAH (Bybee, 2002; Phillips, 1984),[31] which draws different frequency effects based on the differentiation between articulatory biases and lexical analogy. Empirical results also align better with the listener-based model than with a plausible alternative inspired by the FAH (described below), which draws different frequency effects based on the differentiation between articulatory biases and dispersion. In what follows, I work through the logic underlying these three approaches to generate predictions for word-frequency effects in various different kinds of sound change, which I compare schematically in Figure 3.9.

As discussed in Section 3.1, the FAH makes a distinction between physiologically motivated changes at the surface phonetic level, which are driven by articulatory biases and assumed to affect high-frequency words fastest, and non-physiologically motivated changes at the level of phonological grammar, which are driven by lexical analogy and assumed to affect low-frequency words fastest. While intuitively attractive, this distinction does not make clear predictions for all kinds of sound change.

[31] It is difficult to make a direct comparison between the predictions of the listener-based model and those of the FAH, since they concern different properties of change (rate versus actuation). To facilitate comparison, I assume neutral initial conditions, i.e. no relevant differences based on word frequency before the onset of the change.

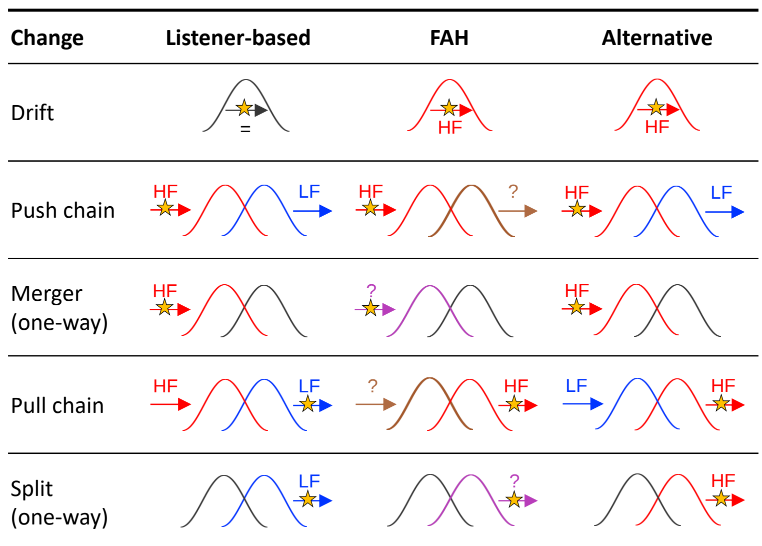

Figure 3.9: General predictions for frequency effects. Comparison of the qualitative predictions of the present listener-based model to those of the Frequency Actuation Hypothesis (FAH; Bybee, 2002; Phillips, 1984) and those of an alternative proposal (see text). Arrows indicate movement over time, and stars indicate movement due to phonetic biases. Red indicates a high-frequency advantage, blue indicates a low-frequency advantage, and black indicates no frequency advantage. In the FAH predictions, not every case is clear-cut (marked by ?; see text), introducing cases where there could be a high-frequency advantage or no frequency advantage (brown, no star) or where there could be a high-frequency advantage or a low-frequency advantage (purple, star). The predictions of the listener-based model and the proposed alternative are identical for push chains and mergers, but different for phonetic drift, pull chains, and splits.

For example, the pushee in a push chain may be argued to move in response to the same articulatory bias that moves the pusher, or in response to dispersive pressures from perception. If the movement is due to articulatory bias, then the FAH predicts a high-frequency advantage, but if it is due to dispersive pressures, then it is unclear what the FAH predicts, as such pressures are based in neither articulation nor analogy. Similarly, sound changes such as merger affect both surface realizations and the phonological grammar and may be driven by articulatory biases or lexical analogy. In such cases, it is unclear whether the FAH predicts a high- or low-frequency advantage.

How can an approach make clearer predictions than the FAH, while maintaining its core claim that high-frequency words change faster than low-frequency words in physiologically motivated changes and slower in non-physiologically motivated changes? The following alternative approach represents one plausible attempt to do so, based on two key assumptions. The first assumption is that a phoneme category may participate in regular sound change only due to articulatory biases or due to dispersive pressures, and that the number of categories subjected to articulatory biases should be minimized. Thus, for example, the pusher in a push chain moves due to articulatory biases, but the pushee moves due to dispersive pressures. The second assumption is that articulatory bias – a physiological motivator of sound change – affects high-frequency words fastest, while dispersive pressures – non-physiological motivators of sound change – affect low-frequency words fastest. The result is an approach that assumes that sound change occurs in response to pressures from both the speaker and the listener, like the listener-based model, but that prioritizes the speaker over the listener, unlike the listener-based model. Because any sort of sound change must be motivated *somehow* – whether by biases in production or by pressures from perception – this approach would predict that every sound change should show some word frequency effect. It would predict high-frequency advantages whenever a phoneme is subject to a production bias, and low-frequency advantages in all other movements.

As shown in Figure 3.9, the alternative approach makes the same predictions as the listener-based model for push chains and mergers, and both are able to account for the observed word frequency effects in /t/-tapping (Hay & Foulkes, 2016) and /ɛ/-raising (Hay et al., 2015) in New Zealand English. I take this observation as support for the general usage-based framework that both approaches take, in which regular sound change occurs in response to forces from both speakers and listeners, due to both articulatory biases and dispersive pressures. However, the predictions differ for phonetic drift, and the empirical observations from /t/-glottaling in Manchester English align with the prediction of the listener-based model. I take this observation as initial support for the listener-based model over the alternative, and correspondingly for treating the listener rather than the speaker as the driver of word frequency effects in regular sound change. Finally, the predictions completely oppose one another for pull chains and splits, but I lack any data at present that can adjudicate between them. Of course, many more empirical results are required to test and compare the predictions in full, given their probabilistic nature (see Section 3.3.3).

3.6.2 Computational modeling and sound change

The computational model extended in this chapter is a significant contribution to the field, both as crucial support for the centrality of the listener to regular sound change, and as an object in its own right. It has allowed me to remove gaps and uncertainties in predictions by ensuring that they are holistic and internally consistent, to hold intuition up to scrutiny by putting it on a formal foundation, and to clarify the causal relations between assumptions and predictions.

Computational modeling ensures that predictions are holistic and internally consistent. In the present model, all sound change is underpinned by movement of exemplars due to forces based in the speaker and the listener, and word frequency effects are driven by listener-based forces, in entirely general ways. Since every kind of sound change can be conceived of in relation to these forces, the model makes clear (probabilistic) predictions for every kind of sound change (modulo the limitations discussed in Section 3.3.3). The same is not true of previous hypotheses that have not been implemented in computational models, such as the FAH (see Section 3.6.1). Because the FAH only addresses sound changes caused by articulatory biases or lexical analogy, it does not extend to changes which do not seem to be caused by either (e.g. the pushee in a push chain), and it makes unclear predictions for changes which could plausibly be caused by both (e.g. merger).

Furthermore, because the present model is a formal implementation, it allows intuition to be held up to to scrutiny. For example, in Section 3.4.1, I showed that the model does not support the widespread intuition that exemplar-based models always predict high-frequency words to change fastest in response to production biases (Abramowicz, 2007; Bermúdez-Otero et al., 2015; Dinkin, 2008; Tamminga, 2014). This intuition relies in part on a conflation of type frequency and category frequency. The present model represents types and categories at separate levels, making it clear that type frequency affects not only rate of production, but also density of exemplar distribution. These two effects counteract one another in determining how quickly the type moves in response to bias (see also Appendix D for discussion from a mathematical point of view).

Finally, the present model has allowed me to clarify the causal relations between assumptions and predictions. By showing that the model predicts different frequency effects with the assumption of a high-frequency discriminability advantage than without it (Section 3.4.2), I have established that word frequency effects on rate of change can be causally related to word frequency effects on perception. The model also allows for investigation of the influence that changing certain assumptions would have on predictions. For example, there are at least two possible mechanisms through which the existence of asymmetries in discriminability evaluation can follow from the exemplar-based architecture of the model (i.e. under a constant discriminability threshold, δ), which make different predictions for asymmetries in typicality evaluation (Section 3.3.2). In the development of ever-more-sophisticated models that approximate the complexities of natural language use and change, it is crucial to understand fully the implications of assumptions, both for the accurate identification of the cause of predicted behaviors, and for the potential unification of different observed behaviors under common assumptions.

The study of regular sound change is becoming increasingly rigorous with the availability of collections of speech recordings that span long periods of (real or apparent) time. The combination of this empirical data with appropriate computational modeling will be central to testing predictions and hypotheses about the connections between speech perception and regular sound change.

3.7 Summary

In this chapter, I have shown how a listener-based approach to language change allows a single perceptual bias to give rise to different effects of word frequency in different kinds of sound change. I have extended the listener-based computational model described in Chapter 2 with an experimentally-established frequency-based asymmetry in discriminability thresholds, and I have shown that this extension allows the model to generate word frequency effects that match all of the empirical results that exist at the time of writing (Bermúdez-Otero et al., 2015; Hay & Foulkes, 2016; Hay et al., 2015). In addition, I have shown that this extension predicts a typology of word frequency effects across many different kinds of sound change, which opens doors to many further empirical studies.

The listener-based approach is both powerful and flexible. The success of the approach in predicting empirically-observed word frequency effects on rates of sound change provides key support to the claim that the listener is central to sound change. I do not intend this claim to imply that the speaker is unimportant, but rather that the speaker cannot be the sole primary influence on sound change. I acknowledge that aspects of production are widely attested and accepted to vary with word frequency, at least in the case of reduction (e.g. Bell et al., 2009). However, given an exemplar-based production-perception loop, it is not necessary to assume that word frequency-based biases in production are responsible for generating effects of word frequency on rate of change, as such effects can follow straightforwardly from experimentally-established, passive but powerful biases in perception.

Chapter 4

Social attitudes in lexical adoption

In Chapter 3, I showed how a listener-based approach to language change predicts asymmetries in the rate at which words of different frequencies are affected by sound change, based on asymmetries in how words of different frequencies are processed in speech perception. But language change is bigger than just sound change, and is affected by more than just frequency. A comprehensive approach to language change must address changes in discrete linguistic elements as well as continuous ones, and must address the large role of social factors in structuring language use and change (Weinreich et al., 1968). In this chapter, I adopt a broader focus on the spread of lexical items across different social groups, and I show how the principles of a listener-based approach extend to situations of this type. My aim is to demonstrate how a listener-based approach provides a parsimonious explanation of effects of social attitudes on language change, through reference to known biases in speech perception, in a parallel fashion to the explanation of word frequency effects in sound change offered in Chapter 3.

This chapter is structured as follows. First, I lay out the general focus of extending the listener-based approach to socially-structured changes in discrete linguistic elements, as well as the specific empirical focus of this chapter on the spread of the word *eh* across ethnic groups in New Zealand (Section 4.1). Next, I ground the extension of the listener-based approach to the realm of social factors with demonstrations of the relevance of social factors to language change, and with psycholinguistic evidence that social information and attitudes affect speech perception (Section 4.2). Using this grounding, I show in general how a listener-based approach predicts socially-based asymmetries in language change (Section 4.3). Then, I zoom in to the specific case of inter-ethnic adoption of *eh* in New Zealand (Section 4.4), for which I outline a corpus study (Section 4.5). I show that the results of the corpus study are consistent with the listener-based approach and support the hypothesis that the adoption of *eh* has increased over time due to improvements of social attitudes (Section 4.6). Finally, I discuss the general advantages of a listener-based approach to social factors in language change, together with the predictions the approach makes for other cases of language change (Section 4.7).

4.1 Focus

The general focus of this chapter is on extending the listener-based approach described in Chapters 2 and 3 in two ways, in order to demonstrate its applicability to language change more generally. Firstly, I extend the application of the approach from sound change, where the linguistic elements in question are continuous (i.e. embedded within a continuous acoustic space), to other kinds of language change that involve the addition, removal, or substitution of a discrete linguistic element, such as lexical adoption. Secondly, I extend the perceptual biases that underpin the approach from language-internal factors, such as word frequency, to language-external factors, such as social information and attitudes. Since all levels of linguistic structure above the phonetic signal are discrete, and since (almost) all language use involves interaction in a social environment, this extension aims to make the listener-based approach much more widely applicable, in a way that doesn't betray its strong basis in empirically-demonstrable perceptual biases (see Section 4.2 for further discussion).

At first glance, this extension seems to take the model in a completely different direction; however, much of the conceptual framework carries over from the study of sound change (Chapter 2) and word frequency effects (Chapter 3), underscoring the potential of a listener-based approach to explain vastly different elements of language change through a single, unified system. For example, variationist sociolinguistics has long associated variation in the use of discrete linguistic elements with probabilistic constraints or *variable rules* (Cedergren & Sankoff, 1974; Labov, 1969). Variable rules allow the use of a particular linguistic element in conveying a particular (social and/or linguistic) meaning to be thought of as the outcome of sampling from a repertoire of elements associated with that meaning (Benor, 2010; Gumperz, 1964), just as the phonetic realization of a phoneme in a particular instance of a word may be thought of as the outcome of sampling from an acoustic distribution. Moreover, variable rules are partly determined by social and stylistic factors, which implies that the linguistic elements to which they refer can be considered to be embedded within a continuous social space, grouped by their meanings, much like the phonetic realizations in Chapter 2 are embedded within a continuous acoustic space, grouped by their phoneme representations. Similarly, a host of experimental studies have found that social information and attitudes affect speech perception in very similar ways to things like word frequency (see Section 4.2.2). Thus, social information and attitudes induce perceptual biases, and these biases plausibly operate in similar ways to those already established in Chapter 3.

To demonstrate the viability of the listener-based approach in the realm of social effects on change in the use of discrete linguistic elements, I focus on the word *eh* (/ɛi/) in New Zealand. *Eh* is a discourse tag particle, typically present only in spoken language; for this reason, I continue to focus on biases induced through speech perception, as in Chapter 3. I focus on *eh* for both empirical and theoretical reasons evoked by previous work (see Section 4.4.1 for in-depth discussion). Empirically, previous work has shown that *eh* was historically used predominantly by Māori, the indigenous people of New Zealand, but is now also used by young Pākehā, the descendants of European settlers. Importantly, however, there remains a strong association between *eh* and stereotypes of Māori speech. Theoretically, the use of *eh* by young Pākehā has been interpreted as consistent with two different hypotheses. The *change-in-progress hypothesis* claims that *eh* is being adopted by all Pākehā, from below the level of consciousness. The *age-grading hypothesis* claims that *eh* is being adopted only by young Pākehā, as a 'young' way of expressing alignment with an interlocutor. Among these two hypotheses, the change-in-progress hypothesis is most consonant with a listener-based approach to language change, as restricting the adoption of *eh* to young Pākehā in the age-grading hypothesis may require appealing to explicit awareness and intentions of the speaker.

Under a listener-based approach, the hypothesis that *eh* has spread from Māori to all Pākehā in recent years has a principled underpinning that does not require any assumptions about Pākehā intention. Recent years have seen notable improvements in the social attitudes of Pākehā toward Māori (see Section 4.4.2). Results in speech perception have shown that social attitudes induce perceptual biases, altering the influence that experiences have on linguistic representations (Section 4.2.2). A listener-based approach predicts that improvements in attitudes toward Māori cause Māori instances of *eh* to have greater influence on the representations that Pākehā draw upon as part of their own linguistic repertoire, thereby increasing the number of Pākehā that have *eh* in their repertoire (see Section 4.3).[32] To test this prediction, I study the use of *eh* by Pākehā in the large, diachronic ONZE Corpus (Gordon, Maclagan, & Hay, 2007). The quantitative results confirm that Pākehā have increasingly added *eh* to their linguistic repertoire over time, consistent with the change-in-progress hypothesis, and qualitative investigation of historical metalinguistic commentaries highlights the role of social attitudes in this change. I argue that the adoption of *eh* by Pākehā was initially blocked by social stigma for Māori, and was only allowed due to recent destigmatization and a shift toward an inclusive attitude that views New Zealand as a bicultural fusion.

What makes *eh* so interesting is that its perceptual-ideological association with Māori has remained even though some Pākehā now use it. The same may not be true of other cases where one social group adopts a lexical item from another; for example, though *cool* was adopted into American English from African American English, it has since been adopted into English more generally (e.g. Reyes, 2005), and its free use by (young) English speakers of various ethnicities all over the world implies that it no longer necessarily carries strong associations with African Americans. It is the maintenance of strong associations between *eh* and Māori, against a backdrop of changing Pākehā attitudes toward Māori, that makes *eh* a particularly appropriate case study for a listener-based approach (see Section 4.3), and one that can illustrate how changes in attitude facilitate changes in language.

[32]Other, speaker-based approaches may share the prediction that Pākehā adopt *eh* as attitudes improve, as a way of signaling an affiliation with Māori. While such signaling behavior *can* and *does* happen, it arguably requires an intention that is unlikely to be explicit in every produced instance of *eh*. As in previous chapters, I focus on the listener-based approach here for reasons of parsimony, as it is relatively more likely that implicit attitude-related perceptual biases are activated with every perceived instance of *eh*.

4.2 Social effects

A great deal of work in the field of sociolinguistics has shown that variation in language use is socially structured. It is therefore no surprise that language change is shaped by social factors and attitudes, just as it is shaped by language-internal factors such as word frequency. And it is also no surprise that listeners are sensitive to social information conveyed through speech, which induces perceptual biases just like language-internal factors. A listener-based approach connects these otherwise independent observations to each other and to language-internal factors, leveraging socially-based perceptual biases to offer a parsimonious explanation of why social effects in language change exist and how they operate.

In this section, I describe insights gained from empirical studies of social effects in language change and in speech perception. I cast a wide net to establish the parallels between social effects and the effects of word frequency discussed in Chapter 3, thereby illustrating the potential for the application of a listener-based approach.

4.2.1 Social effects in language change

A vast literature has demonstrated that language change progresses in a socially-structured manner, with implications for variation in how different people use the same language at any given time (e.g. Labov, 2001, and references therein). Here, I focus on four high-level insights from this literature that are relevant to the present investigation of the role of social attitudes in Pākehā *eh*-adoption: methods for inferring change-in-progress from synchronic linguistic variation; the nature of evidence that can support such inferences; the mechanism by which language change is understood to spread throughout a population; and the way that this spread of language change is mediated by social attitudes.

The first insight to be gained from the study of social effects in language change concerns how change-in-progress can be inferred on the basis of synchronic linguistic variation. The *apparent time construct* holds that differential degrees of use of a particular linguistic feature according to age is a possible indicator of change-in-progress (Labov, 1963). This indication follows from the assumption that, even though speakers may change their linguistic system throughout the lifespan (cf. Harrington, 2006; Harrington et al., 2000; Sankoff & Blondeau, 2007), they exhibit less flexibility in doing so as adults than they do as children (Baxter & Croft, 2016). In most of the cases that have been explicitly tested, the changes inferred through apparent time analysis – i.e. by comparing multiple age-groups a single point in time – have been confirmed through real time analysis – i.e. by comparing a single age-cohort at multiple points in time (Cukor-Avila & Bailey, 2013). Nevertheless, since it remains possible that age-based patterns are stable over time (a situation known as *age-grading*), the apparent time construct is not *diagnostic* of change-in-progress; it is ideally supported by other forms of evidence.

The second insight concerns the nature of these other forms of evidence that can be used to support the claim of change-in-progress. Just as sound change affects some words faster than others (Chapter 3), language change is often observed to be more advanced among some social groups than others. That is, certain social groups are *leaders* of certain changes. Additional evidence for a particular claim of language change-in-progress is therefore provided by consistency between the identity of the leaders of the (apparent) change and the assumed social mechanisms supporting the change. For example, Labov (1972) found that the leaders of /r/-rhoticization in New York City were members (or aspiring members) of the upper middle class, consistent with the assumption that /r/-rhoticization was motivated by concerns of prestige. Similarly, Eckert (1988) found that the backing and lowering of /ʌ/ spread differentially through locally-defined social groups in a Detroit-area high school. It was more advanced among *Burnouts* (rebellious, working-class, oriented toward the metropolitan urban environment) than among *Jocks* (clean-cut, middle-class, oriented toward the local suburban environment), consistent both with the wider patterns of the change and with local ideologies toward it as part of the Northern Cities Shift.

The third insight concerns how language change spreads from the innovators or leaders of the change through an entire community or population. Since adults remain participants in language change throughout the lifetime (Harrington, 2006; Harrington et al., 2000; Sankoff & Blondeau, 2007), and since language change is often socially structured (e.g. Eckert, 1988; Labov, 1972, 2001; Weinreich et al., 1968), it is little surprise that the social network structure of adult linguistic interaction plays an important role in the spread of language change (e.g. Milroy & Milroy, 1985). Within a social network, one of the primary determinants of the spread of linguistic change is frequency of interaction (Labov's (2001) *principle of density*) – who talks to whom, and how often. This statement captures Bloomfield's (1933, p. 326) insight that "every speaker is constantly adapting his speech-habits to those of his interlocutors"; an innovative variant will spread easily to a speaker for whom a large proportion of interactions are with interlocutors who use that variant.[33] Crucially, change is assumed to spread in this way automatically, without intention or even awareness on the part of the speaker (Trudgill, 2008).

The fourth and final insight concerns the mediation of the spread of language change by social attitudes. The spread of language change is sometimes assumed to be 'social' only to the extent that it is governed by social networks, through frequency of interaction (Labov, 2001; Trudgill, 2008). However, this assumption is rejected by approaches that foreground the role of social meaning-making in linguistic variation and change (e.g. Eckert, 2012, and references therein). It has also been shown to be

[33] The reverse implication is also true: an innovative variant will be inhibited from spreading to a speaker for whom a large proportion of interactions are with interlocutors who *do not* use that variant. In this way, a social network can serve to enforce pre-existing norms, in the resistance of a change-in-progress.

insufficient to account for the formulation of New Zealand English, for which simulations have demonstrated that frequency of interaction alone generates rates of change that are slower than those observed empirically (Baxter, Blythe, Croft, & McKane, 2009). Just as the generation of empirically-attested word frequency effects on rates of sound change require the assumption of asymmetric treatment of words (Chapter 3), Baxter et al. (2009) demonstrate that the generation of empirically-attested rates of change across speakers requires the assumption of asymmetric treatment of speakers. A natural source of such asymmetric treatment is provided by social attitudes, i.e. by how someone feels about their place in a social group and about other social groups, which have been shown to be crucial to some instances of language change. For example, Labov (1963) shows that various features of the Martha's Vineyard accent were retained by those who felt positively about the island community and negatively about the mainland community (as represented by tourists), but lost by those who felt the opposite. Similarly, Maegaard, Jensen, Kristiansen, and Jørgensen (2013) show that the spread from Copenhagen of two changes in regional varieties of Dutch was associated with speakers' positive subconscious attitudes toward Copenhagen, and Stausland Johnsen (2015) shows that the spread of features from the upper-class dialect in Oslo to local dialects in South-East Norway is blocked by strong negative attitudes. Thus, social attitudes constitute one of the many forces involved in language change (cf. Chapter 2). Furthermore, since attitudes of individuals and groups are dynamic, the relationship between language and attitudes is bidirectional: changes in either one may trigger changes in the other (e.g. Coupland, 2014; Zhang, 2018).

The insights gained from the study of social effects in language change establish the potential for a listener-based approach, in parallel with the approach to word frequency effects demonstrated in Chapter 3. In the following section, I draw on results concerning social effects in speech perception to further establish this potential, which I apply to puzzles in language change in Section 4.3. The insights described here also establish some practical means through which language change-in-progress can be assessed synchronically, in connection with socially-structured linguistic variation and social attitudes. In Sections 4.5–4.6, I draw on these insights to test the hypothesis that Pākehā have increasingly adopted *eh* over time, in association with changes in attitudes toward Māori.

4.2.2 Social effects in speech perception

When perceiving spoken language, listeners cannot help but infer social information about the speaker alongside information about the linguistic units (phonemes, words, etc.) they are using (Sumner et al., 2014). Listeners are highly adept at drawing such social inferences; though they are often supported by visual and situational context, experimental results have shown that they are also drawn in the absence of context, even with severely impoverished signals. For example, Szakay (2012) used low-pass filtered speech to show that New Zealand listeners are able to tell whether a speaker is Māori or Pākehā just from low-level cues such as rhythm and intonation.

The social information that listeners infer plays a crucial role in inducing perceptual biases. The literature has empirically established numerous cases of these socially-induced perceptual biases, which fall into two main kinds. The first kind of socially-induced perceptual bias, which I will call *expectation bias*, uses the social information inferred about the speaker to adjust the listener's expectations. Expectation bias draws on the listener's experiences with others who are socially similar to the speaker and typically serves to facilitate in-the-moment communication, in the same way as biases based on lexical status and frequency (Chapter 3). The second kind of socially-induced perceptual bias, which I will call *evaluative bias*, uses the inferred social information to evaluate the speaker relative to the listener. Evaluative bias influences the way the listener categorizes and treats the speaker on a longer-term basis, both in terms of their actions in the world and in terms of influences on their linguistic representations.

One of the most common demonstrations of expectation bias involves listeners shifting the perceptual boundary between different phonemes. For example, Strand and Johnson (1996) showed that listeners retract the perceptual boundary between /s/ and /ʃ/ in nonwords when shown a picture of a male face relative to when shown a female face, consistent with the fact that experiences of male speech tend to contain more retracted /s/ than experiences of female speech. This result is parallel to results showing that listeners will adjust their perception of an acoustically ambiguous phone so as to hear a signal as a real word rather than a nonword (Ganong, 1980). Similarly, Hay et al. (2006) showed that New Zealand listeners perceive tokens of the recently-merged NEAR (/iə/) and SQUARE (/ɛə/) vowels more distinctly when looking at a photo of an old or middle-class-looking person than when looking at a photo of a young or working-class-looking person, exploiting the fact that listeners are more likely to have experienced the distinction from older and/or higher-class speakers. This result is parallel to results showing that listeners are biased to perceive a signal that is acoustically ambiguous between two words (like *best* and *pest*) as the higher-frequency word (Connine et al., 1993; de Marneffe et al., 2011).

Expectation bias also extends beyond adjustments of a single phoneme boundary. For example, McGowan (2015) showed that listeners can transcribe Chinese-accented speech in noise more accurately when shown a photo of an Asian person than when shown a photo of a White person, parallel to results showing that listeners favor high-frequency words as interpretations of degraded signals (Howes, 1957; Savin, 1963) or of speech from a dialect other than their own (Clopper et al., 2010). As well as making speech perception more robust to confusion, expectation bias can make it faster. For example, Staum Casasanto (2008) showed that Californian listeners are faster to identify a token such as [mæs] as having a deleted *t/d* (corresponding in this case to *mast*) when shown a photo of a Black man than when shown a photo of a White man, consistent with the higher rate of *t/d*-deletion in African American Vernacular English than in General American English. This result is parallel to results showing that listeners classify high-frequency words as real words faster than low-frequency words (Forster & Chambers, 1973).

The studies described above use visual cues to generate expectation bias, but expectation bias does not only come from vision. For example, it is possible to induce expectation bias by explicitly telling the listener about the speaker (Niedzielski, 1999), or even by situationally evoking social information (Hay & Drager, 2010; Hay, Nolan, & Drager, 2006). Expectation bias can also emerge from cues intrinsic to the speech signal itself (e.g. Szakay, Babel, & King, 2016). Thus, expectation bias is extremely powerful and highly prevalent; strong biases are induced in listeners very readily.

Evaluative bias is induced in listeners just as readily as expectation bias, sometimes from extremely small features of the linguistic signal. For example, Campbell-Kibler (2007) found that speakers are perceived as country-oriented and unintelligent when using the alveolar variant *-in* of (ING) (e.g. *walkin'* for *walking*) and as city-oriented and less masculine when using the velar variant *-ing*. In this way, evaluative bias is a means through which listeners build a comprehensive social image of the speaker, evaluating them in a much more nuanced fashion than is permitted by inferred sociodemographic information.

In mediating the construction of a social image of a speaker, evaluative bias often evokes an attitude toward the speaker in the listener. In this way, evaluative bias can affect the process through which a listener decides which action to take in response to the speaker. For example, Purnell, Idsardi, and Baugh (1999) found that the exact same speaker was less successful in telephone applications for housing when using African American Vernacular English or Chicano English than when using General American English. That is, listeners inferred the speaker's ethnicity, and inferences of non-white ethnicity evoked social attitudes that biased them (consciously or unconsciously) toward deciding not to rent to him.

Evaluative biases also affect linguistic decision-making, as can be seen through the mediation of phonetic accommodation by attitudes, both implicit and explicit. A host of research under the umbrella of Communication Accommodation Theory (Giles, Coupland, & Coupland, 1991) has shown that listeners are more likely to imitate phonetic features of speakers they feel positively inclined toward than speakers they feel negatively inclined toward. For example, Babel (2010) recorded New Zealanders' speech before and after listening to an Australian speaker, and found that those with implicit pro-Australian attitudes adjusted their vowel pronunciations in the direction of the Australian speaker more than those with implicit anti-Australian attitudes. Similarly, Yu, Abrego-Collier, and Sonderegger (2013) recorded subjects before and after listening to a story containing artificially-lengthened voice onset time (VOT) of voiceless stops, and found that those with explicit positive attitudes toward the narrator lengthened their own VOT the most. Interestingly, not only did attitude predict the degree of VOT-imitation, but it was also the strongest subject-based factor to do so, indicating the immense power of evaluative biases.

The effect of evaluative bias on linguistic decision-making is mostly assumed to be unconscious (Giles et al., 1991), through the mediation of the activation and/or updating of linguistic representations. However, it is possible that listeners could be consciously choosing to imitate phonetic features of speakers they feel positively inclined toward. To demonstrate that evaluative bias affects linguistic representations, it is necessary to consider its effects on perceptual tasks.

In perceptual tasks, negative attitudes yield evaluative bias through inciting perceptual othering. This evaluative bias mediates the process by which a token activates existing linguistic representations. For example, Nguyen, Shaw, Tyler, Pinkus, and Best (2015) found that (non-Asian) listeners with negative attitudes toward Asians are worse at categorizing Vietnamese English-accented vowels than listeners with positive attitudes toward Asians. The evaluative bias also mediates the process by which a token is stored in memory and thus updates existing linguistic representations. For example, Sumner and Kataoka (2013) found across semantic priming and false-memory tasks that General American English listeners have weaker encoding in memory of words spoken by a New York City English speaker (NYC) than of words spoken by a General American English (GA) or British English (BE) speaker. These results arguably reflect evaluative biases, since similar listeners generally have more negative attitudes toward NYC (as a low-prestige accent) than they do toward GA and BE. In general, the perceptual othering incited through negative attitudes can weaken both the resonance of a new linguistic experience with existing representations and the influence that such an experience has in updating those representations.

Evaluative bias can also come from perceptual othering without explicitly going through the route of attitudes, based on the extent to which a listener perceives the speaker to be similar to themselves. Bestelmeyer, Belin, and Ladd (2015) present fMRI evidence that the perception of similarity underlying evaluative bias is based on social relevance rather than on (asocial) familiarity. Thus, listeners compare the social information they infer about speakers to their own social identity, which generates negative evaluative bias in the case of a mismatch. The effect of this bias can be observed in both phonetic accommodation and in perceptual tasks.

Listeners appear to phonetically accommodate more with speakers that they sound similar to than with speakers that they sound different from (in a socially relevant way). For example, Kim, Horton, and Bradlow (2011) showed that American English and Korean listeners accommodate more with speakers of the same regional dialect of their language than with speakers of a different dialect or a different language. Similarly, there are indications that listeners accommodate more with speakers of the same sex than with members of the opposite sex (Babel, 2012; Babel & Bulatov, 2012; Babel, McAuliffe, & Haber, 2013), though such indications are not always present and are rarely explicitly tested (Pardo, Urmanche, Wilman, & Wiener, 2017).

Listeners also appear to encode spoken utterances in memory better – i.e. with greater strength or fidelity – if they come from a speaker that they sound similar to. For example, Sumner and Samuel (2009) found that listeners who grew up in the New York City area and speak non-rhotic NYC English exhibit long-term repetition priming of non-rhotic variants from a NYC speaker, but listeners who grew up in the area and speak a rhotic (non-NYC) variety do not. Similarly, Perrachione, Chiao, and Wong (2010) found that Black listeners were better at individuating and remembering the names of speakers who were perceived as sounding Black than they were at individuating and remembering the names of speakers who were perceived as sounding White, and vice versa for White listeners.

The literature reviewed in this section has made it clear that social information and attitudes play a large role in speech perception, and, by extension, in language perception more generally. Listeners are adept at inferring social information about speakers, and this information is readily used by the perceptual system. The resultant perceptual biases create asymmetries in the way that listeners represent and respond to language. In Section 4.3, I show how these perceptual biases give rise to socially-based asymmetries in language change under a listener-based approach.

4.3 A listener-based approach

As discussed in Section 4.2.1, language change often displays social effects. A listener-based approach predicts that such social effects on language change may be explained by socially-based asymmetries in the way that listeners experience, perceive and respond to language. In this section, I sketch[34] some of the ways in which a listener-based approach accounts for social effects on language change. In keeping with the broader focus of this chapter, I show how the listener-based approach addresses three general puzzles that arise in consideration of social effects on language change: the maintenance of leaders of change; the progression of change in spite of increasing variation; and the influence of social attitudes on change. Then, in keeping with the empirical focus of this chapter, I develop the account of the influence of social attitudes with respect to interethnic lexical adoption of *eh* in New Zealand.

In many language changes, one social group may lead all others in experiencing the change, for all times while the change is in progress. The basis of a listener-based approach in the listener's experience provides a possible explanation for why such leaders are maintained once a change begins, instead of the rest of the population catching up and experiencing the change in parallel. If the use of a linguistic feature is changing, with the change most advanced among members of a particular group, then not only will members of that group use the feature more than others as *speakers*, but they will also experience the feature more as *listeners*, since within-group interaction is more common than cross-group interaction. For example, the merger between the NEAR (/iə/) and SQUARE (/ɛə/) vowels in New Zealand English was led by working-class speakers, meaning that working-class listeners heard near-merged or fully-merged tokens more often than middle-class listeners simply because they interacted with working-class speakers more (see discussion in Hay et al., 2006). In this situation, though middle-class speaker-listeners did begin to merge as the merger became more prevalent, they did not have the chance to catch up in the merger because they were not exposed to merged speech as often as working-class listeners.

[34] I leave aside the formalization of any of these accounts in the modeling framework presented in Chapter 2, but note that a route for extending the model to incorporate social information is provided by Johnson (2006).

A listener-based approach explains the maintenance of leaders of change as a simple consequence of the fact that linguistic knowledge is built from experience, which varies across social groups.

The fact that different social groups may experience language change at different times or rates implies that linguistic variation may increase during intermediate periods of change. Conceivably, this increase could create problems for communication; in cases like the NEAR–SQUARE merger, listeners who maintain a distinction between the vowels may not readily be able to understand speakers who do not. In the extreme, a communicative breakdown of this sort could lead different groups of speakers to dissociate, so that the change never progresses through the entire community. However, increasing variation with change typically does not create communicative breakdowns, and a listener-based approach can explain why not. Just like perceptual biases based on lexical frequency make it possible for listeners to understand acoustically ambiguous words (Chapter 3), socially-induced expectation biases make it possible for listeners to understand – and thus store in memory and eventually reproduce – the speech of those from other social groups (Section 4.2.2). For example, though at one point middle-class speakers of New Zealand English were unlikely to merge NEAR and SQUARE, they could use various cues to infer that a merged speaker was working-class, and then alter their perception to make sense of merged tokens. Over time, their stored experiences with merged tokens would grow, making them increasingly likely to produce merged tokens themselves and thus follow in the change. In this way, a listener-based approach recruits socially-based expectation biases to ensure that changes progress in spite of increasing variation.

Of course, listeners do not simply imitate all they hear. Sometimes, a listener may consciously choose *not* to imitate a linguistic feature from a different social group, because of a negative attitude toward members of that group. In this way, social attitudes can bias the decisions underlying linguistic action, just as they can bias the decisions underlying non-linguistic action (Purnell et al., 1999). But the use of linguistic features is often unconscious, and thus the imitation of linguistic features is presumably also often unconscious. Even in these unconscious cases, imitation is affected by social attitudes, through evaluative biases (Giles et al., 1991, and other research under the umbrella of Communication Accommodation Theory). Since a listener-based approach assumes that language changes through a chain of listeners (approximately) imitating speakers, the fact that imitation is mediated by social attitudes leads to the prediction that language change is also mediated by social attitudes.

In general, a listener-based approach predicts that the spread of a linguistic feature from one social group to another will be facilitated if the adopting social group hold positive attitudes toward the other group and inhibited if they hold negative attitudes (at least as long as the feature maintains a perceived association with the other group). Crucially, since a listener-based approach holds that the spread of a feature draws upon evaluative bias in the listener, it does not require speakers to actively try to imitate or avoid the feature, just as perceptual biases based on lexical frequency can cause high- and low-frequency words to change at different rates without requiring speakers to try to speak clearly (Chapter 3).

When social attitudes change over time, as do Pākehā attitudes toward Māori, the corresponding change in evaluative bias is predicted to cause changes in language use. In the framework of the model described in Chapter 2, changes in evaluative bias can be understood as akin to changes in typicality threshold.

For concrete illustration, consider Figure 4.1, in which there is a category-level representation corresponding to discourse tags, containing type-level representations[35]

[35] As in the model from Chapter 2, types within a category remain functionally distinct; thus, while *eh* and *you know* are used similarly, and access similar representations, they need not be considered alternants in direct competition with one another.

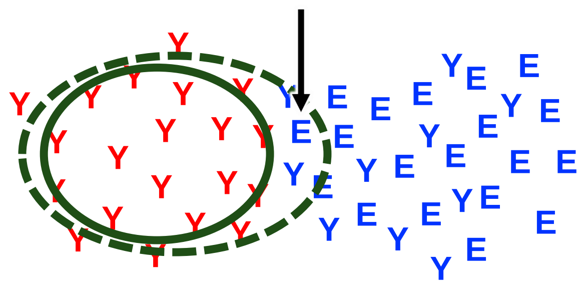

Figure 4.1: Illustration of a listener-based approach to *eh*-adoption. Tokens of the types corresponding to *eh* (E) and *you know* (Y) are located in a socio-perceptual space. Each token carries information about whether the speaker who produced it was (perceived as) Pākehā (red) or Māori (blue). For Pākehā with a strong negative attitude toward Māori, the high typicality threshold causes the boundaries of the discourse tag category representation (solid green ellipse) to exclude all tokens of *eh* as atypical. Consequently, experienced tokens of *eh* are blocked from affecting production. As attitudes improve and become more inclusive, the typicality threshold lowers and the boundaries expand along the ethnicity dimension (dashed green ellipse), allowing some tokens of *eh* to be perceived as typical (black arrow). When the first token of *eh* is perceived as typical, the exemplar that it deposits in the space becomes part of the representation that the Pākehā speaker draws upon for production, thereby adding *eh* to their linguistic repertoire. It may also trigger a knock-on effect, expanding the boundaries of typicality even further to include more tokens of *eh* and thereby solidify its place in the repertoire.

for different tags such as *eh* and *you know*.[36] The exemplars for these types are located within a space partly defined by social factors; for concreteness, assume that one dimension corresponds to the perceived ethnicity of the speaker. At an early point in time, tokens of *eh* are almost exclusively concentrated toward the Māori end of the ethnicity dimension, while tokens of *you know* are spread all along the dimension, across both ethnicities. The category-level discourse tag representation that a Pākehā speaker draws upon for production is centered over the Pākehā end of the dimension and thus excludes *eh*. Pākehā can recognize *eh* as belonging to the category, but perceive it as atypical because of its association with Māori.

[36]Acknowledging the fact that listeners may *adopt* a type from another group necessitates a further extension of the model from Chapter 2, whereby type frequencies can change over time. Since I am not focused on constructing a model of *eh*-adoption in this chapter, I assume such an extension without developing the details of how it might work.

In this situation, strong negative attitudes will behave like a high typicality threshold, making it highly unlikely that tokens of *eh* will be stored in the category-level discourse tag representation. Effectively, a negative attitude increases the distance between the Māori speaker's token and the Pākehā listener's representation, giving the former little influence in updating the latter. However, as attitudes become more positive, losing stigma and shifting toward a sense of bicultural fusion, Māori will be perceived as less distant from Pākehā. Correspondingly, the typicality threshold will lower and atypical instances of *eh* from Māori will begin to be stored in the representation that the Pākehā listener draws upon for their own production.

Once the perceptual barrier of the negative attitude has been removed and the Pākehā listener has added weak traces of *eh* to their representation, successive instances of *eh* will have more influence, as they have a foothold that further increases their typicality. Thus, an individual is expected to reach full fluency with *eh* quickly after adopting it, meaning that the change will be more strongly reflected in *who* uses *eh* rather than in *how often* they use it. In fact, given that the discourse licensing of *eh* is tied to interpersonal and situational factors (see Section 4.4.1), its quantity of use by an individual is expected to be highly variable, and thus may not strongly reflect *any* patterns of change. While distinguishing *who* and *how often* would be prudent for any analysis of language change, the parallel distinction between interactions with representations in perception and production offered by a listener-based approach gives this distinction new salience, as it motivates different predictions for each question.

As demonstrated in this section, a listener-based approach provides solutions to puzzles that emerge from considering social effects on language change. One of the core advantages of a listener-based approach is that it predicts that effects of social information and attitudes ought to be the default in language change. Listeners experience language in a social environment and readily infer social information and evoke social attitudes in this environment, which give rise to expectation and evaluative biases. These perceptual biases affect listeners' productions and perceptions without the need for effort or even conscious awareness on their part, enabling and maintaining social effects on language change over time.

4.4 *Eh* in New Zealand

The first part of this chapter has established that a listener-based approach predicts social attitudes to play a role in language change. In the second part of the chapter, I zoom in to focus on a specific case of this prediction, concerning the use of *eh* by Pākehā in New Zealand. In this section, I fill in the gaps that I have left in earlier discussion of this case, about past studies of *eh*, their conclusions, and changes in the attitudinal situation in New Zealand. In doing so, I lay the groundwork for a corpus-based study that tests whether Pākehā have increasingly adopted *eh* over time, as expected from the improvement of social attitudes under a listener-based approach.

4.4.1 Background

Eh is a tag particle, meaning it is constrained syntactically to be used at the end of a clause, but has no semantic constraints and adds no at-issue semantic content to the proposition expressed in that clause. Rather, it adds pragmatic content, acting as a positive politeness marker (Brown & Levinson, 1987). Its functions are numerous and nuanced, but it most commonly functions to establish or signal common ground and to facilitate the listener's continued involvement in a narrative (Meyerhoff, 1992), and it is most commonly used in evaluative contexts, such as after an opinion or in emphatic situations (Columbus, 2010). In this way, *eh* often emphasizes a speaker's values, beliefs, and opinions in a way that invites or expects the listener to share them. For example, its use in the utterance "I couldn't imagine living anywhere else eh" (excerpt from ONZE) serves to help the speaker convey her appreciation of her home and reflect her belief that everyone should appreciate his or her own home in this way. Since *eh* promotes alignment of interlocutors with respect to values, beliefs, and opinions, it is argued to be used most often when such alignment is a priori most likely, in cases where the perceived "likeness" of the interlocutors is high (Meyerhoff, 1994).

In popular perception, *eh* appears to be more strongly associated with the indigenous Māori than it is with White Pākehā, leading Meyerhoff (1994, p. 375) to label it an "in-group marker for Māori speakers of New Zealand English". However, in spite of this strong association with Māori, *eh* is commonly used by young speakers of both ethnicities. For example, O'Flaherty (2015) claims in *The Guardian* that it is "tacked on to the end of every spoken sentence", and Meyerhoff (1994, p. 367) writes that "few features of New Zealand English are more eagerly recognized by New Zealanders as a marker of their identity" than it. These observations raise the question: if *eh* is strongly associated with Māori – who, historically, have been socially stigmatized – then why do young Pākehā use it?

Two hypotheses for the use of *eh* among young Pākehā have been put forth in the literature: *age-grading* and *change-in-progress*. The age-grading hypothesis, suggested by Meyerhoff (1994), holds that young Pākehā use *eh* simply because they are young. Age-grading could be reflective of properties of the community – with Pākehā actively suppressing the use of *eh* as they age – or of the data sample – with young Pākehā using *eh* to emphasize the values, beliefs, and opinions they share with young interviewers by virtue of their closeness in age. The change-in-progress hypothesis, defended by Stubbe and Holmes (1995), holds that young Pākehā use *eh* because they are leading a change across *all* Pākehā, in which they adopt *eh* from Māori. Because both hypotheses predict the same effects of age on Pākehā *eh*-usage, the literature adduces quantitative arguments for one hypothesis over the other by considering how

Pākehā *eh*-usage differs according to factors other than age.

Meyerhoff (1994) considers how patterns of *eh*-usage among Pākehā relate to patterns of *eh*-usage among Māori. Her data are taken from sociolinguistic interviews with working-class New Zealanders living in Porirua City between 1989 and 1990, where the interviewer matched the interviewee in terms of ethnicity and sex, but not age. Overall, she finds that Māori use *eh* more than Pākehā on average, consistent with the change-in-progress hypothesis. However, she also finds that Māori as well as Pākehā exhibit less *eh*-usage by old speakers than by young speakers, consistent with the age-grading hypothesis. Furthermore, she finds a reversal of sex-based patterns across ethnic groups: Māori men use *eh* more than Māori women, but Pākehā men use *eh* less than Pākehā women. She argues that this reversal at least complicates the change-in-progress hypothesis, since it shows that *eh* is not spreading even-handedly from Māori to Pākehā. Observing that the most prolific young female Pākehā *eh*-users lived with or were married to Māori (or Pacific Island) men, she suggests that the uptake of *eh* by Pākehā may be influenced by degree of exposure to, or sense of affiliation with, Māori. Finally, among young female speakers, she also finds a reversal of the predominant ethnicity-based pattern: though Māori in general use *eh* more than Pākehā in general, young Māori women use it less than young Pākehā women. She argues that this reversal doesn't make sense under the basic change-in-progress hypothesis, because Pākehā would not be expected to overtake Māori in light of the strong popular associations between *eh* and ethnicity, and because there is no principled reason why such overtaking should be conditioned on sex. Given this argument, she suggests that there is little evidence for interpreting age-based differences in *eh*-usage among Pākehā as supporting the change-in-progress hypothesis over the age-grading hypothesis. However, she notes that her results may be affected by the inclusion of speakers who use *eh* to an extreme degree and thus are not representative of the general population. Representativeness is a more general issue of Meyerhoff's (1994) study, as it is based on relatively few speakers (5 per age $\times$ ethnicity $\times$ sex cell).

Stubbe and Holmes (1995) investigate how *eh*-usage by Pākehā differs according to the interaction of age with class and sex. Their data are taken from conversations between friends (matched for age, sex, and class) living in Wellington City (the main metropolitan center of which Porirua City is a constituent) between 1988 and 1994. They find that young speakers use *eh* more than middle-aged speakers, and since their speakers are all talking to a friend of the same age, they argue that this pattern contradicts the age-grading hypothesis and supports the change-in-progress hypothesis. They adduce further support for the change-in-progress hypothesis from the fact that *eh* is used more by working-class speakers than by middle-class speakers, and, among working-class speakers, more by men than by women. This pattern echoes that of other well-studied discourse variables in New Zealand English which are widely accepted to be undergoing *change from below* among Pākehā, whereby they are increasingly being used subconsciously to build solidarity. Because the use of *eh* is subconscious, it can occur in spite of any negative association with Māori. Finally, they note that young working-class *men* use *eh* approximately ten times more often than young working-class women, which stands in strong opposition to the effects identified by Meyerhoff (1994) that presented complications for the change-in-progress hypothesis. Free of these complications, they argue that there are no barriers to rejecting the age-grading hypothesis in favor of the change-in-progress hypothesis. Stubbe and Holmes (1995) employ statistical testing to support their results, but may still be influenced by issues of representativeness, as their data also include relatively few speakers (6 per age $\times$ class $\times$ sex cell, with young speakers only among the working class).

4.4.2 Reasons for change: attitudes

In her argument for why high *eh*-usage among young Pākehā may *not* reflect change-in-progress, Meyerhoff (1994, p. 375) makes the point that use of *eh* is likely to evoke associations with Māori, with negative consequences:

> We would not expect [young Pākehā women] to accommodate to the speech norms of Māori men when talking to another young Pākehā woman. Identifying oneself with Māori speakers of English outside the Māori community is not without cost for Pākehās [*sic*]. Instead of being associated with the ethnic group which (in the wider NZ community) has most power and status, it associates the speaker with an out-of-power and less statusful social group.

If the change-in-progress hypothesis is correct, then it must answer the question of why Pākehā would adopt a feature from a stigmatized group. In other words, what triggered the change? A listener-based approach suggests that the answer may lie in changes in Pākehā attitudes toward Māori.

There is good independent evidence to show both that Māori have historically been treated as inferior in New Zealand and that that treatment has recently begun improving. The Treaty of Waitangi, the document which brokered the colonization of New Zealand in 1840 in exchange for protection of Māori people and assets, was flagrantly abused shortly after being signed, with Pākehā confiscating Māori land through war. Pākehā held linguistic biases against both the Māori language and Māori features of English. For example, children were routinely disciplined for speaking Te Reo Māori at school, and controversy arose when telephone operator Naida Glavish was nearly fired in 1984 for answering the toll switchboard with the Māori greeting *kia ora* (Boshier, 2015). In recent years, however, Māori have become more respected and accepted by the New Zealand government, which has reached major settlements with tribes over injustices under the Treaty, established devoted Māori seats in parliament, and launched a series of initiatives to revitalize Māori culture and language. For example, the Māori Language Act of 1987 gave Te Reo Māori official and protected status in New Zealand. Follow-up initiatives have seen the entrance of Te Reo Māori in the public school curriculum as something of which "all students have the opportunity to acquire knowledge" (Ministry of Education, 2007, p. 9), as well as the adoption of Māori greetings and vocabulary on all news broadcasts on public television. These recent actions establish direct contrasts with earlier discriminatory behaviors.

The shift in treatment of Māori is not just behavioral, at the institutional level, but also attitudinal, at the individual level. Pākehā now seem to value the place of Māori in New Zealand society; for example, Holmes (2009) reports studies from the early 1990s showing that most New Zealanders support the Māori language, and Hashimoto (2019) reports similar results from the late 2010s. Similarly, Sibley and Liu (2007) present experimental evidence showing that Pākehā endorse a bicultural view of New Zealand both implicitly and explicitly. Such attitudinal changes seem to have redefined what it means to be a New Zealander; Liu, Wilson, McClure, and Higgins (1999) show that Pākehā university students ascribe the label "European" to their ancestors who committed injustices against the Māori, but do not choose the same label for themselves, and Cormack and Robson (2010) present census data showing the number of New Zealanders identifying ethnically as "New Zealander" (as opposed to the racially-exclusive "New Zealand European", or similar) has risen from 0.6% of the population in 1986 to 11.1% of the population in 2006.

Thus, a change in cultural attitudes toward Māori and Māori forms of language appears to have been occurring for the past 30 years or so. Māori are increasingly identified as part of the same group as Pākehā, weakening the stigma that enforced separationist attitudes previously. This attitudinal change provides a plausible motivation for the increasing adoption of *eh* by Pākehā that is proposed by the change-in-progress hypothesis, and a listener-based approach provides a parsimonious mechanism through which this motivation can be realized without the intention or even awareness of speakers.

Of course, the motivation of language change by changes in Pākehā attitudes toward Māori is not restricted to *eh*. The attitudes in question – and the perceptual biases they engender – are based on Māori *speakers*, not (just) the linguistic features that they use. Consequently, an explanatory focus on attitudes provides a way to connect the hypothesized change-in-progress in *eh* with the adoption of other features of Māori English, including phonological features (e.g. final /z/-devoicing; Holmes, 1996), prosodic features (e.g. syllable-timed rhythm; Nokes & Hay, 2012; Szakay, 2006), and lexical borrowings (e.g. Macalister, 2006).[37] Indeed, previous studies have highlighted the role of attitudes in the adoption of multiple features of Māori English, as Pākehā who report positive attitudes toward Māori language and culture are more likely to use a Māori word in an English sentence (Thompson, 1990) and to faithfully realize /r/ as [ɾ] in Māori loanwords (Hashimoto, 2019). The present study represents an attempt to link the hypothesized adoption of *eh* by Pākehā with the adoption of these other features of Māori English, and to provide a framework in which it can be understood *how* and *why* changes in attitudes give rise to such adoptions.

[37] I do not mean to suggest that improvements in Pākehā attitudes will necessarily lead to the wholesale adoption of *all* features of Māori English. The linguistic influences are not unidirectional: over time, the English spoken by Pākehā has also influenced Māori English, as well as the Māori language (see e.g. Maclagan, Watson, Harlow, King, & Keegan, 2009). More importantly, the influences are not absolute: even as attitudes change, it remains the case that Pākehā and Māori are different – with different social networks and language exposure profiles, different cultural practices, and a different sense of self – and these differences will continue to be reflected both perceptually and in terms of linguistic behavior. See Section 4.7 for further discussion.

4.5 A corpus study

Since the results in the literature on *eh* in New Zealand (Section 4.4.1) disagree with one another on the socio-demographic patterns of *eh*-usage among Pākehā, and thus on whether high *eh*-usage by young Pākehā reflects age-grading or change-in-progress, it is clear that more data are needed. In particular, there is a need for data from many more speakers, to counteract issues of representativeness. In this section, I introduce data drawn from a much larger (and hence, presumably, much more representative) set of interviews than previously investigated (Section 4.5.1). I apply rigorous quantitative methods to these data (Section 4.5.2), to test precise predictions of each hypothesis (Section 4.5.3).

4.5.1 Data

The data used for this study come from the ONZE Corpus (Gordon et al., 2007), which is composed of 3 sub-corpora of interviews collected at different times with participants whose birthdates span almost the entire lifetime of New Zealand English. The main data I report on come from the contemporary Canterbury Corpus, which features informal one-on-one sociolinguistic interviews with speakers born between 1926 and 1987, conducted between 1994 and 2007 by university undergraduate students that usually knew the speaker well. No controls were placed on the socio-demographic categorization of the interviewers, for example by matching their sex with that of the speaker. The speakers in the corpus are balanced with respect to age (*young*, born after 1960 and aged approximately 20–30 at time of interview, and *old*, born prior to 1960 and aged approximately 45–60 at time of interview), sex (*male* and *female*), and class (*professional*, scoring highly on both socio-economic scale and level of education, and *nonprofessional*, scoring lowly on these scales). In the dataset, there are 394 speakers in total, with approximately 50 speakers from each of the 8 cells defined by intersecting the three socio-demographic categories (range 44–53).

It is not recorded in the corpus whether speakers identify as Māori; however, since most speakers were born in the Canterbury region, where only 7% of the population identified as Māori by the end of the interview collection period (Statistics New Zealand, 2006b), it is likely that most are Pākehā. To coarsely guard against the possibility that the results would contain Māori uses of *eh*, I searched the corpus for instances of the word Māori, under the assumption that Māori speakers would talk about their ethnicity at some point. I found a few speakers who actively identified as Māori, but none of them used *eh*. This may be because they are not part of a Māori community, in which case there would be little expectation that they use Māori in-group markers. Comments from one speaker confirmed the relevance of this factor, indicating that, though she was Māori by heritage, she was disconnected from the Māori community and thus from the culture. Given this disconnection, I kept speakers that were Māori by heritage but not by culture in the dataset; the results remain qualitatively unchanged if they are removed.

I extracted all instances of *eh* from the Canterbury Corpus, excluding those that were clearly requests for clarification, meta-linguistic commentaries, or fillers (with the same orthographic transcription but different phonetic realization, i.e. [ɛ] or [ʌ]). Cases for which an inclusion decision could not confidently be made on the basis of the transcript were listened to for clarification and excluded if they contained rising intonation, which signals a factual verification-seeking rather than positive politeness function (Meyerhoff, 1994).

For each speaker, I recorded their age, sex and class, the number of times they used *eh* in the interview, and the total number of words they uttered in the interview. From this information, I calculated a variant of the *eh-index* proposed by Meyerhoff (1994), which reflects the expected number of occurrences of *eh* in approximately

100 minutes of interacting with a speaker or group, on average.[38] The formula for calculating this index is given in Equation (4.1).

$$eh\text{-index for group } G = 12000 \times \frac{\sum_{s \in G} \text{uses of } eh \text{ by speaker } s}{\sum_{s \in G} \text{number of words uttered by speaker } s} \quad (4.1)$$

4.5.2 Methods

To assess the two hypotheses about *eh*-usage among Pākehā – age-grading and change-in-progress – it is necessary to reliably establish the quantitative patterns in this *eh*-usage. A first indication will be given by considering the overall picture of *eh*-usage across each socio-demographic group as a whole, mirroring the analyses by Meyerhoff (1994) and Stubbe and Holmes (1995). However, this overall picture will conflate two important questions, which are expected to have different answers under a listener-based approach (Section 4.3). Firstly, which socio-demographic groups use *eh* at all, and which don't? Secondly, how often do *eh*-users of different socio-demographic groups use *eh*? In this section, I describe robust quantitative methods that establish an overall picture of *eh*-usage among Pākehā and then break it down to answer the questions of who uses *eh*, and how often.

Since the overall picture of *eh*-usage by Pākehā conflates the questions of who uses *eh*, and how often, I do not use it to compare the age-grading and change-in-progress hypotheses. Instead, I use the overall picture to compare the present data to those of Meyerhoff (1994) and Stubbe and Holmes (1995), to get a sense of the reliability of *eh*-indices and thus how the results of previous studies may generalize. For this, I use resampling methods to calculate bias-corrected and accelerated (BCa) bootstrap

[38] Meyerhoff's (1994) original formulation of the *eh*-index used a denominator measuring the duration of the interview with the speaker; here, I use the number of words uttered by the speaker, as it is independent of speech rate and restricts the calculation to the period in which the speaker is talking. The multiplication by 12000 puts this new formulation on a similar scale to that of Meyerhoff's (1994) original formulation, under the assumption that the speaker talks for half of the total interview and says an average of 4 words per second.

confidence intervals.[39] Under this method, a distribution of *eh*-indices is generated for a group by randomly resampling its members with replacement (so that the same member may be included multiple times, while another may be excluded outright), calculating the *eh*-index for this resampled group, and repeating 5000 times. To ensure that the results are not confounded by membership in control groups that are not in question, members may not be reassigned between control groups; for example, in calculating the confidence intervals based on class, a member that is originally in both the professional and male groups may be reassigned to the nonprofessional group, but not to the female group. BCa confidence intervals calculated from the distribution of resampled *eh*-indices incorporate adjustments for the bias and skewness of the distribution, effectively controlling for external constraints (such as the fact that a speaker cannot utter negative instances of *eh*) and the presence of outliers and high-leverage points (such as speakers who use *eh* disproportionately rarely or often) (Haukoos & Lewis, 2005).

My main focus in the present analysis is on the questions conflated by the overall picture: who uses *eh*, and how often. In order to answer these questions appropriately, it is necessary to confront two fundamental issues raised by the distribution of *eh* over speakers, which is both sparse and bursty. The distribution of *eh* is *sparse* because many people do not use it in sociolinguistic interviews; in the present data, 79% of speakers (311 of 394) do not use *eh* at all. This failure to use *eh* could be either because the speaker does not have it in their linguistic repertoire – consistent with the change-in-progress hypothesis – or because the right circumstances (discourse, social, pragmatic, etc.) for using it did not arise in the interview – consistent with the age-grading hypothesis. Similarly, the distribution of *eh* is *bursty* (or *overdispersed*)

[39] I also use bootstrapped confidence intervals when considering how often *eh*-users use *eh*, but there are not enough datapoints to estimate appropriate BCa intervals. Instead, I use percentile confidence intervals, whose endpoints are the 2.5% and 97.5% percentiles of the bootstrap distribution.

because many people who do use *eh* in sociolinguistic interviews use it sparingly, while a few people use it excessively; in the present data, 52% of *eh*-users (43 of 83) have *eh*-indices less than 15, while the *eh*-indices of remaining *eh*-users range all the way up to 140 (variance: 309.94).

To handle these issues, a robust quantitative analysis must do two things. Firstly, it must consider both possible sources of failure of *eh*-usage equally, so as not to induce artificial bias that gives a priori favor to one hypothesis over the other. Secondly, it must be conservative in the face of extreme outliers, who are unlikely to represent *eh*-usage among their socio-demographic group in general. To accomplish both of these things while separating the questions of who uses *eh* and how often, I introduce a new tool into the quantitative (socio)linguistics toolbox: zero-inflated negative binomial (ZINB) regression.[40] A ZINB regression model combines two distinct components: a binary logistic regression component, which estimates the probability that a speaker has *eh* in their linguistic repertoire, and a negative binomial regression component, which estimates the individual-level *eh*-index for a speaker who has *eh* in their repertoire. The binary logistic regression component answers the question of which socio-demographic groups use *eh*. This answer does not exhibit a priori model bias because the two components work together to split speakers who don't use *eh* in the interview, between those that don't have it in their repertoire and those that do but did not use it on that occasion (for whatever reason).[41] The negative binomial

[40]The ZINB regression model that I use is instantiated in the `pscl` package in *R* (Zeileis, Kleiber, & Jackman, 2008).

[41]Though the ZINB regression analysis makes a distinction between speakers who don't use *eh* because they lack it in their repertoire and those who have it in their repertoire but don't use it in the interview, it is ultimately still conducted based on observed instances of *eh*. Thus, the results of the binary logistic regression component may be affected by certain speakers not using *eh with the interviewers*. Nevertheless, in the absence of data that are perfectly controlled for content and similarity / relationship between interlocutors, the ZINB regression analysis represents the best way to separate the question of presence in the linguistic repertoire from the question of usage.

regression component answers the question of how often *eh*-users of different socio-demographic groups use *eh*. This answer is appropriately conservative because the regression effectively downweights the influence of extreme outliers (for discussion, see Appendix E).

I assume that different socio-demographic factors can influence each component of the regression model. Thus, I consider the fit of each component separately: first, the binary logistic regression component, and second, the negative binomial regression component. In each component, I start from a maximal model containing all socio-demographic categories and their interactions. I remove interactions in a stepwise fashion if they do not significantly improve model fit, as assessed by model comparison likelihood ratio tests. I do not remove any non-interaction terms, so as to maintain a full picture of the influence of socio-demographic categories. In the negative binomial regression component, I also include an offset term of $\log(\text{wordcount}/12000)$, which transforms the dependent variable of the regression from raw *eh* counts to *eh*-indices and thus controls for the fact that different speakers utter different numbers of words in total in their interview. Because the binary logistic regression component is initially selected conditioned on a maximal negative binomial regression component, I reconsider it after completing the selection process for the latter component, adding interaction terms back into it in a stepwise fashion if they significantly improve model fit. Though the two components are fit as part of the same model, I report them separately as they bear on different questions of the analysis.

Finally, to assess how the questions of who uses *eh* and how often compare to one another in their capacity to account for the corpus data, I compare the independent contributions of the corresponding components of the ZINB regression model (the binary logistic regression component and the negative binomial regression component, respectively). To conduct this comparison, I consider two weaker versions of the model, in each of which one component is reduced to a constant and the other component is preserved. I consider each weaker model in comparison to the full model in which both components are preserved, using both likelihood ratio tests and AIC. For comparison based on likelihood ratio tests, I consider a statistically significant difference between the full model and a weaker model to indicate that the component that has been reduced in the weaker model makes a substantial contribution to the full model. For comparison based on AIC, I calculate the difference in AIC between the full model and each weaker model. Following Burnham and Anderson (2002, p. 70), I interpret a difference of 10 or more to indicate that the component that has been reduced in the corresponding weaker model makes a substantial contribution to the full model, and a difference of 2 or less to indicate that it does not.

4.5.3 Predictions

Both the age-grading hypothesis and the change-in-progress hypothesis predict old Pākehā to use *eh* less often than young Pākehā in sociolinguistic interviews, when considering the overall general picture. However, their predictions differ in important ways once the general picture is broken down into the questions of who uses *eh*, and how often. Here, I walk through the different predictions for each question, and I show how they can be tested by the corpus analysis.

In understanding these predictions, it is important to note that the notion of *repertoire* is distinct from the notion of *representation*. All speakers have a representation for *eh*, reflecting their experience of others using it. However, only those who have the potential to use *eh* themselves have it in their linguistic repertoire. In other words, a speaker will have *eh* in their linguistic repertoire only if their representation of *eh* is located within the area of representational space that they draw upon for their own production.[42] For practical purposes, I assume that, once *eh* builds

[42] This approach is analogous to that of the ZINB regression model, where the speaker draws samples from a bag of representations in production; *eh* is present in the repertoire if it is contained

a sufficient foundation in the linguistic repertoire, it takes a long time to exit, even if it is rarely used. I also assume that, while a speaker with *eh* in their repertoire can actively choose to avoid or suppress it, it may nevertheless slip out occasionally. Thus, a speaker who uses *eh* fluently at one point in time will likely have it in their repertoire and still occasionally use it years later, even if they intend not to.[43]

The question of who uses *eh* can be interpreted as asking who has *eh* as a potential resource in their linguistic repertoire and who does not. The age-grading hypothesis does not predict any differences of this sort between different socio-demographic groups, while the change-in-progress hypothesis (under a listener-based approach) predicts several.

The age-grading hypothesis does not necessarily predict any socio-demographic group to be more likely to have *eh* in their repertoire than any other. This prediction follows because age-grading is concerned with the use or avoidance of a resource, which implies that the resource must be available to be used or avoided in the first place. As such, it holds regardless of whether age-grading is reflective of the community or of the data sample. If age-grading is reflective of the community, then all speakers are assumed to have used it at some point in their lives, which means that it is in their linguistic repertoire, even if it is currently suppressed. Similarly, if age-grading is reflective of the data sample, then all speakers are assumed to have the capacity to use it, meaning that it is in the repertoire of all speakers, even though only some speakers are observed to use it within the data due to properties of the data collection.

The change-in-progress hypothesis, on the other hand, predicts minimally that young speakers should be more likely to have *eh* in their repertoire than old speakers.

---

in that bag.

[43]These assumptions have implications for the ZINB regression model, because it assesses the probability that an individual has *eh* in their repertoire in part by considering the use of *eh* across all demographically similar individuals. Thus, the isolated use of *eh* by a single member of a demographic group is taken to indicate that multiple members of that group likely have *eh* in their repertoire, even if they have suppressed it (or otherwise didn't use it in their interview).

This prediction follows from the apparent time construct (Cukor-Avila & Bailey, 2013), as it implies that change-in-progress – which, here, is the adoption of *eh* into the linguistic repertoire – is most advanced among young speakers. A listener-based approach provides further motivation for this predicted age-based difference, through the assumption that the adoption of *eh* is driven by changes in attitudes, which are most strongly evident among young Pākehā.

A listener-based approach also predicts that *eh* should be more likely in the repertoire of male and nonprofessional Pākehā than in that of female and professional Pākehā. This prediction follows from the fact that *eh*-usage in Māori is highest among males and nonprofessionals (Meyerhoff, 1994), in two different ways. Firstly, if Pākehā interactions with Māori are skewed toward members the same sex and/or class, then male and nonprofessional Pākehā will experience *eh* more as listeners, and even weak influences of those experiences will add up over time.[44] Secondly, even if all Pākehā have similar degrees of experience with *eh* from Māori, not all experiences will carry equal weight. According to a listener-based approach, listeners assign more weight to an experienced token the more similar they perceive the speaker to be to them. Since *eh* is produced most by male and nonprofessional Māori, the corresponding Pākehā listeners stand to gain the most from such differential weighting.

Since the two hypotheses make different predictions with respect to the question of who uses *eh*, they can be tested. In particular, these predictions can be tested by the binary logistic regression component of the ZINB regression analysis, which measures the probability that a given Pākehā has *eh* in their linguistic repertoire, according to their membership in socio-demographic groups. Under the age-grading hypothesis,

[44]The differences in degrees of experience are particularly pronounced for class, as Māori are overrepresented among nonprofessionals. For example, in the New Zealand Census in 2006 (close to time at which the last interviews in ONZE were conducted), the median Māori income was approximately 20% lower than the median Pākehā income, and the proportion of Māori on a governmental benefit was more than twice the proportion of Pākehā on a governmental benefit (Statistics New Zealand, 2006a).

this analysis is predicted to show no significant effects; under a listener-based approach to the change-in-progress hypothesis, it is predicted to show significant effects of age (young > old), class (nonprofessional > professional), and sex (male > female). Further support for the comparison of predictions about who uses *eh* can be gained from considering the historic component of the ONZE corpus. If a similar number of people (within the same age group) use *eh* at earlier points in (real) time as in the Canterbury Corpus, then the age-grading hypothesis will gain further support; conversely, if fewer people use *eh* at earlier points in time, then the change-in-progress hypothesis will gain further support.

The question of how often Pākehā use *eh* must be restricted to those who are assessed to have *eh* in their linguistic repertoire, whom I will refer to as *eh*-users. Among these *eh*-users, the age-grading hypothesis predicts there will be differences based on socio-demographic group, but the change-in-progress hypothesis does not.

The age-grading hypothesis minimally predicts that young *eh*-users should use *eh* more than old *eh*-users. This prediction follows from the fact that age-grading is concerned with differences in usage, not differences in underlying capacity. The age-grading hypothesis also makes further predictions if it is due to properties of the data sample rather than properties of the community. In this case, the hypothesis holds quite generally that *eh*-usage is facilitated amongst interlocutors from similar socio-demographic groups, because it allows them to emphasize aspects that may be expected to be shared on the basis of group membership (see also Bell, 2001). Since the interviewers in the present data are university students, and therefore fairly homogeneous with respect to age and class, this prediction can only be applied on the basis of sex. Specifically, then, the age-grading hypothesis predicts that *eh*-usage among males should be highest with male interviewers, and *eh*-usage among females should be highest with female interviewers.

Unlike the age-grading hypothesis, the change-in-progress hypothesis makes no predictions about differences in *eh*-usage among *eh*-users, following its focus on differences in underlying capacity rather than usage. If *eh* is spreading through all Pākehā, then it is not expected to function as a demographic marker, with degree of usage depending on the demographics of the speaker. Instead, in line with its discourse function, its degree of usage is expected to depend upon interpersonal and situational factors, which vary widely independent of who the speaker is. Furthermore, the degree of *eh*-usage among *eh*-users can reasonably be expected to show different patterns than the distribution of *eh*-users themselves, as a repetition of the same patterns would suggest that they are indicative of a demographic marker rather than change-in-progress.

A listener-based approach to the change-in-progress hypothesis goes one step further, predicting that all *eh*-users will use *eh* to similar extents.[45] This prediction is motivated by the idea that an individual will build the strength of *eh* in their representations quickly after adopting it, due to the concomitant increases in perceived typicality (see Section 4.3). If there are isolated groups of *eh*-users with lower *eh*-usage than others, a listener-based approach to the change-in-progress hypothesis expects that these *eh*-users have only recently added *eh* to their linguistic repertoire and are still building up experience with it.

The different predictions of the two hypotheses with respect to the question of how often Pākehā use *eh* can be tested by the negative binomial regression component of the ZINB regression analysis. The two-component nature of ZINB regression ensures that this analysis is restricted to *eh*-users, unlike the overall general picture. Under the age-grading hypothesis, this analysis is predicted to show a significant effect of age

[45]The suggestion that all *eh*-users should use *eh* to similar extents is not a contradiction of the insight that change-in-progress typically creates *structured heterogeneity* (Weinreich et al., 1968; see also Labov, 2001, and Section 4.2.1). Change-in-progress creates structured heterogeneity with respect to the question of *who* uses (or has the potential to use) *eh*, but not *how often* they do so. *Eh* has nuanced functional (and distributional) differences from other discourse tags, so its adoption need not introduce new alternations in which frequency of use would be crucially implicated.

(young > old); under a listener-based approach to the change-in-progress hypothesis, it is predicted to show no significant effects. Further (indirect) testing of the age-grading predictions can be gained from considering how *eh*-usage varies depending on the sex of the interviewee and the interviewer. The ideas underlying (one version of) the age-grading hypothesis predict higher *eh*-usage when both interlocutors are the same sex.

Finally, the two hypotheses can be compared to one another in terms of their exclusive capacity to explain the data by comparing the independent contributions of the two components of the ZINB regression model. The change-in-progress hypothesis alone makes non-null predictions for the question of who uses *eh*, reflected in the binary logistic regression component of the model. The age-grading hypothesis alone makes non-null predictions for the question of how often *eh*-users use *eh*, reflected in the negative binomial regression component of the model. Correspondingly, the change-in-progress hypothesis predicts that the binary logistic regression component makes a substantial contribution to the model but the negative binomial regression component does not, and the age-grading hypothesis predicts the converse.

4.6 Results

In this section, I analyze what the ONZE data introduced in Section 4.5.1 have to say about the use of *eh* by young Pākehā. First, in Section 4.6.1, I establish the overall general picture of the data and compare it with the picture obtained from previous studies (Meyerhoff, 1994; Stubbe & Holmes, 1995). Next, in Sections 4.6.2 and 4.6.3, I present the results of the robust quantitative analysis described in Section 4.5.2, to test the predictions of the age-grading and change-in-progress hypotheses described in Section 4.5.3. These results yield support for the change-in-progress hypothesis. Finally, in Section 4.6.4, I turn to a qualitative approach in order to

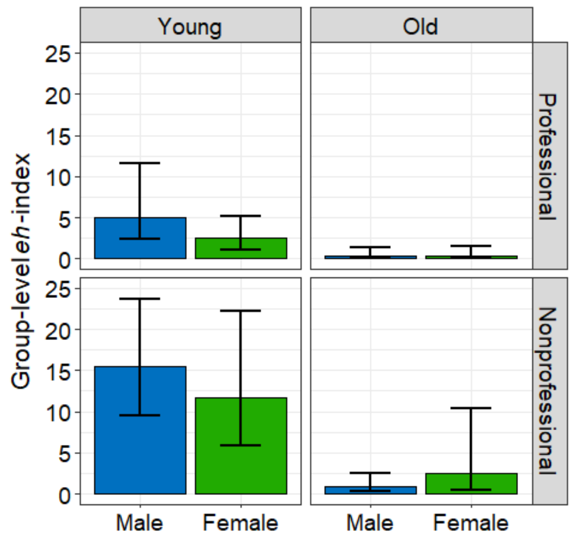

Figure 4.2: Group-level *eh*-indices across age, class, and sex. Error bars indicate bootstrapped BCa 95% confidence intervals. Calculations include all speakers, regardless of whether they use *eh* or not.

assess whether there is any evidence that Pākehā adoption of *eh* is associated with changes in attitudes toward Māori, as expected under a listener-based approach.

4.6.1 General picture

Figure 4.2 shows the overall general picture obtained from comparing group-level *eh*-indices across all socio-demographic groups in the ONZE data. Since this picture conflates the questions of who uses *eh* and how often, I do not use it to perform statistical tests. Rather, I compare it impressionistically to the pictures obtained from the previous studies by Meyerhoff (1994) and Stubbe and Holmes (1995), and I use it to gauge the reliability with which generalizations can be made from this kind of data.

For age and class, the ONZE data yield overall general patterns that resemble those seen in both previous studies. In the data of both Meyerhoff (1994) and Stubbe and Holmes (1995), group-level *eh*-indices were higher for young Pākehā than for old Pākehā, and for working-class (nonprofessional) Pākehā than for middle-class (professional) Pākehā. Figure 4.2 shows the same patterns in the ONZE data. Since these patterns are observed across all three datasets, and since, in the ONZE data, the corresponding confidence intervals (i.e. for groups distinguished only by age or by class) are almost all non-overlapping, it is likely that they are reliable in general.

For sex, however, the patterns are not so clear. In the data of Meyerhoff (1994), group-level *eh*-indices were highest for female Pākehā, particularly among young speakers, but in the data of Stubbe and Holmes (1995), they were highest for (young) male Pākehā. In the ONZE data (Figure 4.2), *eh*-indices are numerically highest for males among young speakers, but for females among old speakers. However, the large degrees of overlap in confidence intervals indicates that these numerical differences are likely a reflection of the data sample, not of the underlying population. The ONZE data therefore indicate that there may be no reliable and generalizable effect of sex on overall general *eh*-index.

It was differences between Māori and Pākehā in the effect of sex that led Meyerhoff (1994) to suggest the age-grading hypothesis as an alternative to the change-in-progress hypothesis (see Section 4.4.1). The indication from the ONZE data that such differences may not be reliable does not nullify this suggestion, but it does put the two hypotheses on even ground. More generally, the wide confidence intervals observed in the ONZE data indicate a substantial amount of uncertainty, in spite of the large number of speakers per cell. Given that previous studies contained approximately ten times fewer speakers per cell, the ONZE data suggest that conclusions drawn from these studies may be driven by non-representative data samples, and thus may not be appropriately generalizable.

4.6.2 Who has *eh* in their repertoire?

Who uses *eh*? Or, more precisely, who has the potential to use *eh*, through having it in their linguistic repertoire? Recall from Section 4.5.3 that the age-grading

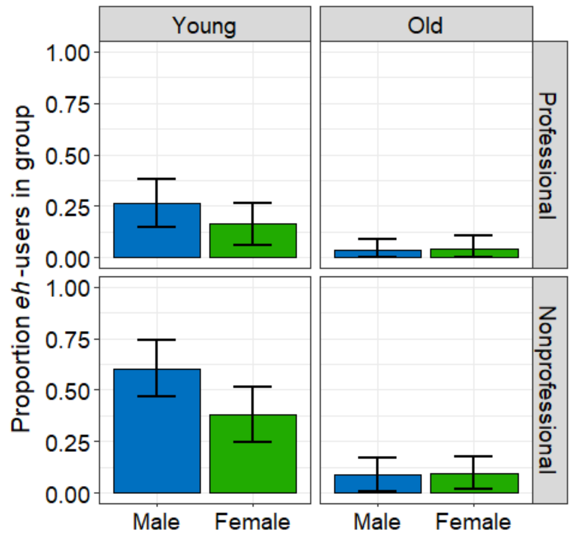

Figure 4.3: Proportion of speakers who use *eh* in their interview. Proportions and 95% confidence intervals are calculated from raw data. Calculations are by age × class × sex cell and count any speaker who used *eh* (at least once) in the interview, regardless of the number of times they used it.

hypothesis does not predict any differences in the number of *eh*-users from different socio-demographic groups, while a listener-based approach to the change-in-progress hypothesis predicts there to be more young *eh*-users than old, more nonprofessional *eh*-users than professional, and more male *eh*-users than female.

A first indication can be gained from the raw data, by considering the proportion of speakers in each socio-demographic cell who use *eh* at least once in their interview. These raw values are presented in Figure 4.3. As can be seen, the proportion of *eh*-users appears to be larger for nonprofessionals than for professionals, for young speakers than for old speakers, and for males than for females.

The differences in the raw data presented in Figure 4.3 could be misleading, as it is possible that some speakers have *eh* in their linguistic repertoire but nevertheless failed to use it in the interview (for whatever reason). The ZINB regression analysis provides a way to test the statistical significance of differences in *eh*-users across socio-demographic categories while guarding against this possibility. The binary logistic

Table 4.1: Binary logistic regression component of ZINB analysis of *eh*. The model predicts the probability that a speaker has *eh* in their linguistic repertoire. Coefficient estimates indicate how much more likely members of the corresponding social group are to have *eh* in their repertoire, relative to members of the relevant reference group (log-odds scale). The reference groups (incorporated in the Intercept) are professional, old, and female.

	Estimate	Std. Error	z-value	Pr($>\|z\|$)	
(Intercept)	−2.7313	0.8085	−4.232	0.0359	***
Class = Nonprofessional	1.4322	0.5606	2.555	0.0106	*
Age = Young	2.1900	0.6223	3.519	0.0004	***
Sex = Male	1.0346	0.5929	−1.745	0.0810	.
Overall ZINB model: AIC = 710.28; BIC = 750.05; logLik = −345.14; df = 10; N = 394					

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$, $.p < 0.1$

regression component of the analysis, which estimates the probability that a speaker has *eh* in their linguistic repertoire, is reported in Table 4.1.

The model shows that nonprofessional and young Pākehā are significantly more likely to have *eh* in their linguistic repertoire than professional and old Pākehā, respectively. It also suggests that male Pākehā may be more likely to have *eh* in their linguistic repertoire than female Pākehā, though this difference does not quite reach statistical significance. These results are not expected from the age-grading hypothesis, but they closely match the predictions of the change-in-progress hypothesis. Thus, consideration of who uses *eh* lends support to the change-in-progress hypothesis.

To further test the change-in-progress hypothesis, it is useful to consider earlier points in time. The hypothesis predicts that fewer Pākehā used *eh* at earlier points in time, regardless of age. To test this prediction, I searched the historical component of the ONZE corpus that immediately precedes the Canterbury Corpus, the Intermediate Archives. This component of the corpus consists of interviews with old speakers (born between 1890 and 1930; aged 60 and upwards at the time of interview) recorded in the early 1990s, up to 15 years before the interviews in the Canterbury Corpus. Out of 87 speakers, I found just one who used *eh* in their interview, and then only as a metalinguistic commentary (discussed in detail in Section 4.6.4). This number is significantly smaller than what would be expected based on the Canterbury Corpus, where 13 out of 194 old speakers used *eh* at least once in their interview ($p = 0.037$, one-tailed Fisher's exact test). While this result doesn't preclude the possibility of an age-based bias against *eh*-usage – since the old speakers in the Intermediate Archives are older than those in the Canterbury Corpus – the complete absence of naturally-occurring *eh* in the Intermediate Archives is striking and reinforces the support for the change-in-progress obtained from the Canterbury Corpus.

4.6.3 How often do *eh*-users use *eh*?

How often do the *eh*-users in each socio-demographic group use *eh*? Again, the two hypotheses predict different answers to this question (Section 4.5.3). The age-grading hypothesis predicts that young *eh*-users will use *eh* more than old *eh*-users, while the change-in-progress hypothesis does not predict any differences in *eh*-usage between *eh*-users from different socio-demographic groups.

An indication of rates of *eh*-usage by *eh*-users in different socio-demographic groups can be obtained from the raw data. Figure 4.4 shows the mean individual-level *eh*-indices for each cell of the corpus design, restricted to speakers who use *eh* at least once in their interview. By comparing Figure 4.4 to Figure 4.3, it is clear that the differences in use of *eh* among members of different groups are much more modest than the differences in the number of *eh*-users in different groups. For example, though nearly 14 times more young male nonprofessionals than old female professionals use *eh* at least once in their interview (60.4% vs. 4.5%), the average young male nonprofessional *eh*-user uses *eh* only approximately twice as often as the average old female professional *eh*-user (mean individual-level *eh*-index of 28.8 vs. 13.0).

However, again, the differences in the raw data (Figure 4.4) could be misleading – a possibility that is highlighted by the large, overlapping confidence intervals. To gain a reliable indication of any differences in *eh*-usage among *eh*-users, an analysis

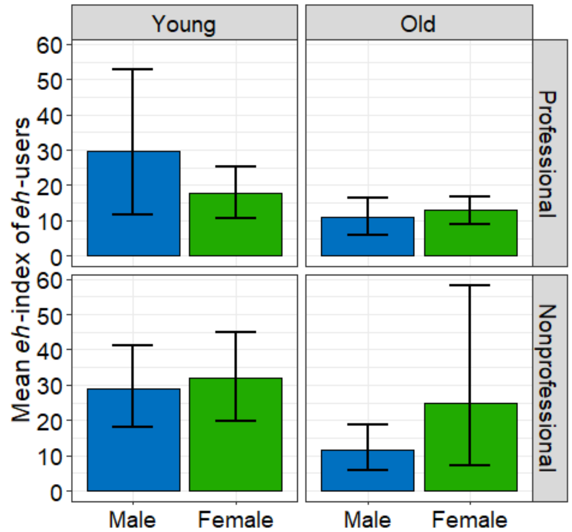

Figure 4.4: Mean individual-level *eh*-indices of *eh*-users. Calculations are based on bootstrapping raw data, grouped by age × class × sex cell, and include all speakers who used *eh* (at least once) in the interview. Error bars indicate 95% bootstrap percentile confidence intervals.

must account for speakers who have *eh* in their linguistic repertoire but did not use it in the interview, and for the presence of extreme outliers. The ZINB regression analysis accounts for these things, allowing it to test the statistical significance of the differences in *eh*-usage. The negative binomial regression component of the analysis, which estimates the *eh*-index of a speaker who has *eh* in their linguistic repertoire, is reported in Table 4.2.

The model shows that most *eh*-users use *eh* to similar extents (at least, in the interview setting represented in the corpus). There is no statistically significant difference in *eh*-usage by *eh*-users based on class (nonprofessionals vs. professionals). Among female *eh*-users, there is also no difference based on age: young and old female *eh*-users use *eh* to similar extents. However, among male *eh*-users, there is a significant difference based on age: young male *eh*-users use *eh* more than old male *eh*-users. The model suggests that this effect is driven by the lower use of *eh* among old male *eh*-users than among old female *eh*-users; young *eh*-users all use *eh* to a

Table 4.2: Negative binomial regression component of ZINB analysis of *eh*. The model predicts the average individual-level *eh*-index of a speaker with *eh* in their linguistic repertoire. Coefficient estimates indicate how much higher the *eh*-index is expected to be for *eh*-users in the corresponding social group, relative to members of the relevant reference group (log scale). The reference groups (incorporated in the Intercept) are professional, old, and female.

	Estimate	Std. Error	z-value	$\Pr(>\|z\|)$	
(Intercept)	1.9518	0.6760	2.887	0.0039	**
Class = Nonprofessional	0.5315	0.3867	1.374	0.1694	
Age = Young	0.3002	0.6203	0.484	0.6284	
Sex = Male	−1.4322	0.7803	−1.836	0.0664	.
Age = Young & Sex = Male	1.6241	0.8031	2.022	0.0431	*

Overdispersion parameter: $\log(\theta) = -0.9058$, std. error = 0.3084, $z = -2.937$, $p = 0.0033$
Overall ZINB model: AIC = 710.28; BIC = 750.05; logLik = −345.14; df = 10; N = 394

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$, $.p < 0.1$

similar extent, regardless of sex.

The model results do not align exactly with the predictions of either hypothesis. The age-grading hypothesis predicted an effect of age, but this effect was only partially observed, for male *eh*-users but not female *eh*-users. The question for the age-grading hypothesis then becomes why old female *eh*-users use *eh* so often. There are two possible answers to this question, depending on whether age-grading reflects a property of the community or the data sample. If age-grading reflects a property of the community, then it could be that the reasons that cause speakers to suppress *eh* as they age – for example, overt stigma – are weak among females. Alternatively, if age-grading reflects a property of the data sample, then it could be that the old female speakers share many values, beliefs, and opinions with their interviewer that can be emphasized by *eh*,[46] or at least attempt to convey such sharing as a positive politeness strategy (Holmes, 1995). The change-in-progress hypothesis, on the other hand, did not predict any effects – though, as noted in Section 4.5.3, it is not necessarily incompatible with isolated effects. The question for the change-in-progress

[46]The possibility that old female *eh*-users share more values, beliefs, and opinions with the interviewer than their male counterparts would not be surprising, given that more than 75% of the interviewers are young female undergraduates.

hypothesis then becomes why old male *eh*-users use *eh* so little. The answer suggested by a listener-based approach is that old male *eh*-users may be relatively new to the adoption of *eh*, and may still be accumulating experience to solidify its place in their linguistic representations. An alternative answer is the flipside of the answer given for the age-grading hypothesis as a reflection of a property of the data sample: old male *eh*-users may share few values, beliefs, and opinions with their interviewer that can be emphasized by *eh*.[47]

It is clear that further investigation is necessary in order to assess which of the hypotheses better align with the effects on *eh*-usage by *eh*-users identified by the ZINB regression analysis. As a first step toward such investigation, I consider how the socio-demographic similarity between the speaker and the interviewer affects the *eh*-usage of an *eh*-user. This approach allows me to test the fundamental idea behind the age-grading hypothesis, viewed as reflecting a property of the data sample, according to which *eh*-usage should be highest between interlocutors who are socio-demographically similar. Because the interviewers are all undergraduate students, I group speakers and interviewers by sex; I collapse across age and class groups in order to counter data sparsity stemming from the fact that more than 75% of the interviewers are female. The mean individual-level *eh*-indices calculated from this grouping in the raw data are illustrated in Figure 4.5.

Although, numerically, the *eh*-indices for *eh*-users do appear higher when the speaker and interviewer are the same sex, the large, overlapping confidence intervals suggest that any differences are unlikely to be reliable. In line with this suggestion, including interviewer sex and its interaction with speaker sex in the ZINB regression analysis did not yield statistically significant effects or result in a statistically

[47]The notion that old male *eh*-users use *eh* so little because they share few values, beliefs, and opinions with their interviewer does not follow from the details of a listener-based approach, but rather from the socio-pragmatic discourse function of *eh*. The *usage* of *eh* according to its function is not inconsistent with the principle that the *adoption* of *eh* is driven by processes in perception.

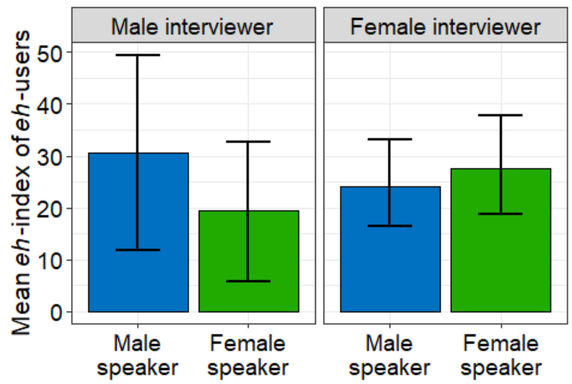

Figure 4.5: Mean individual-level *eh*-indices of *eh*-users, by interviewer sex. Calculations are based on bootstrapping raw data, grouped by age × class × sex cell, and include all speakers who used *eh* (at least once) in the interview. Error bars indicate 95% bootstrap percentile confidence intervals.

significant improvement in model fit ($\chi^2(2) = 3.46$, $p = 0.18$). Though little can be concluded from this null result, it suggests that the effect identified by the ZINB regression analysis is limited to a difference in *eh*-usage between old male *eh*-users and old female *eh*-users, rather than reflecting a larger pattern. Since the age-grading hypothesis depends upon the existence of such a larger pattern in *eh*-usage, whereas the change-in-progress hypothesis does not, I argue that consideration of the question of how often *eh*-users use *eh* lends support to the change-in-progress hypothesis.

Finally, additional support for the change-in-progress hypothesis over the age-grading hypothesis is provided by comparing the independent contributions of the two components of the ZINB regression model. According to likelihood ratio tests, a weakened model in which the binary logistic regression component is reduced to a constant and the negative binomial regression component is preserved fits the data significantly worse than the full model ($\chi^2(3) = 17.31$, $p < 0.001$), while one in which the negative binomial regression component is reduced to a constant and the binary logistic regression component is preserved fits the data only marginally worse than the full model ($\chi^2(4) = 9.08$, $p = 0.059$). Similarly, a model in which just the binary logistic regression component is reduced yields a increase in AIC of 11.3

relative to the full model (721.6 vs. 710.3), while a model in which just the negative binomial regression component is reduced yields an increase of just 1.1 (711.4 vs. 710.3). All of these results match the predictions of the change-in-progress hypothesis and oppose those of the age-grading hypothesis, as they indicate that patterns in who uses *eh* – as captured by the binary logistic regression component – make a substantial contribution in accounting for the data, while patterns in how often *eh*-users use *eh* – as captured by the negative binomial regression component – do not.

4.6.4 Hints of attitudes: qualitative analysis

All things considered, the quantitative results from the ONZE Corpus suggest that the use of *eh* by young Pākehā is more likely a reflection of change-in-progress than of age-grading. However, they can give no answers to the question of *why* this change is occurring. A listener-based approach claims that changes in social attitudes toward Māori should play a role in increasing Pākehā adoption of *eh* (provided *eh* retains perceptual associations with Māori over time). In this section, I pursue evidence for this claim, by conducting a qualitative analysis that looks closer at the data against the general backdrop of social change in New Zealand.

As described in Section 4.4.2, the last 30 years or so have seen improvements in the way that Māori are treated in New Zealand society, together with corresponding shifts in the attitudes of Pākehā, from stigma to bicultural fusion. These shifts are also evidenced in two naturally-occurring metalinguistic commentaries from the ONZE corpus, referencing times that are 50 years apart.

In (C1), an older Pākehā man, recalling events from approximately 1940, states that an association between *eh* and Māori caused him to have to abandon an *eh*-using habit in order to avoid being bullied as a child. Not only did the other children in this excerpt view *eh* as a negative feature due to its association with Māori, as evidenced by the fact that they bullied him for it, but the speaker also soon learned to view it as a negative feature, as he "very quickly" stopped using it and now refers to it as "the worst thing" about his childhood.

(C1) but . the **worst thing** was that
. **we used to talk a little bit like the Māoris** and we used to say
"What do you think of that, eh?"
. um . I never . I never did say . "I seen this" I used to say "I saw this"
e- and yet a lot of kids in in Rotorua did say "I seen this" and "I done that"
um
and those were very common in Rotorua
. I didn't do that but at . I used to say . "What do you think of that, eh?"
"Good, eh?" everything was . with an eh and of course
. it was pretty **hard to stop myself doing that and I had to very quickly**
but
. for the first few days at school
. **it made me a bit of a butt of their humor** and um
. as I say I had a quick temper and I didn't . take insults
. so **I got into a lot of fights**

In (C2), a younger Pākehā man (S), talking about attitudes in 1994 (50 years since (C1)), also identifies an association between *eh* and Māori, but does not ascribe stigma to *eh* as a result. In fact, his suggestion that *eh* could be interpreted as "common" (and thus stigmatized) is rejected by the interviewer (I), and he makes no attempt to appeal this rejection. It is only after the rejection of *eh* being "common" that he presents the association with Māori as an alternative, making it clear that Māori are not negatively evaluated.

(C2) S: Does eh **make me sound co- common?**
I: Eh?

S: Eh.

I: ***I don't think so.***

S: Or is it just because I've been ha- .

S: um . **got on a course with lots of Māori students** at the moment.

I: ***Yeah they would say it.***

These two commentaries clearly show differences in Pākehā attitudes toward Māori over time, in a way that is explicitly evoked through the use of *eh*. These attitudinal differences are demonstrated explicitly through the content of the commentaries, as previously discussed, as well as implicitly through their forms. For example, the word *Māori* in Te Reo Māori is pronounced [mɑːɔɾi] and does not inflect when plural. The old speaker's treatment of this word in (C1) shows little respect for, or awareness of, Māori norms: he anglicizes it, pronouncing it [mæːɹi] and inflecting it with English plural morphology. Conversely, the young speaker's treatment in (C2) shows respect and awareness: he pronounces it [mɑɔɾi] and chooses an alternative phrasing that does not require it to be inflected. Similarly, the old speaker's use of the definite article in reference to "the Māoris" betrays a monolithic attitude of social othering (Acton, 2019), while the young speaker's reference to "Māori students" suggests a more individuated attitude, in which Māori are put on even ground by sharing his status of being a student.

The two commentaries analyzed here provide support for key components of a listener-based approach to Pākehā adoption of *eh*. They show the prerequisite for a listener-based approach, which is that the perception of an association between *eh* and Māori has been maintained over time, even as Pākehā begin to adopt *eh*. They show the main ingredient for a listener-based approach, which is that Pākehā attitudes toward Māori – and toward the use of *eh*, in recognition of its association with Māori – have improved over time, both explicitly and implicitly. And, finally, they show the outcome of a listener-based approach, which is that, while *eh* used to be blocked in Pākehā speech due to strong social stigma, its widespread usage now may be because such stigma is no longer prevalent.

4.7 Discussion

The first part of this chapter established parallel effects on speech perception and language change from language-internal factors, such as word frequency, and language-external factors, such as social information and attitudes. In doing so, it laid out an extension of the listener-based approach developed in previous chapters to social effects in language change. The second part of this chapter picked up that extension and applied it to an open question about English in New Zealand, which asks why young Pākehā use the Māori-associated tag particle *eh*. The listener-based approach not only allowed the formulation of precise predictions for testing whether Pākehā patterns of *eh*-usage reflect age-grading or change-in-progress, but also provided a reason for *why* change-in-progress may be occurring, which integrates the wider socio-cultural and ideological landscape of New Zealand with established passive but powerful perceptual biases. A large corpus study supported the hypothesis that the use of *eh* by young Pākehā reflects change-in-progress, as well as the hypothesis that such change is related to changes in social attitudes.

The story told by a listener-based approach to Pākehā *eh*-adoption is as follows. More than 50 years ago, speech from Māori evoked strong negative social attitudes among Pākehā. As discussed in Section 4.3, this caused features of Māori speech such as the word *eh* to be perceived as atypical with respect to Pākehā speech and thus to have little influence on the representations that Pākehā drew upon for their own speech production. That is, the negative social attitudes toward Māori effectively blocked the uptake of *eh* among Pākehā. As Pākehā started to view Māori more positively, however, they started to view Māori speech as more typical, causing Māori features to become more strongly represented in the linguistic knowledge underlying their own productions. Thus, the improvement in social attitudes toward Māori allowed Pākehā to begin using *eh* in their own speech.[48]

Crucially, under a listener-based approach, the explanation for Pākehā *eh*-adoption does not depend upon intentions or awareness on the part of the speaker, which may be practically impossible to identify as an external observer. While it is possible that Pākehā originally actively avoided sounding like Māori because of social stigma and then began actively trying to sound like Māori when this stigma was weakened, such explicit avoidance and imitation is not necessary to explain the widespread pattern of change.[49] A listener-based approach provides an alternative explanation in which the Pākehā use of *eh* increased passively and subconsciously, drawing parsimonious connections between the change and known processes of speech perception.

The shift in attitudes in New Zealand is not limited to a simple improvement in the way that Māori are perceived; it is also changing the socio-cultural ideological landscape of New Zealand. Māori are going from being viewed as separate from Pākehā to being viewed as key partners in a bicultural fusion embodying the notion of *New Zealander*, or *Kiwi*[50]. A listener-based approach predicts that this change in what it means to be a New Zealander has repercussions for the use of Māori features such as *eh* in Pākehā speech above and beyond those of a simple improvement in attitudes. If both Māori and Pākehā are perceived as *Kiwi*, then the speech of both

[48]The story of Pākehā *eh*-adoption told by a listener-based approach is, of course, a generalization of default behavioral tendencies at the population level. It remains possible for individuals to stray from this default behavior. For example, a speaker may intentionally refrain from using *eh* even with a positive attitude toward Māori for any number of independent reasons, such as not thinking it appropriate to identify with Māori, or disliking the word itself.

[49]A listener-based approach does not rule out the possibility that some speakers may (have) intentionally use(d) *eh* on some occasions in order to express an affiliation with Māori. It simply does not require such intentional usage to be widespread in order to derive the widespread patterns of change.

[50]The word *kiwi* is from the Māori language. It originally referred to the national bird of New Zealand, and is now commonly used with pride to refer to the people of New Zealand, irrespective of ethnicity (see e.g. Onysko & Calude, 2013).

groups yields a common social inference. This common underlying social inference allows the socially-induced perceptual biases for one group to be partially activated by speech from the other group (cf. Szakay et al., 2016). The result is that listeners come to expect – and thus accept – *eh* in the speech of Pākehā more, so that when it does occur, it isn't downweighted in memory as much by residual negativity toward Māori. In other words, if listeners reduce their separation of Māori and Pākehā in terms of socio-cultural ideology, then they should also reduce their separation in terms of perception. By consequence, a listener-based approach predicts that there should be fewer barriers that prevent speakers of either group from successfully using features from the other group, making it easier for *eh* to spread throughout the Pākehā population.

Of course, though the degree of ideological and perceptual separation between Pākehā and Māori may reduce, it is likely that Pākehā and Māori will never be ideologized or perceived as identical. Pākehā and Māori may both be *Kiwi*, but Pākehā are not Māori, and vice-versa; the two groups have different cultural practices and a different sense of self. Consequently, it is not a necessary consequence of a listener-based approach that the two groups will linguistically assimilate to each other totally in the hypothetical eventuality that listeners come to hold identical evaluative attitudes toward them. Rather, given the maintenance of some degree of ideological and perceptual separation, a listener-based approach predicts that the two groups should continue to exhibit some linguistic differences. This prediction can even be seen from simple differences in quantity of experience: if – for whatever reason – Māori generally have more interaction with Māori English speakers than do Pākehā, then changes originating in Māori English will spread faster through the Māori community than they will through the Pākehā community, providing a seed for continued linguistic differentiation between the groups.

Stepping back, the fact that a listener-based approach derives effects of social attitudes on language change through perceptual biases opens the door to new kinds of evidence in studying language change. As attitudes change, not only is language production expected to change, but so is language perception. Thus, differences across age groups in perceptual tasks can be used as evidence for change under the apparent time construct. For example, with the adoption of *eh* by Pākehā over time, representations of *eh* as a feature of Pākehā speech are expected to be stronger among young listeners than among old listeners. Correspondingly, young listeners are expected to exhibit more expectation bias that facilitates the processing of *eh* when uttered by a Pākehā speaker. Such differences in expectation bias can be measured through perceptual tasks such as word monitoring (Marslen-Wilson & Tyler, 1980), and can help to support the claim of change-in-progress.

Furthermore, in the case where language change is related to attitudinal change, the listener-based approach can use perception experiments to recruit evidence for language change that is independent of age-graded patterns. Since the effects of social attitudes on perception are not limited to differences between listeners of different ages, the listener-based approach predicts that variation within an age group in perceptual tasks should be correlated with variation in related social attitudes. For example, the more positive and inclusive a Pākehā listener's attitude toward Māori, the stronger their representation of *eh* as a feature of Pākehā speech, and the more they will exhibit expectation bias toward it from a Pākehā speaker. The quality of the attitude is therefore predicted to be correlated with performance in tasks such as word monitoring, even within a particular age group. If this prediction holds, it provides support for the hypothesized causal mechanism behind change-in-progress, thereby indirectly supporting the claim of change-in-progress itself.

By drawing on additional evidence from perception experiments, a listener-based approach brings new perspectives to the question of whether an age-based difference in production reflects age-grading or change-in-progress. In addition, a listener-based approach makes extremely salient the distinction of *who* uses a certain linguistic feature from *how often* they use it, by emphasizing the parallel distinction between the way linguistic representations are built and updated in perception and the way they are drawn upon in production. Though age-grading and change-in-progress have the same general appearance in overview, they make different predictions from each other for these questions. By using zero-inflated negative binomial regression, which this chapter has added to the quantitative linguistics toolbox, these predictions can be tested on corpus data. Since the specters of age-grading and change-in-progress loom in all age-based analyses of language variation, a listener-based approach promises to be of broad and important influence.

4.7.1 General predictions

Because the listener-based approach is based on general influences of social information and attitudes on speech perception, it can be applied to much more than just the case of Pākehā *eh*-adoption. In particular, it makes some very general predictions about the social circumstances under which language change will be more likely to take off or will spread faster throughout a population. For example, it predicts that the probability or rate with which a change spreads will be influenced by social network structure. Specifically, it predicts that a speaker will be more likely to participate in a change the more they experience tokens reflecting the change as a listener, as that experience is what builds up the representations that they draw upon for production. It further predicts that this process will be facilitated if the experience comes from a variety of speakers, since then the listener is unlikely to attribute the change to an idiosyncrasy and perceptually normalize for it. Moreover, as indicated in Section 4.3, a listener-based approach predicts that a change will spread more quickly and easily the more listeners perceive themselves to be similar to speakers who have undergone the change, and the more positive and inclusive listeners' social attitudes toward speakers who have undergone the change.

These general ideas coalesce in addressing the question of which members of a societally dominant group (e.g. based on ethnicity) will readily pick up linguistic features of a non-dominant group, thereby potentially introducing or facilitating language change across the population of the dominant group. As in the case of lexical adoption of *eh* by Pākehā, the prediction is that the working class constitutes a hotbed for introducing such changes. First and foremost, this prediction follows from social network structure. Since non-dominant groups are typically overrepresented in the working class, members of the dominant group who are in the working class will experience the feature most as listeners, from a wide variety of speakers. As described above, this high degree of exposure will facilitate the strengthening of the feature in the representations they drawn upon for their own productions. The prediction gains further support from the idea that members of the same class are more likely to share experiences in work and life than members of different classes. Members of the dominant group in the working class are thus more likely than those in the middle and upper class to perceive members of the non-dominant group as similar to themselves, and may also be more likely to develop positive and inclusive attitudes toward them, both of which decrease perceptual barriers to adopting their linguistic features.

4.8 Summary

Language change is affected by social information and attitudes. This chapter has focused on how the listener-based approach developed in previous chapters can extend to explain the role of social information and attitudes in language change. In Section 4.1, I grounded this extension by reviewing evidence that social information and attitudes play a major role in language change and yield perceptual asymmetries, just like lexical information such as frequency (Chapter 3). In Section 4.3, I discussed how socially-based asymmetries in experience and perception create asymmetries in socially-based language change, using mechanisms familiar from the model developed in Chapter 2. In Sections 4.4–4.6, I presented new evidence for the spread of the tag particle *eh* in New Zealand from Māori to Pākehā, and I showed how a listener-based approach explains this spread to have been facilitated by improvements in Pākehā attitudes toward Māori.

The extension of the listener-based approach to socially-based language change is a natural one, because linguistic experience – which is taken to underpin linguistic knowledge – occurs within a social context. Consequently, listeners excel at inferring social information and evoking social attitudes about speakers. This chapter has shown how a listener-based approach links the social nature of linguistic experience to the social basis of many linguistic changes in a straightforward, independently-motivated fashion. In this way, it has built on the previous chapters to show that viewing the listener as central to language change has potential to offer a unified explanation of many different patterns in language change.

Chapter 5

Conclusions

This dissertation started from a big-picture question: what is the process through which language changes? More precisely, how do many different influences – both language-internal (e.g. word frequency) and language-external (e.g. social attitudes) – fit together in the complex system of language change? Noting that individuals participate in language change throughout the lifetime, I proposed that language change emerges (at least in part) from the iteration of processes involved in everyday linguistic interactions. In particular, I argued for the centrality of passive but powerful perceptual biases *in the listener*, which may result from language-internal or language-external factors. Through three core chapters, I have developed and applied a framework for this listener-based approach, showing that it solves notable empirical and theoretical puzzles.

In Chapter 2, I grounded the approach in a formal computational model of regular sound change, in which perceptual biases mediate the updating of phoneme representations that listeners draw upon when they come to talk. This model paves the way for testing causal hypotheses that use asymmetries observed in speech perception to explain asymmetries observed in language change. It is also the first model of its kind to successfully generate movement of phoneme categories in an acoustic space while maintaining their shapes and large degrees of overlap, as observed in empirical sound changes. It therefore constitutes a valuable contribution to the field, both theoretically and practically.

In Chapter 3, I applied the model to the effects of word frequency on rates of sound change. Only in recent years have empirical studies of such effects been possible, but they have already raised a major puzzle for linguistic theory, with different effects of frequency observable in different kinds of sound change. I showed that the listener-based model provides a solution to this puzzle: from a single perceptual bias, which is supported by numerous psycholinguistic experiments, it generated the effects seen in all existing empirical studies. This result underscores the strong theoretical contribution that a listener-based approach can make to the study of language change. It also makes an important contribution by vindicating usage-based approaches more generally, which have been criticized in the literature under the mistaken assumption that they always predict high-frequency words to change fastest.

Finally, in Chapter 4, I sketched an extension of the listener-based approach to language change beyond sound change, under the influence of language-external factors. I applied this extension to the adoption of the Māori-associated discourse tag *eh* by young Pākehā, where it motivated the new analytical technique of zero-inflated negative binomial regression, a major practical contribution to the quantitative linguistics toolbox. Using this technique, I solved an established puzzle in the literature by providing good evidence that Pākehā *eh*-adoption reflects change-in-progress rather than age-grading. Furthermore, I showed that the listener-based approach explains why the change is occurring only now, as social attitudes toward Māori have only recently improved, bringing associated changes in experimentally-established perceptual biases. This finding contributes to (socio)linguistic theory by confirming that social attitudes affect language change, and by highlighting the interrelated dynamicity of attitudes and language.

There are three big-picture theoretical takeaways from this dissertation. Firstly, since an individual's use of language changes throughout their lifetime, a primary mechanism of language change concerns the cycle between language use and language representations. Secondly, the listener plays an integral role in this cycle, and is therefore central to language change. And thirdly, as a result of this centrality, perceptual biases in the listener can explain patterns in language change; it is not necessary to rely heavily on the speaker.

There are also three big-picture methodological takeaways from this dissertation. Firstly, usage-based approaches to language change have typically focused heavily on word frequency effects, but word frequency effects draw upon just one kind of perceptual bias; many other kinds also exist, stemming from both language-internal and language-external factors, and they remain a largely untapped potential. Secondly, when testing hypotheses about the *cause* of language change, computational modeling is an extremely useful tool; indeed, it is one of the few ways in which empirically-observed language change can be 'rerun' under different conditions, giving rise to controlled investigation. And thirdly, the combination of computational modeling with corpus studies and experiments provides a cyclic approach to the study of language change, with each method making predictions for, and testing predictions from, the others.

By weaving together multiple methodologies, a listener-based approach allows us to triangulate on answers to longstanding questions about language change. I have begun to answer some of these questions in this dissertation, but many more remain for future research, and many new questions exist as well. For example, to what extent do empirical word frequency effects on rates of sound change match those predicted by the typology in Chapter 3? Does the explanatory value of perceptual biases remain when more of the complexity of language change is taken into account, or when speaker-based but listener-oriented production biases are also considered?

And how can a listener-based approach explain phenomena in language change other than those described in this dissertation? Each of these questions provides a different test for the listener-based approach.

One way in which future work may test the listener-based approach is through its predictions. For example, it may address how well empirical word frequency effects on rates of sound change match those predicted by the typology in Chapter 3. Since this typology is probabilistic, testing it fully will require the accumulation of many corpus studies. Some of these studies may assess the replicability of word frequency effects in additional examples of the kinds of sound change that have already been analyzed in the literature, such as the fronting of /a/ (a component of a push chain) and merger of /ɛə/ with /iə/ in New Zealand English. Others may look for word frequency effects in kinds of sound change that have not yet been analyzed with respect to rate of change, such as the tense-lax /æ/ split in Philadelphia English. Studies of this latter sort will be particularly valuable, as they test situations in which a listener-based approach makes different predictions to an alternative approach that also draws heavily on the speaker.

Another way in which future work may test the listener-based approach is by assessing its robustness to complexity. In this dissertation, I have demonstrated the explanatory value of perceptual biases for language change under many simplifying assumptions, but it remains to be seen if this value holds for a more complex and realistic picture of language change. For example, does the frequency-based perceptual bias discussed in Chapter 3 still make viable predictions for kinds of sound change that are not limited to interactions between at most two phoneme categories along a single acoustic dimension, or that include a role for articulatory limitations? Similarly, do the socially-based perceptual biases discussed in Chapter 4 accurately predict observed social influences on language change, given the complexity of the social world? Finally, does the explanatory value of perceptual biases remain when the speaker has access to biases that modulate their productions for the listener's benefit? Answering these questions will require revisions to the model from Chapter 2, but doing so will unlock much broader empirical predictions than are presently available.

A final way in which future work may test the listener-based approach is by assessing how readily it extends to phenomena in language change beyond the cases presented in this dissertation. One promising avenue for such extension is the Functional Load Hypothesis (Martinet, 1952), which states that the probability of a phonological distinction being lost over time is related to the amount of 'work' that distinction does in preventing ambiguity between minimal pairs. The listener-based approach provides a mechanism through which the Functional Load Hypothesis can operate, by downweighting the influence of ambiguous tokens on the listener's linguistic representations (formally represented in the model of Chapter 2 by the discriminability evaluation; see Appendix C). Crucially, this mechanism is the same as the one that generated word frequency effects on rates of sound change in Chapter 3. Extending the listener-based approach to account for new phenomena using the same mechanisms will underscore its potential as a comprehensive theory of language change, and will also offer a new perspective on how different phenomena in language change may be related to each other.

Taking a listener-based approach to language change provides new perspectives not just on language change, but also on the relationship between language change and other areas. For example, it may lead future work both to draw new insights for and from speech perception, and to revise assumptions about language evolution.

One way in which a listener-based approach to language change may offer new insights is through bringing together speech perception and language change. Creating a close relationship between these two areas allows each to offer new insights and predictions for the other. For example, recent work in speech perception has suggested that social stereotypes are recruited in the construction of linguistic memories, somewhat independently of actual linguistic experience (Mengesha, Todd, & Sumner, 2019). Under a listener-based approach to language change, this finding makes the novel suggestion that language change may be constrained by social stereotypes. Crucially, this suggestion is empirically testable, as it predicts that a linguistic variable that is strongly associated with a social stereotype should resist being spread across social groups moreso than a variable that is not associated with stereotypes. In the other direction, from the existence of word frequency effects in sound change that do not depend upon minimal pairs, the listener-based approach derives the existence of a corresponding bias on recognition that should apply even in the absence of minimal pairs (Chapter 3). Such a bias is overlooked by most existing models of speech perception, in which frequency effects serve to modulate the competition in recognition between real words only. The existence of this bias is also testable, through the prediction that the Ganong effect – the degree to which an acoustically ambiguous token is biased toward being interpreted as a real word over a nonword (Ganong, 1980) – should be sensitive to word frequency. By making connections between speech perception and language change, the listener-based approach raises new, testable predictions for the theory of each.

The listener-based approach to language change may also lead future work to revise assumptions about language evolution, by analogy to language change. In recent years, research programs have begun to systematically explore the factors that may have influenced the way in which linguistic structure arose from pre-linguistic communication signals. A tacit assumption underlying the computational (e.g. Nowak & Krakauer, 1999) and experimental (e.g. Kirby et al., 2008) methods used to conduct this exploration is that structure is imbued by the language *learner*. According to this assumption, *inductive* biases cause a child's communication system to deviate in systematic ways from that which they are exposed to by their parents, to better fit with principles of cognitive organization. These biases are similar to those that have previously been assumed to underlie language change, through repeated language acquisition, and thus it has been proposed that the study of language change may to some extent inform the study of language evolution (e.g. Kiparsky, 1976). If we are to take this proposal seriously, then the notion that the listener plays a large role in change implies that we might also expect the listener to play a large role in evolution, opening exciting new methodological and theoretical possibilities for the study of language evolution.

Taken together, this dissertation lays a solid foundation for a theory of language change in which the listener is central. In proposing that the listener plays a large role in the process of language change, it provides a system in which many influences on language change fit together, from factors both language-internal (e.g. word frequency) and language-external (e.g. social attitudes). It opens the door to new ways of studying language change, drawing together methods and insights from speech perception experiments, computational modeling, and diachronic corpus studies. By fusing these insights in a cyclic listener-based approach, this dissertation provides a way to continually triangulate on an ever-stronger theory of the complex system of language change.

Appendix A

Model initialization details

The model in Chapter 2 is initialized by providing an initial exemplar cloud for each category, i.e. a distribution of exemplars across a set of types. There are three components to the initial exemplar cloud for each category:

- A set of types, whose frequencies follow a specified type-frequency distribution.
- A set of exemplars, whose acoustic values follow a specified acoustic distribution. The total number of exemplars is equal to the sum of type frequencies for the category.
- An assignment of exemplars to types, such that a given type of frequency f has f exemplars, in accordance with the multiple-trace hypothesis (Hintzman & Block, 1971).

I generated a distinct set of 92 types for each category, so that the system did not include minimal pairs. I used the same type-frequency distribution for each category. This distribution was based on the distribution of word log-frequencies in COCA: The Corpus of Contemporary American English (Davies, 2008–) (see Figure A.1).

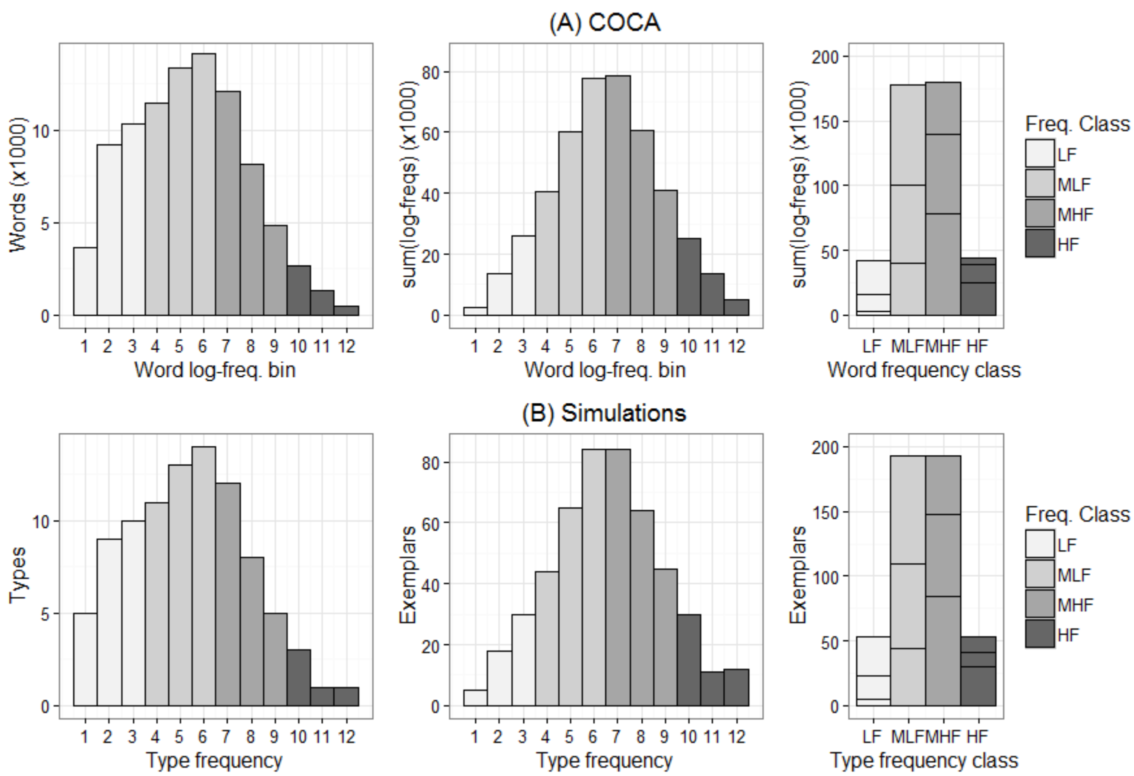

Figure A.1: Frequency in a corpus and in the model. (A) The distribution of word frequencies in COCA (Davies, 2008–). Panels show the number of unique words (left) and the sum of word log-frequencies (middle) in each word log-frequency bin, and the sum of word log-frequencies in each of four word frequency classes (right), from *low-frequency* (bins 1–3; lightest gray) to *high-frequency* (bins 10–12; darkest gray). I calculated log-frequencies using the natural logarithm and binned them by rounding up, then excluded words occurring extremely infrequently (log-frequency 0) or extremely frequently (log-frequency > 12). (B) I modeled the distribution of type frequencies in the simulations on the distribution observed in COCA, equating words with types, and type frequency (and thus number of exemplars per type, following the multiple-trace hypothesis (Hintzman & Block, 1971)) with word log-frequency. This yielded notable symmetry across frequency classes (right panel): there are as many exemplars of low-frequency types as there are of high-frequency types.

As a consequence of using corpus log-frequencies, I obtained the same number of exemplars of high-frequency types (with few types and many exemplars per type) as exemplars of low-frequency types (with many types and few exemplars per type). This means that, in the simulations presented here, the set of exemplars of all high-frequency types and the set of exemplars of all low-frequency types are both updated at the same rate;[51] any observed effects of type frequency are thus based on differences in the way that types are processed, not on differences in sheer quantity of exemplars.

For each category, I generated a set of 492 exemplars with acoustic values following a raised-cosine distribution. I used a raised-cosine distribution because its short tails are less susceptible to iterated sampling error than the long tails of a normal distribution, and thus are more robust to the effects of discriminability and typicality evaluation. I sampled from this distribution in such a way that the sample of exemplars of high-frequency types was identical to the sample of low-frequency types (and similarly for mid-high- and mid-low-frequency types). This sampling strategy means that there is no frequency asymmetry in the initial conditions of the model; any observed effects of type frequency over time are thus based on differences in the way that types are processed, not on differences in initial conditions. Finally, I assigned identical (but displaced) initial exemplar sets to each category, to ensure that any observed differences between categories are due to category interaction and not initial conditions.

The process I used to sample the initial exemplar sets was as follows. First, I drew 246 values from the raised cosine distribution whose probability density function is

[51]In a given period of time, a given high-frequency type is produced and perceived more than a given low-frequency type, but high-frequency types *in aggregate* are produced and perceived the same number of times as low-frequency types *in aggregate*, since there are many fewer high-frequency types than low-frequency types (Zipf, 1935).

Table A.1: Initial exemplar distribution statistics for type frequency classes.

Freq.	N	Mean	SD	Skew	Ex. Kurtosis
(All)	492	−0.013	0.998	0.109	−0.717
HF	53	−0.009	0.999	0.097	−0.770
MHF	193	−0.013	0.998	0.112	−0.702
MLF	193	−0.013	0.998	0.112	−0.702
LF	53	−0.009	0.999	0.097	−0.770

given in Equation (A.1) and rounded them to the nearest 0.1.

$$r(v) = \begin{cases} \frac{c}{2}\left(1+\cos\left(c\pi v\right)\right) & \text{if } v \in \left[\frac{-1}{c}, \frac{1}{c}\right] \\ 0 & \text{otherwise} \end{cases} \tag{A.1}$$

where

$$c = \sqrt{\frac{1}{3} - \frac{2}{\pi^2}} \tag{A.2}$$

I re-drew values until I obtained a sample with mean approximately equal to zero, standard deviation approximately equal to 1, and low skewness. I then split this sample into two subsets of 53 and 193, ensuring that the statistics for each subset were approximately equal to the statistics across the entire sample. I made two copies of each subset and designated them to classes of types on the basis of type frequency; I designated a copy of the first subset (with 53 exemplars) for each of the high- and low-frequency classes, and a copy of the second subset (with 193 exemplars) for each of the mid-high- and mid-low-frequency classes. The statistics for the sample and frequency-class subsets are given in Table A.1.

I used this distribution of acoustic values as the basis of all simulations with the model, adjusting it as required by the parameters of the simulation. To create initial categories of width σ, I scaled the acoustic value of every exemplar by σ. To create an initial category distance of μ between the categories, I subtracted μ from the (scaled) acoustic value of each of the exemplars in the Pusher. Following any adjustments, I re-rounded the acoustic values to the nearest 0.1.

In each run of the model, I assigned the exemplars for a given frequency class to types in that frequency class at random, ensuring that a given type of frequency f had f exemplars. This random assignment means that the results of many runs of the model with a given set of parameter values reflect the dynamics expected on average under that set of parameter values, independent of the effects of initial allocation of exemplars to types.

Appendix B

Model tuning details

The model tuning process described in Chapter 2 (Section 2.4.2) required pre-determining values for a number of parameters and exploring a range of values for other parameters. In this appendix, I describe the parameter values involved in the tuning process, as well as the category-level properties that resulted from successful tuning.

B.1 Parameter values

For both phonetic drift and push chains, I pre-determined three values for σ: $\sigma = 0.6$ (narrow categories), $\sigma = 0.8$ (medium-width categories), and $\sigma = 1.0$ (wide categories). In each case, I set $\alpha = \sigma/2$: for $\sigma = 0.6$, I set $\alpha = 0.3$; for $\sigma = 0.8$, I set $\alpha = 0.4$; and for $\sigma = 1.0$, I set $\alpha = 0.5$.

For phonetic drift, I also pre-determined three values for β: $\beta = 0.05$ (weak bias), $\beta = 0.15$ (medium-strength bias), and $\beta = 0.25$ (strong bias). I then explored 10 values for each of τ and ι. The values of τ ranged from 0.02 to 0.20 in steps of 0.02, representing a requirement for the activation incited by a token to be 2–20% of the maximum possible in order to be stored with probability 0.5. The values of ι ranged from 0.1 to 1.0 in steps of 0.1, representing a degree of imprecision that could shift the target by up to 6–100% of the span of a category in either direction (depending also on the width of the category, σ).

For push chains, in addition to the previously-described pre-determination of σ and α, I pre-determined two values for μ for each value of σ. For $\sigma = 0.6$, I set $\mu \in \{2.1, 1.9\}$; for $\sigma = 0.8$, I set $\mu \in \{3.0, 2.8\}$; and for $\sigma = 1.0$, I set $\mu \in \{3.9, 3.7\}$. I also pre-determined three values for δ: $\delta = 0.25$ (weak discriminability force), $\delta = 0.50$ (medium-strength discriminability force), and $\delta = 0.75$ (strong discriminability force). I retained the value of τ selected by tuning for phonetic drift and explored 4 values for ι for each value of σ, which yielded between 0–50% increase in category width during phonetic drift. For $\sigma = 0.6$, I chose $\iota \in \{0.4, 0.5, 0.6, 0.7\}$; for $\sigma = 0.8$, I chose $\iota \in \{0.5, 0.6, 0.7, 0.8\}$; and for $\sigma = 1.0$, I chose $\iota \in \{0.6, 0.7, 0.8, 0.9\}$. Finally, I explored 25 values for β, which ranged from 0.01 to 0.25 in steps of 0.01 and represented a consistent bias approximately equal to 0.2–8.5% of the span of a category (depending also on the width of the category, σ).

B.2 Results

The parameter values chosen by the tuning process for phonetic drift are summarized in Table B.1, together with the category-level properties they gave rise to (compare to the initial category properties in Table A.1).

The parameter values chosen by the tuning process for push chains are summarized in Table B.2, and the category-level properties they gave rise to are summarized in Table B.3 (compare to the initial category properties in Table A.1).

Table B.1: Phonetic drift tuning results. Tuned parameter values and average category properties for a single category after 5000 iterations. Displacement measures the distance traveled by the category centroid.

Parameters					Category properties			
σ	β	ι	α	τ	Displacement	Width	Skew	Ex. Kurtosis
0.6	0.05	0.3	0.3	0.10	0.35	0.59	−0.12	−0.48
0.8	0.05	0.4	0.4	0.10	0.35	0.79	−0.09	−0.46
1.0	0.05	0.5	0.5	0.10	0.32	0.97	−0.07	−0.49
0.6	0.15	0.3	0.3	0.10	1.03	0.61	−0.33	−0.20
0.8	0.15	0.4	0.4	0.10	1.06	0.79	−0.23	−0.34
1.0	0.15	0.5	0.5	0.10	1.04	0.99	−0.21	−0.37
0.6	0.25	0.3	0.3	0.10	1.68	0.64	−0.47	0.21
0.8	0.25	0.4	0.4	0.10	1.69	0.83	−0.35	−0.08
1.0	0.25	0.5	0.5	0.10	1.72	1.01	−0.28	−0.26

Table B.2: Tuned parameter values for push chains.

Set	σ	μ	β	ι	α	δ	τ
(1)	0.6	2.1	0.08	0.5	0.3	0.25	0.10
(2)	0.6	1.9	0.10	0.5	0.3	0.25	0.10
(3)	0.8	3.0	0.06	0.6	0.4	0.25	0.10
(4)	0.8	2.8	0.08	0.6	0.4	0.25	0.10
(5)	1.0	3.9	0.05	0.7	0.5	0.25	0.10
(6)	1.0	3.7	0.07	0.7	0.5	0.25	0.10
(7)	0.6	2.1	0.12	0.5	0.3	0.50	0.10
(8)	0.6	1.9	0.16	0.5	0.3	0.50	0.10
(9)	0.8	3.0	0.08	0.6	0.4	0.50	0.10
(10)	0.8	2.8	0.12	0.6	0.4	0.50	0.10
(11)	1.0	3.9	0.12	0.8	0.5	0.50	0.10
(12)	1.0	3.7	0.15	0.8	0.5	0.50	0.10
(13)	0.6	2.1	0.20	0.6	0.3	0.75	0.10
(14)	0.6	1.9	0.25	0.6	0.3	0.75	0.10
(15)	0.8	3.0	0.16	0.7	0.4	0.75	0.10
(16)	0.8	2.8	0.21	0.7	0.4	0.75	0.10
(17)	1.0	3.9	0.14	0.8	0.5	0.75	0.10
(18)	1.0	3.7	0.19	0.8	0.5	0.75	0.10

Table B.3: Push chain tuning results. Average properties of the interactions obtained under the sets of parameter values in Table B.2 after 5000 iterations. Overlap measures the span of the overlapping region between categories (i.e. the distance between the most advanced Pusher exemplar and the least advanced Pushee exemplar). Pushee displacement measures the distance traveled by the Pushee centroid (i.e. the size of the push).

			Pushee				Pusher		
Set	Category dist.	Overlap	Displacement	Width	Skew	Ex. Kurtosis	Width	Skew	Ex. Kurtosis
(1)	2.09	0.80	0.08	0.62	0.12	−0.56	0.61	−0.17	−0.54
(2)	1.91	0.96	0.11	0.61	0.11	−0.54	0.60	−0.18	−0.52
(3)	3.01	0.85	0.07	0.81	0.10	−0.57	0.81	−0.12	−0.56
(4)	2.81	1.00	0.10	0.80	0.11	−0.55	0.80	−0.15	−0.55
(5)	3.91	0.88	0.05	1.01	0.10	−0.57	1.00	−0.09	−0.57
(6)	3.69	1.03	0.07	1.00	0.11	−0.57	0.99	−0.12	−0.56
(7)	2.08	0.68	0.12	0.60	0.15	−0.55	0.58	−0.23	−0.50
(8)	1.90	0.83	0.17	0.59	0.14	−0.53	0.57	−0.26	−0.45
(9)	3.02	0.67	0.11	0.80	0.13	−0.57	0.79	−0.17	−0.56
(10)	2.80	0.82	0.14	0.78	0.14	−0.55	0.77	−0.21	−0.51
(11)	3.87	0.80	0.11	1.02	0.14	−0.57	1.00	−0.18	−0.55
(12)	3.68	0.95	0.15	1.01	0.14	−0.57	0.99	−0.21	−0.53
(13)	2.10	0.67	0.18	0.62	0.17	−0.54	0.59	−0.30	−0.45
(14)	1.94	0.83	0.23	0.61	0.15	−0.51	0.57	−0.32	−0.39
(15)	2.99	0.68	0.17	0.81	0.16	−0.55	0.79	−0.25	−0.50
(16)	2.79	0.83	0.21	0.80	0.16	−0.54	0.77	−0.28	−0.46
(17)	3.89	0.67	0.14	1.00	0.14	−0.58	0.99	−0.21	−0.54
(18)	3.67	0.81	0.19	0.99	0.15	−0.57	0.96	−0.25	−0.50

Appendix C

Minimal pairs in the model

The simulations presented in Chapters 2 and 3 do not include minimal pairs, under the claim that minimal pairs alone are unlikely to drive the effects that I am concerned with modeling. In this appendix, I present evidence to support this claim.

The claim that minimal pairs alone are unlikely to drive regular sound change takes empirical support from the interaction of /æ/ and /ɛ/ in the New Zealand short front vowel shift. In the ONZE corpus (Gordon et al., 2007), only 8.2% of words containing /æ/ or /ɛ/ (164 of 2,000 unique wordforms) have a relevant non-proper-noun minimal partner that also appears in the corpus. These minimal pairs are distributed across the frequency range and account for 21.1% of the total tokens (11,620 of 55,200) analyzed by Hay et al. (2015). Since the vast majority of the New Zealand English /æ/-/ɛ/ data (words and tokens) correspond to words without a relevant minimal partner,[52] the properties of the vowel interaction are likely to be general, holding across words both with and without relevant minimal partners.

The claim that minimal pairs are unnecessary *in the model* can be supported by

[52] The number of words in the dataset with potential relevant non-proper-noun minimal partners (as assessed by the Unisyn lexicon (Fitt, 2000)) that happened not to be mentioned (e.g. through not being of the appropriate register or through being extremely low-frequency) is capped at 15.4% (308), accounting for no more than 37.1% (20,501) of the tokens. Even in this more extreme interpretation of the data, minimal pairs are in a minority.

comparing the results of simulations with and without minimal pairs. The claim can be broken into two sub-claims: firstly, that minimal pairs do not make necessary contributions to any part of modeling empirically-observed category movements and frequency effects; and secondly, that minimal pairs alone are not sufficient to generate these movements and effects. In the following sections, I present an extension of the model to include minimal pairs, and I use it to derive support for each of these sub-claims.

C.1 Modeling minimal pairs

There are two options for introducing minimal pairs into the model, differentiated based on the assumed influence of higher-level (syntactic, semantic, pragmatic, or discourse) context. Under the first, context-sensitive option, higher-level context has a large influence: it uniquely determines the intended type even when the phonological frame is consistent with multiple types. The listener operates as in the present model, storing the token as an exemplar of the intended type if it passes the discriminability and typicality evaluations, and discarding it otherwise. Consequently, there is no potential for *variant trading* (Blevins & Wedel, 2009), where the listener mistakenly stores a token of one type as an exemplar of another type. Under the second, context-insensitive option, higher-level context has no influence: when the phonological frame is consistent with multiple types, the context can never uniquely determine which type was intended. The listener considers all possible types that are consistent with the phonological frame, including those corresponding to nonwords, and chooses one probabilistically based on category activation and type frequency. The token is stored as an exemplar of the winning type if it passes the typicality evaluation and if the type corresponds to a real word; thus, variant trading is possible.

Within the framework of the model, the context-sensitive option is more conservative because it does not permit variant trading; in every other respect, the two options are mathematically equivalent. Without variant trading, all that introducing minimal pairs does is effectively raise the discriminability threshold. Recall that the threshold was stated to be low ($\delta < 1$) due to the existence of lexical bias (Ganong, 1980) toward the intended type, making it relatively easy to map an acoustically ambiguous token to the intended type. With the introduction of minimal pairs, the unintended type can also have such a lexical bias, introducing another plausible identity for an acoustically ambiguous token, and thus making it harder to map such a token to the intended type. As shown mathematically in Equation (C.4), this counterveiling pressure effectively raises δ. Since I have already explored the role of different discriminability thresholds in the main text without invoking minimal pairs, introducing minimal pairs under the context-sensitive option would not yield any new insight. For this reason, I chose to introduce minimal pairs under the context-insensitive option, allowing me to explore anew the influence of variant trading on the model's results.

The introduction of context-insensitive minimal pairs embraces a view of the discriminability evaluation as a probabilistic recognition process involving competition between the types (from different categories) that are consistent with a given frame. I assume that a token is identified as belonging to the category that wins the discriminability evaluation, and that the discriminability evaluation fails just in case this identification yields a nonword. More precisely, the discriminability evaluation (repeated in Equation (C.1)) is replaced with a process of *recognition* of the token (Equation (C.2)), where a δ value is computed as before for each category according to the corresponding type frequency (Equation (C.3)). The token is passed to the typicality evaluation if and only if it has been recognized as corresponding to a real word.

$$P(\text{pass discriminability evaluation}|A_i, A_o) = \frac{\frac{1}{\delta} \cdot A_i}{\frac{1}{\delta} \cdot A_i + 1 \cdot A_o} \tag{C.1}$$

$$P(\text{token recognized as } T_k) = \frac{\frac{1}{\delta_k} \cdot A_k}{\sum_k \frac{1}{\delta_k} \cdot A_k} \tag{C.2}$$

$$\delta_k = \begin{cases} \left[\lambda + \left(\frac{2(f_k-1)}{M-1} - 1\right)\phi\right]_0^1 & T_k \text{ corresponds to a real word} \\ 1 & T_k \text{ corresponds to a nonword} \end{cases} \tag{C.3}$$

When two competing types both correspond to real words, the recognition equation can be explicitly written out as follows:

$$\begin{aligned} P(\text{token recognized as } T_i) &= \frac{\frac{1}{\delta_i} \cdot A_i}{\frac{1}{\delta_i} \cdot A_i + \frac{1}{\delta_o} \cdot A_o} && \text{(C.4)} \\ &= \frac{\frac{\delta_o}{\delta_i} \cdot A_i}{\frac{\delta_o}{\delta_i} \cdot A_i + 1 \cdot A_o} && \text{(C.5)} \\ &= \frac{\frac{1}{\delta_i/\delta_o} \cdot A_i}{\frac{1}{\delta_i/\delta_o} \cdot A_i + 1 \cdot A_o} && \text{(C.6)} \end{aligned}$$

It can be seen that this form is equivalent to Equation (C.1), with $\delta = \delta_i/\delta_o$. Since $\delta_o < 1$, it follows that this new version of δ is increased from the original value of δ_i it would take were there no real word competitor, and hence that the introduction of minimal pairs effectively raises the discriminability threshold.

C.2 Minimal pairs are not necessary: simulations with a subset of minimal pairs

Given the extension of the model to include minimal pairs, I begin by questioning whether minimal pairs are necessary for generating any desirable pattern in simulations of push chains. I compare simulations with and without minimal pairs to see whether the addition of minimal pairs makes new results possible or existing results impossible.

In each model run, I randomly changed 10% of the types to participate in minimal pairs, as an approximation to the proportion of minimal pairs in the /æ/–/ɛ/ interaction in New Zealand English. To accomplish this, I randomly chose 10 types from each category for each run and created minimal pair relations between them. This random pairing process ensured that there would be no confound between type frequency and minimal pair existence in the results.

The addition of minimal pairs allows categories to stably exist in closer proximity to one another, since tokens that would otherwise fail the discriminability evaluation instead participate in variant trading. To ensure that this decreased stable distance between categories did not disrupt the interpretation of the simulation results, I retuned the initial category distance (μ) using frequency-insensitive discriminability thresholds (δ). Keeping all other parameters as in Table B.2, I obtained the following: for $\sigma = 0.6$, I set $\mu \in \{1.8, 2.0\}$; for $\sigma = 0.8$, I set $\mu \in \{2.6, 2.8\}$; and for $\sigma = 1.0$, I set $\mu \in \{3.4, 3.6\}$.

For each of the parameter settings in Table B.2 (with μ retuned as above), I ran 1000 simulations with 10% minimal pairs for 5000 iterations each. The average properties obtained from these simulations are given in Table C.1. As can be seen, the properties obtained with 10% minimal pairs (Table C.1) are highly similar to those obtained without minimal pairs (Table B.3), indicating that the same kinds of stable category interactions are generated with and without minimal pairs.

It is possible that the similarities between the simulations with and without minimal pairs hold only at the coarse-grained category level and not at the fine-grained type level. To assess this possibility, I repeated the investigation of frequency effects from Chapter 3 with the inclusion of 10% minimal pairs. Using the 15 δ functions from Chapter 3 and the 18 sets of parameter values with retuned μ, I ran the model 1000 times for 5000 iterations each. In Figure C.1, I compare frequency effects in the Pushee for models with and without minimal pairs for parameter sets (4), (10), and

Table C.1: Properties of push chains with 10% minimal pairs. Average properties of the categories in models with 10% minimal pairs under the sets of parameter values in Table B.2 (with μ retuned) after 5000 iterations. Overlap measures the span of the overlapping region between categories (i.e. the distance between the most advanced Pusher exemplar and the least advanced Pushee exemplar). Pushee displacement measures the distance traveled by the Pushee centroid (i.e. the size of the push).

				Pushee				Pusher		
Set	μ	Category dist.	Overlap	Displacement	Width	Skew	Ex. Kurtosis	Width	Skew	Ex. Kurtosis
(1)	2.0	1.99	0.89	0.08	0.62	0.12	−0.55	0.61	−0.18	−0.53
(2)	1.8	1.80	1.07	0.11	0.61	0.12	−0.53	0.59	−0.19	−0.49
(3)	2.8	2.82	1.02	0.07	0.81	0.11	−0.56	0.80	−0.14	−0.56
(4)	2.6	2.62	1.16	0.09	0.80	0.12	−0.55	0.79	−0.16	−0.53
(5)	3.6	3.63	1.12	0.06	1.00	0.11	−0.57	1.00	−0.11	−0.57
(6)	3.4	3.42	1.28	0.08	0.99	0.12	−0.56	0.98	−0.13	−0.55
(7)	2.0	1.99	0.77	0.12	0.60	0.15	−0.54	0.58	−0.25	−0.49
(8)	1.8	1.80	0.94	0.16	0.59	0.13	−0.52	0.57	−0.27	−0.42
(9)	2.8	2.85	0.85	0.11	0.79	0.14	−0.56	0.78	−0.18	−0.54
(10)	2.6	2.64	1.01	0.15	0.78	0.14	−0.54	0.76	−0.22	−0.50
(11)	3.6	3.62	1.04	0.13	1.01	0.14	−0.56	0.99	−0.19	−0.54
(12)	3.4	3.43	1.18	0.17	0.99	0.14	−0.55	0.97	−0.22	−0.52
(13)	2.0	2.01	0.77	0.18	0.61	0.17	−0.54	0.59	−0.31	−0.43
(14)	1.8	1.85	0.94	0.23	0.60	0.15	−0.49	0.57	−0.32	−0.37
(15)	2.8	2.83	0.85	0.17	0.81	0.16	−0.56	0.78	−0.26	−0.49
(16)	2.6	2.63	1.01	0.22	0.79	0.16	−0.52	0.76	−0.28	−0.45
(17)	3.6	3.64	0.94	0.16	1.00	0.15	−0.56	0.97	−0.22	−0.52
(18)	3.4	3.43	1.10	0.22	0.98	0.15	−0.55	0.95	−0.25	−0.48

(16) (with μ retuned as above).

The models with minimal pairs show the same broad patterns in frequency effects as the models without minimal pairs: when high-frequency types are sufficiently perceptually advantaged relative to low-frequency types (with respect to discriminability, δ), they become more likely to cluster in the overlapping region between categories, allowing low-frequency types in the Pushee to change at a faster rate. Wherever a robust frequency effect of this sort exists in the model without minimal pairs, it also exists in the model with minimal pairs.

However, the addition of minimal pairs also exacerbates the existence of reversed frequency effects for some δ functions (lower-left panels of Figure C.1), where high-frequency types in the Pushee change at a faster rate than low-frequency types. As

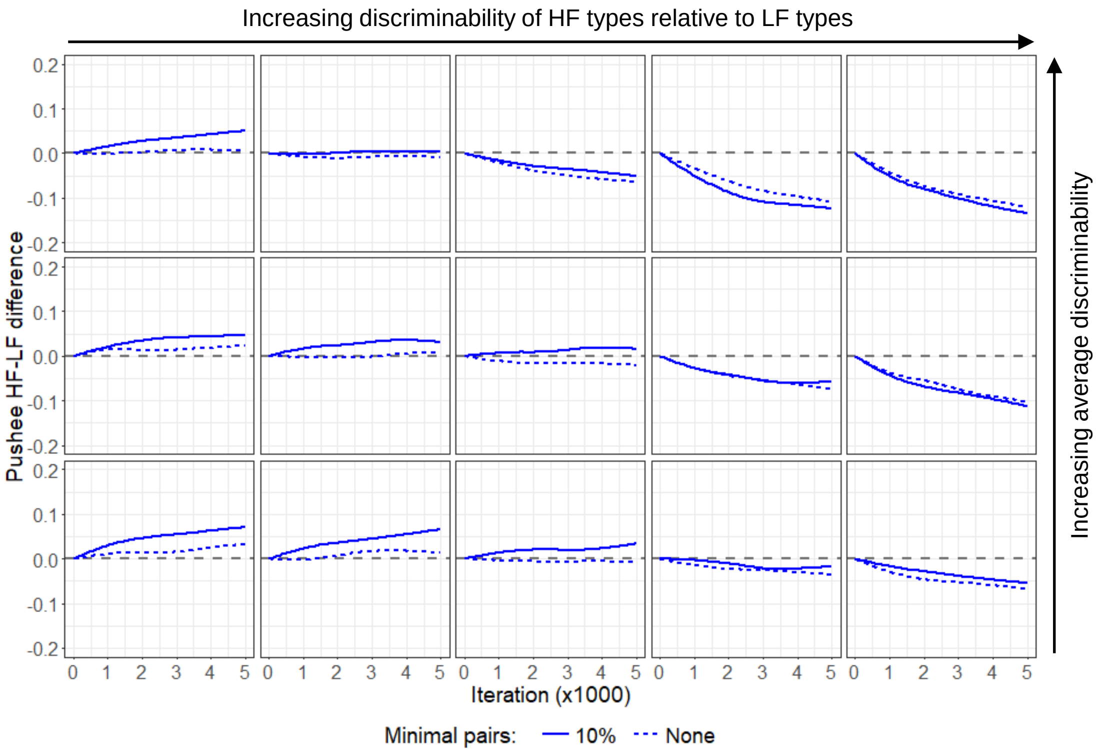

Figure C.1: Frequency effects with and without minimal pairs. Results of varying discriminability threshold (δ) with type frequency for 3 different sets of parameter values (1 per row), comparing a system with 10% minimal pairs (solid lines) to a system with no minimal pairs (dotted lines). The figure is laid out in the same way as Figure 3.1. The system with minimal pairs shows the same patterns as the system without minimal pairs, with increasing discriminability of high-frequency types relative to low-frequency types (movement from left to right across columns) causing slower change of high-frequency types than of low-frequency types in the Pushee (negative-sloping sections). However, it also shows reversed effects when there is little or no difference in discriminability between high- and low-frequency types (left columns), unlike the system without minimal pairs.

discussed in Section 3.4.2, these reversed frequency effects are the result of an interaction between my assumptions about production and storage, which causes high-frequency types to be more sensitive to perceptual forces than low-frequency types. Because the addition of minimal pairs allows for categories to stably exist closer to one another, it increases the discriminability force.[53] High-frequency types are more sensitive to this increased discriminability force, causing them to be pushed apart moreso than low-frequency types in the absence of a countervailing perceptual asymmetry. I consider the size of the reversed frequency effect not to qualify as a meaningful difference between the models with and without minimal pairs, since the effect is an artifact of the simplified assumptions about production and storage, and since it only occurs in situations where an empirically-supported perceptual asymmetry is not present.

The addition of minimal pairs thus does not meaningfully affect the model's results. Models both with and without minimal pairs are equally capable of generating push chains displaying key empirically-observed properties, such as the maintenance of category width and overlap. Furthermore, given sufficiently strong perceptual asymmetries, models both with and without minimal pairs generate word-frequency effects of the kind observed in documented sound changes. I conclude that minimal pairs are not necessary for generating any desirable pattern in simulations of push chains.

[53] To see why decreased category distance results in increased discriminability force, consider a token at the edge of the intended category, in the overlapping region. The discriminability force is a function of the number of exemplars of the other category contained within the activation window around this token; more exemplars from the other category provide more competition during the discriminability evaluation, yielding a larger discriminability force. The closer the categories are, the closer the edge of the intended category will be to the centroid of the other category, and thus the more exemplars from the other category there will be in the activation window.

C.3 Minimal pairs are not sufficient: simulations with only minimal pairs competing

Given that the model's key results can be obtained both with and without minimal pairs, I next ask whether they can be obtained if minimal pairs alone underpin push chains. I conduct simulations varying the degree to which types with and without minimal partners contribute to category interaction, to see whether minimal pairs alone are sufficient for generating desirable category movements and frequency effects.

Since the model assumes that the phonological frame is perfectly perceived, it assumes that recognition of a type involves competition only between types with the same phonological frame, i.e. between two real words in a minimal pair or between a real word and a nonword. Given this assumption, to say that minimal pairs alone underpin category interaction is to say that types corresponding to nonwords do not compete with types corresponding to real words for recognition. This lack of nonword type competition is a tacit assumption in existing models of spoken word understanding (e.g. Norris & McQueen, 2008).[54] In the model presented in the main text, I assume that nonword types compete to the degree that would be expected based on their category activation alone, by setting the default value of δ for nonwords to 1. Here, I relax this assumption by increasing the default value of δ; the larger the value, the less nonword types compete, and thus the more minimal pairs carry the burden of category interaction.

To control the extent to which nonword types compete with real words during the recognition process, I introduce a new parameter, χ. χ is a scale factor that multiplies

[54] Models of spoken word understanding typically assume that competition is between real words that may be phonological neighbors without being minimal pairs in regards to the segment in question (vowel); for example, *bat* competes not just with *bet*, but also with words like *pat* and *back*. This assumption is a consequence of the phonological frame not being perfectly perceived, and would also follow in the present model if it allowed imperfect frame perception. However, extending the model in this way is beyond the scope of the present work, and it is not clear that it would systematically contribute to the interaction between vowel categories.

the activation of a category for the purpose of recognition when the corresponding type is a nonword, just as $\frac{1}{\delta}$ multiplies the activation when the corresponding type is a real word (Equation (C.7)). In this way, $\frac{1}{\chi}$ corresponds to the default δ value assigned to nonword types. χ can be interpreted as (proportional to) the response bias toward a category yielding a nonword type.

$$\delta_k = \begin{cases} \left[\lambda + \left(\frac{2(f_k-1)}{M-1} - 1\right)\phi\right]_0^1 & T_k \text{ corresponds to a real word} \\ \frac{1}{\chi} & T_k \text{ corresponds to a nonword} \end{cases} \tag{C.7}$$

When the unintended ('other') type corresponds to a nonword, the formula underlying recognition as the intended type (Equation (C.2) can be written out as:

$$P(\text{token recognized as } T_i | T_o \text{ corresponds to a nonword}) = \frac{\frac{1}{\delta_i} \cdot A_i}{\frac{1}{\delta_i} \cdot A_i + \chi \cdot A_o} \tag{C.8}$$

When $\chi = 0$, the right-hand side of Equation (C.8) becomes 1, meaning that every type that is not in a minimal pair relation is automatically correctly recognized (because there is only one real word compatible with the perfectly-perceived phonological frame). In other words, nonword types do not compete for recognition. When $\chi = 1$, the recognition process reverts to that explored in the previous section, in which nonword types compete for recognition to the same extent as in the main text, but trigger failure when they win. For intermediate values of χ, nonword types have intermediate degrees of influence on the recognition process.

Note that Equation (C.7) is equivalent to Equation (C.9), where $\delta'_k = \delta_k \chi$. Consequently, introducing the parameter χ is equivalent to multiplying both λ and ϕ by a scale factor. In other words, reducing the extent to which nonword types compete with real word types for recognition is equivalent to increasing the average discriminability of types (lowering the discriminability threshold) and decreasing the discriminability

Table C.2: Properties of push chains with varying nonword competition. Average values and % changes for properties of interactions after 50000 iterations, for models with 10% minimal pairs where nonword types compete during recognition to various degrees (represented by χ). The models use parameter set in Table B.2, retuned to have $\mu = 2.6$, and have no bias ($\beta = 0$). Pushee frequency effect measures the distance between the centroid of the sub-distribution of high-frequency Pushee types and the centroid of the sub-distribution of low-frequency Pushee types.

					Pushee						
χ	Category distance		Category overlap		Width		Skew		Ex. Kurtosis		Freq. Effect
0	2.45	(− 5.8%)	2.26	(+125.8%)	0.90	(+11.3%)	0.02	(−79.9%)	−0.49	(+34.2%)	0.00
0.1	2.97	(+14.3%)	1.36	(+ 36.2%)	0.87	(+ 7.8%)	0.07	(−15.5%)	−0.53	(+28.1%)	−0.03
0.25	3.26	(+25.3%)	0.95	(− 5.2%)	0.86	(+ 6.6%)	0.07	(− 5.8%)	−0.55	(+25.4%)	−0.03
0.5	3.47	(+33.5%)	0.65	(− 34.6%)	0.86	(+ 6.3%)	0.09	(+13.5%)	−0.57	(+23.7%)	−0.03
0.75	3.60	(+38.3%)	0.50	(− 49.8%)	0.86	(+ 5.9%)	0.09	(+16.0%)	−0.57	(+23.6%)	−0.05
1	3.69	(+42.1%)	0.40	(− 59.7%)	0.86	(+ 5.8%)	0.09	(+16.1%)	−0.56	(+24.1%)	−0.04

of high-frequency types relative to low-frequency types.

$$\delta'_k = \begin{cases} \left[\lambda\chi + \left(\frac{2(f_k-1)}{M-1} - 1\right)\phi\chi\right]_0^1 & T_k \text{ corresponds to a real word} \\ 1 & T_k \text{ corresponds to a nonword} \end{cases} \tag{C.9}$$

To test how χ affects category interaction and type frequency effects, I conducted simulations. These simulations involved 10% minimal pairs, as in Section C.2, and used parameter setting (10) from Section C.2, with a value of 0.5 for ϕ. To ensure that I could focus just on category interaction, independent of external effects, I removed Pusher bias by setting $\beta = 0$. I explored 6 values for χ: 0, 0.1, 0.25, 0.5, 0.75, and 1. For each value of χ, I ran 1000 models for 50000 iterations each. I present a summary of the rest of the average results after 50000 iterations in Table C.2. As can be seen, category shape (width, skewness, and excess kurtosis) is approximately maintained for all values of χ, but increases of category distance and Pushee frequency effects are only obtained for $\chi > 0$. I present a summary of how category distance and Pushee frequency effects change over time in Figure C.2.

It is clear from Table C.2 and Figure C.2 that having an extremely large default value of δ – corresponding to no recognition competition from nonword types – is

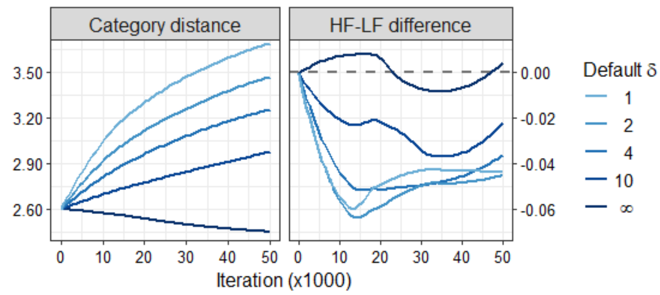

Figure C.2: Frequency effects with varying nonword competition. Results of simulations involving two categories with no bias, where the default value of δ for nonword types is increased to represent less nonword type competition in recognition, from the standard value of 1 (lightest; expected competition) to the largest value possible (darkest; no competition). Left: the distance between the categories grows in all cases except when there is no nonword type competition. Right: high-frequency types change slower than low-frequency types in all cases except when there is no nonword type competition.

not appropriate, for two main reasons. Firstly, the categories drift closer together over time to greatly increase overlap, in spite of the expectation that they should be mutually repellent. Secondly, types of all frequencies change at the same rate in the Pushee, in spite of the expectation that perceptual asymmetries should allow low-frequency types to change faster (as in the previous versions of the model). Both of these results follow from the fact that, when nonword types do not compete for recognition, an intended type without a minimal partner is automatically recognized regardless of its frequency, because it is the only real word type that is compatible with the perfectly-perceived phonological frame. Thus, there is no discriminability force for the 90% of types without minimal partners, meaning that there is insufficient force to keep the categories apart, and there is insufficient potential for perceptual asymmetries to be leveraged in the generation of frequency effects.

Conversely, any default δ that is not extremely large – i.e. any non-zero degree of competition from nonword types – is sufficient to generate mutual category repulsion and frequency effects. While smaller default δ (more nonword type competition) causes greater increase in category distance, it has little impact on category shape, nor on the degree to which low-frequency types change faster than high-frequency types in the Pushee. Consequently, the model reported in the main text – where nonword types compete fully, i.e. as would be expected based on the category activations they incite – yields a qualitative pattern of results that is expected to hold even if the degree of nonword type competition is reduced.

In summary, the model's key results cannot be obtained if minimal pairs alone underpin category interaction (at least, assuming that only a minority of types are in relevant minimal pair relations, as indicated by New Zealand English corpus data). The burden for driving push chains must be extended to types without minimal partners, so that phonotactically plausible nonword types compete for recognition (even to a small degree). Since it is types without minimal partners that are crucial to the model's key results, I conclude that the decision to leave out minimal pairs from the model in the paper had no qualitative effect on the main results.

Appendix D

Exemplar overwriting and decay

The model constructed in Chapter **??** assumes that a stored exemplar overwrites another, and thus that all exemplars have a fixed strength that does not decay over time. This treatment is different to the standard one in exemplar dynamics models, in which there is no overwriting, but exemplar strength decays exponentially over time (Ettlinger, 2007; Pierrehumbert, 2001; Tupper, 2015; Wedel, 2006, 2012; Wedel & Fatkullin, 2017). In this appendix, I show that the difference is superficial; averaged over many runs, the expected behavior of an overwriting model such as the present one is equivalent to that of a special case of a decay model.

For both the overwriting and the decay models, I consider a type with frequency f in a system where the total combined frequency of all types (i.e. total number of exemplars, in the overwriting model) is N. The mathematical analysis assumes that f represents subjective type frequency, but it is not sensitive to the way in which these subjective frequencies are obtained from objective values (though the caveats required for the analysis to hold are sensitive to extreme differences in frequencies). I focus on a single exemplar of this type that was stored at time 0, and I consider its contribution to the behavior of the system at time t (i.e. after t production-storage iterations). For simplicity, I assume that every token is stored; allowing some tokens not to be stored simply slows down the rate of evolution. For the decay model, I assume that the strength of each exemplar is scaled by a factor of $k < 1$ with each iteration. As in the simulations presented in the main text, I also assume that both models consist of a single agent talking to themself, so that the sources of storage and production are identical; the derivations as stated do not apply to situations with multiple interacting agents.

The mathematical derivations presented here assume that there are multiple types in the system, but they make no assumptions about the allocation of those types to categories. The results can be understood to apply equally well to a case with a single category (and multiple types in that category) or to a case with multiple categories (and at least one type per category). Because I assume there are multiple types, the decay-based model I consider is not the same as the one presented by Pierrehumbert (2001), which observed apparent frequency effects using a single type in a single category. In Section D.4, I develop a full comparison with the actual model presented by Pierrehumbert (2001), through which I demonstrate why the apparent frequency effects observed from that model do not hold of exemplar-based models in general.

D.1 Equivalence of memory treatments

I first show that the overwriting and decay models have equivalent treatments of memory. For this purpose, I compare the probability that an exemplar remains after t iterations in the overwriting model with the strength of an exemplar after t iterations in the decay model.

D.1.1 Overwriting model

In the overwriting model, an exemplar stored at time 0 will remain at time t provided that any subsequent tokens of the same type do not overwrite it.

The probability of producing a token of the given type on any iteration is f/N. Given a token of that type, the probability of overwriting the given exemplar with it is $1/f$. Thus, the probability of overwriting the given exemplar on any iteration is $1/N$, so the probability of *not* overwriting it on any iteration is $1-(1/N)$.

For the exemplar still to be present after t iterations, it must not have been overwritten on each iteration. Since each iteration is independent, the probability of this is

$$P(\text{exemplar remains at time } t) = \left(1 - \frac{1}{N}\right)^t \tag{D.1}$$

which is exponentially decreasing with t at a rate given by $1 - 1/N$.

D.1.2 Decay model

In the decay model, an exemplar is stored at time 0 with strength 1, and this strength decays exponentially.

On each iteration, the strength of the exemplar is multiplied by $k < 1$. Thus, the strength of the exemplar at time t is

$$S_x(t) = k^t \tag{D.2}$$

D.1.3 Model comparison

As can be seen, the probability of an exemplar remaining after t iterations in the overwriting model (Equation (D.1)) and the strength of an exemplar after t iterations in the decay model (Equation (D.2)) have the same form. Furthermore, for the particular choice of $k = 1 - (1/N)$, they are identical. Thus, though the two models appear to have very different treatments of memory, they are mathematically equivalent in terms of their expected outcomes (averaged over many runs), for a particular choice of k.

D.2 Equivalence of overall expected behavior

Having established that the two models have equivalent expected treatments of memory (averaged over many runs), I now show that they have equivalent treatments of production, in terms of their expected choice of production targets (averaged over many runs). Since the system evolves by means of producing new tokens to store in memory, these equivalences jointly imply that the two models are equivalent in terms of their overall expected behavior (averaged over many runs).

I consider an exemplar stored at time 0 and compare the probability of choosing that exemplar as production target at time t in both models.

D.2.1 Overwriting model

In the overwriting model, the choice of a given exemplar as production target at time t has three conditions. Firstly, the exemplar must remain in the system at time t. Secondly, the speaker must choose to produce the type of which the exemplar is an instance. Thirdly, the exemplar must be chosen as target from all exemplars of that type.

The probability of the exemplar remaining in the system at time t is $(1-(1/N))^t$ (Equation (D.1)), the probability of the type being chosen is f/N, and the probability of the exemplar being chosen from all f exemplars of that type is $1/f$. Thus, the probability of choosing an exemplar as production target t iterations after it was stored is:

$$P(\text{exemplar chosen as target at } t) = \frac{1}{N}\left(1-\frac{1}{N}\right)^t \tag{D.3}$$

D.2.2 Decay model

In the decay model, the choice of a given exemplar as production target at time t has two conditions: the speaker must choose to produce the corresponding type, and the exemplar must be chosen as target from all exemplars of that type.

The probability of the type being chosen is f/N. The probability of choosing the exemplar from all exemplars of that type is $S_x(T)/\sum_{y\in T} S_y(t)$, where $S_x(t) = k^t$ is the strength of the exemplar at time t and $\sum_{y\in T} S_y(t)$ is the total strength of all exemplars of that type at time t. Thus, the probability of choosing an exemplar as production target t iterations after it was stored is:

$$P(\text{exemplar chosen as target at } t) = \frac{f}{N} \cdot \frac{k^t}{\sum_{y\in T} S_y(t)} \tag{D.4}$$

For the sake of exploring expected behavior (i.e. behavior on average, over many runs), $\sum_{y\in T} S_y(t)$ may be approximated by $S^\star$, the expected total strength at any time. To obtain a value for $S^\star$, I consider r synchronized runs of the model (where r is large) at a particular point in time, with total strengths $\widehat{S_i^\star}$ (for i from 1 to r) for a given type. $S^\star$ is given by the mean of these total strengths.

$$S^\star = \frac{\sum_{i=1}^{r} \widehat{S_i^\star}}{r} \tag{D.5}$$

After a single iteration, each total strength $\widehat{S_i^\star}$ will have been multiplied by k due to decay, and $(f/N)r$ of them are expected to have also grown by 1 due to new productions of the given type. Their mean is still expected to be $S^\star$.

$$S^\star = \frac{\left(\sum_{i=1}^{r} k\widehat{S_i^\star}\right) + \frac{f}{N}r}{r} \tag{D.6}$$

$$= k\frac{\sum_{i=1}^{r} \widehat{S_i^\star}}{r} + \frac{f}{N} \tag{D.7}$$

Substituting Equation (D.5) into Equation (D.7) yields

$$S^\star = kS^\star + \frac{f}{N} \tag{D.8}$$

which can be solved for $S^\star$:

$$S^\star = \frac{f}{N(1-k)} \tag{D.9}$$

Substituting $S^\star$ for $\sum_{y \in T} S_y(t)$ in Equation (D.4) gives an analytic approximation of the expected probability (averaged over many runs) of choosing an exemplar as production target t iterations after it was stored:

$$P(\text{exemplar chosen as target at } t) \approx (1-k)k^t \tag{D.10}$$

Two caveats are required in order for this approximation to be valid. Firstly, the system must not be in its early iterations. Secondly, the decay rate must not be extremely fast relative to the range of type frequencies, such that low-frequency types are expected to have total exemplar strength $S^\star < 1$. I describe the caveats in more detail in Section D.3.

D.2.3 Model comparison

As can be seen, the probability of an exemplar being chosen as a production target t iterations after it was stored has the same form in both the overwriting (Equation (D.3)) and the decay (Equation (D.10)) models. As was the case for memory (Section D.1), these probability expressions are identical for the particular choice of $k = 1 - (1/N)$. Thus, the overwriting model's expected overall behavior (averaged over many runs) is a special case of the decay model's expected overall behavior (averaged over many runs). Given an overwriting model with a total number of exemplars N, it is possible to choose a decay rate k allowing the construction of a decay model with the same expected overall behavior (averaged over many runs). Consequently, any overwriting model is equivalent to some decay model.

Note, however, that the reverse equivalence is not always true: for some decay models, it is not possible to construct an overwriting model showing the same expected overall behavior (averaged over many runs). Because an overwriting model necessarily contains at least one exemplar of each type at every point in time, it requires all types to have expected exemplar strength $S^\star \geq 1$, which is not true in decay models in which the decay rate is extremely fast (relative to the range of type frequencies). See Section D.3 for further discussion.

D.3 Caveats for the decay model

In Section D.2.2, I noted that there are two caveats on the analytical approximation for production target selection in the decay model. Both caveats concern the approximation of the total strength of all exemplars of a particular type T at time t, $\sum_{y \in T} S_y(t)$, by the expected total strength at any time, $S^\star$ (Equation (D.9)).

Firstly, the system must 'burn in' – i.e. be run for sufficiently many iterations – in order for strengths to build to the expected value $S^\star$. In other words, the exemplar distributions for each type must build up to stable densities before the approximation is valid. Thus, the analytical approximation does not hold for the early iterations of a decay-based model that is seeded from sparse exemplar distributions. Consequently, it would not apply to a situation such as a child accumulating experience as they learn a language. However, given that the model addresses regular sound change, which can occur within a lifetime and be reflected in the way that an adult's speech changes (Harrington, 2006), I do not believe this limitation to prevent insight being drawn from the model comparison.

Secondly, the system must be defined in such a way that $S^\star$ is sufficiently greater than 1. When $S^\star$ is close to 1, the total strength $\sum_{y \in T} S_y(t)$ is volatile, as the addition of 1 strength with each new exemplar constitutes a substantial portion of $S^\star$. In this case, $\sum_{y \in T} S_y(t)$ will tend to be above $S^\star$ for small t, meaning that the approximation will tend to overestimate the probability of recent exemplars being selected as target. Consequently, types for which $S^\star \lessapprox 1$ will not have recent exemplars selected as production targets as often – and thus will not advance as rapidly – as expected under the approximation. Since $S^\star$ decreases with type frequency (Equation (D.9)), extreme cases of the decay model (i.e. ones in which $S^\star \approx 1$ for low-frequency types and $S^\star \gg 1$ for high-frequency types) may thus predict low-frequency types to advance at a slower rate than high-frequency types. Such extreme cases would arise in the presence of either an extremely fast decay rate or an extremely long-tailed distribution of type frequencies, where a high-frequency type is presented for storage orders of magnitude more often than a low-frequency type.

In what follows, I illustrate how choices made by the modeler can affect this second caveat, radically altering the qualitative results of a decay-based model (assuming it has been run for sufficient iterations first, as in the first caveat). To facilitate this illustration, I introduce two quantities of interest: the *e-folding time* for the system and the *recurrence time* for different types. The e-folding time, defined in Equation (D.11), is related to the decay rate and represents the number of iterations required for exemplar strength to decay by a factor of e. From the e-folding time, the *exemplar lifespan* can be obtained, representing the number of iterations for which an exemplar persists in memory; for the following discussion, I assume that an exemplar may be removed once its strength depletes by more than 99% (following Wedel & Fatkullin, 2017), giving a lifespan of approximately 5 e-folding times. The recurrence time, defined in Equation (D.13), is the reciprocal of (normalized) type frequency and represents the expected number of iterations between productions of a given type.

$$e\text{-folding time: } E \coloneqq \frac{-1}{\ln(k)} \tag{D.11}$$

$$\approx \frac{1}{1-k} \tag{D.12}$$

$$\text{recurrence time: } R := \frac{N}{f} \tag{D.13}$$

Equation (D.12)[55] and Equation (D.13) can be substituted in Equation (D.9) to yield a definition of expected total exemplar strength, $S^\star$, in terms of e-folding time and recurrence time, given in Equation (D.14).

$$S^\star \approx \frac{E}{R} \tag{D.14}$$

The definition in Equation (D.14) can be used to easily recognize when the second caveat will not hold, and thus when frequency effects will be expected in phonetic drift. Frequency effects can be expected in phonetic drift in a model in which high-frequency types have recurrence times much shorter than the e-folding time and low-frequency types have recurrence times at least as long as the e-folding time.

How can the e-folding and recurrence times be determined? Both are measured in terms of model iterations. An iteration corresponds to the production and perception of a single token that is considered for storage. Thus, some guidance can be provided by consideration of the objective rates of production and perception of words in the real world. After accounting for sampling error (Pierrehumbert & Granell, 2018), objective recurrence times for different words can be estimated to range from 20 (for the word *the*) to more than 100 million (for extremely rare words). To put these numbers in context, Brysbaert, Stevens, Mandera, and Keuleers (2016) calculate that the average person may hear just under 12 million words per year, and a typical psycholinguistic study (e.g. Carreiras, Mechelli, & Price, 2006) defines "high-frequency" words as having recurrence times of approximately 25,000 (40 tokens per million) and

[55] The approximation in Equation (D.12) is obtained from taking the first-order Taylor polynomial of $\ln(k)$ about 1 and holds provided k is sufficiently close to 1. For example, for all cases discussed here ($k > 0.9995$, $E \geq 2000$), the multiplicative error in the estimation is 0.025% or less, which does not substantially impede the ability to identify circumstances in which $S\star \lessapprox 1$ or $S^\star \gg 1$.

"low-frequency" words as having recurrence times between 300,000 and 2 million (3 to 0.5 tokens per million). However, these objective distributions do not translate directly into the model. Since not every actual word token that is uttered need be considered for storage (as discussed in Section 2.3), the representations of type frequency in the model – and thus the determinations of the e-folding and recurrence times – rely on subjective distributions. The modeler is free to choose the function mapping from objective to subjective distributions, giving a large amount of freedom over the choice of e-folding and recurrence times. This freedom of choice can determine model behavior.

For example, if the modeler chooses subjective frequencies that are identical to objective frequencies, then the extremely large range of recurrence times means that there is a correspondingly large range of e-folding times in which a model of phonetic drift will show frequency effects. For example, any e-folding time around 2 million iterations or less – corresponding to an exemplar lifespan of 10 months or more – will generate lag among "low-frequency" words as defined by the psycholinguistic literature. The literature does not contain enough results on the processing of rare words to determine whether this long exemplar lifespan is appropriate for low-frequency words, but the use of objective frequencies implies that it must also apply to high-frequency words, for which it is likely too long. Consequently, any choice of e-folding time that is not too long for high-frequency words will cause some types to change faster than others in a model using objective frequencies.

In such a situation, precisely *which* types change faster will depend upon the e-folding time. For example, with an e-folding time of 2,000 (following Pierrehumbert, 2001), faster change would be observed among types with recurrence times of less than around 2,000. For English, this corresponds to a small set of about 200 extremely common words, which does not have good coverage of the content words defined as "high-frequency" in the prior literature. Consequently, the frequency effects obtained in this situation would not correspond to real effects observed empirically. Under an alternative e-folding time of 30,000, the set of faster-moving types would expand to include the approximately 3,000 English words typically defined as "high-frequency". In this situation, an exemplar would have a lifespan of approximately 5 days, which is extremely fast in comparison to the recurrence times for low-frequency words that occur around once a year (or less). Consequently, a model assuming this e-folding time and distribution of recurrence times would also have to assume that rare words – which encompass a non-negligible proportion of the lexicon, as demonstrated in Figure A.1 – are practically incapable of establishing stable exemplar-based representations in the minds of typical speakers. Such an assumption would raise questions for studies drawing on representations of rare words, such as *mammary* in Bybee's original work adducing a connection between word frequency and leniting changes (Hooper, 1976).

Alternatively, if the modeler chooses subjective frequencies that are nonlinearly 'flattened' from objective frequencies, then the range of recurrence times is likewise compressed, and it becomes easier for models to show no frequency effects. For example, in the present model, recurrence times range from 41 (for the highest-frequency type) to 492 (for the lowest-frequency type).[56] In a corresponding decay model with an e-folding time of 2,000 iterations (again following Pierrehumbert, 2001), no frequency effects would be expected in phonetic drift. To provide an indication of what this e-folding time means on a real-time scale, it is useful to change the interpretation of subjective frequencies. Previously, I interpreted subjective frequencies as reflecting the assumption that some tokens are filtered out in perception before being considered

[56]It is a consequence of the overwriting-based treatment of memory that the recurrence time for the lowest-frequency type can be no greater than the total number of exemplars in the system. For reasons of computational resources, the present simulations assume a small number of types and hence a small number of exemplars, yielding relatively short recurrence times. Scaling up the number of types in the model will also scale up the recurrence times. If the e-folding time is not scaled up commensurately, then there are certain parameter ranges in which frequency effects can be expected in phonetic drift (where the e-folding time falls between the recurrence times for high- and low-frequency types).

for storage, with more filtering for higher-frequency types. Alternatively, it can be maintained that all tokens are considered for storage, and subjective frequencies can be interpreted as reflecting the assumption that exemplars of higher-frequency types are stored with lower initial strength, giving them shorter lifespans. Under this interpretation, an e-folding time of 2,000 means that "high-frequency" words (as defined in the psycholinguistic literature) would have lifespans of around 1–2 months, while "low-frequency" words would have lifespans of around 5–10 months. Much work remains to be done in investigating a real-world-scale decay version of the present model, but such work goes beyond the scope of the present discussion.

To summarize, the question of whether or not a decay-based model meets the second caveat – and thus whether it displays no frequency effects or a high-frequency advantage in phonetic drift – depends on the relationship between the e-folding time, determined by the decay rate, and the distribution of recurrence times, determined by subjective type frequencies. To generate frequency effects, the e-folding time needs to be sufficiently fast and the range of recurrence times sufficiently large. Furthermore, when there are frequency effects, the subset of words that change faster is determined by where the e-folding time falls in the distribution of recurrence times.

The e-folding time and distribution of recurrence times are determined by choices made by the modeler, which concern the decay rate and the function mapping from objective to subjective type frequency. The most appropriate choices have not yet been determined in the literature, as it is unclear precisely how long exemplars may persist in memory – particularly for extremely rare words – and precisely how an incoming stream of tokens is filtered to allow only a subset to be presented for potential storage. Until the appropriate choices are elucidated by the literature, I believe it is reasonable to assume that they meet with the caveats outlined in this section, and thus that the overwriting-based and decay-based models are truly (bidirectionally) equivalent.

D.4 Direct comparison to Pierrehumbert (2001)

In this appendix, I have pointed out that the expected behavior (averaged over many runs) of a model in which memory turnover involves random overwriting of exemplars is equivalent to that of one in which memory turnover involves exponential decay of exemplars, provided the latter meets the caveats in Section D.3. However, the results of the most well-known exemplar model in the literature (Pierrehumbert, 2001), which uses exponential decay, appear to differ from those of the present model, which uses random overwriting. For simulations of phonetic drift, Pierrehumbert (2001) reported that high-frequency words change faster than low-frequency words, whereas the present model yields no frequency effect. Since Pierrehumbert (2001) has been widely cited as a demonstration that exemplar models necessarily predict frequency effects (that favor high-frequency words), it is important to diagnose the reasons for this difference.

Pierrehumbert (2001) represents the first foray into formal modeling of exemplar dynamics in sound change, and lays important groundwork for the present model. However, as pioneering work, Pierrehumbert's model is necessarily very schematic, and has some limitations. The limitation that is primarily responsible for the discrepancy with the results of the present model concerns the model architecture.

The model presented by Pierrehumbert (2001) does not contain a lexical (type) level, meaning that – without frequency-based variation in the decay rate or exemplar lifespan – it is technically incapable of obtaining type frequency effects in simulations involving a single phoneme category. Without separate type representations, each phoneme category effectively contains a single type and the same type is necessarily produced on every iteration. Pierrehumbert (2001) observes that the advancement of the type's exemplars is determined by the number of iterations for which the simulation is run: the more iterations, the more the type is produced with articulatory bias, and thus the more it advances. While it is tempting to interpret this observation as reflecting an effect of frequency, it actually reflects an effect of time. This is because each iteration also corresponds to a single application of decay. After sufficient iterations, a specific exemplar will become so weak that its contribution is negligible, meaning it can effectively be dropped from the system. Thus, exemplars have a lifespan, which corresponds to a certain period of time, and each iteration represents a fixed fraction of this lifespan. Both the decay rate and the lifespan of an exemplar are assumed not to vary with type frequency. Therefore, regardless of type frequency, each iteration corresponds to a fixed period of time, and running a simulation for more iterations corresponds to observing a change over a greater period of time. Having the potential to observe an effect of frequency would require the number of iterations taking place in a given period of time to vary with type frequency. This would only be possible in simulations of a single type if decay rate or exemplar lifespan were assumed to vary with type frequency.

In simulations with multiple types within the same category, by contrast, each type may be produced on different numbers of iterations within the same period of time. The present model, which has type-level representations, shows that phonetic drift is typically unaffected by type frequency. In a decay-based model that meets the caveats laid out in Section D.3, this lack of frequency effect follows because the strengths of exemplars of a given type continue to decay during the gaps between productions of that type. The shorter the gap, the stronger old exemplars of the type will be relative to the most recent exemplar, and thus the more they will compete with it to provide the acoustic target for the next production of the type. Competition from old exemplars holds a category back, since old exemplars represent earlier (less advanced) stages of the change. Since a high-frequency type has shorter gaps between productions than a low-frequency type, it will be held back by competition from old exemplars more, counterbalancing the fact that it will be produced (with articulatory bias) more often. I illustrate this process in Figure D.1, using parameters corresponding to the simulations presented in this dissertation.

The fact that the model presented by Pierrehumbert (2001) is technically unable to display frequency effects in phonetic drift renders moot the question of differences from the present model. While some decay-based models (with type-level representations) *do* display frequency effects in phonetic drift, consistent with the broader suggestions made by Pierrehumbert (2001), these effects are contingent on modeler choices, as discussed in Section D.3. The literature to date has not recognized this contingency and has taken Pierrehumbert's suggestions extremely generally, giving rise to criticisms that exemplar models necessarily over-predict word frequency effects and cannot explain all the patterns found in empirical studies (Abramowicz, 2007; Bermúdez-Otero et al., 2015; Dinkin, 2008; Tamminga, 2014). As I have shown, these criticisms are not applicable to exemplar models as a class, and the new model presented in this dissertation is successful in generating all of the reported patterns.

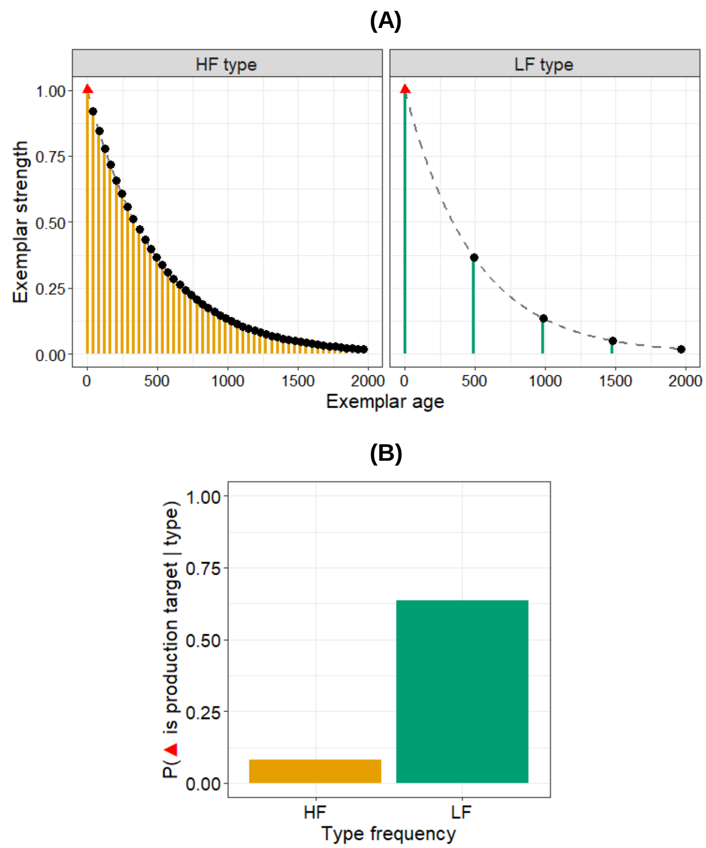

Figure D.1: Counterbalancing frequency effects in phonetic drift. Gaps between productions cause high-frequency types (orange; left) to be held back by competition with old exemplars more than low-frequency types (green; right). (A) Comparison of strength of most recent exemplar (red triangle) to strength of older exemplars of the same type (black circles). A high-frequency type has many more old exemplars than a low-frequency type, with correspondingly greater strengths. (B) Probability of selecting the most recent exemplar (red triangle) as the target for production of the type. Since the aggregate strength of old exemplars is greater for a high-frequency type than for a low-frequency type, they compete much more for selection.

Appendix E

Zero-inflated negative binomial regression

In Chapter 4, I introduced zero-inflated negative binomial (ZINB) regression as an effective way of analyzing count data that are *sparse* (containing few non-zero values) and *bursty* / *overdispersed* (containing extremely variable values). In particular, I argued that ZINB regression was effective for discourse variables such as *eh*, where there are two reasons why a speaker may not use the variable in an interview – because it is not in their linguistic repertoire, or because the circumstances for using it did not arise – and where a small number of speakers who *do* use the variable may use it in excessive, non-representative way. In this appendix, I show the mathematical details of the two properties that make ZINB regression effective in this case, which are conveniently represented by the "zero-inflated" and "negative binomial" components of the name.

E.1 "Zero-inflated"

The "zero-inflated" property of ZINB regression helps to deal with the sparsity of the data. For the application to discourse variables, this means that it attempts to disentangle whether a particular individual's failure to use the variable in an interview is because they don't have it in their linguistic repertoire or because the circumstances for using it did not arise. It does this by assuming the data are generated from a two-stage process, where the first process determines whether a speaker has the variable in their repertoire or not, and the second process determines how often a speaker with the variable in their repertoire will use it. Correspondingly, a ZINB regression consists of two components: a binary logistic regression component, which estimates the probability that a speaker has the variable in their repertoire; and a negative binomial regression component, which estimates the number of times a speaker will use the variable. These components are jointly estimated on the basis of the data, to minimize the errors of the predicted probability distribution over the number of uses of the variable, N, for each cell of speakers:

$$P(N = n) = (1 - P(\text{in repertoire})) \cdot \delta_0(n) + P(\text{in repertoire}) \cdot P(N = n|\text{in repertoire}) \tag{E.1}$$

where $\delta_0(n)$ evaluates to 1 when $n = 0$ and to 0 otherwise. $P(\text{in repertoire})$ is the output of the logistic regression component and $P(N = n|\text{in repertoire})$ is the output of the negative binomial regression component. Either component can be assessed to cause an individual speaker to fail to use the discourse variable (the case $N = 0$); failures caused by the logistic regression component correspond to the speaker not having the variable in their repertoire, and failures caused by the negative binomial regression component correspond to the speaker having the variable in their repertoire but not finding the circumstances required for its use.

E.2 "Negative binomial"

The "negative binomial" property of ZINB regression helps to deal with the burstiness, or overdispersion, of the data. For the application to discourse variables, this means that it attempts to downweight non-representative speakers who use the variable excessively. It does this by extending Poisson regression, which is the standard method for analyzing count data. To illustrate both of these methods, consider a cell containing 10 speakers, each of whom utters 10,000 words in their interview. Assume that one speaker uses the discourse variable of interest ten times, one uses it four times, three speakers use it twice each, and five speakers do not use it, for a total of 20 uses across all speakers in the cell. In the discussion that follows, assume that the discourse variable is in the repertoire of all speakers in the cell, even though some didn't use it in their interview.[57]

Poisson regression takes a 'bag-of-words' approach to analyzing the use of the discourse variable by speakers in the cell: all 100,000 words of the 10 speakers are put in a bag (individually), and each speaker's interview is considered to be a set of 10,000 random draws from this bag (with replacement). The number of times that a speaker uses the discourse variable is therefore modeled by the outcome of 10,000 independent random draws of a word, each of which has a 1 in 5,000 chance of being the discourse variable. Considering all such outcomes generates a probability distribution over the number of uses of the discourse variable, which can be generalized to expectations from a new speaker in the same cell (in an interview where they utter 10,000 words). This probability distribution is characterized by the proportion of words in the bag that are the discourse variable (1 in 5,000), which is referred to as the *rate* parameter, λ; the mathematical form of the distribution is given in Equation (E.2).

[57]Correspondingly, assume that all probability distributions are conditioned on the existence of the variable in the speaker's repertoire – though, for convenience, I do not write this condition throughout.

$$P(N = n) = \frac{e^{-\lambda} \lambda^n}{n!} \tag{E.2}$$

Negative binomial regression also takes a 'bag-of-words' approach to analyzing the use of the discourse variable by speakers in the cell, but it is rather more complex than the Poisson regression approach. Instead of assuming that all speakers draw from the same bag, negative binomial regression allows speakers to draw from different bags. These different bags may contain different numbers of the discourse variable, corresponding to different rate parameters: for example, in one bag, the rate parameter may be 1 in 10,000 (1 instance of the discourse variable in the bag), whilst in another, it may be 1 in 1,000 (10 instances of the discourse variable in the bag). In this way, nonrepresentative speakers who use the variable excessively will draw from a different bag than standard speakers. When a new speaker is encountered, they are assigned a bag at random; however, they are more likely to get a standard bag, containing few instances of the discourse variable, than a nonrepresentative bag, containing many instances of the discourse variable. Accordingly, the predictions for new speakers will not be skewed by the contributions of a few nonrepresentative speakers who use the variable excessively. Formally, the rate parameter is assumed to be distributed according to a random variable Z:

$$P(N = n|z) = \frac{e^{-\lambda z} (\lambda z)^n}{n!}, \quad \text{where } P(Z = z) = g(z) \tag{E.3}$$

Integrating out the condition in Equation (E.3) gives:

$$P(N = n) = \frac{\lambda^n}{n!} \int_0^\infty e^{-\lambda z} z^n g(z) \, \mathrm{d}z, \quad \text{where } P(Z = z) = g(z) \tag{E.4}$$

For negative binomial regression, the rate parameters are assumed to follow a gamma distribution, with the gamma distribution parameters α and β both taking on the value θ.[58] Smaller values of θ correspond to more overdispersion in the data, i.e. more influence of non-representative speakers who use the discourse variable excessively.

$$g(z) = \frac{\theta^\theta}{\Gamma(\theta)} z^{\theta-1} e^{-\theta z} \tag{E.5}$$

Thus, the probability distribution over counts becomes:

$$P(N = n) = \frac{\lambda^n \theta^\theta}{n!\Gamma(\theta)} \int_0^\infty e^{-(\lambda+\theta)z} z^{n+\theta-1} \, \mathrm{d}z \tag{E.6}$$

$$= \frac{\lambda^n \theta^\theta}{n!\Gamma(\theta) \, (\lambda+\theta)^{n+\theta}} \int_0^\infty \left[(\lambda+\theta)\, z\right]^{n+\theta-1} e^{-(\lambda+\theta)z} \, (\lambda+\theta) \, \mathrm{d}z \tag{E.7}$$

Substituting $u = (\lambda + \theta)\, z$, Equation (E.7) becomes:

$$P(N = n) = \frac{\lambda^n \theta^\theta}{n!\Gamma(\theta) \, (\lambda+\theta)^{n+\theta}} \int_0^\infty u^{n+\theta-1} e^{-u} \, \mathrm{d}u \tag{E.8}$$

Now since $\Gamma(x) = \int_0^\infty y^{x-1} e^{-y} \, \mathrm{d}y$ by definition, Equation (E.8) becomes:

$$P(N = n) = \frac{\lambda^n \theta^\theta}{n!\Gamma(\theta) \, (\lambda+\theta)^{n+\theta}} \Gamma(n+\theta) \tag{E.9}$$

$$= \frac{\Gamma(n+\theta)}{n!\Gamma(\theta)} \left(\frac{\lambda}{\lambda+\theta}\right)^n \left(\frac{\theta}{\lambda+\theta}\right)^\theta \tag{E.10}$$

The probability distribution represented in Equation (E.10) is that used for the negative binomial regression component in ZINB regression (Equation (E.1)).

[58] Setting $\alpha = \beta$ is required in order to preserve the expected value of N under the assumption of a distribution over λ.

Bibliography

Abramowicz, Ł. (2007). Sociolinguistics meets Exemplar Theory: Frequency and recency effects in (ing). *University of Pennsylvania Working Papers in Linguistics*, *13*(2), 27–37. Retrieved from http://repository.upenn.edu/pwpl/vol13/iss2/3/

Acton, E. K. (2019). Pragmatics and the social life of the English definite article. *Language*, *95*(1), 37–65. doi: 10.1353/lan.2019.0010

Arnold, J. E., Wasow, T., Asudeh, A., & Alrenga, P. (2004). Avoiding attachment ambiguities: The role of constituent ordering. *Journal of Memory and Language*, *51*(1), 55–70. doi: 10.1016/j.jml.2004.03.006

Ashby, F. G., & Alfonso-Reese, L. A. (1995). Categorization as probability density estimation. *Journal of Mathematical Psychology*, *39*, 216–233. doi: 10.1006/jmps.1995.1021

Babel, M. (2010). Dialect divergence and convergence in New Zealand English. *Language in Society*, *39*(4), 437–456. doi: 10.1017/S0047404510000400

Babel, M. (2012). Evidence for phonetic and social selectivity in spontaneous phonetic imitation. *Journal of Phonetics*, *40*(1), 177–189. doi: 10.1016/j.wocn.2011.09.001

Babel, M., & Bulatov, D. (2012). The role of fundamental frequency in phonetic accommodation. *Language and Speech*, *55*(2), 231–248. doi: 10.1177/0023830911417695

Babel, M., McAuliffe, M., & Haber, G. (2013). Can mergers-in-progress be unmerged in speech accommodation? *Frontiers in Psychology*, *4*, 653. doi: 10.3389/fpsyg.2013.00653

Bard, E. G., Anderson, A. H., Sotillo, C., Aylett, M., Doherty-Sneddon, G., & Newlands, A. (2000). Controlling the intelligibility of referring expressions in dialogue. *Journal of Memory and Language*, *42*(1), 1–22. doi:

10.1006/jmla.1999.2667

Baxter, G. J., Blythe, R. A., Croft, W., & McKane, A. J. (2009). Modeling language change: An evaluation of Trudgill's theory of the emergence of New Zealand English. *Language Variation and Change*, *21*(2), 257–296. doi: 10.1017/S095439450999010X

Baxter, G., & Croft, W. (2016). Modeling language change across the lifespan: Individual trajectories in community change. *Language Variation and Change*, *28*(2), 129–173. doi: 10.1017/s0954394516000077

Beckner, C., Blythe, R., Bybee, J., Christiansen, M. H., Croft, W., Ellis, N. C., ... Schoenemann, T. (2009). Language is a complex adaptive system: Position paper. *Language Learning*, *59*(Suppl. 1), 1–26. doi: 10.1111/j.1467-9922.2009.00533.x

Begg, I. (1974). Estimation of word frequency in continuous and discrete tasks. *Journal of Experimental Psychology*, *102*(6), 1046–1052. doi: 10.1037/h0036356

Bell, A. (2001). Back in style: Reworking audience design. In P. Eckert & J. R. Rickford (Eds.), *Style and sociolinguistic variation* (pp. 139–169). Cambridge, UK: Cambridge University Press. doi: 10.1017/CBO9780511613258.010

Bell, A., Brenier, J. M., Gregory, M., Girand, C., & Jurafsky, D. (2009). Predictability effects on durations of content and function words in conversational English. *Journal of Memory and Language*, *60*(1), 92–111. doi: 10.1016/j.jml.2008.06.003

Benjamin, A. S. (2003). Predicting and postdicting the effects of word frequency on memory. *Memory & Cognition*, *31*(2), 297–305. doi: 10.3758/BF03194388

Benor, S. B. (2010). Ethnolinguistic repertoire: Shifting the analytic focus in language and ethnicity. *Journal of Sociolinguistics*, *14*(2), 159–183. doi: 10.1111/j.1467-9841.2010.00440.x

Bermúdez-Otero, R., Baranowski, M., Bailey, G., & Turton, D. (2015). *A constant rate effect in Manchester /t/-glottalling: High-frequency words are ahead of, but change at the same rate as, low-frequency words.* Paper presented at the 2nd Edinburgh Symposium on Historical Phonology, University of Edinburgh.

Bestelmeyer, P. E., Belin, P., & Ladd, D. R. (2015). A neural marker for social bias toward in-group accents. *Cerebral Cortex*, *25*(10), 3953–3961. doi: 10.1093/cercor/bhu282

Blevins, J. (2006). A theoretical synopsis of Evolutionary Phonology. *Theoretical Linguistics*, *32*(2), 117–166. doi: 10.1515/TL.2006.009

Blevins, J., & Wedel, A. (2009). Inhibited sound change: An evolutionary approach to lexical competition. *Diachronica*, *26*(2), 143–183. doi: 10.1075/dia.26.2.01ble

Bloomfield, L. (1933). *Language*. London: George Allen & Unwin.

Boshier, R. (2015). Learning from the moa: The challenge of Māori language revitalization in Aotearoa/New Zealand. In W. J. Jacob, S. Y. Cheng, & M. K. Porter (Eds.), *Indigenous education: Language, culture and identity* (pp. 207–226). Dordrecht: Springer. doi: 10.1007/978-94-017-9355-1_11

Bowers, J. S. (2000). The modality-specific and -nonspecific components of long-term priming are frequency sensitive. *Memory & Cognition*, *28*(3), 406–414. doi: 10.3758/BF03198556

Bradlow, A. R., & Bent, T. (2008). Perceptual adaptation to non-native speech. *Cognition*, *106*(2), 707–729. doi: 10.1016/j.cognition.2007.04.005

Brown, P., & Levinson, S. (1987). *Politeness: Some universals in language usage.* Cambridge, UK: Cambridge University Press. doi: 10.2307/3587263

Brysbaert, M., Stevens, M., Mandera, P., & Keuleers, E. (2016). How many words do we know? Practical estimates of vocabulary size dependent on word definition, the degree of language input and the participant's age. *Frontiers in Psychology*, *7*, 1116. doi: 10.3389/fpsyg.2016.01116

Burnham, K. P., & Anderson, D. R. (2002). *Model selection and multimodel inference: A practical information-theoretic approach* (2nd ed.). New York, NY: Springer-Verlag.

Busemeyer, J. R., Dewey, G. I., & Medin, D. L. (1984). Evaluation of exemplar-based generalization and the abstraction of categorical information. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *10*(4), 638–648. doi: 10.1037/0278-7393.10.4.638

Buz, E., Tanenhaus, M. K., & Jaeger, T. F. (2016). Dynamically adapted context-specific hyper-articulation: Feedback from interlocutors affects speakers' subsequent pronunciations. *Journal of Memory and Language*, *89*, 68–86. doi: 10.1016/j.jml.2015.12.009

Bybee, J. (2002). Word frequency and context of use in the lexical diffusion of phonetically conditioned sound change. *Language Variation and Change*, *14*, 261–290. doi: 10.1017/S0954394502143018

Campbell-Kibler, K. (2007). Accent, (ING), and the social logic of listener perceptions. *American Speech*, *82*(1), 32–64. doi: 10.1215/00031283-2007-002

Carreiras, M., Mechelli, A., & Price, C. J. (2006). Effect of word and syllable frequency on activation during lexical decision and reading aloud. *Human Brain Mapping*, *27*(12), 963–972. doi: 10.1002/hbm.20236

Cedergren, H. J., & Sankoff, D. (1974). Variable rules: Performance as a statistical reflection of competence. *Language*, *50*(2), 333–355. doi: 10.2307/412441

Chee, M. W. L., Goh, J. O. S., Lim, Y., Graham, S., & Lee, K. (2004). Recognition memory for studied words is determined by cortical activation differences at encoding but not during retrieval. *NeuroImage*, *22*(4), 1456–1465. doi: 10.1016/j.neuroimage.2004.03.046

Chen, M. (1970). Vowel length variation as a function of the voicing of the consonant environment. *Phonetica*, *22*, 129–159. doi: 10.1159/000259312

Clarke-Davidson, C. M., Luce, P. A., & Sawusch, J. R. (2008). Does perceptual learning in speech reflect changes in phonetic category representation or decision bias? *Perception & Psychophysics*, *70*(4), 604–618. doi: 10.3758/PP.70.4.604

Clopper, C. G., Pierrehumbert, J. B., & Tamati, T. N. (2010). Lexical neighborhoods and phonological confusability in cross-dialect word recognition in noise. *Laboratory Phonology*, *10*, 65–92. doi: 10.1515/labphon.2010.005

Clopper, C. G., Tamati, T. N., & Pierrehumbert, J. B. (2016). Variation in the strength of lexical encoding across dialects. *Journal of Phonetics*, *58*, 87–103. doi: 10.1016/j.wocn.2016.06.002

Columbus, G. (2010). A comparative analysis of invariant tags in three varieties of English. *English World-Wide*, *31*(3), 288–310. doi: 10.1075/eww.31.3.03col

Connine, C. M., Blasko, D. G., & Hall, M. (1991). Effects of subsequent sentence context in auditory word recognition: Temporal and linguistic constraints. *Journal of Memory and Language*, *30*, 234–250. doi: 10.1016/0749-596X(91)90005-5

Connine, C. M., Titone, D., & Wang, J. (1993). Auditory word recognition: Extrinsic and intrinsic effects of word frequency. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *19*(1), 81–94. doi: 10.1037/0278-7393.19.1.81

Cormack, D., & Robson, C. (2010). *Ethnicity, national identity, and 'New Zealanders': Considerations for monitoring Māori health and ethnic inequalities* (Tech. Rep.). Wellington: Te Rōpū Rangahau Hauora a Eru Pōmare. Retrieved from http://www.otago.ac.nz/wellington/otago600094.pdf

Coupland, N. (2014). Language change, social change, sociolinguistic change:

A meta-commentary. *Journal of Sociolinguistics*, *18*(2), 277–286. doi: `10.1111/josl.12077`

Cukor-Avila, P., & Bailey, G. (2013). Real time and apparent time. In J. K. Chambers & N. Schilling (Eds.), *The handbook of language variation and change* (2nd ed., pp. 259–282). Malden, MA: Wiley-Blackwell. doi: `10.1002/9781118335598.ch11`

Dahan, D., Drucker, S. J., & Scarborough, R. A. (2008). Talker adaptation in speech perception: Adjusting the signal or the representations? *Cognition*, *108*(3), 710–718. doi: `10.1016/j.cognition.2008.06.003`

Dahan, D., Magnuson, J. S., & Tanenhaus, M. K. (2001). Time course of frequency effects in spoken-word recognition: Evidence from eye movements. *Cognitive Psychology*, *42*(4), 317–367. doi: `10.1006/cogp.2001.0750`

Davies, M. (2008–). *The Corpus of Contemporary American English: 520 million words, 1990–present.* Retrieved July, 2014, from `http://www.english-corpora.org/coca/`.

de Marneffe, M.-C., Tomlinson, J., Tice, M., & Sumner, M. (2011). The interaction of lexical frequency and phonetic variation in the perception of accented speech. In L. Carlson, C. Hoelscher, & T. F. Shipley (Eds.), *Proceedings of the 33rd annual meeting of the Cognitive Science Society* (pp. 3575–3580). Austin, TX: Cognitive Science Society.

Diana, R. A., & Reder, L. M. (2006). The low-frequency encoding disadvantage: Word frequency affects processing demands. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*(4), 805–815. doi: `10.1037/0278-7393.32.4.805`

Dijksterhuis, A., & Bargh, J. A. (2001). The perception-behavior expressway: Automatic effects of social perception on social behavior. *Advances in Exeperimental Social Psychology*, *33*, 1–40. doi: `10.1016/s0065-2601(01)80003-4`

Dinkin, A. J. (2008). The real effect of word frequency on phonetic variation. *University of Pennsylvania Working Papers in Linguistics*, *14*(1), 97–106. Retrieved from `http://repository.upenn.edu/pwpl/vol14/iss1/8/`

Eckert, P. (1988). Adolescent social structure and the spread of linguistic change. *Language in Society*, *17*(2), 183–207. doi: `10.1017/S0047404500012756`

Eckert, P. (2008). Variation and the indexical field. *Journal of Sociolinguistics*, *12*(4), 453–476. doi: `10.1111/j.1467-9841.2008.00374.x`

Eckert, P. (2012). Three waves of variation study: The emergence of meaning in the study of sociolinguistic variation. *Annual Review of Anthropology*, *41*(1), 87–100. doi: 10.1146/annurev-anthro-092611-145828

Emanuel, R., Adams, J., Baker, K., Daufin, E. K., Ellington, C., Fitts, E., ... Okeowo, D. (2008). How college students spend their time communicating. *International Journal of Listening*, *22*(1), 13–28. doi: 10.1080/10904010701802139

Ettlinger, M. (2007). Shifting categories: An exemplar-based computational model of chain shifts. In D. S. McNamara & J. G. Trafton (Eds.), *Proceedings of the 29th annual conference of the Cognitive Science Society* (pp. 239–234). Austin, TX: Cognitive Science Society. Retrieved from http://escholarship.org/uc/item/5c66h9nk

Evans, B. G., & Iverson, P. (2007). Plasticity in speech production and perception: A study of accent change in young adults. *The Journal of the Acoustical Society of America*, *121*(6), 3814–3826. doi: 10.1121/1.2722209

Fitt, S. (2000). *Documentation and user guide to Unisyn lexicon and post-lexical rules.* Centre for Speech Technology Research, University of Edinburgh. Retrieved from http://www.cstr.ed.ac.uk

Flemming, E. (2004). Contrast and perceptual distinctiveness. In B. Hayes, R. Kirchner, & D. Steriade (Eds.), *Phonetically based phonology* (pp. 232–276). Cambridge, UK: Cambridge University Press. doi: 10.1017/CBO9780511486401.008

Forster, K. I., & Chambers, S. M. (1973). Lexical access and naming time. *Journal of Verbal Learning and Verbal Behavior*, *12*(6), 627–635. doi: 10.1016/S0022-5371(73)80042-8

Forster, K. I., & Davis, C. (1984). Repetition priming and frequency attenuation in lexical access. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *10*(4), 680–698. doi: 10.1037/0278-7393.10.4.680

Fowler, C. A., Brown, J. M., Sabadini, L., & Weihing, J. (2003). Rapid access to speech gestures in perception: Evidence from choice and simple response time tasks. *Journal of Memory and Language*, *49*(3), 396–413. doi: 10.1016/S0749-596X(03)00072-X

Fox, R. A. (1984). Effect of lexical status on phonetic categorization. *Journal of Experimental Psychology: Human Perception and Performance*, *10*(4), 526–540. doi: 10.1121/1.2019656

Gahl, S. (2008). Time and thyme are not homophones: The effect of lemma frequency on word durations in spontaneous speech. *Language*, *84*(3), 474–496. doi:

10.1353/lan.0.0035

Ganong, W. F. (1980). Phonetic categorization in auditory word perception. *Journal of Experimental Psychology: Human Perception and Performance*, *6*(1), 110–125. doi: 10.1037/0096-1523.6.1.110

Garrett, A. (2014). Sound change. In C. Bowern & B. Evans (Eds.), *The Routledge handbook of historical linguistics* (pp. 227–248). New York, NY: Routledge. doi: http://doi.org/10.4324/9781315794013.ch9

Giles, H., Coupland, N., & Coupland, J. (1991). Accomodation Theory: Communication, context, and consequence. In H. Giles, N. Coupland, & J. Coupland (Eds.), *Contexts of accomodation: Developments in applied sociolinguistics* (pp. 1–68). Cambridge, UK: Cambridge University Press. doi: 10.1017/CBO9780511663673.001

Goldinger, S. D. (1996). Words and voices: Episodic traces in spoken word identification and recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*(5), 1166–1183. doi: 10.1037/0278-7393.22.5.1166

Goldinger, S. D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological Review*, *105*(2), 251–279. doi: 10.1037/0033-295X.105.2.251

Goldinger, S. D., Luce, P. A., & Pisoni, D. B. (1989). Priming lexical neighbors of spoken words: Effects of competition and inhibition. *Journal of Memory and Language*, *28*(5), 501–518. doi: 10.1016/j.biotechadv.2011.08.021.Secreted

Gordon, E., Campbell, L., Hay, J., Maclagan, M., Sudbury, A., & Trudgill, P. (2004). *New Zealand English: Its origins and evolution*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511486678

Gordon, E., Maclagan, M., & Hay, J. (2007). The ONZE corpus. In J. C. Beal, K. P. Corrigan, & H. Moisl (Eds.), *Creating and digitizing language corpora. Volume 2: Diachronic databases* (pp. 82–104). New York, NY: Palgrave Macmillan. doi: 10.1057/9780230223202_4

Gumperz, J. J. (1964). Linguistic and social interaction in two communities. *American Anthropologist*, *66*(6), 137–153. doi: 10.1525/aa.1964.66.suppl_3.02a00100

Harrington, J. (2006). An acoustic analysis of 'happy-tensing' in the Queen's Christmas broadcasts. *Journal of Phonetics*, *34*(4), 439–457. doi: 10.1016/j.wocn.2005.08.001

Harrington, J., Kleber, F., Reubold, U., Schiel, F., & Stevens, M. (2018). Linking cognitive and social aspects of sound change using agent-based modeling. *Topics in Cognitive Science*, 1–22. doi: 10.1111/tops.12329

Harrington, J., Palethorpe, S., & Watson, C. I. (2000). Does the Queen speak the Queen's English? *Nature*, *408*(6815), 927–928. doi: 10.1038/35050160

Hashimoto, D. (2019). *Loanword phonology in New Zealand English: Exemplar activation and message predictability* (Unpublished doctoral dissertation, University of Canterbury). Retrieved from http://ir.canterbury.ac.nz/handle/10092/16634

Haukoos, J. S., & Lewis, R. J. (2005). Advanced statistics: Bootstrapping confidence intervals for statistics with "difficult" distributions. *Academic Emergency Medicine*, *12*(4), 360–365. doi: 10.1197/j.aem.2004.11.018

Hay, J., & Drager, K. (2010). Stuffed toys and speech perception. *Linguistics*, *4*, 865–892. doi: 10.1515/ling.2010.027

Hay, J., & Foulkes, P. (2016). The evolution of medial /t/ over real and remembered time. *Language*, *92*(2), 298–330. doi: 10.1353/lan.2016.0036

Hay, J., Nolan, A., & Drager, K. (2006). From fush to feesh: Exemplar priming in speech perception. *The Linguistic Review*, *23*(3), 351–379. doi: 10.1515/TLR.2006.014

Hay, J. B., Pierrehumbert, J. B., Walker, A. J., & LaShell, P. (2015). Tracking word frequency effects through 130 years of sound change. *Cognition*, *139*, 83–91. doi: 10.1016/j.cognition.2015.02.012

Hay, J., Warren, P., & Drager, K. (2006). Factors influencing speech perception in the context of a merger-in-progress. *Journal of Phonetics*, *34*(4), 458–484. doi: 10.1016/j.wocn.2005.10.001

Hillenbrand, J., Getty, L. A., Clark, M. J., & Wheeler, K. (1995). Acoustic characteristics of American English vowels. *The Journal of the Acoustical Society of America*, *97*(5), 3099–3111. doi: 10.1121/1.411872

Hintzman, D. L. (1986). "Schema abstraction" in a multiple-trace memory model. *Psychological Review*, *93*(4), 411–428. doi: 10.1037/0033-295X.93.4.411

Hintzman, D. L., & Block, R. A. (1971). Repetition and memory: Evidence for a multiple-trace hypothesis. *Journal of Experimental Psychology*, *88*(3), 297–306. doi: 10.1037/h0030907

Holmes, J. (1995). *Women, men and politeness.* New York, NY: Routledge.

Holmes, J. (1996). Losing voice: Is final /z/ devoicing a feature of Māori English? *World Englishes*, *15*(2), 193–205. doi:

10.1111/j.1467-971X.1996.tb00105.x

Holmes, J. (2009). Māori and Pākehā English: Some New Zealand social dialect data. *Language in Society*, *26*(1), 65. doi: 10.1017/S0047404500019412

Hooper, J. B. (1976). Word frequency in lexical diffusion and the source of morphophonological change. In W. Christie (Ed.), *Current progress in historical linguistics* (pp. 96–105). Amsterdam: North Holland.

Howes, D. (1957). On the relation between the intelligibility and frequency of occurrence of English words. *The Journal of the Acoustical Society of America*, *29*(2), 296–305. doi: 10.1121/1.1908862

Johnson, K. (1997). Speech perception without speaker normalization: An exemplar model. In K. Johnson & J. Mullenix (Eds.), *Talker variability in speech processing* (pp. 145–165). San Diego, CA: Academic Press.

Johnson, K. (2006). Resonance in an exemplar-based lexicon: The emergence of social identity and phonology. *Journal of Phonetics*, *34*, 485–499. doi: 10.1016/j.wocn.2005.08.004

Johnson, K., Flemming, E., & Wright, R. (1993). The hyperspace effect: Phonetic targets are hyperarticulated. *Language*, *69*(3), 505–528. doi: 10.2307/416697

Kim, M., Horton, W. S., & Bradlow, A. R. (2011). Phonetic convergence in spontaneous conversations as a function of interlocutor language distance. *Laboratory Phonology*, *2*(1), 1–30. doi: 10.1515/labphon.2011.004

Kinoshita, S. (1995). The word-frequency effect in recognition memory versus repetition priming. *Memory & Cognition*, *23*(5), 569–580. doi: 10.3758/bf03197259

Kiparsky, P. (1976). Historical linguistics and the origin of language. *Annals of the New York Academy of Sciences*, *280*(1), 97–103. doi: 10.1111/j.1749-6632.1976.tb25475.x

Kirby, S., Cornish, H., & Smith, K. (2008). Cumulative cultural evolution in the laboratory: An experimental approach to the origins of structure in human language. *Proceedings of the National Academy of Sciences*, *105*(31), 10681–10686. doi: 10.1073/pnas.0707835105

Kleinschmidt, D. F., & Jaeger, T. F. (2015). Robust speech perception: Recognize the familiar, generalize to the similar, and adapt to the novel. *Psychological Review*, *122*(2), 148–203. doi: 10.1037/a0038695

Kraljic, T., & Samuel, A. G. (2006). Generalization in perceptual learning for speech. *Psychonomic Bulletin & Review*, *13*(2), 262–268. doi: 10.3758/BF03193841

Kruschke, J. K. (1992). ALCOVE: An exemplar-based connectionist model of category learning. *Psychological Review*, *99*(1), 22–44. doi: 10.1037/0033-295X.99.1.22

Labov, W. (1963). The social motivation of a sound change. *Word*, *19*(3), 273–309. doi: 10.1080/00437956.1963.11659799

Labov, W. (1969). Contraction, deletion, and inherent variability of the English copula. *Language*, *45*(4), 715–762. doi: 10.2307/412333

Labov, W. (1972). The social stratification of (r) in New York City department stores. In *Sociolinguistic patterns* (pp. 43–69). Philadelphia, PA: University of Pennsylvania Press.

Labov, W. (1994). *Principles of linguistic change, Vol. 1: Internal factors.* Oxford: Wiley-Blackwell.

Labov, W. (2001). *Principles of linguistic change, Vol. 2: Social factors.* Oxford: Wiley-Blackwell.

Labov, W. (2010). *Principles of linguistic change, Vol. 3: Cognitive and cultural factors.* Oxford: Wiley-Blackwell.

Landauer, T. K. (1986). How much do people remember? Some estimates of the quantity of learned information in long-term memory. *Cognitive Science*, *10*(4), 477–493. doi: 10.1016/S0364-0213(86)80014-3

Lightfoot, D., & Westergaard, M. (2007). Language acquisition and language change: Inter-relationships. *Language and Linguistics Compass*, *1*(5), 396–415. doi: 10.1111/j.1749-818x.2007.00023.x

Liljencrants, J., & Lindblom, B. (1972). Numerical simulation of vowel quality systems: The role of perceptual contrast. *Language*, *48*(4), 839–862. doi: 10.2307/411991

Lindblom, B. (1990). Explaining phonetic variation: A sketch of the H&H Theory. In W. J. Hardcastle & A. Marchal (Eds.), *Speech production and speech modelling* (pp. 403–439). Dordrecht, The Netherlands: Kluwer Academic Publishers. doi: 10.1007/978-94-009-2037-8_16

Liu, J. H., Wilson, M. S., McClure, J., & Higgins, T. R. (1999). Social identity and the perception of history: Cultural representations of Aotearoa/New Zealand. *European Journal of Social Psychology*, *29*(8), 1021–1047. doi: 10.1002/(SICI)1099-0992(199912)29:8<1021::AID-EJSP975>3.0.CO;2-4

Łubowicz, A. (2011). Chain shifts. In M. van Oostendorp, C. J. Ewen, E. Hume, & K. Rice (Eds.), *The Blackwell companion to phonology* (pp. 1717–1735). Oxford:

Wiley-Blackwell. doi: 10.1002/9781444335262.wbctp0073

Luce, P. A., & Pisoni, D. B. (1998). Recognizing spoken words: The Neighborhood Activation Model. *Ear and Hearing*, *19*(1), 1–36. doi: 10.1097/00003446-199802000-00001

Luce, R. D. (1959). *Individual choice behavior: A theoretical analysis.* New York, NY: Wiley.

Macalister, J. (2006). The Māori presence in the New Zealand English lexicon, 18502000: Evidence from a corpus-based study. *English World-Wide*, *27*(1), 1–24. doi: 10.1075/eww.27.1.02mac

Mack, M. (1982). Voicing-dependent vowel duration in English and French: Monolingual and bilingual production. *Journal of the Acoustical Society of America*, *71*(1), 173–178. doi: 10.1121/1.387344

Maclagan, M., Watson, C. I., Harlow, R., King, J., & Keegan, P. (2009). /u/ fronting and /t/ aspiration in Māori and New Zealand English. *Language Variation and Change*, *21*(2), 175. doi: 10.1017/S095439450999007X

Maegaard, M., Jensen, T. J., Kristiansen, T., & Jørgensen, J. N. (2013). Diffusion of language change: Accommodation to a moving target. *Journal of Sociolinguistics*, *17*(1), 3–36. doi: 10.1111/josl.12002

Marslen-Wilson, W., & Tyler, L. K. (1980). The temporal structure of spoken language understanding. *Cognition*, *8*, 1–71. doi: 10.1016/0010-0277(80)90015-3

Martinet, A. (1952). Function, structure, and sound change. *Word*, *8*(1), 1–32. doi: 10.1080/00437956.1952.11659416

McGowan, K. B. (2015). Social expectation improves speech perception in noise. *Language and Speech*, *58*, 505–521. doi: 10.1177/0023830914565191

Medin, D. L., & Schaffer, M. M. (1978). Context Theory of classification learning. *Psychological Review*, *85*(3), 207–238. doi: 10.1037/0033-295X.85.3.207

Mengesha, Z., Todd, S., & Sumner, M. (2019). *Talker voice and stereotype in the false recall of spoken words.* Poster presented at the 93rd Annual Meeting of the Linguistic Society of America, New York, NY.

Meyerhoff, M. (1992). "We've all got to go one day, eh": Powerlessness and solidarity in the functions of a New Zealand tag. In K. Hall, M. Bucholtz, & B. Moonwomon (Eds.), *Locating power: Proceedings of the second Berkeley women and language conference* (pp. 409–419). Berkeley, CA: Berkeley Women and Language Group.

Meyerhoff, M. (1994). Sounds pretty ethnic, eh?: A pragmatic particle in New Zealand English. *Language in Society*, *23*(3), 367–388. doi: 10.1017/S0047404500018029

Mikolov, T., Sutskever, I., Chen, K., Corrado, G., Dean, J., Chen, K., & Dean, J. (2013). Distributed representations of words and phrases and their compositionality. In C. J. C. Burges, L. Bottou, M. Welling, Z. Ghahramani, & K. Q. Weinberger (Eds.), *Advances in neural information processing systems 26* (pp. 3111–3119). Curran Associates, Inc. doi: 10.1162/jmlr.2003.3.4-5.951

Miller, G. A., & Nicely, P. E. (1955). An analysis of perceptual confusions among some english consonants. *Journal of the Acoustical Society of America*, *27*(2), 338–352. doi: 10.1121/1.1907526

Milroy, J., & Milroy, L. (1985). Linguistic change, social network and speaker innovation. *Journal of Linguistics*, *21*(2), 339–384. doi: 10.1017/S0022226700010306

Milroy, J., Milroy, L., Hartley, S., & Walshaw, D. (1994). Glottal stops and Tyneside glottalization: Competing patterns of variation and change in British English. *Language Variation and Change*, *6*, 327–357. doi: 10.1017/S095439450000171X

Ministry of Education. (2007). *The New Zealand curriculum.* Retrieved from nzcurriculum.tki.org.nz/The-New-Zealand-Curriculum

Morton, J. (1969). Interaction of information in word recognition. *Psychological Review*, *76*(2), 165–178. doi: 10.1037/h0027366

Munro, M. J., Derwing, T. M., & Flege, J. E. (1999). Canadians in Alabama: A perceptual study of dialect acquisition in adults. *Journal of Phonetics*, *27*(4), 385–403. doi: 10.1006/jpho.1999.0101

Murray, W. S., & Forster, K. I. (2004). Serial mechanisms in lexical access: The rank hypothesis. *Psychological Review*, *111*(3), 721–756. doi: 10.1037/0033-295X.111.3.721

Nguyen, N., Shaw, J. A., Tyler, M. D., Pinkus, R. T., & Best, C. T. (2015). Affective attitudes towards Asians influence perception of Asian-accented vowels. In The Scottish Consortium for ICPhS 2015 (Ed.), *Proceedings of the 18th international congress of phonetic sciences.* Glasgow: University of Glasgow. Retrieved from http://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS2015/Papers/ICPHS0561.pdf

Niedzielski, N. (1999). The effect of social information on the perception of sociolinguistic variables. *Journal of Language and Social Psychology*, *18*(1), 62–85. doi: 10.1177/0261927X99018001005

Nielsen, K. (2011). Specificity and abstractness of VOT imitation. *Journal of Phonetics*, *39*(2), 132–142. doi: 10.1016/j.wocn.2010.12.007

Nokes, J., & Hay, J. (2012). Acoustic correlates of rhythm in New Zealand English: A diachronic study. *Language Variation and Change*, *24*, 1–31. doi: 10.1017/S0954394512000051

Norris, D., & McQueen, J. M. (2008). Shortlist B: A Bayesian model of continuous speech recognition. *Psychological Review*, *115*(2), 357–395. doi: 10.1037/0033-295X.115.2.357

Norris, D., McQueen, J. M., & Cutler, A. (2003). Perceptual learning in speech. *Cognitive Psychology*, *47*(2), 204–238. doi: 10.1016/S0010-0285(03)00006-9

Nosofsky, R. M. (1985). Overall similarity and the identification of separable-dimension stimuli: A choice model analysis. *Perception & Psychophysics*, *38*(5), 415–432. doi: 10.3758/BF03207172

Nosofsky, R. M. (1986). Attention, similarity, and the identification-categorization relationship. *Journal of Experimental Psychology: General*, *115*(1), 39–61. doi: 10.1037/0096-3445.115.1.39

Nosofsky, R. M. (1991). Tests of an exemplar model for relating perceptual classification and recognition memory. *Journal of Experimental Psychology: Human Perception and Performance*, *17*(1), 3–27. doi: 10.1037/0096-1523.17.1.3

Nowak, M. A., & Krakauer, D. C. (1999). The evolution of language. *Proceedings of the National Academy of Sciences*, *96*(14), 8028–8033. doi: 10.1073/pnas.96.14.8028

O'Flaherty, S. (2015, March 10). Do you speak Kiwinglish? New Zealand's distinct linguistic identity. *The Guardian*. Retrieved from http://www.theguardian.com/media/mind-your-language/2015/mar/10/do-you-speak-kiwinglish-new-zealands-distinct-linguistic-identity

Ohala, J. J. (1981). The listener as a source of sound change. In C. S. Masek, R. A. Hendrick, & M. F. Miller (Eds.), *Papers from the parasession on language and behavior* (pp. 178–203). Chicago, IL: Chicago Linguistic Society.

Onysko, A., & Calude, A. (2013). Comparing the usage of Māori loans in spoken and written New Zealand English: A case study of *Māori*, *Pākehā*, and *Kiwi*. In E. Zenner & G. Kristiansen (Eds.), *New perspectives on lexical borrowing.* New

York, NY: Mouton de Gruyter. doi: 10.1515/9781614514305.143

Pardo, J. S. (2006). On phonetic convergence during conversational interaction. *The Journal of the Acoustical Society of America*, *119*(4), 2382–2393. doi: 10.1121/1.2178720

Pardo, J. S., Urmanche, A., Wilman, S., & Wiener, J. (2017). Phonetic convergence across multiple measures and model talkers. *Attention, Perception, and Psychophysics*, *79*(2), 637–659. doi: 10.3758/s13414-016-1226-0

Perrachione, T. K., Chiao, J. Y., & Wong, P. C. (2010). Asymmetric cultural effects on perceptual expertise underlie an own-race bias for voices. *Cognition*, *114*(1), 42–55. doi: 10.1016/j.cognition.2009.08.012

Phillips, B. S. (1981). Lexical diffusion and Southern *tune*, *duke*, *news*. *American Speech*, *56*(1), 72–78. doi: 10.2307/454480

Phillips, B. S. (1984). Word frequency and the actuation of sound change. *Language*, *60*(2), 320–342. doi: 10.2307/413643

Pickering, M. J., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences*, *27*(2), 169–225. doi: 10.1017/s0140525x04000056

Pierrehumbert, J. B. (2001). Exemplar dynamics: Word frequency, lenition and contrast. In J. Bybee & P. Hopper (Eds.), *Frequency and the emergence of linguistic structure* (pp. 137–157). Amsterdam: John Benjamins. doi: 10.1075/tsl.45.08pie

Pierrehumbert, J. B. (2002). Word-specific phonetics. In C. Gussenhoven & N. Warner (Eds.), *Laboratory phonology VII* (pp. 101–139). Berlin: Mouton de Gruyter. doi: 10.1515/9783110197105.101

Pierrehumbert, J. B., & Granell, R. (2018). On hapax legomena and morphological productivity. In *Proceedings of SIGMORPHON 2018.* Stroudsburg, PA: Association for Computational Linguistics. Retrieved from http://www.aclweb.org/anthology/W18-5814

Pitt, M. A., & Samuel, A. G. (1993). An empirical and meta-analytic evaluation of the phoneme identification task. *Journal of Experimental Psychology: Human Perception and Performance*, *19*(4), 699–725. doi: 10.1037/0096-1523.19.4.699

Podesva, R. J. (2007). Phonation type as a stylistic variable: The use of falsetto in constructing a persona. *Journal of Sociolinguistics*, *11*(4), 478–504. doi: 10.1111/j.1467-9841.2007.00334.x

Purnell, T., Idsardi, W., & Baugh, J. (1999). Perceptual and phonetic experiments on American English dialect identification. *Journal of Language and Social Psychology*, *18*(10), 10–30. doi: 10.1177/0261927X99018001002

Reyes, A. (2005). Appropriation of African American slang by Asian American youth. *Journal of Sociolinguistics*, *9*(4), 509–533. doi: 10.1111/j.1360-6441.2005.00304.x

Sankoff, G., & Blondeau, H. (2007). Language change across the lifespan: /r/ in Montreal French. *Language*, *83*(3), 560–588. doi: 10.1353/lan.2007.0106

Savin, H. B. (1963). Word-frequency effect and errors in the perception of speech. *The Journal of the Acoustical Society of America*, *35*(2), 200–206. doi: 10.1121/1.1918432

Schulman, A. I. (1967). Word length and rarity in recognition memory. *Psychonomic Science*, *9*(4), 211–212. doi: 10.3758/BF03330834

Shepard, R. N. (1957). Stimulus and response generalization: A stochastic model relating generalization to distance in psychological space. *Psychometrika*, *22*(4), 325–345. doi: 10.1007/BF02288967

Shepard, R. N. (1958). Stimulus and response generalization: Deduction of the generalization gradient from a trace model. *Psychological Review*, *65*(4), 242–256. doi: 10.1037/h0043083

Sibley, C. G., & Liu, J. H. (2007). New Zealand = bicultural? Implicit and explicit associations between ethnicity and nationhood in the New Zealand context. *European Journal of Social Psychology*, *37*, 1222–1243. doi: 10.1002/ejsp.459

Smith, N. J., & Levy, R. (2013). The effect of word predictability on reading time is logarithmic. *Cognition*, *128*(3), 302–319. doi: 10.1016/j.cognition.2013.02.013

Sóskuthy, M. (2013). *Phonetic biases and systemic effects in the actuation of sound change* (Unpublished doctoral dissertation, University of Edinburgh). Retrieved from http://www.era.lib.ed.ac.uk/handle/1842/8946

Sóskuthy, M. (2014). *Explaining lexical frequency effects: A critique and an alternative account.* Paper presented at Sound Change in Interacting Human Systems, 3rd Biennial Workshop on Sound Change, University of California, Berkeley.

Stanford, J. N., & Kenny, L. A. (2013). Revisiting transmission and diffusion: An agent-based model of vowel chain shifts across large communities. *Language Variation and Change*, *25*(2), 119–153. doi: 10.1017/S0954394513000069

Statistics New Zealand. (2006a). *2006 census data – QuickStats about incomes.*

Retrieved from http://www.stats.govt.nz/~/media/Statistics/Census/2006-reports/quickstats-subject/Incomes/2006-census-quickstats-about-incomes-tables.xls

Statistics New Zealand. (2006b). *Canterbury region census statistics.* Retrieved from http://cdm20045.contentdm.oclc.org/digital/collection/p20045coll20/id/45/rec/6

Staum Casasanto, L. (2008). Using social information in language processing. In S. Berson et al. (Eds.), *Proceedings of the Berkeley Linguistics Society 34* (pp. 305–316). doi: 10.3765/bls.v34i1.3578

Stausland Johnsen, S. (2015). Dialect change in South-East Norway and the role of attitude in diffusion. *Journal of Sociolinguistics*, *19*(5), 612–642. doi: 10.1111/josl.12156

Strand, E. A., & Johnson, K. (1996). Gradient and visual speaker normalization in the perception of fricatives. In D. Gibbon (Ed.), *Natural language processing and speech technology: Results of the 3rd KONVENS conference* (pp. 14–26). Berlin: Mouton. doi: 10.1515/9783110821895-003

Stubbe, M., & Holmes, J. (1995). You know, eh and other 'exasperating expressions': An analysis of social and stylistic variation in the use of pragmatic devices in a sample of New Zealand English. *Language & Communication*, *15*(1), 63–88. doi: 10.1016/0271-5309(94)00016-6

Sumner, M., & Kataoka, R. (2013). Effects of phonetically-cued talker variation on semantic encoding. *The Journal of the Acoustical Society of America Express Letters*, *134*(6), 485–491. doi: 10.1121/1.4826151

Sumner, M., Kim, S. K., King, E., & McGowan, K. B. (2014). The socially weighted encoding of spoken words: A dual-route approach to speech perception. *Frontiers in Psychology*, *4*. doi: 10.3389/fpsyg.2013.01015

Sumner, M., & Samuel, A. G. (2009). The effect of experience on the perception and representation of dialect variants. *Journal of Memory and Language*, *60*(4), 487–501. doi: 10.1016/j.jml.2009.01.001

Szakay, A. (2006). Rhythm and pitch as markers of ethnicity in New Zealand English. In P. Warren & C. I. Watson (Eds.), *Proceedings of the 11th Australasian international conference on speech science & technology* (pp. 421–426). Canberra: Australasian Speech Science & Technology Association. Retrieved from http://www.assta.org/sst/2006/sst2006-102.pdf

Szakay, A. (2012). Voice quality as a marker of ethnicity in New Zealand: From acoustics to perception. *Journal of Sociolinguistics*, *16*(3), 382–397. doi: 10.1111/j.1467-9841.2012.00537.x

Szakay, A., Babel, M., & King, J. (2016). Social categories are shared across bilinguals' lexicons. *Journal of Phonetics*, *59*, 92–109. doi: 10.1016/j.wocn.2016.09.005

Tamminga, M. (2014). Sound change without frequency effects: Ramifications for phonological theory. In R. E. Santana-LaBarge (Ed.), *Proceedings of the 31st West Coast conference on formal linguistics* (pp. 457–465). Somerville, MA: Cascadilla Proceedings Project. Retrieved from http://www.lingref.com/cpp/wccfl/31/abstract3049.html

Thompson, W. (1990). Attitudes to Māori and the use of Māori lexical items in English. In J. Holmes (Ed.), *Wellington working papers in linguistics* (Vol. 1, pp. 37–46). Wellington: Victoria University of Wellington.

Todd, S., Pierrehumbert, J. B., & Hay, J. (2019). Word frequency effects in sound change as a consequence of perceptual asymmetries: An exemplar-based model. *Cognition*, *185*, 1–20. doi: 10.1016/j.cognition.2019.01.004

Trudgill, P. (2008). Colonial dialect contact in the history of European languages: On the irrelevance of identity to new-dialect formation. *Language in Society*, *37*(2), 241–254. doi: 10.1017/S0047404508080287

Tupper, P. F. (2015). Exemplar dynamics and sound merger in language. *SIAM Journal on Applied Mathematics*, *75*(4), 1469–1492. doi: 10.1137/140998408

VanDam, M. (2007). *Plasticity of phonological categories* (Unpublished doctoral dissertation, Indiana University). Retrieved from http://search.proquest.com/docview/304855320

Wagenmakers, E.-J., Zeelenberg, R., & Raaijmakers, J. G. (2000). Testing the counter model for perceptual identification: Effects of repetition priming and word frequency. *Psychonomic Bulletin & Review*, *7*(4), 662–667. doi: 10.3758/BF03213004

Walsh, M., Möbius, B., Wade, T., & Schütze, H. (2010). Multilevel Exemplar Theory. *Cognitive Science*, *34*(4), 537–582. doi: 10.1111/j.1551-6709.2010.01099.x

Wedel, A. (2004). Category competition drives contrast maintenance within an exemplar-based production / perception loop. In J. Goldsmith & R. Wicentowski (Eds.), *Proceedings of the seventh meeting of the ACL special interest group in computational phonology* (pp. 1–10). Stroudsburg, PA: Association for Computational Linguistics. doi: 10.3115/1622153.1622154

Wedel, A. (2006). Exemplar models, evolution and language change. *The Linguistic Review*, *23*(3), 247–274. doi: 10.1515/TLR.2006.010

Wedel, A. (2012). Lexical contrast maintenance and the development of sub-lexical contrast systems. *Language and Cognition*, *4*(4), 319–355. doi: 10.1515/langcog-2012-0018

Wedel, A., & Fatkullin, I. (2017). Category competition as a driver of category contrast. *Journal of Language Evolution*, *2*(1), 77–93. doi: 10.1093/jole/lzx009

Weinreich, U., Labov, W., & Herzog, M. I. (1968). Empirical foundations for a theory of language change. In W. P. Lehmann & Y. Malkiel (Eds.), *Directions for historical linguistics* (pp. 95–195). Austin, TX: University of Texas Press.

Wierda, S. M., Taatgen, N. A., van Rijn, H., & Martens, S. (2013). Word frequency and the attentional blink: The effects of target difficulty on retrieval and consolidation processes. *PLoS ONE*, *8*(9), e73415. doi: 10.1371/journal.pone.0073415

Yu, A. C., Abrego-Collier, C., & Sonderegger, M. (2013). Phonetic imitation from an individual-difference perspective: Subjective attitude, personality and "autistic" traits. *PLoS ONE*, *8*(9). doi: 10.1371/journal.pone.0074746

Zeileis, A., Kleiber, C., & Jackman, S. (2008). Regression models for count data in R. *Journal of Statistical Software*, *27*(8). doi: 10.18637/jss.v027.i08

Zhang, Q. (2018). *Language and social change in China*. New York, NY: Routledge.

Zipf, G. K. (1935). *The pscyho-biology of language*. Oxford: Houghton Mifflin.