THE UNIVERSITY of EDINBURGH

This thesis has been submitted in fulfilment of the requirements for a postgraduate degree (e.g. PhD, MPhil, DClinPsychol) at the University of Edinburgh. Please note the following terms and conditions of use:

Direction and directedness in language change
An evolutionary model of selection by trend-amplification

Kevin Stadler

PhD Linguistics & English Language
The University of Edinburgh
2016

Declaration

I hereby declare that this thesis is of my own composition, and that it contains no material previously submitted for the award of any other degree. The work reported in this thesis has been executed by myself, except where due acknowledgement is made in the text.

Kevin Stadler

Abstract

Human languages are not static entities. Linguistic conventions, whose social and communicative meaning are understood by all members of a speech community, are gradually altered or replaced, whether by changing their forms, meanings, or by the loss of or introduction of altogether new distinctions. How do large speech communities go about re-negotiating arbitrary associations in the absence of centralised coordination?

This thesis first provides an overview of the plethora of explanations that have been given for language change. Approaching language change in a quantitative and evolutionary framework, mathematical and computational modelling is put forward as a tool to investigate and compare these different accounts and their purported underlying mechanisms in a rigorous fashion.

The central part of the thesis investigates a relatively recent addition to the pool of mechanisms that have been proposed to influence language change: I will compare previous accounts with a *momentum-based selection* account of language change, a replicator-neutral model where the popularity of a variant is modulated by its *momentum*, i.e. its *change in frequency of use* in the recent past. I will discuss results from a multi-agent model which show that the dynamics of a trend-amplifying mechanism like this are characteristic of language change, in particular by exhibiting spontaneously generated s-shaped transitions. I will also discuss several empirical predictions made by a momentum-based selection account which contrast with those that can be derived from other accounts of language change.

Going beyond theoretical arguments for the role of trends in language change, I will go on to present fieldwork data of speakers' awareness of ongoing syntactic changes in the Shetland dialect of Scots. Data collected using a novel questionnaire methodology show that individuals possess explicit knowledge about the direction as well as current progression of ongoing changes, even for grammatical structures which are very low in frequency. These results complement previous experimental evidence which showed that individuals both possess and make use of implicit knowledge about age-dependent usage differences during ongoing sound changes.

Echoing the literature on evolutionary approaches to language change, the final part of the thesis stresses the importance of explicitly situating different pressures either in the domain of the *innovation* of new or else the *selection* of existing variants. Based on a modification of the Wright-Fisher model from population genetics, I will argue that trend-amplification selection mechanisms provide predictions that neatly match empirical facts, both in terms of the diachronic dynamics of language change, as well as in terms of the synchronic distribution of linguistic traits that we find in the world.

Lay Summary

The languages that we humans speak are constantly undergoing change. Words, sounds, phrases and complex grammatical patterns fall in and out of fashion, with some of them staying in use for centuries, others only for a matter of months. The explanations for these changes which are put forward by laypeople and those by professional linguists are sometimes surprisingly similar, from talk about 'lazy' articulation leading to eroded pronunciation to the fact that changes are often driven by young speakers as an act of demarcating their linguistic identity. What is common to most explanations of this kind is that they are typically 'just so' stories: accounts of specific historical changes that are come up with after the fact. This approach to 'explaining' changes fails to take into account a fundamental feature of language change, namely that it can under most circumstances not be *predicted*.

In this thesis I argue for a framework that can explain why language changes cannot be predicted, while also accounting for the fact that the types of changes that *do* occur are very similar all over the globe. To do this, I follow an evolutionary approach to language change which assumes two separate mechanisms: the first governs the creation or *innovation* of new linguistic forms, such as new words or pronunciations, while the second mechanism is responsible for the *selection* of these new forms which help them spread through a language.

Focussing on the second step of selection, I first study the dynamics of a mathematical model of *trend amplification*. The model shows how the usage of different language forms changes over time given two simple assumptions: firstly, that people can track changes in the popularity of linguistic forms that are used around them, and secondly that individuals prefer to use forms which they think are gaining in popularity. Under these assumptions, we find that the model predicts occasional changes in usage similar to what we find in real world language change, but with no way to predict *when* exactly those changes are going to occur.

Next, I present fieldwork data collected through a linguistic questionnaire filled out by inhabitants of the Shetland Islands, found to the North of Scotland. The data shows that its speakers are aware of changes in the frequency of word order patterns that are currently going on in the dialect of Scots spoken in their community.

Finally, I study a modification of a mathematical model from population biology that combines the trend-based selection mechanism with innovation pressures that favour one linguistic form over the other. While it is still not possible to predict the occurrence of particular changes, the model shows that we can see the effects of the underlying preferences for specific forms, as long as we track the occurrence of changes over long periods of time.

Acknowledgements

Only a few years ago I would have cringed at the idea of writing an Acknowledgements section listing endless swathes of names, but here we go! (This might seem unremarkable but it's another one of those things that fuels my fascination with change for the sake of change, which you have the opportunity to spend the next 200 pages reading about.)

I guess it's only appropriate to start by thanking the people responsible for the fact that the thing in front of you is actually finished: over the past four years my supervisors Simon, Kenny and Richard as well as auxiliary pundit Jenny showed absolutely no tolerance towards my impatience at my own work as well as my curiosity and eagerness to quickly move on to other topics. It was a real joy to be supervised by them, and also to witness the copious amounts of concordant swearing emanating from them, which no doubt even further fortified their authority in my eyes and helped much to keep me on track.

For providing a stimulating atmosphere to work in I want to thank all members of the LEC/CLE, in particular my PhD cohort consisting of Matt, Mark and James who provided good and challenging company while they were around, and distant motivation once they had all left me behind the (presumably s-shaped) curve. For fear of ultimate revenge I have to thank my nemesis Yasamin Motamedi – if you see her roaming around DSB in the near future with a smile on her face even more evil than usual it can only mean that the hardcopy of this thesis never made it to the college office and that I am in deep trouble. If not, she will receive her well-earned Dr. Pepper or personalised Dutch class in due course.

Of all the other postgrads that I had the joy of sharing room 1.15 with I am particularly indebted to my collaborator and future business partner E who was not only in charge of collecting the fieldwork data discussed in Chapter 5, they also tolerated my attempts at writing emails in shoddy Scots and generally kept up the craic while enjoying crisps. I'm looking forward to the time when we're making so much money together that we'll be buying each other Klimt paintings for Christmas.

I also extend my gratitude to everyone who has proofread parts of this thesis, particularly my dad for pointing out the minor detail that I've been misspelling the word *pronunciation* my entire life. On a less linguistically apt note I would like to thank my writing pal Rachael, despite the fact that she utterly failed to teach me any Geordie at all, as well as fellow failed Geordie Vanessa for thesis write-up therapy and sharing my obsession with the New Mexican desert, amongst many other things. (Oh and this entire thesis started off from a paper that she mentioned once in one of our PhD supervisions early on in my first year. No biggie.)
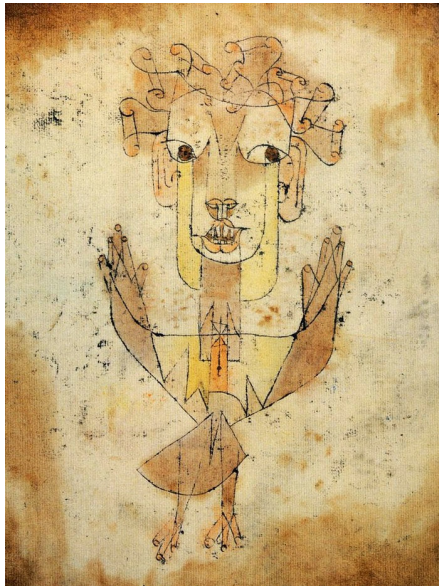
Contents

List of Figures

List of Tables

"Language moves down time in a current of its own making."

(Sapir, 1921, p.160)

Chapter 1

Introduction

he fact that human languages appear to have an inherent propensity to undergo change is well-known to linguists and observant laypeople alike. But while much quantative data on the unfolding and spread of individual changes has been collected over the past two centuries, the underlying mechanisms by which new conventions arise and spread within individual speech communities, as well as how the micro-level dynamics give rise to recurring macro-level patterns across languages, are still not fully understood. Even though changes of a similar kind do appear to re-occur across both related and unrelated languages and language families, and general tendencies in how languages are organised can be found all across the globe, the sporadic and haphazard occurrence of changes leads to the diversification of languages rather than convergence.

One fundamental question regarding language change is why it occurs at all: linguistic conventions draw their power from the fact that they are shared – both actively used and passively understood – by a group of speakers. While the initial emergence of a communication system and the introduction of new meaningful distinctions into an existing one can be of advantage to its users, most instances of language change do not entail such straightforward improvements, but rather exhibit circular patterns that leave the languages as expressive as before the change. Negotiating the replacement of one working convention by another entails some effort for a speech community for no clear functional-communicative gain. The fact that languages *do* undergo change, in combination with what is known about the very *directed* nature with which linguistic innovations spread through speech communities have led to a number of theories, 'explanations' and 'accounts' of language change.

The first central task of this thesis is to give an overview of the many different accounts of and approaches to language change that have been proposed in the literature. Although these different accounts make reference to many different 'factors', 'pressures' and 'biases' that are thought to be 'shaping', 'driving' or 'influencing' the spread of novel linguistic variants at the expense of existing conventions, I will argue that many seemingly different approaches are conceptually very similar in that they all rely on a presumed *asymmetry* between the variants in competition. Much work on language change has focussed on identifying asymmetries, primarily by gathering empirical evidence showing that not all linguistic conventions are equally preferred and that not all language changes are equally likely, both based on the analysis of historical changes in communities as well as the experimental testing of preferences in individuals. But the relative influence of the many different pressures identified and particularly the inconsistencies with which those pressures apply or don't apply at specific points in time is often left unaccounted for.

The fundamental problem is that, although language changes appear to go down the same paths over and over again, changes are not predictable in any strong sense. While many of the identified asymmetries embody strong *universal* constraints on language change, for example by identifying unidirectional patterns of change, conclusions are mostly limited to specifying on the *macro*-level whether a general type of change is more or less likely to happen to one language than another, relative to other changes. The question of why a *particular* change

occurs at all, why it does so exactly when it does (as opposed to earlier, later, or not at all), as well as how the underlying pressures give rise to the *micro*-level patterns of its diffusion cannot easily be explained through universal asymmetries. For decades, linguists have been struggling to bridge the disconnect between the idiosyncracies of individual language changes that result in the vast linguistic variation we see in the world today, and the fact that language changes across the world tend to follow similar trajectories. While it is possible to identify a pool of possible changes that a language is likely to undergo, languages appear to have some arbitrary choices over which of those path to go down, and at what point.

In order to address this problem, this thesis takes an explicit evolutionary approach to language change as *change by replication of concrete linguistic conventions*. In line with Croft (2000) I will argue that both the micro- and macro-level dynamics of language change can be explained by separating out pressures that are responsible for the *innovation* of new linguistic variants, which are in some part universal, from the pressures that drive the *selection* of specific variants at certain points in time. While there is ample empirical evidence for the functional pressures behind the innovation of new variants, the exact nature of the mechanism underlying the selection of variants out of the pool of innovations remains unclear.

The main proposal of this thesis is that the detection and amplification of *trends* in language use by individuals constitutes a concrete mechanism which can account for the second step, the occasional and seemingly arbitrary selection of new linguistic conventions. The thesis will make use of different tools, in particular computational modelling and sociolinguistic fieldwork on the perception of language changes in the individual, to argue that trend amplification is not only a viable candidate for explaining crucial aspects of the dynamics of language change, but that it naturally complements the many functional and communicative pressures which are known to influence language change.

Rather than simply contribute another model of language change to an existing pool of explanations, this thesis is equally concerned with the bigger question of how the task of 'explaining' or accounting for language change(s) is thought about in the scientific literature. In particular, I will argue that if it is not possible to predict the occurrence of specific changes, a complete theory of language change should not just limit itself to capturing the more or less predictable aspects of language change, but also provide an account of just *why language change is unpredictable*. The desire for an account of language change which predicts the unpredictability of changes might seem strange at first, but it does not imply that 'anything goes': as I will show based on a model combining the asymmetric innovation of variants with the symmetric selection mechanism based on the amplification of linguistic trends, a theory that leaves the temporal prediction of changes at the idiosyncratic micro-level underspecified can nevertheless allow for concrete, testable predictions at the macro-level.

Outline of the thesis

The remainder of this thesis is structured as follows: Chapter 2 provides an overview of the vast literature on language change. It covers both the nature of language change as well as the history of how language change has been studied up until the present day.

Chapter 3 is dedicated to the topic of mathematical and computational models of language

change. It offers a critical perspective on the subject of modelling, before presenting in-depth replications of two existing models of language change: the Utterance Selection Model on one hand, and a Markov chain model of Bayesian Iterated Learning on the other.

Chapter 4 presents a novel model of trend amplification as an as of yet understudied factor in language change. Here I augment the Utterance Selection Model with a mechanism for *momentum-based selection*, a model of trend detection and amplification in the individual that was originally proposed to account for cycles in cultural change more generally. Based on multi-agent modelling I will argue that a mechanism like momentum-based selection can account for the spontaneous and sporadic nature of the actuation of language change.

Having explored the theoretical dynamics of momentum-based selection, Chapter 5 sets out to contribute to ongoing quantitative research into individuals' awareness of ongoing changes in their community. Beyond anecdotal data, evidence to this end is currently limited to a few experimental studies on the use of implicit knowledge of sound changes during speech perception. This chapter provides further evidence by testing individuals' explicit knowledge about the direction and progress of three related changes to low frequency syntactic variables in the Shetland dialect of Scots.

Chapter 6 takes a step back to look at the bigger picture of the many pressures and biases that have been attested (or at least posited) to influence language change. Using a basic mathematical model from population genetics and augmenting it with a simple trend-amplification mechanism, I will show how the interaction between asymmetric, functionally-driven innovation pressures and a symmetric selection bias like momentum-based selection can account for both the micro- and macro-level dynamics of language change that we observe empirically.

Finally, Chapter 7 provides a summary, recapitulating the main arguments as well as pointing to potential future work, in particular relating to new research questions raised by the thesis.

Chapter 2

Studying language change

2.1   Defining language change

Most language users, whether they speak or sign, are aware of the variation found in language.
While linguistic communication is a central (if not defining) part of what makes us human, if
we look at specific conventions such as sounds, words or grammatical patterns, we find vast
variation in the different accents, dialects and languages within a community, across a country,
or across the entire globe (Evans, 2009).

The same variation has not always existed, nor did it come about through abrupt divine
intervention, as the tale of the tower of Babel would suggest. Rather, linguistic diversity is the
result of small incremental changes to a community's linguistic system which slowly accumulate
over time, first within speakers, then across generations, until eventually descendants of the
same linguistic community speak a variety that would be unintelligible to speakers from a few
generations earlier.

Language itself is a multi-layered phenomenon which can be approached and characterised
in many different ways. While there is no single definition of what exactly language is that
even a fraction of linguistic researchers would agree on, for the purposes of this thesis it will
suffice to define language as the collection of learned, conventional behaviours that members
of a given speech community use to directly communicate with each other. In its most general
sense, language change then is any change to the makeup or distribution of this collection of
conventions, as well as significant changes in their usage frequencies.

Linguistic conventions can be characterised at several levels of analysis, and languages un-
dergo change at virtually all levels of description: the inventory of physical articulation gestures
that form the basis of linguistic communication can change, whether by slightly altering the pro-
duction of existing gestures in a gradual fashion (phonetic change), replacing or dropping them
from the articulatory inventory (phonological changes such as merger and splits), or through
the introduction of new gestures and distinctions altogether (phonologisation). On the level
of meaningful signs, i.e. associations between articulation patterns and specific communicative
functions, both sides of the association can undergo change, with a form becoming associated
to a different meaning, or otherwise the same meaning coming to be expressed using a dif-
ferent form. Going beyond simple meaningful associations, the morphosyntactic patterns in
which a language augments or combines lexical items, such as word order patterns or complex
inflectional paradigms, can change as well.

While most humans are to some degree aware of the changes going on in their communities
on many of these levels without the need for scientific measuring devices of any kind (Labov,
2001; Tagliamonte, 2012), the systematic study of language changes has yielded not just many
different approaches to capturing and describing changes, but also many theories as to how the
simple fact that languages undergo change at all should be *explained*. The main purpose of the
present chapter is to argue for the relevance of the evolutionary framework of language change
adopted by this thesis, which will be presented in Section 2.5. But in order to get a better
understanding of the motivation behind the approach as well as the underlying phenomenon,
it is worthwhile to have a closer look at the history of both theorical and empirical work on
language change.

2.2 Explaining language change

Like any empirical science, linguistics (*historical* linguistics in particular) is not just concerned with *describing* natural social phenomena. The goal of scientific enquiry is to gather insights into how the phenomena at hand can be explained or accounted for, typically by making plausible claims as to their underlying mechanisms and testing falsifiable predictions made by those accounts against empirical data. Given the breadth of ways in which language changes there is a substantial amount of literature on the topic, much of which is concerned with both the documentation or *description* of language changes, as well as with providing *explanations* for the changes at hand. In this section, I will attempt to provide a structured, historical overview of the main strands of thinking about how language change is explained. While much has been written about the scientific status of different types of 'explanations' in the epistemological sense (see especially Lass 1980; McMahon 1994 and Newmeyer 1998, ch.3), I want to draw particular attention to how different approaches to language change and different types of evidence lead to divergent priorities and no consensus on how (and at what level) language change can or should be explained.

2.2.1 Early accounts

While descriptions of languages (in the sense of *grammars*) are already attested in early antiquity, documentation relating to language *changes* are far more recent (see Jespersen, 1922, ch.I). In the absence of recordings from earlier stages of a language or a written record of any reasonable time depth, early linguists had little to work with that would have allowed for a systematic study of language change. What was required to start investigations in any meaningful way was to make the inference that *dialectal* variation that can be observed for many languages was really a matter of variation that first occurred *temporally* through local change, and which had then diffused across geographical space at different rates and in different directions (Chambers and Trudgill, 1998).

While several earlier sources show that many disconnected scholars inferred this based on their study of ancient languages like Greek, Roman and Sanskrit (Jespersen, 1922, ch.II), the insight is today often credited to the orientalist Sir William Jones (1799), mainly due to the impact his particular publication had on the quickly expanding field of Indo-European studies. The diverse varieties of the then still to be established Indo-European language family, which covered (and still covers) almost the entirety of Europe's linguistic landscape, were already thoroughly documented at the time, forming a perfect testbed for the emergent methodology of linguistic reconstruction.

As soon as historical linguistics had moved from describing superficial similarities between lexical items or morphosyntactic constructions to positing concrete sound correspondences based on the idea of the regularity of sound changes (Paul, 1880), the first theories attempting to explain the correspondences, or rather the changes presumed to have led to the later correspondences, arose. Before the idea that all languages served some universal functions and therefore had to obey particular constraints had taken hold within linguistics, the first such ac-

counts were primarily characterised by creative psychological or even intentionalistic reasoning behind changes (see also Ross and Durie, 1996, p.14). A particular example is Grimm (1848)'s attempt to explain the Germanic consonant shift, a regular sound change describing a change in the articulation of the stop consonants reconstructed for Proto-Indo-European which posits staged processes of de-aspiration, devoicing and spiration:

> By no means do I want to claim that this change occurred without any disadvantage, in some sense sound shifting appears to be barbarism which other, quieter people abstained from, but this has to do with the immense advancements and desire of freedom of the Germans …they ventured into the innermost sounds of their language (p.417, own translation)

From such fanciful ad-hoc accounts which were merely limited by the creative abilities of the linguistic researcher, explanations moved to the somewhat more accountable domain of physiological factors. Trying to account for the High German consonant shift, Grimm finds that "experience teaches us that mountain air makes sounds sharp and rough, where the flat land makes them soft and dull" (ibid, p.828), leading to a theory that traced the sound shift back to the increased expiration that befell the High German tribes as they settled in mountainous regions. As more and more descriptions and evidence for similar changes across different languages began to accumulate, linguists started appealing to even more general, mechanistic explanations. With the field still very much focussed on sound changes but before the advent of reliable tools for acoustic analysis and therefore reliant on what could be gathered from written records, Jespersen (1922) contemplated that

> at one moment, for some reason or other, in a particular mood, in order to lend authority or distinction to our words, we may happen to lower the jaw a little more, or to thrust the tongue a little more forward than usual, or inversely, under the influence of fatigue or laziness, or to sneer at someone else, or because we have a cigar or potato in our mouth, the movements of the jaw or of the tongue may fall short of what they usually are. We have all the while a sort of conception of an average pronunciation, of a normal degree of opening or of protrusion, which we aim at, but it is nothing very fixed, and the only measure at our disposal is that we are or are not understood. (p.166)

While the particular depiction of what might today be thought of as underspecification or coarticulation effects is rather baroque, the quote exemplifies a general trend in the field to abandon particular explanations for particular changes in favour of more general *linguistic* pressures that apply across languages (see also Bloomfield, 1933, ch.21). This period gave rise to what is probably the biggest framework for explaining changes whose legacy extends well into the present day, in what can be referred to as *language-internal* accounts of change.

2.2.2   Language-internal accounts

The defining feature of language-internal accounts of change is that they explain the occurrence of changes based on an *asymmetry* between the language states before and after the change that is *internal* to the linguistic system and its acquisition. While there is no strict definition of exactly what makes an account language-internal, one finds several attempts to delineate

the approach in the literature, e.g. that "specific changes are thought to be internally caused when there is no evidence for external causation, i.e. for language contact" (Luraghi, 2010, p.366). While the matter of distinguishing 'internal' from 'external' causation is far from simple when investigated in detail, we find similar definitions across the literature, e.g. handbook definitions of internal change as "[a]ny change which can be traced to structural considerations in a language and which is independent of sociolinguistic factors can be classified as internally-motivated" (Hickey, 2012, p.388). These accounts generally differ from earlier accounts by not attempting to ground language changes in non-linguistic features of their specific speaker groups, nor in other external inter-individual or political dynamics that form the basis of *social accounts* to be discussed below.

In terms of the scientific treatment of language change, language-internal accounts signified a transition from the study of *particular* changes to that of general, *universal* principles of change. This move was primarily supported by the methodology of and results obtained from the systematic study of sound changes. On one hand, the success of the new dogma of the *exceptionlessness* of sound changes indicated that language change was not completely erratic, but followed rule-like patterns. In terms of scientific thinking about language change, this suggested that it might be possible to discover the underlying mechanical principles of change, which would turn linguistics from a historical social science to one of more or less deterministic explanation like in the physical sciences (Sapir, 1929).

The idea that it should be possible to move beyond the mere post-hoc description of changes to their prediction was not just due to the theoretical framework behind linguistic reconstruction, but also due to the empirical results obtained from the method. The systematic documentation of sound changes across many languages and language families revealed *unidirectional* developments, the catalogisation of which forms part and parcel of any historical linguistics textbook to this today (see e.g. Campbell, 2013). A prime example are changes in the realisation of phonemes which were copiously attested in one direction, but hardly the other, such as the frequent frication of voiceless stops, as opposed to the more unusual stopping of fricatives. While the identification of these universal *asymmetries* was perceived as an improvement over the earlier ad-hoc accounts of individual changes, the lack of a well-defined framework for describing languages meant that researchers were left with plenty explanatory leeway in their classification of individual changes, so as to account for them as an instance of some universal tendency.

In a recursive manner, the observation of general trends or constraints on the *direction* of changes gave rise to the postulation of general principles which were then invoked to explain the very changes they were derived from (Lass, 1980; Haspelmath, 2006). While this approach certainly helped to identify paths that sound changes were unlikely to go down, it allowed the occurrence of any individual change to be 'explained' by picking from a broad inventory of possible or 'likely' changes that even included ones going in opposite directions, for example the processes of phonetic assimilation and dissimilation, both of which would be considered 'natural' developments.

While the comparatively 'simple' matter of sound change was often maintained to be due to the gradual *accumulation of errors* in pronunciation already hinted at above (see e.g. Hockett,

1958, for a textbook account of the general principle), the rising interest in syntax during the advent of generativism also resulted in more precise theories of how syntactic changes unfold over time. While syntax constituted an arguably more complex domain of inquiry than phonetics or phonology, the basic idea of change as a matter of *mislearning*, i.e. failing to acquire the 'correct' target convention from ones linguistic environment, was transferred from sound to morphosyntactic changes (Salmons and Honeybone, 2013). The main difference between the fields was that, with its underlying assumption of an idealised language capacity and particular focus on the *language acquisition device*, generativist explanations sought mislearning mainly in children's acquisition of a language's syntactic rules. Given the purpose-specific nature of said acquisition device, child-based language change was largely explained to be due to changes in the 'primary linguistic data' from which the device is meant to derive a language's underlying grammar.

The relative ease with which categorical morphosyntactic changes could be described also gave rise to a large body of quantitative work related to grammatical parameter setting (Lightfoot, 1991) with concrete claims about its reliance on specific learning cues in linguistic data (Gibson and Wexler, 1994). In contrast to earlier work on sound changes which still struggled with the lack of precise quantitative description of its subject matter, the exact description of specific cases of syntactic change as well as concrete postulated thresholds for the amount of external data required to correctly acquire their underlying grammars led to specific models that were meant to recapitulate specific historical changes, e.g. Yang (2002).

It is interesting to note how the 'success' or correctness of an *explanation* of language change is evaluated in language-internal accounts of particular changes, as opposed to those oriented towards the identification of general pressures mentioned above. Given that the causal triggering of the change is framed to be internal to the language, rather than due to some arbitrary external, historical factor, the particular language changes under investigation are more or less explicitly framed as inevitable, rather than accidental. Whether the asymmetry is based on universal articulation biases or on changes to the learning data (in particular the frequency of use of morphosyntactic constructions or lexical items, which are for some reason also often regarded as inevitable language-internal facts), an account (and especially a quantitative model) of a specific change would be considered 'wrong' if it failed to predict the occurrence of the change.

The fact that there are many models that 'successfully' predict the occurence of particular changes anticipates a criticism of language-internal accounts that will be elaborated on later, namely that it is always possible to find some asymmetry between the competing language states that biases acquisition towards the later stage of the language. Although language-internal accounts appear to be in principle falsifiable, the fact that the methodology is geared towards the identification of successful particular explanations is not merely a theoretical concern (Lass, 1980): the vast majority of published accounts aimed at identifying specific language-internal causes or triggers for a change do so 'successfully' in the sense of correctly predicting particular historical changes (e.g. Troutman et al., 2008; Ritt and Baumann, 2012; van Trijp, 2013, among many others). While there are also publications which consider unsuccessful, i.e. historically unattested (competitor) models (e.g. Sonderegger and Niyogi, 2010) or those which

also consider a model successful even if it underdetermines the direction of changes to some degree (e.g. Lau, 2016) these are the exception rather than the rule. The construction of quantitative language-internal models of specific changes provides a precise account of the supposed underlying mechanisms responsible for those particular changes, but it offers no generalisability beyond that one historical event.

Alongside the generative work focussed on explaining grammatical change as internal to syntax, cross-linguistic typological approaches to language change gave rise to an explanatory framework similar to that used in sound change. In parallel to the directional tendencies that had been revealed to guide sound changes, the study of *grammaticalisation* suggested a similar pattern where closed-group, bound grammatical morphemes can be traced back to formerly independent lexical items with specific referential content (Hopper and Traugott, 1993). In contrast, barely any examples of the development of (bound) grammatical morphemes becoming standalone lexical items can be found across the world's languages (English 'ish' being one of the rare exceptions). This apparent *irreversibility* of grammaticalisation (Haspelmath, 2004) lent support to a new set of universal asymmetries due to the unidirectionality of both semantic and phonetic erosion that is thought to guide morphosyntactic changes. The eternal cycle of wear of morphosyntactic material and consequent need for replenishment suggested by grammaticalisation theory also fit neatly with the universal circular nature of language changes suggested by more descriptive level work by typologists (Hodge, 1970).

The various approaches to language change mentioned so far come from very different camps of linguistic thinking and cover very different aspects of language change. However, the emerging unifying principle of language-internal explanations is the conceptually similar approach of accounting for changes in a *mechanistic* way due to them being based on some *inherent asymmetry*, either between competing linguistic variants, or between the language state before and after the change. The underlying assumption of all explanations so far is that the acquisition or use of the original language state is prone to lead to the accidental acquisition or realisation of the target state. Exactly how this asymmetry is framed depends mostly on the particular theoretical approach taken, ranging from talk about *pressures* (e.g. Nedergaard Thomsen, 2006), a nod towards deterministic explanation in the natural sciences, over *motivations* (MacWhinney et al., 2014) and *preferences* (Fedzechkina et al., 2016) in psychologically-grounded work, all the way to *biases*, the latter term used both generally as well as in reference to learning and inference frameworks in particular. Putting forward such a mechanistic language-internal account for a particular historical change often implies that that change was in some sense inevitable, an approach to explaining – in the sense of deterministically predicting – the occurrence of specific changes that is still most widespread today in the domain of morphosyntactic changes (Lightfoot, 2010).

Simiarly, in the study of sound change mechanistic error-based accounts continue to be popular (see e.g. Garrett and Johnson, 2013, who make specific reference to the importance of *asymmetries* between phonetic variants to explain change), now encompassing not just phonetic changes but also exhibiting an increasing interest in how *phonologisation* of formerly non-segmental features gives rise to novel phonemic contrasts (e.g. Kirby, 2013).

While providing insight into the ways and trajectories in which languages change, language-

internal accounts have long been criticised for underspecifying *when* changes occur, a conceptual and scientific issue that has come to be known to historical linguists as the *actuation problem*. But before turning to the actuation problem, I will turn to language-internal accounts' bigger and bolder modern sibling, based on the characterisation of language as a *complex adaptive system*.

2.2.3 Competing pressures and language as a complex adaptive system

The turn towards novel research methods and quest for generalisations across languages has seen the rise of a new research moniker that frames *language as a complex adaptive system* (*LCAS* for short, see e.g. Steels, 1998, 2000; Smith, 2003; Situngkir, 2004). This general term provides a catchy name for a wide range of branches of linguistic research which do not necessarily share a common methodology or evaluation criteria. Rather, it covers approaches within specific fields that acknowledge the important role of competing factors from a variety of sources (cognitive, physical or social) while stressing the fact that languages are continuously adapting based on the needs and requirements of everyday language use in interaction (Beckner et al., 2009). Just as the transition from early ad-hoc accounts to more theoretically motivated language-internal explanations was gradual, there is also no clear divide between the LCAS approach and language-internal accounts, many of which also developed an interest in the intricate *interacting factors* and *competing motivations* at play in language change (Vachek, 1962; Berg, 1998; Nedergaard Thomsen, 2006; Wedel, 2009; MacWhinney et al., 2014). Crucially, when it comes to the study of language *change* as a natural outcome of language's continuous adaptation, the approach shares with language-internal accounts that its focus is explicitly on the *universal* patterns and direction of change (Beckner et al., 2009, p.4-5).

Rather than representing a tight methodological framework, LCAS also stands for an openness towards novel and experimental tools to further our understanding of language use and change, in particular approaches that go beyond explanations that could be based on mere 'arm chair reasoning'. In order to get to grips with the kinds of dynamics implied by the word 'complex', mathematical and particularly computational modelling have become a popular tool to investigate emerging and evolving communication systems. Since the entirety of Chapter 3 is dedicated to both the topic of modelling in general as well as replications of some specific models, I will for now focus on the results obtained from empirical research that is thought to speak to the adaptive nature of linguistic changes.

Inferring universal pressures from cross-linguistic data

While diachronically-minded linguists have long been grouping and categorising similar changes to uncover general trends in the paths taken by language change, typologists have arrived at a similar point but from the opposite direction. Based on comparative work on the synchronic distribution of linguistic features such as Greenberg (1963), the preponderance of individual features (or co-occurrence of features) was taken as evidence for some sort of adaptive advantage to those features or feature constellations (Haspelmath, 2008).

In line with LCAS principles, the adaptive pressures are assumed to be associated with

language *use*, which would over time influence the development of languages and consequent distribution of grammatical properties across the world. One contrast between such typological approaches and traditional language-internal accounts is the assumed *specificity* of the pressures that are thought to steer language change. While language-internal accounts often refer to linguistic concepts and mechanisms that are theory- or language-specific, typological work is based on more *descriptive* measures that lend themselves to cross-linguistic comparison (Haspelmath, 2010).

Rather than embodying a simple, qualitative asymmetry between specific competing variants, typological pressures are typically described in *quantitative* terms. For example, Zipf (1935, 1949) popularised (and sought to explain) the finding that, across many languages, the frequency of a word is inversely proportional to its rank in the frequency table of all words of that language. Having such a distribution is indicative of some sort of optimisation since, if more frequently used words are shorter, the average length of signalling decreases overall. While it has since been shown experimentally that individuals will preferentially produce abbreviated forms when it is more effective in information-theoretic terms (Mahowald et al., 2013; Kanwal et al., 2017), a bias for the selective propagation of changes towards such advantageous forms through larger speech communities has not been demonstrated yet.

An early example of a general, quantative pressure whose effect was postulated to affect language change directly is that of a phoneme's *functional load* (Martinet, 1955). The phoneme, typically defined to be the largest non-meaning-bearing unit of linguistic analysis, gains its function as a linguistic category by providing meaning distinctions at the higher level of *combinatorially created* meaningful signals, or *morphemes*. In this way, a phoneme owes its existence to linguistic items at a higher level of organisation which allow it to stand in contrast to other phonemes, typically in the form of lexical minimal pairs. In the absence of any such minimal pairs, upholding a phonemic contrast is inefficient in the sense that it would move the lexicon away from being a productive system built on re-using a limited set of signals towards an unstructured inventory of individual holistic lexical items, with as many distinct articulation gestures in the language as there are morphemes (Spike et al., 2016). In practice, however, it is possible to find several minimal pairs for most pairs of phonemes in any given human language. Combining this observation with studies of the reorganisation of sound systems, especially the *merger* or *contrast loss* of formerly distinct phonemes, led Martinet (1955) to propose that a phoneme's *functional load*, i.e. the amount of contrastive function it performs at the level of the lexicon, would have an influence on whether or not it would be maintained as a contrastive element relative to other phonemes. The concept fits well under the modern LCAS umbrella, seeing as it describes a delicate trade-off between competing pressures of reducing the number of distinct signals maintained by a language against the pressure of keeping distinct phonemes to allow combinatorial signalling to occur at all. While the direct link between number of minimal pairs and contrast maintenance has since seen confirmation based on population-level diachronic data (Wedel et al., 2013), the question of whether the observed pattern satisfies some externally-optimal equilibrium or whether it is simply based on the peculiarities of individual human phonemic acquisition still needs to be addressed (Spike, 2016).

More recently, theories speaking to the efficient organisation of language have also taken a

firm hold in thinking about linguistic structures and inventories at other levels. Based on the functional typology work referred to above, the availability of large-scale syntactic corpora has allowed quantitative investigations into the syntactic organisation of languages based on actual usage data, rather than abstract principles of syntactic description. In particular, studies of the average syntactic dependency length of data for various languages, which is taken as a measure of an utterance's processing complexity, have revealed that many of the correlations between constituency orders for different syntactic constructions can be explained based on simple preferences or biases towards efficient processing (Futrell et al., 2015) or language 'utility' (Jaeger and Tily, 2010).

While the pressures investigated in all these approaches are typically motivated by general cognitive principles at the level of the individual, most of the evidence for their existence is gathered at the very macro-level, often from cross-linguistic data, where those biases are thought to have expressed themselves over time. The question of whether (and how) individual biases become manifest in languages on the inter-individual population level is itself a non-trivial question (Kirby, 1999). The necessity to go beyond macro-level descriptive data to address this question has long been recognised, and given rise to an experimental framework in the LCAS spirit, known as *iterated learning*.

Universal pressures and the Iterated Learning Model

As the name suggests, the Iterated Learning Model (ILM for short) actually started off as a computational model, primarily to investigate the question of innate biases related to the compositionality of human languages (Kirby, 2000; Brighton, 2002; Kirby, 2002). Counter the idea of a strongly innate language capacity that had been dominating most thinking about morphosyntax since Chomsky's 'poverty of the stimulus' argument, these first computational efforts at modelling the emergence of languages suggested that weak, domain-general biases in the individual were enough to yield strong linguistic universals at the population level when learning and production were *iterated*, i.e. repeated over generations of learners (Kirby et al., 2004).

While the iterated learning paradigm has undergone gradual transformation in its scope and goals since these early models (to the point of encompassing any and all experiments or models of repeated communication according to some definitions, see Scott-Phillips and Kirby 2010, Box 3), one aspect that still features strongly in most of the model's incarnations is the *bottleneck*. The original conception of the role of the bottleneck was that limiting the amount of data learners received would force them to generalise from the limited data they received, causing linguistic signalling systems to become more *structured* through repeated production and learning. The effectiveness of the bottleneck was first shown in computational simulations to lead to the emergence of compositionality (Kirby, 2000) as well as recursion (Kirby, 2002), and its role in triggering the emergence of compositionality was later confirmed experimentally with human participants in controlled laboratory experiments (e.g. Kirby et al., 2008; Cornish et al., 2009; Smith and Wonnacott, 2010). The same effect was also taken to show that, in contrast to the assumptions of strongly nativist theories of language, subtle biases in the individual can give rise to strong universal patterns when the learning and production of linguistic systems is

iterated (Kirby, 1999).

This latter idea has undergone implicit generalisation to other biases not necessarily directly related to language *structure*, at the latest with Kirby et al. (2004) who reframed the acquisition by individual learners as a Bayesian inference process. In the Bayesian framework, the task of a learner is to infer which of the possible grammars h out of a given grammar hypothesis space underlie the linguistic data d that they observe in learning. To do so rationally, the learner computes the *posterior probability* of all possible languages (hypotheses) given the learning data,

$$p(h|d) = \frac{p(d|h) \cdot p(h)}{p(d)} \ . \tag{2.1}$$

The Bayesian learning framework, which has come to dominate much thinking about symbolic cognitive science in the past few decades, offered a natural grounding for iterated learning. It suggested a direct correspondence between the aforementioned 'biases', previously defined as "everything that the learner brings to the task independent of the data" (Kirby et al., 2004, p.590), to the *prior* probabilities $p(h)$ of the competing hypotheses in Bayesian inference. In this sense, biases could be made explicit simply by skewing the prior probability distribution over all hypotheses away from a uniform distribution corresponding to no a priori preference for any particular language. (It should be noted that, just in terms of *asymmetries* between competing languages, a bias towards specific hypotheses h can also be introduced by adjusting the different hypotheses' data production structures $p(d|h)$.)

Combining several such Bayesian learners in an iterated learning *chain* would mean providing the first learner with some input data d from which they first try to infer the underlying hypothesis by computing their posterior distribution $p(h|d)$. The learner would then become a 'teacher' by producing some of their own data d, either based directly on their posterior, or from a specific hypothesis chosen based on the posterior, passing this data on to the next learner and so forth.

Based on interpreting such chains of Bayesian learners as a stochastic Markov model (a modelling approach that will be discussed in-depth in Section 3.4.1), it is possible to analytically derive the expected probability of observing any of the different hypotheses at a given point in time. Griffiths and Kalish (2007) provided a mathematical proof that a chain's probability of exhibiting a hypothesis in this Bayesian inference-based version of the Iterated Learning Model (BILM) should be identical to the prior distribution $p(h)$, a finding which seems to run counter the empirical evidence which showed an *amplification* of prior biases. Kirby et al. (2007) reconciled the two results by showing that, when learners adopt more deterministic strategies of selecting a hypothesis from the posterior such as choosing the hypothesis with the highest posterior probability, this will lead prior biases to become exaggerated as learning is iterated, especially when there are only small amounts of learning data available. While the question of which hypothesis selection strategy should be used can in theory be derived from experimental results for specific learning tasks, there is still no conclusive evidence on the matter (Reali and Griffiths, 2009; Ferdinand, 2015).

The iterated learning paradigm offers an individual-level complement to the macro-level analyses based on cross-linguistic data discussed above, and suggests a way in which the two

levels of individual biases and typological distributions can be linked directly (Kirby, 1999; Kirby et al., 2008). However, especially due to their origin in studies of the emergence of linguistic systems from scratch, iterated learning models are based on small population sizes, typically pairs of individuals arranged in chains or dyads, who have to learn from very limited and highly variable data in which learning biases are thought to express themselves most strongly (Fedzechkina, 2014). As such, there are still two missing intermediate links between the two levels, namely the synchronic diffusion of traits through larger populations that already possess an established communication system, as well as their consequent spread through diachrony (ibid, p.26). While the role of individual biases on the latter has received support through studies such as Wedel et al. (2013), what is known about the spread of new linguistic variants at the micro-level of the population appears to be at odds with the idea of a straightforward expression of individual biases across all levels, as I will argue below.

Before I turn to a critique of adaptationist accounts based on universal pressures, however, it is also necessary to acknowledge another branch of research situated within the LCAS paradigm dedicated to uncovering the effects of general but not quite universal pressures, and their impact on language and language change.

Adaptation to environmental pressures

Due to its general nature, the LCAS moniker does not just cover the study of *universal* pressures associated with the iterated learning model, but also the investigation of non-universal (yet still generalisable) pressures and how they affect otherwise unrelated languages and language families in similar ways. With the advent of large-scale databases on languages such as the World Atlas of Language Structures (WALS Online, 2013), statistical studies have identified correlational relationships between linguistic structures and non-universal external, non-linguistic factors – whether biological or environmental (Ladd et al., 2015).

Among the first and still thoroughly established results of this kind is Dediu and Ladd's finding that the distribution of tonal languages in the world shows an unexpectedly large degree of overlap with the geographical distribution of two genes related to brain size (Dediu and Ladd, 2007). The authors posit that there might be a causal link between the two, with the existence of a certain biological substrate biasing language acquisition or processing towards certain language structures, in this case ones incorporating tone, a bias that would become exacerbated through iterated cultural transmission.

Many other suggestive correlational patterns have since joined this result, with the purported causality going either from environmental and social factors triggering changes or adaptations in linguistic structures (Hay and Bauer, 2007; Lupyan and Dale, 2010), or conversely from linguistic features to other social behaviours (Chen, 2013). While the lack of concrete evidence for such effects beyond the stating of statistical macro-level correlations has received much criticism, it has also been argued to be an opportunity for both deriving and testing concrete mechanisms by which social and other external factors influence linguistic structure (Roberts and Winters, 2012; Nettle, 2012; Roberts and Winters, 2013). Specific claims as to how social scale can affect structural complexity indirectly, for example by changing the degree of input variability that individual learners receive (Wray and Grace, 2007), have only just begun to be

tested experimentally, lending no consistent support to any of the hypotheses that have been proposed so far (Atkinson et al., 2015; Atkinson, 2016).

With this newest branch of research on factors that are believed to be driving language change we come to the end of one long, more or less continuous arc of approaches that seek the source of language change in general, mechanistic pressures that are thought to be acting on languages at all times. Winding back to the early days of language-internal accounts I will discuss a criticism that has long been hauled at general explanations of language change and that is equally applicable to their modern, adaptively-minded incarnations: the actuation problem.

2.2.4 The actuation problem & sporadic language change

Returning to an earlier period of linguistic study, the advent of structuralism that is typically associated with the publication of Saussure's *Cours de linguistique générale* (1916) reset the focus from establishing historical relationships to the description of individual languages for their own sake. Where in 19th century linguistics research diachronic concerns still played a central role, the systematic identification of the various functions served by *synchronic* language structures under the new paradigm cast additional doubts on why languages should change at all:

> the more linguists became impressed with the existence of structure of language, and the more they bolstered this observation with deductive arguments about the functional advantages of structure, the more mysterious became the transition of a language from state to state. After all, if a language has to be structured in order to function efficiently, how do people continue to talk while the language changes, that is, while it passes through periods of lessened systematicity? (Weinreich et al., 1968, p.100)

The criticism embodied by the actuation problem is mainly directed at language-internal accounts, or any other account that attempts to explain changes through inherent asymmetries between competing variants. While the move from the (over-)explanation of particular language changes to their framing in the light of *universal* pressures that allow generalisation meant progress on the question of *why* or *in which direction* languages change, the approach was said to create "the opposite problem – of explaining why language fails to change." (p.112).

> Why do changes in a structural feature take place in a particular language at a given time, but not in other languages with the same feature, or in the same language at other times? This *actuation problem* can be regarded as the very heart of the matter. (Weinreich et al., 1968, p.102)

Although there is no precise definition of 'actuation' in Weinreich et al.'s seminal paper and no consistent usage of the term in the literature has emerged since, it is generally understood to mark *the point in time* when an incoming linguistic variant starts to see a consistent increase in its usage at the expense of established competitor variants, i.e. the detectable *onset* of the diffusion of a change[1].

---

[1] The term 'actualization' (with a specific meaning in grammaticalisation theory, see Traugott, 2011, p.24) is

While language-internal accounts went a long way in terms of identifying general, inherent instabilities in or asymmetries between two language states, they typically fail to explain the inactivity of a pressure until its point of actuation, i.e. they offer no theory of why a change occurs *exactly when it does*. This concern is aggravated by the fact that the myriad pressures that have been proposed to act on language provide a large amount of explanatory freedom by allowing researchers to choose from a whole pool of 'explanations' more or less at will. This fact is also acknowledged by some adherents of the paradigm which make explicit reference to the fact that the approach makes it difficult to connect specific pressures to particular changes:

> No rule or constraint has a motivation in and of itself, but only within the total system in which it occurs, and crucially, in the history of that system. …the interplay of explanatory factors is vastly too complex to allow individual motivations to be attached to individual grammatical elements. (Newmeyer, 2014, p.313)

Formulated almost 50 years ago, the fact that the actuation problem is still frequently referred to even in contemporary literature on language change indicates that it is not seen as resolved. Importantly, the argument applies equally to accounts referring to domain-general or even external adaptive pressures that are the interest of LCAS approaches, which were also shown to be geared towards very general explanations of change. Particularly in iterated learning models we see a strong focus on universal pressures giving rise to *universal features* such as compositionality, recursion, expressivity etc. (Brighton, 2002; Kirby, 2002; Cornish et al., 2009; Smith et al., 2013; Kirby et al., 2015). When there is room for variable expression of biases, the evolution towards any particular solution is triggered from the outside by experimental manipulation, whether due to the introduction of a pressure for social marking (Roberts, 2013) or by varying the relative importance of different language-internal pressures to solve the communicative problems of the respective experimental tasks (Winters et al., 2015). The dependence of such changes on external actuation offers no explanation of why any particular bias might suddenly come to outweigh another one and thus trigger a change in an established, functioning language.

Within work on particular historical changes, some researchers have alluded to the 'solving' of the actuation problem by pointing to currently understudied (and difficult to study) aspects of language use that might explain why the triggering of changes is so hard to pin down, for example the role of individual differences in articulation (Baker et al., 2011; Stevens and Harrington, 2013). Macro-level work on the general principles of linguistic change acknowledges the issue of actuation much less. This might not be surprising, given that the general thrust of the argument often seems to speak against the possibility of any explanation of language change that goes beyond the account of one individual change and its particular circumstances of actuation. The macro- vs. micro-level approaches to explaining language change have led to a separation of concerns, where researchers are either dedicated to explaining *which* changes actuate, generally, or otherwise to the question of when *specific* changes actuate in particular Bynon (1977). I will propose an approach that simultaneously addresses both those concerns in Section 2.5. But first we need to discuss another strand of explanations of language change, one primarily

---

sometimes used interchangably (Andersen, 2008; Kiparsky, 2014).

informed by fieldwork on the dynamics of the spread and diffusion of linguistic changes through speech communities.

2.2.5 Social accounts

In the same period of the second half of the 20th century in which typological and generative work on morphosyntactic change indicated universal trends and directions of changes, sociolinguistic research painted a wholly different picture of the spread of linguistic changes in parallel (or rather orthogonal) to language-internal accounts. Studies in what is sometimes also referred to as the *variationist* tradition (Tagliamonte, 2015) have shown that there is a vast amount of variation in language use not just between but even *within* individuals of one and the same speech community. What these investigations across different languages and cultures have confirmed is that linguistic variation is not distributed randomly but in what Weinreich et al. (1968) call *structured heterogeneity*, often reflecting the underlying social or political structure of the community with the usage of specific linguistic variants stratified according to social characteristics of the speakers such as their age, ethnicity, or socio-economic status (Foulkes and Docherty, 2006; Tagliamonte, 2012). The importance of linguistic variation for the study of language change is again based on extrapolating the observed synchronic variation into time, by recognising that linguistic innovations do not spread evenly across geographical space, or even within speaker groups.

As was the case with language-internal accounts, it is difficult to find or give a precise definition of what exactly makes a 'social account'. In contrast to language-internal accounts, social accounts are characterised by seeking the 'reason' for the adoption or diffusion of a novel linguistic variant not in (features of) the variant itself but in some external, social factors. While social accounts of language change are primarily informed by empirical research of the *micro-level* of linguistic variation and individual changes, sociolinguistic thinking goes beyond individual case studies. Similar to the historical development of language-internal accounts, sociolinguistic research has moved from the documentation and ad-hoc explanation of the social stratification of particular changes to the postulation of generalisable pressures that are presumed to hold across languages and cultures.

For the present purpose I will characterise and discuss two types of accounts that encompass most of the thinking on social influences on language change, the first one referring to *mechanics* of social interactions, the other on the more fuzzy topic of the social *meaning* of linguistic variation, and how it affects the spread of language change.

Social network accounts

Although it is again hard to find explicit definitions for different types of social accounts, there exists a relatively clear subcategory of social explanations that stresses the important effects of social interactions. Inspired by the quantitative study of social networks and work such as the idea of "weak ties" (Granovetter, 1973), researchers have attempted to reduce the diffusion of language changes to the underlying distribution of social interactions in a speech community, an idea clearly formulated very early on by Bloomfield (1933, p.476):

> The inhabitants of a settlement [...] talk much more to each other than to persons who live elsewhere. When any innovation in the way of speaking spreads over a district, the limit of this spread is sure to be along some lines of weakness in the network of oral communication.

The quantitative, rule-like accountability of this "principle of density" of communication can again be seen to hark back to the ideal of mechanistic causal explanation, this time due to external, interactional factors (Labov, 2001, p.19). The idea of reducing linguistic diversification to discontinuities in the social structures of speech communities has inspired both empirical (Milroy and Milroy, 1985; Herold, 1997; Trudgill, 2008) as well as theoretical and modelling work (Nettle, 1999; Silva and de Oliveira, 2008; Stadler, 2009; Gong et al., 2012; Blythe and Croft, 2012; Pierrehumbert et al., 2014; Renton, 2016; Kauhanen, 2017), some of which I will return to in the next chapter. However, much empirical data suggests that the relevance of such a simple mechanism is limited. Labov (2010, ch.6-10) in particular presents overwhelming evidence from work on the Atlas of North American English as well as the so-called Northern Cities vowel shift, showing how the linguistic systems of neighbouring communities with plentiful cultural and economic contact between them are in fact often diverging in opposite directions rather than converging. Rather than reducing changes to a simple mechanism of interaction, these results indicate that the adoption of competing linguistic variants is not just a matter of automatic mutual accommodation, but that each individual's choice of change is a question of the variants' social meaning in use or, more concisely, a question of *prestige*.

Prestige accounts

Once again, terminological caution is advised, this time in relation to the many possible pressures referred to by *prestige*. While prestige is intuitively understood to stand for a social bias *towards* something, the word leaves open whether this bias is towards the speech of a high-status *individual*, however they may be speaking at the time, or instead towards specific linguistic *variants* which might not actually be used by any prestigious individuals, but merely believed to be (for a more formal distinction, see Sections 3.3.3 and 3.3.4). While early uses of the word more typically refer to the (linguistic) status of individuals (Tarde, 1903; Fries and Pike, 1949), today the term is almost exclusively used to refer to *variant prestige*, a development that can be traced back to the influence of the idea of the 'linguistic marketplace' (Bourdieu, 1977) that has taken a strong hold in sociolinguistic thinking (Le Page and Tabouret-Keller, 1985; Cedergren, 1987; Tagliamonte, 2015).

Unlike the intuitive reading of 'prestige' in the sense of an explicit, positive social valuation of a linguistic variant, the term has gathered additional meanings. The lack of specificity of the concept and potential to be a catch-all term becomes most evident in the case of "covert prestige", taken to be a shared social force that makes (some) individuals inclined to adopt a linguistic linguistic despite the absence of explicit positive evaluation (or even overt rejection) of it (Labov, 1966a, p.108, see also Labov 1972). While the gist of prestige is still that linguistic changes are adopted (and in this sense *caused* or *triggered*) based on underlying changes in the social structure or beliefs of a speech community, Labov (2001) acknowledges that

> the force of Tarde's explanation [of reducing linguistic changes to underlying social pressures] may be considerably weakened if the term "prestige" is allowed to apply to any property of a linguistic trait that would lead people to imitate it. Thus the fact that a linguistic form has prestige would be shown by the fact that it was adopted by others. (p.24)

The exact explanatory role of *prestige* in sociolinguistics today is hard to pin down, not just due to the many different strands of research approaches present in the field (Tagliamonte, 2015). At least the mainstream approach laid out by Labov (2001), however, is clearly dedicated to some sort of sociolinguistic *reductionism* of linguistic choices to underlying social institutions. Labov explicitly embraces the standpoint of Meillet (1926):

> From the fact that language is a social institution, it follows that linguistics is a social science, and the only variable element that we can resort to in accounting for linguistic change is social change, of which linguistic variations are only consequences, sometimes immediate and direct, more often mediated and indirect. …We must determine which social structure corresponds to a given linguistic structure, and how in general changes in social structure are translated into changes in linguistic structure. (translation from Labov 2001, p.22-23)

The fact that this position is not taken by all sociolinguists is reflected in debates regarding whether social valuation exists prior to language use at all, an issue prevalent in the study of language and identity from which the sociolinguistic account of language change draws. Bucholtz and Hall for example note that "much work within variationist sociolinguistics assumes not only that language use is distinctive at some level but that such practices are reflective, not constitutive, of social identities", whereas in much of the linguistic anthropology work "identity is better understood as an outcome of language use rather than as an analytic prime" (Bucholtz and Hall, 2004, p.376).

While sociolinguistic research has equally contributed to the identification of general, universal pressures that seem to be guiding language changes (see Labov, 1994, for an extensive summary in relation to sound changes in particular), most contemporary population-level fieldwork on language variation and change is trying to come to terms with the idiosyncracy of individual changes. The fact that changes sometimes spread from 'above' and sometimes from 'below' the level of conscious awareness (ibid.), from females to males or the other way around (Trudgill, 1972; Milroy and Milroy, 1985; Eckert, 1989; Labov, 2001; Sundgren, 2001) challenges the idea that a purely mechanistic theory of language change is possible. But no matter along which social dimension innovations spread first, a change is only a change if it is increasingly adopted by a speech community across the board. While research has shown that language change (or at least the adjustment of variable usage rates) within an individual's life span is indeed possible (Sankoff and Blondeau, 2007; Buchstaller, 2015), the canonical case of change is still regarded to be the *incrementation* of changes by younger speakers relative to their parental generation (Labov, 2001; Tagliamonte and D'Arcy, 2009). The fact that this incrementation occurs along the same social lines from generation to generation has raised a particular interest in how the *structured heterogeneity* of sociolinguistic variation is acquired by children and adolescents.

Incrementation and age vectors

Most traditional accounts of language change are based on the assumption that linguistic divergence occurs during language acquisition, mostly based on language-internal factors that make learners 'mislearn' or 'reanalyse' their linguistic input, which causes them to end up with a different target language than that spoken by their caretakers (see e.g. Salmons and Honeybone, 2013). But quantitative research on infant and adolescent speech has painted a much more refined picture of the *target* of child language acquisition (Labov, 1989, 2012). Of particular relevance is the question of how individuals acquire sociolinguistic variation, and how this acquisition develops over time. Quantitative studies of the linguistic patterns of different pre-adolescent age cohorts has shown that, while children's usage patterns might mirror the language use of their caretakers up until about age three or four, learners then exhibit a pronounced "outward-orientation": shedding most of the influence of their caretaker speech, learners instead turn not just towards their peers, but towards the usage patterns in the wider speech community as a whole (Labov, 2014).

An exemplary case of this behaviour is the data collected from children in the new town of Milton Keynes, England, which provided a natural testbed for the study of the acquisition of a local dialect against the backdrop of massive individual variation: the settlement expanded massively in the 1970s and 1980s, with most residents moving in from other dialect regions (Kerswill and Williams, 1994; Williams and Kerswill, 1999). Figure 2.1 shows the distribution of variable realisations of the GOAT lexical set (the vowel in English 'goat', 'boat', 'fold' etc., see Wells, 1982) by children of different ages growing up in Milton Keynes, as well as the usage rates of their caretakers. What is striking about this data is not just the fact that children appear to switch from imitating their caretakers to imitating the wider community usage at some point after age 4, but also the accuracy with which children manage to replicate the usage distributions of the relevant target group.

This same pattern of acquisition is found in a slightly refined fashion in Sankoff and Laberge (1973)'s study of the acquisition of the future tense marker 'bai' in Tok Pisin, which underwent a change from being stressed to unstressed as it was grammaticalised during the development of the creole. Figure 2.2 shows rates of secondary stress on 'bai' by children, alongside the stress rates exhibited by their respective caretaker(s), connected by lines. What is evident is not just that children are producing fewer stressed tokens, but that their stress rates are lowered at similar rates *relative to the stress rates of their caretakers*. This has led researchers to propose that adolescent learners do not just acquire the variable elements of language use according to social and stylistic constraints to a high degree of precision, but that they also advance changes in variable usage along their respective "vector of change":

> In the incrementation of change, children learn to talk differently from their parents and in the same direction in each successive generation. This can happen only if children align the variants heard in the community with the vector of age: that is, they grasp the relationship: the younger the speaker, the more advanced the change. (Labov, 2001, p.344)

While this quote does not speak to the initial triggering or actuation of a change, it does
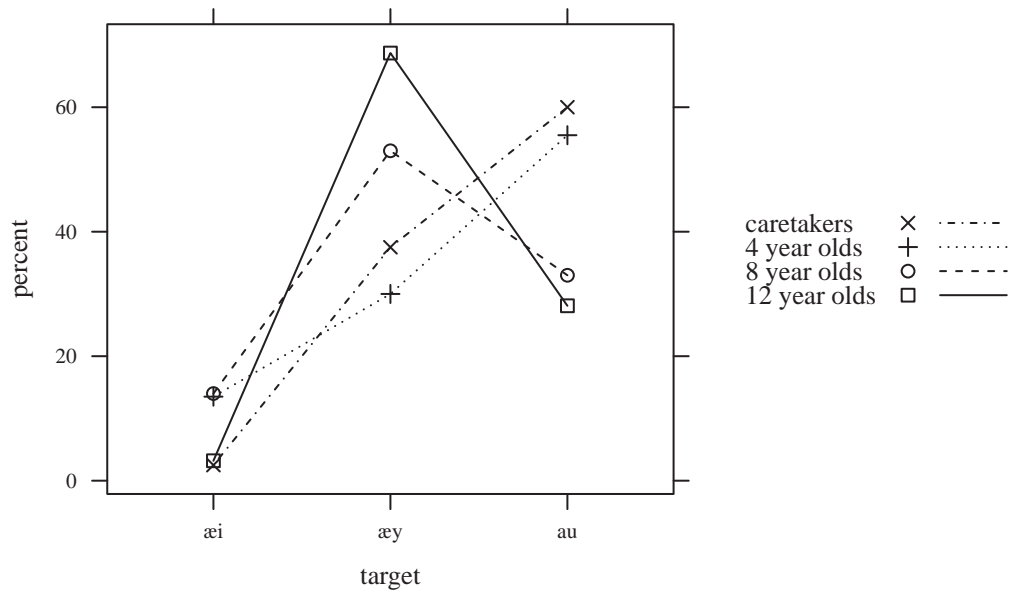
Figure 2.1: Phonetic targets of the `GOAT` vowel for Milton Keynes children by age exhibit increasing alignment to community rather than caretaker speech (data from Kerswill and Williams 1994).

suggest that being able to detect the direction (and rate) of a change could be fundamental to a mechanism which allows speakers to *advance* language changes systematically across generations. Even though the concept of age vectors has been taken up as an explanatory device to analyse and account for trends in macro-level data (Labov, 2012; Sankoff, 2013; Stanford et al., 2014; Driscoll and Lape, 2014), direct testing of the underlying assumption,i.e. that "youth who are engaged in the incrementation of a sound change have some perception of the age vector" (Labov, 2010, p.369), has been more limited. The most compelling evidence comes from experimental studies of listener adaptation based on sociophonetic knowledge about ongoing changes: Drager (2005) and Hay et al. (2006) showed experimentally that listeners use their implicit knowledge about age-specific speech patterns to adjust phoneme boundaries when classifying vowels depending on the purported age of the speaker, which was manipulated experimentally.

Even though the concept of age vectors was originally conceived of as applying to both continuous and categorical changes (Labov, 2001, p.346), in contemporary research it is now mainly applied to (continuous) phonetic changes. Here, the 'age vector' can quite literally be taken to be a vector in phonetic space. This interpretation can be intuitively derived from traditional representations of phonetic changes in progress, where arrows are drawn in acoustic (particular vowel) space to indicate the difference in pronunciations between older and younger speakers of a community (see e.g. Labov, 2001).

This specialisation of the notion of age vectors is merely a consequence of the fact that most sociolinguistic research is based on *sound changes*, particularly ones of a gradual (rather than categorical) nature, such as the canonical example of vowel shifts. While there is consequently also a rich research tradition on individuals' sociophonetic knowledge (Foulkes and Docherty,
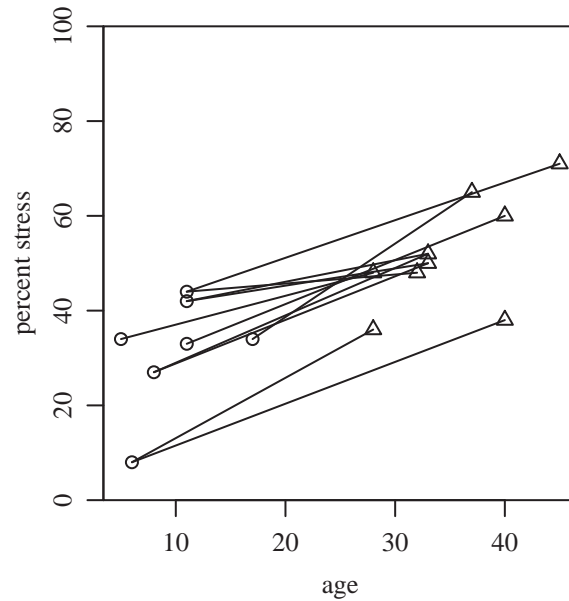
Figure 2.2: Rate of secondary stress on the 'bai' future tense marker in Tok Pisin. The relative difference between the rates of children (circles) and their respective caretakers (triangles) exhibit a similar pattern of increase across pairs (data from Labov 2001, p.425).

2006), work on sociolinguistic awareness (or indeed sociolinguistic indexicality) in the domain of necessarily categorical *syntactic* change is generally underrepresented, a fact that I will address in Chapter 5.

While a more detailed analysis of this very idea of the *amplification of linguistic trends* forms the central part of this thesis, the overview of competing 'explanations' of change presented thus far has painted a rather incoherent and divergent picture of research on language change. Although the diversity of approaches and results makes it difficult to provide anything like a coherent or comprehensive overview of current opinions, I have already suggested the existence of something like a fault line dividing the field, characterised by two opposing views of language change based on distinct research methodologies. While mechanistic and adaptively-minded accounts draw evidence both from macro-level typological and individual-level experimental data, sociolinguistic work based on social population-level phenomena stress the arbitrary, haphazard selection and diffusion of changes.

Rather than simply pick one side and more or less implicitly disregard the other, this thesis is in pursuit of another higher-level goal, namely to unify these seemingly contradictory approaches and show that they are in fact compatible. Before pointing to the evolutionary framework in which the apparent clashes can be reconciled, I will make a final effort to convince even the staunchest optimist that there truly is a schism in the field that warrants talk about the need of unification and reconciliation, by discussing a recurring issue on the question of language change where the differences are as stark as nowhere else: the adaptive nature (or not) of language change.

2.3 Language change and language adaptation

In parallel to the specific theories and accounts of language change presented so far one finds discussion of a more conceptual concern, namely whether language change constitutes language *improvement* of some kind. This question of whether language change is in some sense *adaptive*, whether universally (by constant adjustment to general pressures on communication systems or particular constraints of the human language capacity) or with respect to its current contingencies of usage (such as social or environmental factors), is far from settled, or rather it is assumed to be settled in different ways within different subfields studying language change. Interestingly, the fault line does not coincide with the traditional boundaries of linguistic disciplines, such as phonetics, phonology, syntax, etc. Rather, as I have argued, it is largely a question of research methodology as well as the *scope* of generalisation aimed for by different language change researchers.

The opinions held by the two camps – those who hope to gain insight from framing language change as *optimisation* towards some goal vs. those sceptical of or categorically opposed to such a viewpoint – go back far (see Jespersen, 1922, ch.XIV). In light of the necessity to argue for some kind of *asymmetry* to explain the shift from one linguistic variant to another, it is comparatively easier to find explicit arguments for change as improvement or adaptation of the lingustic system, such as Zipf (1949)'s *principle of least effort* discussed above. The orthogonal view that change really is *just change*, with no particular goal or direction, was in a sense the default view of many early Indo-Europeanists who, if anything, saw the beauty and perfection of ancient languages tainted and eroded by unsystematic, haphazard changes (Jespersen, 1922, ch.II). As such, anti- (or at least non-)adaptationist arguments per se are found most explicitly in direct refutations or criticisms of adaptationist claims:

> Taking linguistic change as a whole, there seems to be no discernible movement towards greater efficiency such as might be expected if in fact there were a continuous struggle in which superior linguistic innovations won out as a general rule. (Greenberg, 1959, p.69)

This point of view was taken even more strongly by researchers not working on cross-linguistic typology, such as in Postal's well-known, somewhat more colloquial statement on syntactic change:

> There is no more reason for languages to change than there is for automobiles to add fins one year and remove them the next, for jackets to have three buttons one year and two the next, etc. (Postal, 1968, p.283)

Despite very different methodological approaches and research agendas at the time, similar thinking prevailed at the other end of linguistic inquiry into phonetics:

> I do not think sound change creates any significant improvement or defect in language. There is sufficient redundancy in language that the message which speech encodes gets through as well (or no worse) before and after sound change. ...All human languages manifest asymmetry or disequilibrium in some part of their phonology but seem, nevertheless, to function adequately for communication. ...There is

> no need to appeal to ill-defined notions such as "system pressure", "pattern symme-
> try", "equilibrium", and the like, nor to maintain that the language is any better or
> any more "fit" as a result of these sounds changes. ...It seems languages are never
> satisfied! More to the point, it seems that languages are not seeking the satisfaction
> of some "ideal" configuration. (Ohala, 1989, p.191-192)

Particularly within the study of sound change, this point of view is now widely accepted
and essentially become textbook knowledge. For example, the entire introductory chapter of
Labov (2001) is dedicated to what he calls the *Darwinian paradox*:

> There is general agreement among 20th-century linguists that language does not
> show an evolutionary pattern in the sense of progressive adaptation to communica-
> tive needs. ...The almost universal view of linguists is the reverse: that the major
> agent of linguistic change – sound change – is actually maladaptive, in that it leads
> to the loss of the information that the original forms were designed to carry. (p.10)

The same view is also adopted by Croft (2006), who concludes that "all of the empirical
evidence in language change indicates that social factors, not functional ones, are the causal
mechanisms for the propagation of a change" (p.116).

The list of supporters of the viewpoint that language change is at least to some extent
adaptive is just as long and illustrious, with dedicated arguments for the cause by Jespersen
(1949), Vennemann (1993) and Haspelmath (1999, 2008). Especially within typology one finds
hints of a conclusion on the question of language adaptation counter to that arrived at in
sociolingustics, for example when Wichmann (2015) finds that "Presumably language change
is only to some extent random, and to a larger extent is adaptive" (p.221).

A striking difference of focus in these last two quotes can be found in the 'causal mechanism'
for the propagation of particular changes, which is evidenced at the micro-level of the diffusion
of changes, versus the 'adaptive' nature of language changes that is observed as a synchronic
fact. What becomes evident here is that the idea of an 'explanation' of a change depends on
the level at which it is investigated:

> There is general agreement that the heart of the study of language change is the
> search for causes. It is what we generally mean by the explanation of change. While
> we would like to apply to this search the universal principles that govern grammar as
> a whole, it is also understood, following Meillet (1921), that no universal principles
> can account for the sporadic course of change, in which particular changes begin
> and end at a given time in history. The actuation problem demands that we search
> for universals in particulars. (Labov, 2010, p.90)

This quote succinctly summarises two different approaches to the question of causation,
or 'why' languages change. Firstly, Labov acknowledges that there are in fact two senses
of 'why', implying separate research interests with diverging priorities regarding the locus of
'explanation'. On one hand, for researchers seeking to generalise over language changes, the
'why' question is not actually asked of individual changes in particular, but is first posed as
a general cross-linguistic 'what' or 'which', in order to identify universal guiding principles of
change, such as the unidirectional patterns that have been shown to hold across languages. The
simple 'why' then becomes a 'why *that*, over and over again?' which, once a universal pressure

has been identified, offers an explanation that is taken to generalise to individual instances of the change, without the need to refer to particular features of those instances. The second sense of 'why' is the one most pressing for the researcher looking at particular changes, and is best rephrased as 'why there and then'? This question has to tackle the actuation problem head on, and can only be answered with particular reference to the specific conditions for the change, so as to explain why it actuated exactly when it did.

While diverging in their focus of explanation (to the degree of incompatibility), the two approaches share a common scientific standard: both seek to explain changes in a deterministic fashion, in the sense of reducing changes to accountable, rule-like development, merely diverging on whether the underlying mechanism of change is based on micro- or macro-level descriptions of the change. What I want to propose here is that the two approaches can in fact be unified, but only by abandoning that very standard of explanation. Counter to Labov (2010, ch.5), who argues for the existence of language-internal 'triggering events' that can account for the actuation of particular changes, I want to suggest that one has to acknowledge the fact that individual changes occur at points in time that are, at least to the level of description available to the linguist, arbitrary and thus unpredictable.

Although this point of view would presumably resonate with many a researcher who is content with revealing generalisable patterns of linguistic changes, my goal is not to discard or even play down the actuation problem. What I suggest is to shift the focus away from the negative, seemingly irrefutable criticism of universal pressures to fail to account for the idiosyncratic nature of particular changes and their seemingly arbitrary onset. Instead, the criticism can be transformed into a challenge of constructively explaining the very feature of language change that is the underlying source of the actuation problem, namely its sporadic nature. A complete theory of language change should not just offer an explanation of the macro-level dynamics and general direction of language change, but also provide a concise account of exactly *why* the actuation of particular changes is unpredictable.

As I showed above, much headway has been made on the question of how and in what way languages change in general, generalisations which are based on universal pressures and asymmetries. If one accepts those pressures which account for the strict uni-directionality of many changes and that seem to dictate the prevalence of linguistic traits across the globe, how can there still be room for systematic uncertainty and indeterminacy in language change? Even though adaptive accounts of language change today are well-motivated and grounded not just in typological observations but also in theories of individual processing (Kirby, 1999; Jaeger and Tily, 2010), neither of these two levels are sufficient to conclude that specific changes spread *because* they are adaptive. So when Wichmann (2015) writes that language change is to a large extent adaptive, this statement can be re-read somewhat less intuitively as 'most changes that spread through populations happen to be adaptive (but that is not necessarily the causal reason why they *do* in fact spread)'.

The crucial point is that the mere fact of adaptation, or even evidence for the preferred spread of adaptive traits, should not be equated with evidence for *selection for the adaptiveness* of that trait or structure (Henrich et al., 2008). The intention here is of course not to discard functional and adaptive pressures completely, but to raise the question of whether they should

be construed as the pressures which drive the diffusion of linguistic variants through a speech community in the individual cases. Instead, I want to stress how functional factors actually play an important role *within* social accounts of language change, and that by taking an evolutionary approach that distinguishes separate pressures behind the *innovation* of linguistic variants and their subsequent *selection*, the two viewpoints can in fact be unified (Croft, 2000, 2006).

Before elaborating on the importance of an evolutionary approach in Section 2.5, I will have to discuss quantitative approaches to language change more generally, as well as the insights into the dynamics of changes which are relatively well understood thanks to centuries of quantitative research on historically attested changes.

2.4   Language change: a quantitative framework

So far, the language changes referred to as brief examples for different accounts above were primarily characterised through qualitative descriptions. Given the vast range of different types of changes – lexical, morphosyntactic, phonological, phonetic, semantic – it might even appear ludicrous to speak of 'language change' as one phenomenon that should be subsumed by a single framework. However, quantitative approaches to many different types of historical language change have made use of similar descriptions to characterise the unfolding of changes. While historical linguists were the first to be concerned with the *variation* exhibited by languages as they undergo change from one state to another, the sociolinguistic study of synchronic variation within the individual set the idea on a more solid footing by introducing the concept of the *sociolinguistic variable.*

2.4.1   The sociolinguistic variable

Forming part and parcel of the sociolinguistic or general *variationist* research framework today, "A linguistic variable in its most basic definition is two or more ways of saying the same thing" (Tagliamonte, 2012, p.4). In more formal terms, different *variants* of one of the same *variable* have the property that they are "identical in truth value, but opposed in their social and/or stylistic significance" (Labov, 1972, p.271). The identification of a sociolinguistic variable is always based on one particular language or vernacular, where you can often find one and the same speaker to be using and mixing the variable's different realisations, referred to as *variants*, in some probabilistic fashion. A canonical example in English is the variable `(ing)`, which refers to the variation in how the final nasal in the present participle suffix *-ing* is pronounced, with the two variants [n] and [ŋ].

While this particular example concerns a phonetic variable which has seen stable variation in its usage for centuries (Labov, 1989), sociolinguistic variables can equally be constructed for variable usage in other linguistic domains, and with a particular eye on capturing the unfolding of language *change.* In Chapter 5 we will be concerned with syntactic changes that have been unfolding in Shetland Scots over the past century. One of the variables, referred to as `(ynq)`, describes the syntactic realisation of yes/no questions. The variable can be realised by either of two variants, one being the use of the older, verb-initial sentence order (e.g. "See you him?"),

the other a relatively newer, Standard English-like formulation employing a sentence-initial periphrastic 'do' (i.e. "Do you see him?"), with both formulations being otherwise identical in their interrogative linguistic function. By tracking the usage rates of different variants of a variable over time, for example by counting their relative occurrence in recordings or corpora at different points, one can capture the temporal dynamics of a change as the novel, *incoming* variant (or variants) gradually replaces the *outgoing* form.

Although language change can be studied on levels other than that of the sociolinguistic variable (see in particular Croft, 2006, p.98), it is chiefly changes at this lowest level, namely in the realisation of individual sociolinguistic variables, that will be covered in this thesis. And, since I have set out to delineate a theory of language change with a particular eye on reproducing the micro-level dynamics of the diffusion of novel linguistic variants, we should have a closer look at what is known (or believed) about how individual language changes spread within communities.

2.4.2 S-shaped curves in language change

When it comes to the question of how language changes unfold quantitatively, it is common wisdom in the field of historical linguistics (and beyond) that the incoming variant gradually replaces the outgoing one(s) along an *s-shaped* trajectory. It should be noted at this point that the notion of s-shaped curves derives from the study of changes in *categorical* variants, where the s-shape describes the development of the *frequency* of the incoming variants. This should not be confused with the study of how *continuous* language changes, i.e. gradual changes to *phonetic* realisations, unfold. Simply as a matter of technical feasibility, quantitative investigations of continuous changes are necessarily much more recent than that of categorical variables which are largely measurable without technological aids. As a consequence, the work in this thesis will be limited to studying changes in categorical variables. (For an empirical investigation of the complex dynamics that can be exhibited by changes in a continuous dimension see e.g. Fruehwald 2013.)

Even for categorical variants, however, capturing the time course of a change is no simple matter, since any claim about precise dynamics is inherently limited by the number of time points throughout the change for which distinct samples can be taken, as well as the respective size of those samples. The earliest descriptions of language changes were typically not based on historical data of the same linguistic variety over time but rather on a *comparative* approach based on analysing the relationships between relatively idealised and most of all non-variable grammars of related languages. This method lay at the heart of the first identification of sound changes through the positing of sound correspondences between languages (Jespersen, 1922, ch.II), but the simple rule-like description of the changes (e.g. by stating the phonetic realisation of a phoneme before and after a change) left open how exactly a community switched from unanimous adoption of one sound or construction to another. As quantitative studies with an increasing resolution of the time course became more common, the idea of the adoption of changes along a relatively directed 's-shaped curve' became firmly established in the field:

The process of change in the community would most probably be represented by an

> S-curve. The rate of change would probably be slow at first, appearing in the speech
> of innovators, or more likely young children; become relatively rapid as these young
> people become the agents of differential reinforcement; and taper off as fewer and
> fewer older and more marginal individuals remain to continue the old forms. On an
> empirical level, it should be possible to make a comparative study of forms used as
> a function of age and other sociological variables. (Greenberg et al., 1954, p.155)

Similar sentiments, again mostly backed up by verbal rather than precise quantitative arguments, are repeated in Lindgren (1953, pp.181+186) as well as later in Weinreich et al. (1968) and Bailey (1973). Today, it is often assumed that the increased usage levels of incoming variants from non-existence to complete adoption follow the trajectory of the s-shaped *logistic* growth curve (Denison, 2003). The origin of the logistic curve as the idealised, underlying can be traced back to Altmann et al. (1983), from where it was perpetuated primarily by Anthony Kroch (Kroch et al., 1982; Kroch, 1989a,b). It should be noted that the amount of empirical data that motivated Altmann et al.'s use of the logistic function would probably not stand the test of the quantitative standards expected of empirical research in the present day. The authors' main evidence is a data set capturing the decrease in usage of Russian genitive markers on unit of measurement terms between 1881 and 1910. Taking the 28 data points measuring the relative rates of dropping the genitive marker from Piotrovskaja and Piotrovskij (1974), who themselves proposed to capture the shape of the curve by use of some trigonometric functions, Altmann et al. show good fit of the trajectory by the logistic function which they argue is also more justified, being an explicit model of growth known from other domains, particularly ecology (Kingsland, 1982), and not just an arbitrarily chosen function that happens to fit the empirical data (Altmann et al., 1983, p.106).

While their particular proposition achieved their goal of setting the quantitative study of language change on a more firm theoretical, mathematical basis, the empirical side of their argument remains relatively understudied to this day. Figure 2.3 shows the two data sets from the paper which motivated the use of the logistic function. Firstly it should be noted that no absolute token counts are provided for either trajectory, neither could any be obtained from the first publication of the Russian genitive data in Graudina (1964). This means that is unclear whether some of the strong downward outliers along the trajectory are merely due to sampling effects or whether they require additional explanation. The second data set capturing the proportion of Arabic loanwords in Persian (data whose origin is unknown and uncredited) which, if anything, only encompasses the very early stages of 'incipient' growth, is presumably not expected to ever reach 'completion', in the sense of Persian vocabulary being 100% replaced by Arabic terms. Nevertheless, both data sets receive a superposed fit of a logistic growth curve that is set to saturate only upon complete diffusion, i.e. ultimate convergence to a proportion of 1.0 of the incoming variant, by Altmann et al. (1983, p.110+113).

The quantitative study and comparison of growth *rates* of linguistic changes can be traced back to Kroch (1989b) who set out to investigate whether the rise of a novel linguistic variant, in particular the rise of periphrastic 'do' in Early Modern English (Kroch, 1989a), differed across related but different grammatical contexts. Kroch determined the relative rate of diffusion of the novel variant for different subsets of the data by performing a logistic regression
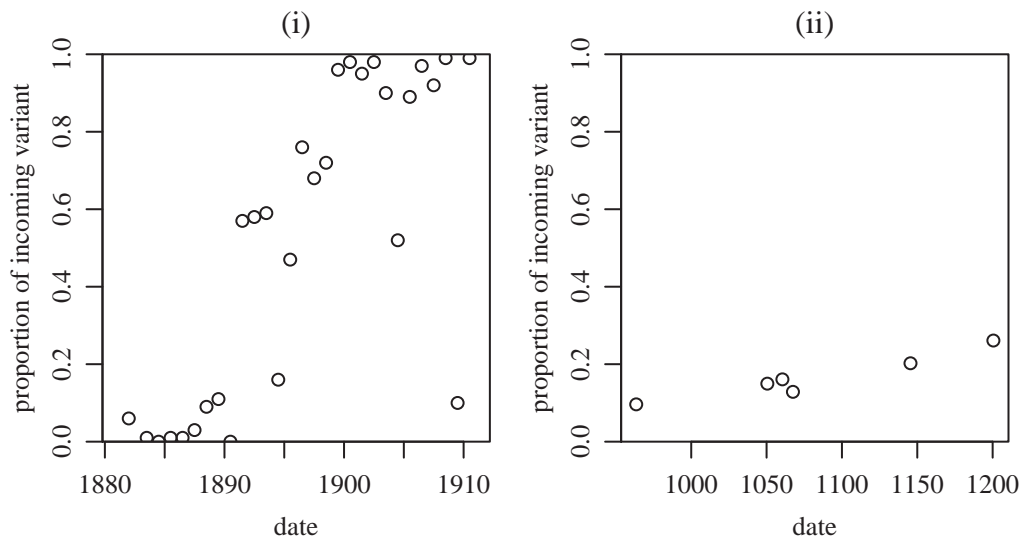
Figure 2.3: The two data sets which provided Altmann et al. (1983) the empirical basis for interpreting the s-shaped curves found in language change as logistic growth. *(i)* increased dropping of genitive markers from Russian measurement units, data originally due to Graudina (1964) *(ii)* proportion of Arabic loanwords in Persian, unknown source.

which provides an estimate of the *slope* of the growth curve as a function of time (Kroch, 1989b, p.215). Comparing the slopes obtained for different grammatical contexts documented in Ellegård (1953)'s extensive data set on the prevalence of 'do', he found no statistical difference between the growth rates of the novel variant for the different contexts. The contexts did however differ in their relative timing of actuation, which he took to imply that contextual effects on changes, including external, stylistic ones, should be constant across time and not interact with the time course of the change, which is primarily dictated by language-internal dynamics (Kroch, 1989b, p.206)[2].

This original formulation and testing of the 'constant rate hypothesis' is still among the most exhaustive efforts at measuring and comparing different rates of changes quantitatively, and the idea that the rollout of categorical changes across different linguistic contexts follows a consistent rule-like pattern has also found tentative confirmation in the study of phonological change (Fruehwald et al., 2009). Only very recently, with the advent of larger corpora of parsed written texts, has the question of determining the rates of individual language change trajectories received new attention, with novel and more sophisticated quantitative methods yielding contradictory results concerning both old established data sets like Kroch's (Vulanović and Baayen, 2007; Vulanović, 2007; Ahern et al., 2016) as well as newly investigated changes (e.g. Lieberman et al., 2007; Cuskley et al., 2014; Ahern et al., 2016).

But the fact that historical linguistics research is, at least on a *quantitative* level, largely not cumulative (Nevalainen et al., 2014, 2016) means that the empirical basis of the logistic growth assumption has rarely been challenged. The lack of systematic and detailed study

---

[2]It should be noted that this conclusion relies on the fact that context effects are measured by the coefficients of a logistic regression which capture a *relative* rather than absolute influence on the usage rates, an assumption that is to my knowledge not motivated beyond the fact that it can easily be captured by logistic regression.
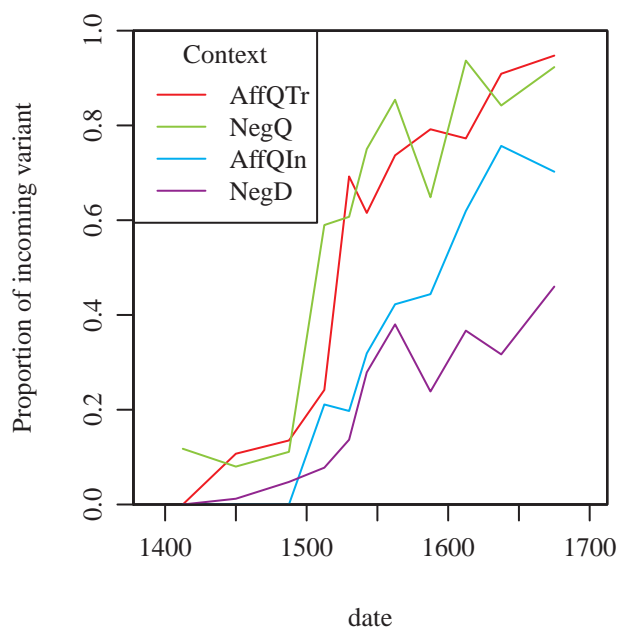
Figure 2.4: Ellegård (1953)'s data set on the rise of periphrastic 'do' which formed the basis of the *Constant Rate Hypothesis* (Kroch, 1989b). Data is shown for contexts in which 'do' eventually became mandatory, in particular affirmative questions whether transitive (`AffQTr`) or intransitive (`AffQIn`) as well as negative questions (`NegQ`) and negative declarative sentences (`NegD`).

of its quantitative basis means that there are many very basic open questions regarding the nature and origin of s-shaped transitions, for example whether the particular shape of changes in usage levels can also be found on the level of the individual, or whether they are actually only an artefact of linguistic analyses which are typically based on pooled population-level data (Denison, 2003). Equally, thinking about ways to determine growth rates that are adequate for the data, e.g. by accounting for the fact that data points sampled from an underlying quantity that is undergoing incremental growth are not in fact independent as is assumed by logistic regression, is relatively underdeveloped, and concerted efforts to go beyond the traditional statistical methods of the field are fairly recent (Kauhanen and Walkden, 2015; Winter and Wieling, 2016).

The fundamental bottleneck in the study of the micro-level diffusion of historical changes, however, appears to be the difficulty of collecting adequate amounts of longitudinal data, with Blythe and Croft (2012) identifying only 18 data sets of language changes that covered complete s-shaped curves in a large literature survey. Consequently there has, on the whole, not been a lot of interest in (or rather potential for) the *cross-linguistic* comparison between individual trajectories of unrelated changes, a notable exception being Ghanbarnejad et al. (2014)'s study of exponential vs. logistic growth features in the adoption curves for a number of orthographic conventions obtained from written texts. Cross-linguistic generalisations, whose results were covered extensively above, are typically limited to determining *what types of changes* occur, but they do not consider the micro-level of how those particular changes unfold over time. I will

return to this issue in Section 4.4.1 to argue that investigating the unfolding of similar changes in different languages can provide useful insights into the different accounts of change that were discussed.

To summarise, despite the fact that the theoretical motivation for the s-shaped pattern of language changes and the underlying assumption of logistic growth in particular stand on shaky foundations on closer inspection, both have firmly established themselves as cornerstones of research on language change on the micro-level. By providing a generally accepted, quantitative 'gold standard' which different accounts of language change can in principle be tested against, the s-shaped curve has also become a de-facto criterion to argue for the adequacy of models of language change on a quantitative level (e.g. Kirby, 1999; Yang, 2002; Blythe and Croft, 2012; Kauhanen, 2017). Given that historical language changes are one off events which do not typically allow for repetition in controlled laboratory conditions on the same scale, computational modelling has emerged as a tool to make explicit the quantitative predictions made by different accounts. Since this thesis makes extensive use of modelling, Chapter 3 will not just present the two modelling frameworks used in great detail, but also discuss the role and relevance of mathematical and computational modelling for thinking about language change more generally. Before dedicating ourselves to the study of specific quantitative models, however, I will round off this general review of what is known about language change by elaborating on the general formal, evolutionary framework which will be pursued for the remainder of this thesis.

2.5 Language change as language evolution

It has often been noted that Darwin's idea of the gradual evolution and diversification of biological species was inspired by philological work of the time which described the historical relationship between different languages (Darwin, 1871). Since then, the field of biological evolution has developed a large theoretical toolkit for capturing and evaluating the replication and unfolding of biological populations in a rigorous, quantitative fashion. These scientific developments have caused researchers in some social and cultural sciences, but especially linguistics, to adopt similar approaches to describe and explain language change (Croft, 2000; Ritt, 2004; McMahon and McMahon, 2005, to list just a few monographs). Crucially, 'evolutionary' approaches to language and language change differ immensely in what aspects of biological evolution they try to import or emulate, as well as how (Croft, 2000). While sometimes biological concepts are merely used as a metaphorical basis to explain linguistic processes that they share some superficial similarity with (e.g. Lass, 1990), others are mainly focussed on applying existing mathematical methodology to a new field (Borgerhoff Mulder, 2001; Atkinson et al., 2005; Jäger, 2008; Castellano et al., 2009).

The evolutionary approach to language change adopted in this thesis follows the one outlined by Croft (2000), which is in turn based on Hull's *generalised analysis of selection*. The fundamental principle of this generalsed analysis is that of defining evolution as *change by replication* (Hull, 1988, p.410). As I already alluded to above, language evolution in this sense is characterised not by any 'inherent' change of an entity (such as a 'language'), but by changes to the type and distribution of a population of entities that are replicated individually. Whereas in

biological evolution the *replicator* (typically assumed to be the genome) and the *interactor* (the corresponding phenotype that is central to the replication of the genome) are united in one and the same biological entity, this is just one special case of 'evolution' in Hull's general framework. The generalised analysed allows for the replicators and interactors to be separated out, which is the stance taken by Croft (2000) who characterises language change as a process in which linguistic structures are replicated by human *interactors* whenever they engage in linguistic behaviour.

Croft calls his theory the 'utterance selection model of language change', stressing the fact that language change should not be sought in individual minds or abstract linguistic identities, but instead in the linguistic *utterance* as the individual speech act used in a specific context. The paradigm *replicator* in linguistic evolution for him are linguistic structures that are contained within utterances, which he calls *linguemes*, and which he identifies with the linguistic variant as studied by variationist sociolinguists (Croft, 2006, p.104).

This new concept of the *lingueme* as the fundamental entity of linguistic evolution also entails a specific theory of its *replication.* Parallel to the concepts of *selection* and *mutation* in biological evolution, Croft distinguishes between two distinct evolutionary pressures, namely those of *differential replication* and *altered replication* (Croft, 2000). Differential replication describes changes where the distribution of existing variants changes due to preferential but otherwise 'correct' or near-identical copying of existing variants by a human interactor, a process that I will call *selection.* Altered replication on the other hand results in the spontaneous production or *innovation* of a new variant that is either derived from or otherwise completely independent of other known variants.

This characterisation of language change as consisting of two separate mechanisms that can at least to some degree be described independently is not completely new in linguistics. Jespersen (1922, ch.XV §11) makes an explicit distinction between the question of how sound changes *originate* in individual speech as opposed to how they *spread* to other individuals, which he frames a matter of imitation, i.e. replication of existing variants. A similar sentiment is expressed by Weinreich et al. (1968), who trace replicator-based thinking in language change as far back as Paul (1880). That same micro-level approach is by far not shared by all linguists working on language change, and it should be noted that this is just as true of research done under the *language as a complex adaptive system* moniker, where the term 'evolution' is still often used simply as a synonym for 'adaptation' (Croft, 2000).

To highlight the relevance of this replicator-based evolutionary take in the light of the research results on language change presented so far, it is worthwhile to re-iterate a basic insight from evolutionary biology regarding the relationship between synchronic distributions of traits and the supposed underlying pressures that led to those distributions, namely that "the prevalence of a particular species in a habitat does not necessarily imply that it is any better adapted to that habitat than its competitors" (Blythe, 2012b, p.2). This quote is a word of caution against not one but two simplistic assumptions that are often made when extrapolating from observed adaptations to the supposed selection of those adaptations. The first is the influence of *neutral evolution*, i.e. the effect that the random fluctuations inherent in the replication of a finite number of replicators can have on causing the widespread adoption

of a trait in the absence of any replicative advantage of that trait. But, given the directed nature of the adoption of language changes suggested by their s-shaped trajectories, neutral theories have played only a very minor role in linguistics, a point that I will return to below. Instead, I want to highlight a second, less obvious concern regarding different possible sources of *asymmetry* that might underlie the skewed synchronic distribution of linguistic traits.

From the evidence discussed above we can conclude that language change looks very *directed*, but it does so on two distinct levels: firstly, we find unidirectional patterns in the way languages change, suggesting some consistent asymmetry that, when described at the cross-linguistic level, makes changes go in similar *directions* over and over again. At the macro-level we are therefore concerned with a temporally reduced characterisation of a change that encompasses both the emergence as well as ultimate adoption of an *innovation*. Crucially, this sense of direction should not necessarily be equated with the *directedness* of the individual trajectories of change which exhibit the aforementioned s-shaped pattern as the individual competing variants undergo *differential replication*.

This distinction might seem far-fetched at first, but it is a central argument of this thesis that the macro-level sense of direction that one gets from the re-occurrence of similar changes towards similar adaptive goals and the directedness of the micro-level diffusion of individual variants that forms the building block of language change are due to fundamentally different pressures. As already alluded to above, I will argue that the influence of adaptive and functional factors should be primarily thought of as affecting the domain of *innovation*, where it is responsible for asymmetries in the generation of new variants. The *selection* pressures which are the driving force behind the directed propagation of individual changes, on the other hand, will be identified with arbitrary social biases which do not in principle distinguish between linguistic novelties that are adaptive or maladaptive. Simply due to the fact that selection pressures can only apply to variants which are already part of the pool of synchronic variation (Ohala, 1989), the reliable re-occurrence of similar changes can be explained due to their relatively higher likelihood of being innovated (Joseph, 2013). In the words of Croft (2006),

> How is it possible that one can observe language changes that are propagated by a social mechanism and yet they consistently display unidirectional patterns in cross-linguistic comparison that are presumably motivated by functional/phonetic factors in innovation? All that is necessary is that the mechanisms for innovation and propagation be independent. (p.116)

Importantly, this approach does not completely rule out changes that go in the opposite direction of any pattern of uni-directionality that might have been identified, but merely predict that those changes would be much less frequent. Attributing the selection of changes to arbitrary social pressures simplifies the matter of accounting for such particular, unusual changes that run counter to universal trends, which would otherwise not just require explanation of why a universal pressure did not hold in a particular instance, but also the origin of an oddball selection pressure in the opposite direction.

This separation of concerns in a strict evolutionary account solves the problem of accounting for both universal trends and particular, idiosyncratic changes, and thus provides a natural solution for the scientific conundrum encountered not just in linguistics but also in other social

historical sciences (Blute, 1997). Even though the explanatory power of this approach has been promoted heavily in particular by Croft (2000, 2006, 2008), the systematic distinction between innovation and selection pressures has not widely caught on in the field. The clear evidence for functional asymmetries as presented above in combination with a lack of candidate mechanisms that could explain the arbitrary and sporadic actuation of particular changes has meant that this second selection step of language change has either been left unaccounted for, or otherwise assumed to be driven by the very same language-internal asymmetries (Croft, 2006, p.111). The main goal of this thesis is to give additional support to the two-step model of language change by filling in the current gap regarding the second, arbitrary selection step with a symmetric selection mechanism that nevertheless produces strongly *directed* transitions whose actuation, however, is sporadic and temporally underspecified. While thinking about such mechanisms is rather unusual for the domain of biological evolution, evolutionary approaches to cultural change more generally have given rise to a wider range of selection mechanisms that are worth exploring.

2.5.1   Tackling the actuation problem

Virtually all approaches to explaining linguistic changes discussed so far struggle with one and the same aspect of its dynamics, namely the sudden incrementation or transmission of change to a part of its grammar following long periods of stagnation or stability. The inability to account for this *sporadic* nature of language change is not always acknowledged, and in fact often outsourced to triggers that fall outside the scope of the specific account of change itself, whether implicitly in the case of language-internal (as criticised by the *actuation problem*), or explicitly in the case of social accounts. While the LCAS paradigm recognises the complexity of the many interacting pressures that are thought to underlie the spread of specific changes, the exact nature and dynamic of the transitions from stability to change are hardly ever accounted for or explained explicitly.

2.5.2   Beyond biological metaphors: regulatory pressures

Beyond the attempts to explain language change in relatively simple mechanistic terms akin to the straightforward causal explanation of events that is perceived to prevail in some natural sciences, researchers have also acknowledged the limitations of such models. In particular, it has been argued that approaches based on the identification or positing of static biases or inherent asymmetries fail to account for the dynamics of linguistic systems that are in many respects irregular and non-linear (see e.g. Larsen-Freeman, 1997; Fortescue, 2006). These very same effects have also been the subject of the quantitative study of cultural change more generally, particularly in relation to traits of fashion and other arbitrary conventions whose change cannot be easily attributed to external adaptive pressures (Acerbi et al., 2012). Despite formal accounts of pressures and mechanisms in which *social learning* from other individuals plays a crucial role (see e.g. Boyd and Richerson, 1985), Bikhchandani et al. argue that none of the well-known social mechanisms such as punishment, conformity or explicit negotiation of conventions can explain "why mass behavior is often fragile in the sense that small shocks can

frequently lead to large shifts in behavior" (p.993).

Beyond simple adaptive selection biases that are often associated with the concept of the biological fitness landscape (Kaplan, 2008; Gerlee, 2015), researchers have been looking for mechanisms which can reproduce the rapid non-linear transitions found in many cultural domains, where the sudden adoption of novel traits appears to be driven by social bandwagon effects rather than by the nature of the adopted traits themselves. Rather than relying on external triggers, these fads and fashions can be described as spontaneous cascades of behavioural change that emerge solely from individual interactions (Hirshleifer, 1995; Bikhchandani et al., 1998; Goldstone et al., 2008). A specific set of candidate mechanisms behind such cascades which this thesis will hone in on is that of *regulatory pressures* (Acerbi et al., 2014), i.e. culturally acquired traits which themselves regulate or steer the acquisition or selection of (other) cultural traits. Of particular interest here is the idea that features of the *temporal dynamics* of the generation and adoption of cultural variants can affect their spread, in a self-reinforcing fashion. At first sight this type of explanation – that a variant is going to be adopted more precisely because it is being adopted more – might seem cyclical or tautological, but a mechanism based on the detection and amplification of trends exhibits a quantitative behaviour that matches well onto the dynamics of language change, as I will argue in much more detail in Chapters 4 and 7. Before we turn to computational modelling as a tool to investigate the predictions of such a dynamical model of change, however, we shall have a final look at the origin of the idea that the historical aspect of cultural traits itself can affect their dynamics of adoption, with respect to linguistic change in particular.

2.5.3 The time dimension in linguistic thinking

While evolutionary approaches as described above as well as the general spirit of reusing concepts and methodologies originally devised to study biological evolution to thinking about language change have become widespread (Atkinson and Gray, 2005), only few efforts have been made to move beyond the straightforward models of simple mutation and selection which are typically known to laypersons outside the field of mathematical and evolutionary biology. Dixon (1997) represents an early attempt at explaining sporadic change by transferring the idea of *punctuated equilibria* to language change, but in the absence of a concrete quantitative model applicable to linguists, the idea failed to grow beyond its metaphorical basis.

While current work on the *incrementation* of language changes (see Labov, 2001, ch.14 in particular) by means of amplifying linguistic trends can be traced back to the concept of *age vectors*, the idea that the *history* of linguistic systems beyond just their present state can influence how individuals change the language can also be found in much earlier work. That the temporal stratification of a language constitutes a fundamental aspect of a grammar was for example argued by Fries and Pike (1949):

> It is impossible to give a purely synchronic description of a complex mixed system, at one point of time, which shows the pertinent facts of that system; direction of change is a pertinent characteristic of the system and must also be known if one wishes to have a complete description of the language as it is structurally constituted. (p.42)

While there are isolated attempts at overcoming the strict dichotomy between synchronic and diachronic linguistics on a formal basis (e.g. Bailey, 1970), the "static paradigm" (p.161) underlying generativist approaches to language largely eradicated thinking about language changes that went beyond the transition between language states that are described purely synchronically. Likewise, in evolutionary approaches to language change inspired by work in population biology, one typically talks about the diffusion of the incoming variants which are, beyond their initially low frequency, not explicitly assumed to be in any way marked relative to the established majority one. Particularly in the sociophonetic literature though, one finds evidence of the notion of the diffusion of a *change* (rather than a variant), possibly to be construed as an incoming variant that is defined *relative to* an established one. This becomes particularly apparent in Labov (2001)'s notion of *age vectors*. That this notion survived (or re-emerged) in the study of sound change is maybe not surprising, since the continuous nature of phonetic changes lends itself to thinking about changes as incrementing *relative to* an average pronunciation. In this sense it is also possible for individual language users to produce even 'more incoming' pronunciations by extending the age vector beyond its current level of advancement. This differs from the incrementation of categorical variants, where the only way for an individual to advance a change is by increasing the relative usage frequency of the incoming variant.

Showing that the same principle of incrementation can also be effective for categorical variables is one goal of this thesis, and I will return to this issue both theoretically (in Chapter 4) and empirically (Chapter 5). Having gained a stronger sense of the historical origin and subtle continuity of the idea within linguistic thinking, we are now ready to face forwards and tackle the question of how the many competing accounts and their underlying *biases*, *pressures* and *explanations* for language change can be evaluated quantitatively. Before doing so, however, we must have a closer look at the methodology with which the study of such complex phenomena can be achieved. The next chapter is therefore dedicated to the role that computational modelling can play in elucidating the validity of different accounts beyond verbal argumentation.

Chapter 3

Modelling language change

3.1   Why model?

Within the field of language evolution (here meant to encompass both research on the evolution of the human language capacity, as well as modelling of the type of cultural, linguistic changes described above), computational models have undergone pronounced trends. During the hayday of computational modelling, the field first produced a plethora of qualitative and quantitative models of the evolution of the language faculty (Kirby, 1999; Nowak et al., 2001) as well as the emergence of linguistic conventions, such as the 'Naming Game' (Baronchelli et al., 2008). Around the same time, similar methodologies became popular to study the dynamics of language *change*, i.e. the replacement of already established conventions, both in general (Niyogi and Berwick, 1995, 1997; Arita and Koyama, 1998; Nettle, 1999; Kataoka and Kaneko, 2000; Livingstone, 2000; Ritt, 2004; De Oliveira et al., 2005; Niyogi, 2006; Wedel, 2006; Baxter et al., 2006; Wedel, 2007; Ettlinger, 2007a,b; Fagyal et al., 2010; Blythe and Croft, 2012; Gong et al., 2012; Otero-Espinar et al., 2013; Sóskuthy, 2013; Pierrehumbert et al., 2014; Enke et al., 2016; Kauhanen and Walkden, 2015) as well as for some specific historical changes in particular (Yang, 2002; Choudhury et al., 2006, 2007; Pearl and Weinberg, 2007; Troutman et al., 2008; Baxter et al., 2009; Sonderegger and Niyogi, 2010; Swarup and McCarthy, 2012; Ritt and Baumann, 2012; Kirby, 2013; Kirby and Sonderegger, 2013).

In terms of external contributions to the study of language change, mathematicians and physicists in particular have brought the formal tools from their own domains to bear on questions of interest to linguists (Castellano et al., 2009; Blythe, 2015). Given the complexity of the methods involved, such contributions often fail to have a lasting impact on thinking in the field if they do not form part of a broader linguistically-motivated research programme which makes it accessible to linguists. Given the relatively abrupt rise of this new methodology, it is not surprising that this hype was followed by several meta-scientific and review papers advocating and/or defending the use of computational models (Cangelosi and Parisi, 2002; Wang et al., 2004; de Boer, 2006; Baker, 2008; Jaeger et al., 2009; Hruschka et al., 2009; Vogt and Lieven, 2010; de Boer, 2012; Smith, 2016). Since most parts of this thesis are going to be concerned with computational modelling, it is worth asking: what *is* the point of having a computational model?

The primary advantage of a formal model is that it allows (or rather forces) one to step away from pure arm-chair theorising, which can be difficult to treacherous when applied to complex phenomena such as languages, which are affected by the interplay of many interacting parts or language users. Instead of guessing at the effects of micro-level assumptions on the macro-level dynamics of the system, a computational approach forces the researcher to explicitly lay out their assumptions about the individual, interacting parts in a quantitatively measurable (and ideally also well-motivated) way. From there, computational methods take the lead by determining in an objective way how the transparent assumptions about individual behaviour culminate in (potentially) complex interactional phenomena in the population.

Probably the earliest example of computational work on the emergence of communication systems is the so-called *Naming Game* (Steels, 1995; Steels and McIntyre, 1998; Baronchelli et al., 2006). In its simplest form a population of agents, each starting off with an empty

lexicon, has to come to agree on a unique 'name' (linguistic form) for a referent or speech act. In-depth study of the Naming Game and its dynamics showed not only how a population could come to agree on a shared convention in the absence of any centralised coordination, it also helped shed light on the types of linguistic preferences or mechanisms that individual agents should have to enable de-centralised coordination to unfold seamlessly (Wellens, 2012; Spike et al., 2016). Crucially, given the very simple problem description of the minimal version of the Naming Game which lacks any risk of referential ambiguity, its outcome is predicted (and in some cases even proven De Vylder and Tuyls, 2006; Skyrms, 2010) to be the emergence of a stable communication system. In the absence of any noise or stochasticity, this simple model does not exhibit continuous, ongoing change that is so characteristic of human language.

Going beyond the initial emergence of a symbolic communication system and closer to more realistic cases of language change under noisy transmission conditions, Wedel (2004, 2006) offered a computational investigation of how simple mechanisms of replication can lead to, amongst other things, phonological category formation as well as contrast maintenance through change in the phonetic dimension. Inspired by models of evolutionary pressures taken from biology, the models were again chiefly a study of general mechanisms from which the aforementioned universal dynamics of language organisation emerged. When idiosyncratic factors or triggers of particular changes were concerned, such as in the study of contrast maintenance under the threat of contrast loss in Wedel (2006), the sudden onset of contrast loss in one dimension is again applied externally, falling outside the scope of the theory of general, universal mechanisms that forms the basis of demonstrably adaptive changes whose actuation is reactive to an external trigger.

While computational modelling has also become a standard technique in related empirical fields such cognitive science, the models of individual behaviour that are of interest to psychologists are necessarily of a very different character than the multi-agent models typically employed to study language as a distributed population-level phenomenon. The differences between the fields are not just limited to the *types* of models though, they also extend to the exact goals of modelling and consequently to how models are evaluated. While cognitive models are often assessed on a quantitative basis (Busemeyer and Diederich, 2010), many models of language evolution and change have retained a *proof of concept* like character. This tradition harks back to original work on the evolution of shared communication systems (Steels, 1995; Kirby, 2000) where models are primarily judged based on their exhibiting some *qualitative* feature such as compositionality, rather than by quantitative comparison to other models or to empirical data. This development can be attributed to the often close (and sometimes confusing; Haspelmath, 2016) interlacing of questions regarding the evolution of *Language*, in the sense of the language *capacity*, the emergence of universal features of languages (such as duality of patterning) by cultural evolution, as well as evolutionary approaches to 'mere' language *change*, possibly in combination with the lack of established corpora of historical changes mentioned previously.

The promise that explicitly spelling out the quantitative assumptions of different models would bring clarity to the field is consequently not as straightforward as it might seem. Depending on the precise framing of the same phenomenon, such as the establishment of a shared communication system in the case of the Naming Game, or the emergence of compo-

sitional language from repeated interactions, superficially different mechanisms which actually have very similar effects on a behavioural level can be considered competing explanations for years (Wellens, 2012; Spike et al., 2016). Especially when models are explicitly dedicated to comparing the effect of different parameter settings *within* them, the general dynamics (such as the basic learning rules or other parameters like population turnover employed in virtually every social learning model) are often taken for granted, although it is important to note that much of the dynamics are implicit in these basic assumptions themselves. Without a dedicated effort to replicate existing models and bring them in direct relation to each other, even computational models can run risk of becoming ideological 'blackboxes', counter to their original intention to make underlying assumptions more transparent. The rhetoric with which computational models are presented can further aggravate this situation. Particularly when it is of interest to make computational models more appealing and convincing to the non-modeller, as is the case when the methodology first spreads to a new field, efforts to portray models as a tool for revelation and enlightenment (rather than as obfuscating black magic) run the risk of trivialising either the models themselves, or at least the analyses presented.

The computational models which have stood the test of time are therefore those which are not one-offs, such as many of the early models which employed bespoke ad-hoc learning rules that are often not grounded in the general learning literature, but models which have undergone intense study and analysis from the ground up. For the case of the evolution of novel inventions, the Naming Game is a case in point (Baronchelli et al., 2008). For the case of language *change*, i.e. the continuous replacement of established conventions, probably the most extensive and well-explored model is the Utterance Selection Model (USM) of language change, which forms the basis of most of the modelling work in this thesis.

3.2   The Utterance Selection Model

The version of the Utterance Selection Model (USM) discussed here grew out of Croft's more general formulation of language change as evolutionary competition between utterances. While in its original, theoretical formulation in Croft (2000) it is truly full *utterances* which are undergoing replication, in its mathematical-computational incarnation the USM is best understood as a quantitative model of the competition between different variants of one sociolinguistic variable, as described in Section 2.4.1.

At its core, every agent in the USM is completely characterised by its variable use over variants, specified by the proportions with which each variant is used, all of which together sum to 1. For sake of simplicity we will limit ourselves to the canonical case of two competing variants, where the behaviour of an agent i can be captured by a single variable $0 < x_i < 1$ representing its relative usage level of the incoming variant, with that of the competing variant taken to be $1 - x_i$.

The primary contribution of the computational USM is that it provides a well-defined and rich framework to study the dynamics of these internal usage levels as they are influenced by observing realisations of the same linguistic variable in interactions with other speakers in a population.

3.2.1 Model parameters of the USM

Learning rate λ

Following an interaction, the agents update their internal frequency according to the following USM update rule, which is again applied for both agents (Baxter et al., 2006, p.4):

$$x_i' = \frac{x_i + \lambda \cdot y_i}{1 + \lambda} \ , \tag{3.1}$$

where y_i is the subjective *perceived frequency* of the variable usage rate, whose computation will be discussed below.

Perhaps the most important model parameter is the agents' learning rate λ, which is by default assumed to be the same for all agents. What the USM's update rule in Equation 3.1 does is change an agent's internal frequency x_i by shifting it a small step towards the relative perceived frequency that it observed in its most recent interaction. The higher the learning rate, the larger the step towards this target frequency: at $\lambda = 0$ there is no learning and the agent remains at their initial frequency forever, as $\lambda \to \infty$, the agent approaches a regime in which they instantly adopt exactly those usage frequencies observed in their last interaction. While there are instantiations of the USM in which the learning rate for individual agents is not constant but *decreases* over time to imitate the effect of increasing rigidity of language use with age (Baxter and Croft, 2016), this thesis will be concerned with the simpler case of a constant learning rate that is identical for all agents in the population. Since we are mostly interested in reliable model behaviour that exhibits gradual assimilation rather than abrupt and erratic changes in individual usage levels, like most investigations of the USM we will limit ourselves to low values in the range of $\lambda \leq 0.01$).

It should be acknowledged that the particular form of the learning rule was partly chosen due to its mathematical properties, which make it amenable to analysis using tools from statistical physics (see in particular Baxter et al., 2006). To get a more intuitive understanding of what the update rule does in terms of agents' learning dynamics, it is worth noting that it is equivalent to defining an agent's usage levels as an exponentially weighted moving average (EWMA) over its learning input data series of perceived frequencies $\vec{y}$. EWMAs themselves are a generalisation of Bush-Mosteller learning (Bush and Mosteller, 1955) for non-discrete input data points which, rather than employing a fixed time window to average over, always gives relatively more weight to the most recent data points, with the absolute contributions of individual learning samples decaying over time. Upon receiving a new data point y indicating a certain usage level observed in an interaction, the agent updates their own usage level x according to

$$x' = (1 - \alpha) \cdot x + \alpha \cdot y \ . \tag{3.2}$$

This representation of the learning rule makes it clear that the agent's own usage level is simply a moving overage over the perceived frequencies it observes in interactions, where α controls the relative weight of the newest data point toward that moving average. This

formulation is equivalent to the original USM updating rule in Equation 3.1 given

$$\alpha = \frac{\lambda}{1+\lambda} \tag{3.3}$$

$$\lambda = \frac{\alpha}{1-\alpha} \ , \tag{3.4}$$

the only difference being a rescaling of the parameter space from $\lambda \in [0, \infty)$ to $\alpha \in [0, 1]$, as shown in Figure 3.1.
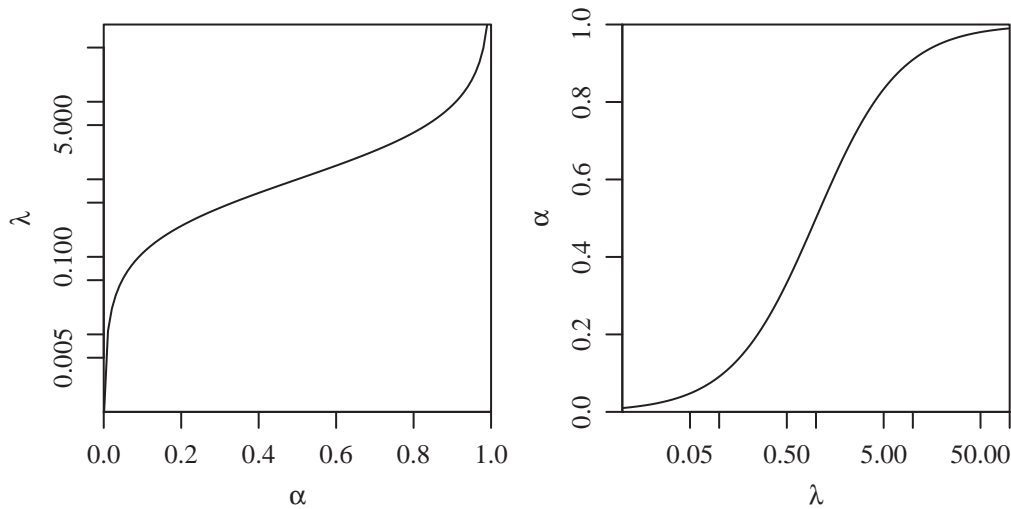
Figure 3.1: Mapping between the α and λ parameter spaces, $\alpha = \frac{\lambda}{1+\lambda}$ or $\lambda = \frac{\alpha}{1-\alpha}$, respectively. $\lambda = 0$ corresponds to $\alpha = 0$, $\lambda = 1$ to $\alpha = 0.5$, and $\alpha = 1$ to the limit of $\lambda \to \infty$.

The USM's dynamics beyond the simple update rule are controlled by a number of other parameters which will be briefly introduced here, before their individual effects are explained in more detail in the following Sections. Firstly, at every point in time a new pair of distinct agents i, j has to be chosen from the population, which consists of a fixed number of N agents total. Interacting agents are randomly drawn based on a matrix G which specifies the probabilities of interacting for all pairs of agents. Whenever an agent i with an internal frequency of use x_i is chosen to engage in an interaction with another speaker j, they each produce and exchange T tokens of the variable under investigation by taking a sample from the corresponding Binomial distributions $Bin(T, x_i)$ and $Bin(T, x_j)$ respectively. Based on the samples n_i and n_j taken from each of the distributions, the agents combine the relative frequencies $\frac{n_i}{T}$ and $\frac{n_j}{T}$ into *perceived frequencies* y_i, y_j according to the following formula:

$$y_i = (1 - H_{ij}) \cdot f(\frac{n_i}{T}) + H_{ij} \cdot f(\frac{n_j}{T}) \ . \tag{3.5}$$

In other words, the perceived frequency is based on a weighted sum of the agent's own productions and that of their interlocutors, and is calculated separately for the other agent j by exchanging all the indices i, j. Here the high degree of modularity of the model becomes evident in the number of parameters, only some of which will be of interest to us here, but

which it is worth going through in turn.

Population size N

Like virtually all models of language change, the USM is a *multi*-agent model, i.e. it simulates a *population* of agents that engages in interactions. While a dynamic population with changing population size would be possible, most investigations are limited to assuming a fixed number of agents N that remain in the population the entire time (again see Baxter and Croft, 2016, for an exception). This simplifying assumption lends the USM to more general analysis and enables to connect it to evolutionary models from other domains. In particular, Blythe (2007) showed the USM's equivalence to Wright's island model (1931), where the population size N corresponds to the number of biological subpopulations or 'islands' between which only limited exchange of replicators takes place. The effect of different values of N on the dynamics of the USM depend on several of the other model parameters, and will be explored in more detail below.

Social network structure/interaction probability matrix G

The parameter G is a square matrix of size $N \times N$ which specifies the probabilities for every pair of agents to be chosen to interact with each other, so that the sum over *distinct* pairs $\sum_{\langle i,j \rangle} G_{ij} = 1$. This parameter can not just gradually alter the frequency or density of interactions between different agents or agent groups. By setting a specific $G_{ij} = 0$ one can completely 'disconnect' two agents i, j in the interaction network, thereby creating the same effect that *social network structure* has in many other multi-agent models of language change. As I discussed above, the exact role that networks of social interactions have on the diffusion of language changes is still debated, with equally conflicting results over whether network structure matters fundamentally (Blythe, 2007; Fagyal et al., 2010; Gong et al., 2012; Pierrehumbert et al., 2014; Kauhanen, 2017) or only marginally (Nettle, 1999; Baxter et al., 2008; Blythe and Croft, 2009; Stadler, 2009), with the results obtained from computational models again largely dependent on many other underlying assumptions and the particular learning models used.

Since this thesis will not investigate the effect of either network structures or nonuniform interaction probabilities, we will abdicate the many degrees of freedom bestowed by the this parameter matrix by always assuming a fully connected network of N agents with equal interaction probabilities, setting $G_{ij} = \frac{1}{N-1}$ for all $i \neq j$.

Accommodation/alignment matrix H

The parameter H in Equation 3.5 above is a square matrix which specifies the weights that all ordered pairs of individual agents give to each others' productions, with $H_{ij} \in [0, 1]$. At the extreme of $H_{ij} = 0$, agent i completely discards any input it receives from agent j and its perceived frequency y_i is consequently completely determined by its own productions. A value of $H_{ij} = 0.5$ would give equal weight to both the speaker's and the listener's production in an interaction. By employing different values in the cells of H (in particular by setting pairs of agents' mutual accommodation parameters H_{ij}, H_{ji} unequal), the matrix can be used to model

asymmetries in adoption structures in a population, as well as increased influence of some individuals' usage levels as a form of individual (rather than variant) prestige, a mechanism that will be explored below.

Beyond using H to introduce individual differences, it is also possible to set uniform accommodation behaviour by setting all matrix values H_{ij} to the same fixed constant $h \in [0, 1]$. The degree of accommodation only affects the USM's dynamics when there are systematic differences in usage levels within the population, which could be due to inter-individual differences such as age-stratified populations or differing variant selection biases (Baxter and Croft, 2016) or otherwise due to clusters or differences in the degree of connectivity in a social network, cases which have only seen limited investigation so far (Blythe, 2007; Michaud, 2017). Since this thesis will not be concerned with inter-individual differences or stratified network structures, all simulations will be performed so that agents are set to only align with their interlocutor and not to their own productions, equivalent to $h = H_{ij} = 1$ for all i, j.

Production sample resolution T

T, a positive integer, is the aforementioned sample size which determines the 'resolution' with which agents can observe the variable use of different variants of an agent with usage rate x in an interaction by randomly sampling from a binomial distribution $Bin(T, x)$. For sake of simplicity we will only be concerned with the case of two competing variants, but the definition generalises to $k \geq 3$ variants in which samples are taken from the multinomial distribution $Mult(T, \vec{x})$, where an agent's usage probabilities over the k variants are specified by a vector $\vec{x}$ of length $k-1$.

The parameter T is rather unusual, in the sense that no comparable parameter features in most other computational models of language change. Among the many models referenced above, most can be assigned to one of two groups based on when and how learning, in the sense of inferring or updating a linguistic property or system from data, occurs. One group, in which agents remain in the population and learning occurs *incrementally*, agents typically receive one data point at a time for each learning event they are sampled to partake in, for example in the Naming Game. In the other group of models there is explicit reference to a *sample size* of the learning data, but there is typically only one learning event at the beginning of an agent's lifetime, such as in the case of the Iterated Learning model. A combination of both, multiple learning events throughout an agent's life time each of which with a learning sample of more than one data point, is not normally considered, but turns out to be a crucial aspect of an evolutionary model of language change as described above. This has to do with the fact that the continued differential *selection* of linguistic variants relies on the existence of variation in the population, variation which can only be attested in learning samples of sizes $T > 1$. This point will become more apparent when we discuss the core parameter that determines how agents derive the *perceived* usage frequencies of variants in interactions, through the bias function $f(.)$.

Bias function $f(.)$

While the numerical parameters so far all control some aspect of the population or interactions, the bias function $f(.)$ is where *selection* of specific variants by individuals comes into play. Its

role is to alter the *objective* relative frequency of tokens produced in the interaction, $\frac{n}{T}$, to an agent's subjective *perceived* frequency. $f(.)$ is simply a function that maps from the frequency interval $(0, 1)$ to $(0, 1)$. While in principle any arbitrary function could be plugged in here, most analyses are limited to mappings that obey some reasonable criteria, in particular that they are *monotonically increasing* within the interval, so that relatively higher objective frequencies are always mapped to higher (or equal) perceived frequencies (Blythe and Croft, 2012).

Special attention should be drawn to the fact that the bias function is only defined for $(0, 1)$ and that, per definition, $f(0) = 0$ and $f(1) = 1$. These two equivalences are imposed because $f(.)$ embodies the *differential selection* mechanism of the USM. While the bias function can alter the variation observed in individual samples in one way or another, the function must not indicate the presence of a variant when it is not attested in the sample. More generally, this constraint also stops the bias function from introducing new variants into the population, and is thus a strict requirement for any evolutionary model that distinguishes the selection of existing variants from pressures of innovation through altered replication.

While the USM's original definition in Baxter et al. (2006) also incorporated parameters for the spontanous generation of unattested variants, most studies of the model so far have been concerned with the analysis of the diffusion and selection of traits that are already established at a low level across the population. With the exception of the final chapter, this thesis will also primarily be concerned with *selection* mechanisms, of which many different ones can be implemented through the function $f(.)$.

3.3 Comparing accounts with the Utterance Selection Model

Having covered the general mechanism of the USM, we can now investigate the predicted dynamics under the presence (or absence) of different biases. This section recapitulates the in-depth study of several different USM biases by Blythe and Croft (2012) while contributing an additional model of asymmetric replicator selection in Subsection 3.3.4. The motivation for the present analysis is to address the question of which accounts or presumed pressures would predict s-shaped transitions of variant use (and under which conditions) when compared in one unified framework, which necessarily also includes a detailed study of the model's baseline behaviour in the absence of any pressures.

3.3.1 Neutral evolution

While the USM's updating rule given in Equation 3.1 is very general and allows for a vast number of modifications through the additional parameters, it is interesting to analyse the model's learning dynamics in the absence of any pressures of either innovation or differential replication. This *neutral evolution* condition, so-called because it is based on completely neutral replication of existing traits from the population according to their current prevalence (Blythe, 2012a), is achieved by using the identity function

$$f(u) = u \tag{3.6}$$

as the USM's bias function, which means that the agents' *perceived* frequency y in an interaction can be directly derived from their interlocutor's productions, i.e.

$$y_i = \frac{n_j}{T} \; . \tag{3.7}$$

Using this simple assumption, we can investigate when the agents' internal x value changes most. Since a lot of the dynamics stem from the basic learning rules, the exact roles of the basic parameters and their behaviour at different moments in the model should be studied in detail. Firstly, Figure 3.2 shows the point change away from a internal usage proportion $x = 0$ for different input data points y as a function of the agent's learning rate (plots are provided for both the α as well as the λ formulation of the learning rate). The equal spacing between the curves for different y means that the impact of different input data points is proportional to their difference to the agent's internal value x.
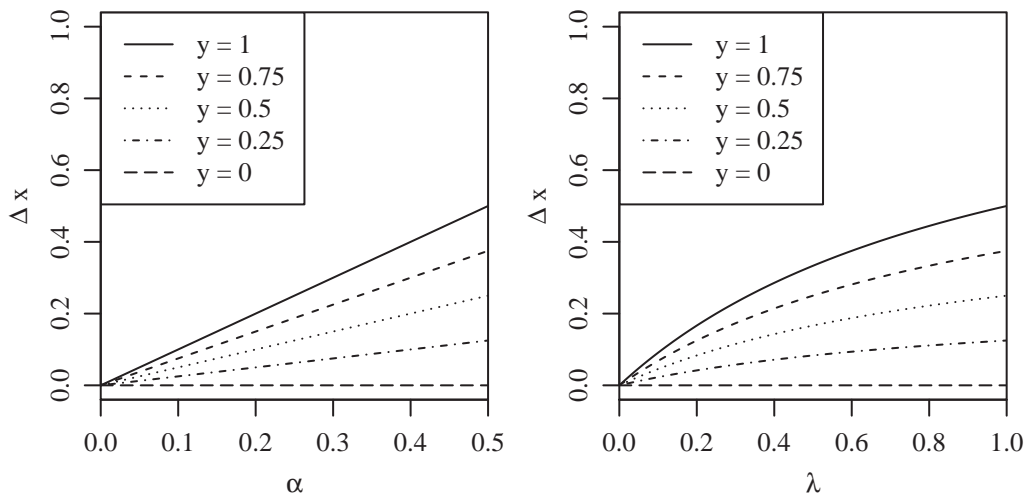
Figure 3.2: Absolute point change to the agent's usage rate for different learning rates with the same initial value $x = 0$ given different input datapoints y. Left: absolute point change as a function of the learning rate α. Given an x value at one extreme and input data at the other, the maximum change to x is equal to α. Right: absolute point change as a function of the learning rate λ.

In fact, an identical picture emerges in the case of a fixed input data point $y = 1$ that is incorporated into different internal values x, as shown in Figure 3.3. Generally, given our EWMA update rule we find that

$$\Delta x = x' - x = \alpha \cdot n + (1 - \alpha) \cdot x - x = \alpha \cdot (n - x) \,, \tag{3.8}$$

i.e. the point change to x is always directly proportional to the difference between the agent's current usage level x and the input data y. The USM's individual agent update dynamics therefore follow a general learning framework that is free from nonlinearities and which, in the absence of any biasing, has been shown to be equivalent to several other models of neutral evolution in biology (Blythe, 2007), and whose dynamics are not significantly affected by setting
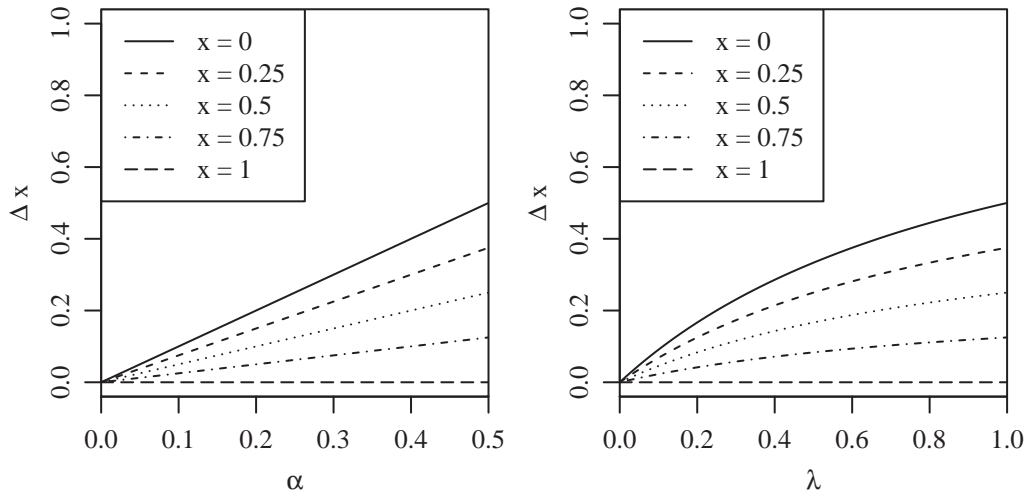
Figure 3.3: Absolute changes for the same input data $y = 1$ for different values of x given a range of learning rates α (left) and λ (right). Given an x value of 0 at one extreme and input data $y = 1$ at the other, the maximum change to x is equal to α.

the neutrally copying agents in many different types of structured networks (Blythe, 2010, 2012b, but see Kauhanen 2017).

Due to the complete lack of asymmetries in the neutral evolution condition its underlying mechanism, often referred to as "random copying", has scarcely been proposed to be the underlying force of language change, or even cultural change more generally (Mesoudi and Lycett, 2009). Upon numerical inspection, neutral evolution exhibits its characteristic dynamics which include "large fluctuations and a tendency for an upward or downward trend to reverse one or more times before an innovative variant goes extinct or wins out" (Blythe and Croft, 2012, p.285). Blythe (2012a) in particular argues that neutral evolution should in fact be taken as a *null model* against which competing accounts of language change should be compared, as a baseline similar to those underlying the neutral theory of molecular evolution in biology (Kimura, 1983). Apart from using it as such a null condition for the model presented in Chapter 4, I will also return to an in-depth study of the dynamics of neutral evolution in Section 3.4. But for now we will focus on replication mechanisms in the USM that actually implement *selection* of some kind.

3.3.2 Replicator-neutral selection

While the term selection is often associated with a preference for particular variants, it really covers *differential replication* of any kind, and can therefore also be used to implement *symmetric* selection functions which are neutral regarding the different variants or replicators. One example is the case of *frequency-dependent selection* biases which can be used to systematically favour variants not based any inherent *a priori* property but based on their current attested frequency in the population, a feature which itself changes over time (Boyd and Richerson, 1985). As the simplest symmetric, frequency-dependent selection mechanism Blythe and Croft (2012) suggest

the function

$$f(u) = u + a \cdot u \cdot (1 - u) \cdot (2u - 1) \ , \tag{3.9}$$

where values of the parameter $a > 0$ lead to a *boost* of variants whose relative frequency is already greater than 50%, while settings of $a < 0$ implement selection in favour of any variants currently in the minority. The two regimes of this non-linear selection function are displayed in Figure 3.4a, giving a visual sense of the function's symmetry: mirroring the plots along the centre point of both axes yields an identical curve, an indication that, mathematically, the function satisfies the symmetry criterion $f(u) = 1 - f(1 - u)$ (Blythe and Croft, 2012).

What is crucial to understand about the USM is that these mapping functions $f(.)$ affect the selection dynamics only somewhat indirectly. First of all, according to the definitions above the choice of resolution parameter T constrains the points in the $[0, 1]$ range at which $f(.)$ is actually ever evaluated, namely only at the fractions which can be sampled from the underlying binomial distribution, i.e. $\{\frac{n}{T} \mid n = 0 \ldots T\}$. Particularly at low values, T can therefore have a drastic impact on the dynamics. To visualise the effect of this parameter, we can determine the *typical change* to an agent's usage rate x by a specific selection function $f(.)$ for different values of x, T. To do so we first calculate the mean perceived frequency $\bar{y}$ over all possible sample outcomes $n \sim Bin(T, x)$,

$$\bar{y} = \sum_{n=0}^{T} P(n; T, x) \cdot f(\frac{n}{T}) \ . \tag{3.10}$$

Using Equation 3.8 we can then determine the average change to an agent's usage level for some constant learning rate, which is plotted in Figure 3.4b. The first striking observation is that, for $T = 2$, the mean expected change Δx is 0 for the entire range of values of x, i.e. the model is equivalent to the random copying of the neutral evolution condition. This result can be explained by inspecting the mapping functions directly above: with $T = 2$, the functions are only ever evaluated at $0, 0.5$ and 1, all values for which they are identical to neutral copying, i.e. where $f(u) = u$. For higher values of T, however, both regimes exhibit the expected influence on the agent's usage rate, with conformity-copying (left panel) decreasing the usage of infrequent variants while further boosting the frequency of those which are already used more than 50% of the time, and the opposite for anti-confirmity copying (right panel). Generally, the higher T, the stronger the impact of the function $f(.)$, since bigger samples allow more evidence for variation[1], a necessary ingredient for differential replication.

Blythe and Croft (2012) report that with values of $a > 0$ any minority variants are rapidly eliminated from the population leading to the fixation of just one variant, while values $a < 0$ give rise to stable co-existence of all different variants, with the stability of co-existence dependent on the community size and learning rate of the individual agents (p.286). These two regimes show that a frequency-dependent selection mechanism of this kind can be used to implement a bias for *regularisation* (as well as *de-regularisation*) of competing linguistic variants. However, neither of these two scenarios lead to directed transitions from the introduction of a novel variant to

---

[1]In particular, less probability mass on the two homogeneous samples $n = 0$ and $n = T$ to which $f(.)$ does not apply.
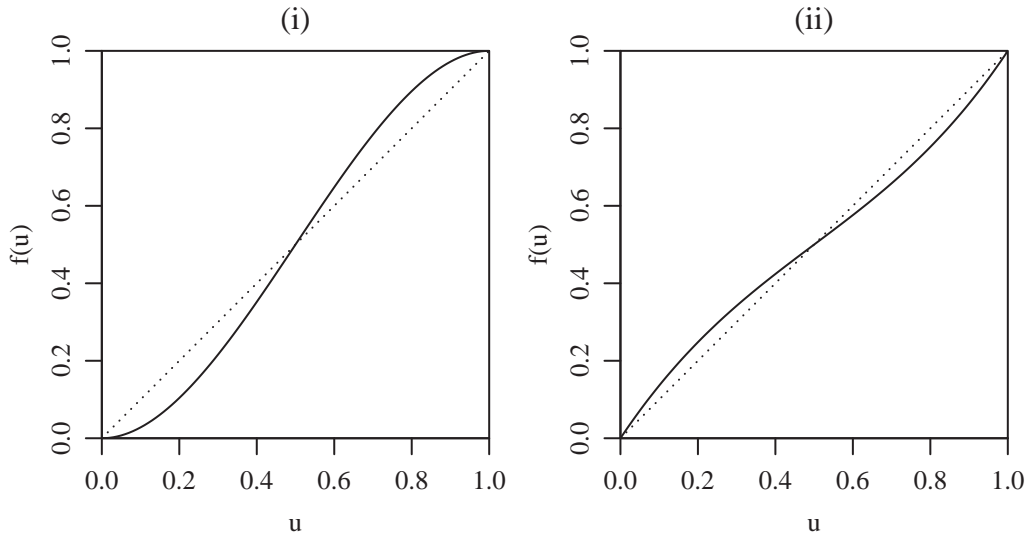
(a) The frequency-dependent selection mapping function from Equation 3.9 against the baseline of neutral evolution, $f(u) = u$, indicated by the dotted line. *(i)* conformity copying with $a = 1$: all input frequencies $u > .5$ are mapped to even higher perceived frequencies so that $f(u) > u$. *(ii)* anti-conformity copying with $a = -.5$, leading to an effect in the opposite direction.
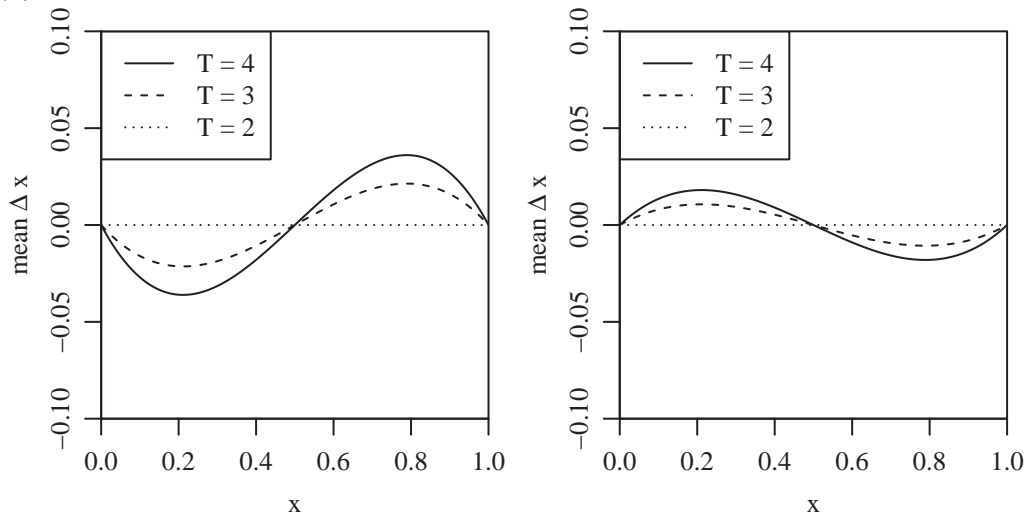
(b) Mean expected change to an agent's usage level x for different values of T, assuming learning rate $\lambda = 1$.

Figure 3.4: The dynamics of frequency-dependent selection as implemented in the utterance selection model through Equation 3.9.

its complete adoption, as is the case in language change. While we will revisit the mechanism of regularisation based on *innovation* rather than selection in a different evolutionary model in Section 3.4, we now turn to the dynamics of USM configurations that implement *asymmetries* of some kind.

3.3.3 Weighted interactor selection

Rather than rely on asymmetries in the *replicators* (the linguistic variants) that is implied by language-internal and variant prestige accounts, the more *mechanical* social accounts discussed in Section 2.2.5 seek the cause for the preferential spread of linguistic innovations in features of the social networks, such as differential interaction densities or the skewed influence of specific individuals or nodes in the network. As discussed above, at least under the assumption of pure random copying the social network structure alone is not sufficient to alter the dynamics of neutral evolution to yield reliable directed transitions. The idea of differential *individual prestige* or influence however corresponds to a wholly separate mechanism, namely that of *interactor selection*, where the asymmetry leading to the differential replication of variants is not due to the bias function $f(.)$ but instead determined by the matrix H which controls the weight given to the samples obtained from different individuals in their interactions with others.

By adjusting the values in H accordingly, one can thus create an interaction structure where the production levels of some group of linguistically 'leading' individuals is preferentially imitated by another group of 'followers', but not vice versa. Proposals of this kind have been suggested to influence the diffusion of language changes through different *adopter group* structures of varying complexity (Rogers, 1962; Milroy and Milroy, 1985; Labov, 2001; Nevalainen et al., 2011; Kauhanen, 2017). Blythe and Croft (2012) use the utterance selection model to investigate the quantitative predictions made by some common assumptions regarding the structure of adopter groups. Initiating only the 'leading' group with high usage rates of an otherwise not established variant, they find that only a highly unrealistic staging of an entire chain of adopter groups, with the respective sizes of the groups following an exponential pattern, lead to the adoption of the incoming variant. Having exhausted all other mechanisms of differential replication, they finally turn to the most direct way in which to affect the model dynamics, namely through a direct asymmetry between the linguistic replicators.

3.3.4 Replicator selection

The most straightforward way to achieve a directed increase of one variant at the expense of the other is by implementing a bias function which consistently boosts the perceived frequency of that variant at the expense of all others. In this way, "$f(u) > u$ for all frequencies u between zero and one, and hence the listener perceives the innovation to be at a higher frequency than it was actually produced at, and overproduces accordingly" (Blythe and Croft, 2012, p.291). The simplest asymmetric linear function used by Blythe and Croft,

$$f(u) = u \cdot (1 + b_m) \,, \tag{3.11}$$

with a selection bias applied for all $b_m > 0$, has an interesting property, namely that it is asymmetric in two ways: not only does it skew the likelihood of adoption towards one of the two competing variants, the strength with which this bias is applied also increases for higher frequencies of the innovative variant, as can be seen in Figure 3.5a(i). To confirm their finding that "an S-shape is easily obtained through replicator selection" (p.291) with little sensitivity to the precise selection function used, I will also be investigating a second model of replicator selection that employs an *additive* instead of a multiplicative bias which exerts an equally strong bias across the entire trajectory, i.e.

$$f(u) = u + b_a \ , \tag{3.12}$$

again capped at the maximum value of 1.0, as can be seen in Figure 3.5a(ii). In order to achieve a comparable bias strength for the two types of replicator selection, for all later figures the respective bias values will satisfy $b_a = \frac{b_m}{2}$, a choice which results in equivalent strength selection at the midpoint as well as (roughly) equal amplification relative to the neutral evolution condition $f(u) = u$ across the entire trajectory.

From these bias functions $f(.)$ we can again derive the average perceived frequency for different settings of T as well as the consequent expected change to the agents' usage frequencies, which are shown in Figure 3.5b. The asymmetry of replicator selection is immediately evident from the fact that the expected change is always greater than zero, meaning that the incoming variant is boosted across the entire frequency range. By solving the differential equations defined by these functions we can also calculate the average trajectories that would be produced as the selected for variant spreads through the population, which are shown in Figure 3.5c (for an in-depth explanation of the approach see the appendix to Blythe and Croft, 2012, mathematical derivations of the results for both models of replicator selection are provided in Appendix A).

While for both models of selection the average changes to x at $T = 2$ result in growth patterns that are equivalent to the canonical s-shaped logistic function, the relative intensity of selection across the trajectory diverges for higher values of T. For the original, multiplicative bias model shown in Figure 3.5c(i), increasing the sample resolution leads to an extended period of initial growth in which the rate of change increases past the halfway mark at which simple logistic growth characteristically starts to slow down again, making the growth pattern more and more exponential as the sample resolution $T \to \infty$. Consequently, this selection function predicts the maximum rate of growth, indicated by the peak in Figure 3.5b(i), to occur later during the trajectory for higher settings of T, along with relatively greater levels of selection and thus faster transitions given the same bias strength b_m.

A different picture emerges for the additive bias, as shown in Figure 3.5b(ii). While the overall selection pressure imposed by the bias function also increases for higher T, the pattern of selection stays symmetric around the mid-point, with a relative acceleration of transitions achieved by a greater degree of selection in the other frequency regions. For $T = 2$ as well as $T = 3$ the selection dynamics can be shown to be identical to logistic growth as displayed in Figure 3.5c(ii), with growth rates of $2b_a$ and $3b_a$ respectively (see Appendix A). No general solution for the average expected trajectory is provided here but, since the function describing

the mean change to x shown in Figure 3.5b(ii) approaches a constant value throughout the interval $(0,1)$ as T increases, we should expect the dynamics to start resembling those of steady linear growth as $T \to \infty$.

With this analysis we come to the end of an overview of some of the most relevant results that have been obtained from the computational version of the Utterance Selection Model originally due to Baxter et al. (2006). Starting from an investigation of its baseline learning dynamics, which were shown to implement a model of *neutral evolution*, we recapitulated the survey of pressures provided in Blythe and Croft (2012) to determine which evolutionary replication mechanisms could lead to directed, s-shaped trajectories. Blythe and Croft concluded that only an inherent asymmetry between variants, implemented as a preference for the innovative variant that is shared by the majority of interacting agents, can reliably produce directed transitions. This conclusion can help us rule out some of the proposed pressures which, while not completely ineffective, do not appear to have the necessary leverage to be the main driving force behind the adoption of individual language changes, at least as far as one accepts the model's underlying assumptions regarding the behaviour of individual agents and their interactions. But, curiously, the result does not speak to the two biggest contenders to explaining language change: both language-internal and social, *variant prestige* accounts are based on an underlying preference for an innovative over an outgoing variant, as implemented by a replicator selection bias. While this last Section showed how slightly different assumptions regarding the strength of the asymmetry and force of the selection pressure across different attested frequencies of the variants can yield slightly different predictions regarding the resulting trajectories, a look at the quantitative investigations of language changes presented in Chapter 2 indicated that the empirical data might be too sparse to make any strong claims or perform conclusive comparisons between models and data (although see Altmann et al., 2013; Ghanbarnejad et al., 2014).

Despite the fact that the USM features a larger number of parameters than most other models of language evolution, every single one is both transparent and grounded in the USM's dedication to a concrete evolutionary framework (Croft, 2000). The sample resolution parameter T, for example, while unusual in terms of its absence from other models of language change, is well-motivated by the model's strict separation of innovation and selection pressures. The persistence with which the model has been analysed, both via analytical methods and numerical simulation, means that the meaningfulness of its dynamics (or lack thereof) with regard to different parameters and parameter combinations are much better understood than for most other models (Blythe, 2007; Baxter et al., 2009; Blythe and Croft, 2009; Blythe, 2012a; Baxter and Croft, 2016; Michaud, 2017).

Before I go on to expand the studies of the USM through another selection mechanism intended to tackle the question of how asymmetries between variants might emerge out of the replication dynamics itself, I will first introduce a simpler, yet related, modelling framework that has also been used to investigate s-shaped curves, and which will also resurface again in Chapter 7.
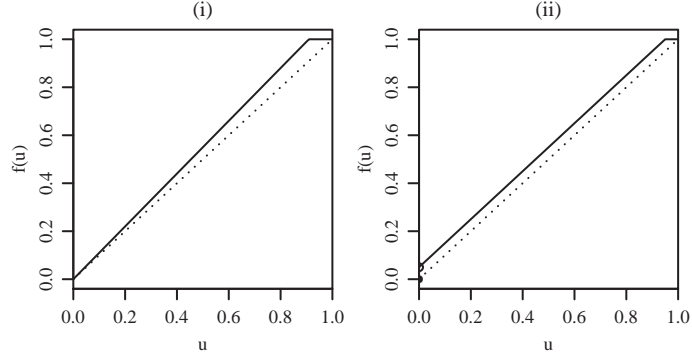
(a) Objective to perceived frequency mapping under replicator selection. *(i)* multiplicative replicator bias presented in Blythe and Croft (2012), with $b_m = 0.1$. *(ii)* additive bias with approximately equivalent bias strength $b_a = b_m/2 = 0.05$. All mapping functions are capped at 0 and 1 and impose $f(0) = f(1) = 0$. The dotted line indicates the unbiased mapping characteristic of neutral evolution, i.e. $f(u) = u$.
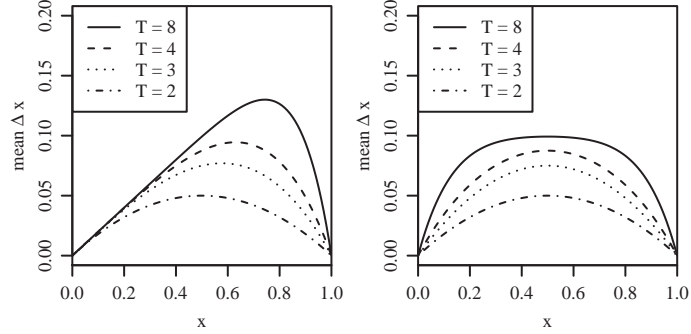
(b) Average bias applied to input frequence x with multiplicative and additive replicator selection for different values of T, assuming $\lambda = 1$. The asymmetry of the perceived frequency mapping functions above (relative to the neutral evolution condition indicated by the diagonal) leads to an increase of the innovative variant across all frequency ranges, with generally stronger selection for higher T. Mathematical derivations for the functions can be found in Appendix A.
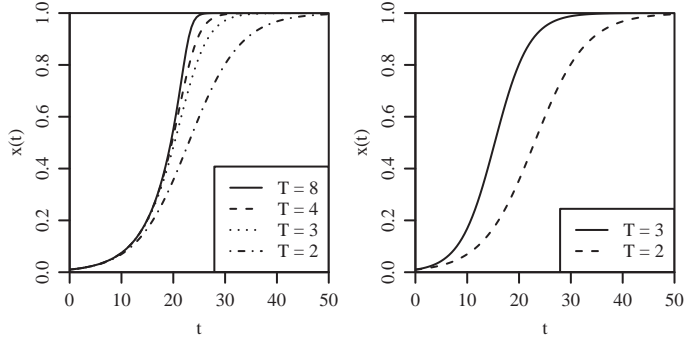
(c) Typical average trajectories resulting from applying the multiplicative and additive replicator biases. For higher T the multiplicative bias extends the initial period of exponential growth, while the additive bias remains symmetric around the mid-point with fast growth from the very start of the trajectory.

Figure 3.5: Analysis of the selection dynamics for the original, multiplicative (left column) as well as additive (right column) replicator selection bias.

3.4   A Markov chain model of neutral evolution

One formal model of neutral evolution (i.e. copying of linguistic traits in the absence of any replicator or interactor selection) that makes particular reference to the temporal dynamics of changes is Reali & Griffiths model of regularisation by Bayesian learners (2009; 2010).

At its core, Reali and Griffiths present a model of frequency learning by Bayesian inference. In their particular framing, an individual is trying to infer the relative frequencies $\theta_i \in [0, 1]$ of different variants $i = 1 \dots n$ based on some input data as well as prior beliefs about what the true values of θ_i are likely to be. These prior beliefs act as *inductive biases* and are captured by the *prior*, represented by a probability distribution $f(\vec{\theta})$ defined over all possible values of $\vec{\theta}$.

For the simple case of two competing variants, even though the individual is technically inferring two complementary relative frequencies θ_1, θ_2, we can limit our analysis to the problem of inferring θ_1, since trivially $\theta_2 = 1 - \theta_1$. The model can easily be extended from the *binomial* (two-variant) outcome to *multinomial* outcomes, i.e. with three or more competing variants but, without loss of generality, I will limit the demonstration to the case of two competing variants. To simplify notation I will henceforth also simply write θ to refer to θ_1.

While any continuous probability distribution over the interval $[0, 1]$ could serve as a prior, the authors choose the *Beta distribution*, whose probability density function is defined as

$$f(x; \alpha, \beta) = \frac{1}{B(\alpha, \beta)} x^{\alpha - 1} (1 - x)^{\beta - 1} \; , \tag{3.13}$$

where $B(.)$ is the *Beta function*.

Because we are interested in a *neutral* model that is not a priori biased in favour or against either of the competing variants, the shape of the prior distribution over the support will have to be *symmetric*: the prior probability density of θ taking a certain value, $f(\theta)$, should be the same as its probability of taking the complementary value $f(1 - \theta)$. This can be achieved by setting the Beta distribution's two shape parameters α, β to the same value. Consequently the authors use prior distributions of the form

$$\Theta \sim Beta(\frac{\alpha}{2}, \frac{\alpha}{2}) \; . \tag{3.14}$$

with just a single parameter, α, which controls the degree of *regularisation*. Figure 3.6 shows the effect of this parameter on the prior distribution. For a value of $\frac{\alpha}{2} = 1$ the prior distribution is uniform: not only is the individual not biased towards any of the variants (the distribution is symmetric), their estimate of the underlying frequency θ is not biased towards any particular frequency region in $[0, 1]$ either. This isn't the case when $\frac{\alpha}{2} \neq 1$: for values < 1, the inference of θ is explicitly geared towards more extreme relative frequencies closer to 0 or 100% usage – the model implements a *regularisation bias*. The opposite is the case when > 1 which favours values of θ that are closer to the 0.5 mark. Agents employing such a setting are inclined to infer more mixed usage of the competing variants than suggested by their learning data alone.

The particular choice of prior distribution (Beta or Dirichlet for the multinomial case) has elegant mathematical properties: when a learner receives an input sample of size N, where
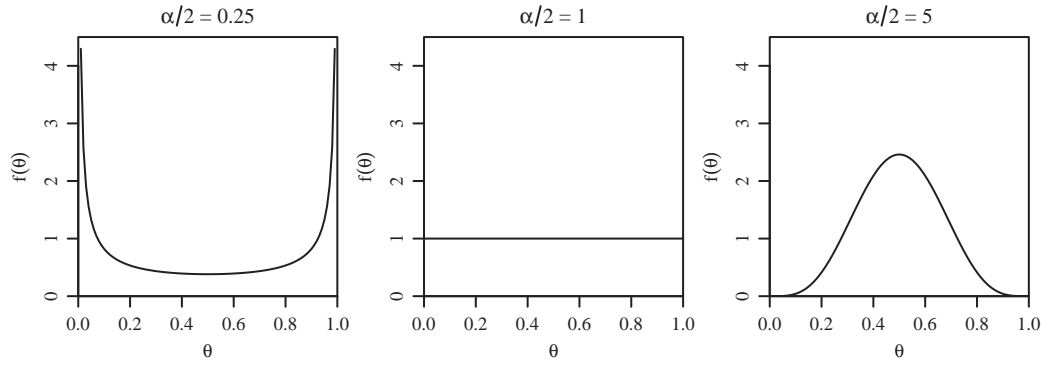
(a) Prior distributions for three different levels of α.
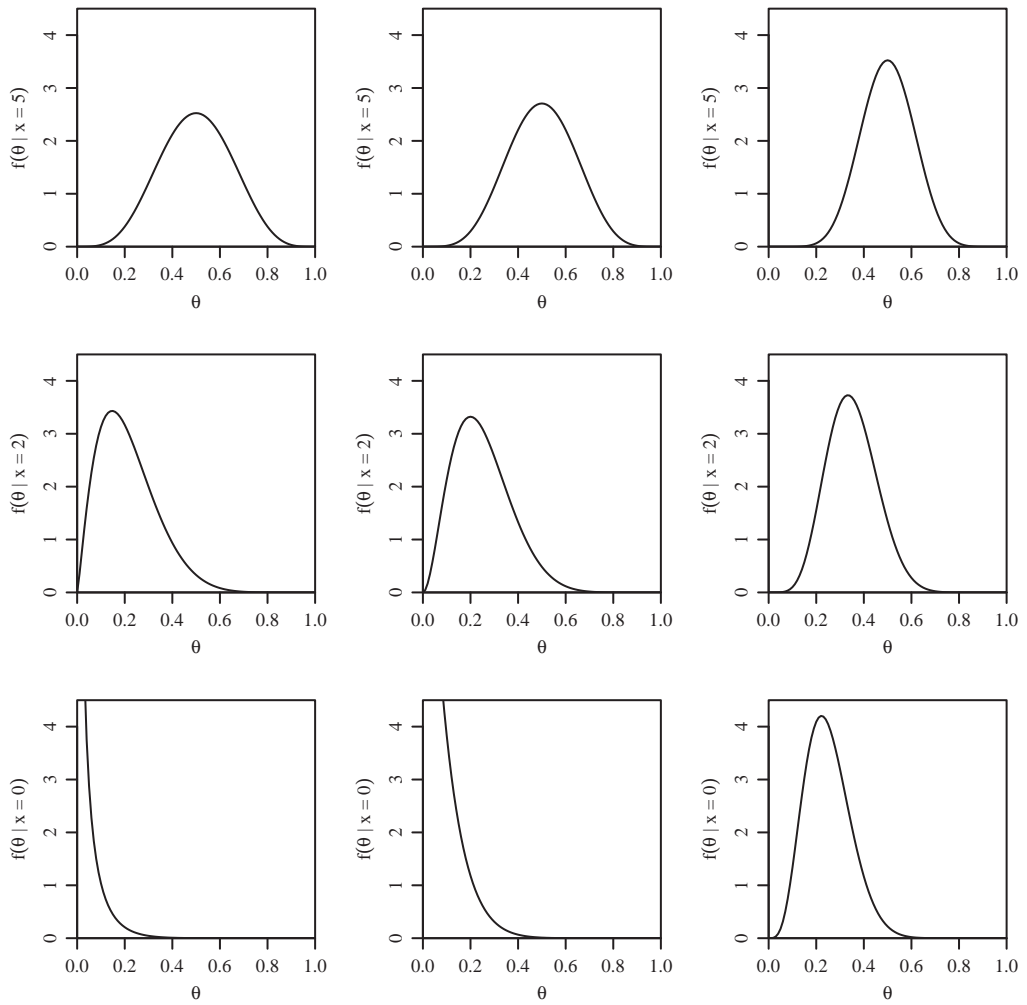
(b) Posterior distributions after observing $N = 10$ data points with various input distributions $(x = 5, 2, 0)$, regularisation parameters as in (a).

Figure 3.6: Examples of Beta distribution priors and posteriors with three different levels of the regularisation parameter $\alpha/2$.

$0 \leq x \leq N$ of the tokens were instances of the variant whose frequency θ they are trying to infer, then the posterior is again a Beta distribution, namely

$$\Theta|x \sim Beta(x + \frac{\alpha}{2}, N - x + \frac{\alpha}{2}) \ . \tag{3.15}$$

Following this inference step, there is still the question of how the posterior distribution is translated into actual production behaviour, which provides us with testable predictions of the model. Here, I will consider three different ways for an individual to generate their own productions x' based on the learning sample x that they themselves received. The first two were also treated by Reali and Griffiths (2009), the third covered by Ferdinand (2015, p.156):

Sampling from the posterior: when generating new productions directly from the posterior, the probability that a *sampling learner* produces a particular variant x' times out of a total of N productions is distributed according to a *Betabinomial distribution* with the same parameters as the posterior distribution in Equation 3.15, i.e.

$$X'|x \sim BB(x + \frac{\alpha}{2}, N - x + \frac{\alpha}{2}, N) \ . \tag{3.16}$$

Adopting the *average* of the posterior: instead of sampling from the posterior for every production, an individual could deterministically select the mean of the posterior distribution, which is

$$\hat{\theta} = \frac{x + \frac{\alpha}{2}}{N + \alpha} \ . \tag{3.17}$$

The productions of a Bayesian learner who deterministally chooses the parameter $\hat{\theta}$ are then distributed according to a Binomial distribution,

$$X'|x \sim Bin(N, \hat{\theta}) \ . \tag{3.18}$$

While Reali and Griffiths (2009) call this a 'MAP' learner, I will refer to this mechanism of selecting a hypothesis as the *averager* strategy.

Adopting the *mode* of the posterior (*maximum a posteriori*): The posterior distribution's mode, where the probability density function is highest, can be found at

$$\theta_{MAP} = \arg \max_{\theta} f(\theta|x) = \frac{x + \frac{\alpha}{2} - 1}{N + \alpha - 2} \ , \tag{3.19}$$

except when $x = 0$ or $x = N$, in which case the resulting posterior Beta distribution is *j-shaped*, with the mode falling on 0 or 1, respectively. When such a *MAP learner* has adopted the mode as their production probability then their own productions are distributed according to a Binomial distribution with $p = \theta_{MAP}$, i.e.

$$X'|x \sim Bin(N, \theta_{MAP}) \ . \tag{3.20}$$

One way in which the impact of these different ways of sampling data (either directly from

the posterior or by first deterministically selecting a θ) can be exemplified is by visualising the *average production* of the different learners based on the input sample they just observed. This data is shown in Figure 3.7, which maps the different possible input distributions (along the x-axis) to the average output productions $\pm$ their standard deviation. The identity function $x = y$, equivalent to pure probability matching, is shown for reference. In this graphical representation, a mapping function that leads to increased *regularisation* should map input proportions between 0 and 50% to even *lower* output proportions, while input proportions $> 50\%$ should yield output proportions even closer to 100%.

What is evident from Figure 3.7 is that the only method which on average leads to regularisation at every iteration is the *maximum a posteriori* method with $\alpha \leq 1$. None of the other mapping functions are consistently regularising. Rather, as was pointed out by Ferdinand (2015, p.176) both data production methods discussed by Reali and Griffiths rely on mechanisms that merely increase the sample variability *in either direction*, until the system drifts into a state of categorical presence of one variant only. This contrasts with the regularising mapping functions of the Utterance Selection Model shown in Section 3.3.2, which systematically increase the proportion of whichever variant is currently more prevalent. We will return to a critique of the present regularisation model in the next section.

3.4.1 Representing Bayesian Iterated Learning as a Markov chain

While the model presented above captures frequency learning by Bayesian inference within one individual, it is interesting to ask how the productions of a sequence of such learners would develop over time when one individual's output serves as the learning input of another. To do this, we can analyse the interactions between repeated learning input and production output as a *Markov chain*, a simple modelling tool for understanding systems which can be in one of a finite number of states that they switch between probabilistically.

More formally, a Markov model can be defined by specifying conditional transition probabilities $P(X_{t+1} = x'|X_t = x)$ between a number of discrete states $x, x' \in \mathcal{S}$, which we call the Markov model's *state space*. The Markov model is completely described by a function $P : \mathcal{S} \times \mathcal{S} \to [0,1]$ where the transition probabilities *out* of any given state have to sum to one, i.e.

$$\sum_{x' \in \mathcal{S}} P(X_{t+1} = x'|X_t = x) = 1 \qquad \forall \, x \in \mathcal{S} \, . \tag{3.21}$$

In the case of the Bayesian inference model above, there are two equally valid ways in which it could be translated into a Markov model, based on how the state space $\mathcal{S}$ is construed. The logical alternation between learning parameter θ and production of x tokens of a specific variant out of N total productions allows for both a characterisation of the Markov model as transitioning from one individual's posterior distribution $f(\theta|x)$ to another or, alternatively, from one individual's number of productions x to the next.

To define the state space, we have to set a fixed size of productions N, from which a new learner has to infer the underlying production frequency θ.

An example of such a transition matrix for $N = 10, \alpha = 0.5$ is found in Table 3.1. This
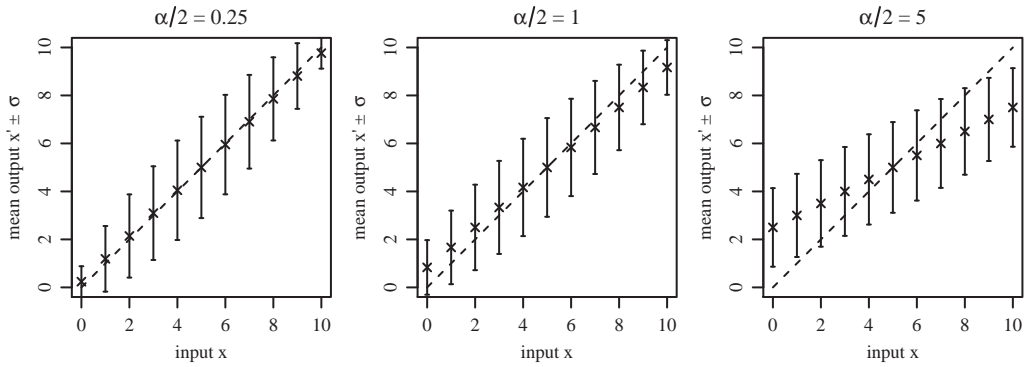
(a) Input-to-mean output mapping when sampling from the posterior.
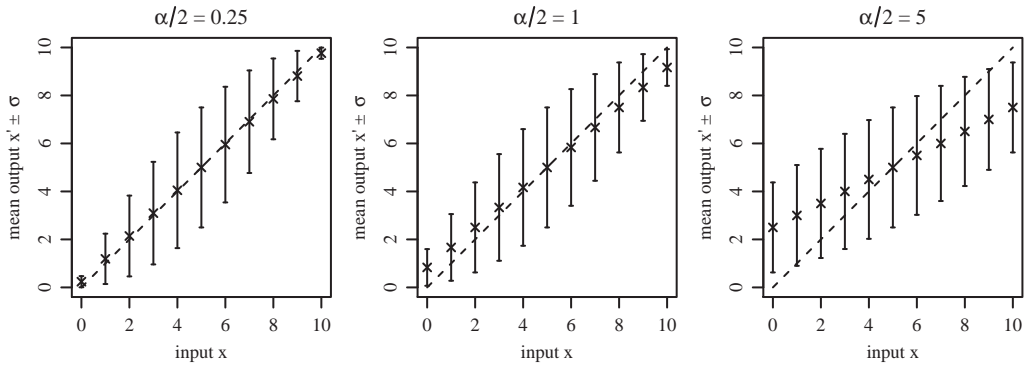
(b) Input-to-mean output mapping when selecting the *average* of the posterior as the hypothesis. As pointed out by Ferdinand (2015), the *mean* output of this model is identical to that of the sampler shown above, only that the averager exhibits different amounts of sampling error about this mean, depending on the input frequency.
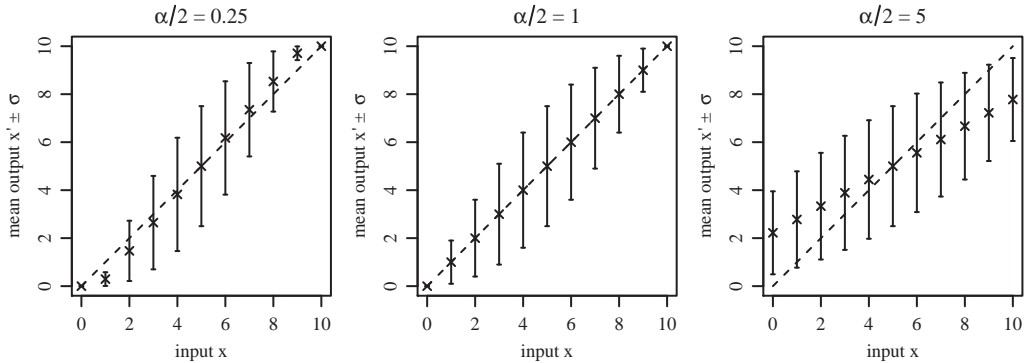
(c) Input-to-mean output mapping when selecting the *maximum* of the posterior as the hypothesis (MAP). With $\frac{\alpha}{2} = 1$ (middle panel) this strategy is identical to pure frequency matching, while MAP with $\frac{\alpha}{2} < 1$ (left panel) is the only strategy that, *on average*, leads to regularisation in one iteration.

Figure 3.7: Input to mean-output mapping for the three ways of producing data from the posterior and three levels of the regularisation parameter. The three settings of α capture inductive biases ranging from regularisation ($\frac{\alpha}{2} = .25$, left column) to de-regularisation ($\frac{\alpha}{2} = 5$, right column).

particular matrix is created based on the assumption that learners sample their data directly from the posterior distribution they computed from the input they received.

Compare this to Table 3.2, which is based on a chain of learners that deterministically select the mode θ_{MAP} of the posterior distribution $f(\theta|x)$ as their estimate of θ. Their data production probabilities are consequently distributed according to a Binomial distribution with $p = \theta_{MAP}$, so the rows of this transition matrix are equivalent to this Binomial distribution.

	$x' = 0$	$x' = 1$	$x' = 2$	$x' = 3$	$x' = 4$
$x = 0$	0.8379	0.1156	0.0347	0.0099	0.0019
$x = 1$	0.3756	0.3005	0.1932	0.0985	0.0322
$x = 2$	0.1352	0.2318	0.2659	0.2318	0.1352
$x = 3$	0.0322	0.0985	0.1932	0.3005	0.3756
$x = 4$	0.0019	0.0099	0.0347	0.1156	0.8379

Table 3.1: Markov chain transition matrix for the Bayesian Iterated Learning model with $N = 4$ and $\alpha/2 = 0.25$. The rows represent the probabilities of producing any of the given samples, assuming that the production is sampled from the posterior.

	$x' = 0$	$x' = 1$	$x' = 2$	$x' = 3$	$x' = 4$
$x = 0$	1.0000	0.0000	0.0000	0.0000	0.0000
$x = 1$	0.6561	0.2916	0.0486	0.0036	0.0001
$x = 2$	0.0625	0.2500	0.3750	0.2500	0.0625
$x = 3$	0.0001	0.0036	0.0486	0.2916	0.6561
$x = 4$	0.0000	0.0000	0.0000	0.0000	1.0000

Table 3.2: Markov chain transition matrix for the Bayesian Iterated Learning model with $N = 4$ and $\alpha/2 = 0.25$. The rows represent the probabilities of producing any of the given samples, equivalent to $Bin(x'; N, p = \theta_{MAP})$

The system that we describe by specifying the transition probabilities between individual states is a random process called a *Markov chain*. Stochastic systems of this kind are said to obey the *Markov property*, which means that the probability of entering a particular state only depends on the system's *current* state, but not on any other prior states or state sequences that preceded the current one. This image of a *chain* maps neatly onto the Iterated Learning model, where every new learner receives input from their parent generation who they then replace.

Importantly for us, the characterisation of a stochastic system as a Markov chain allows for straightforward analyses of different kinds. For example, assuming that our system would run indefinitely, we can calculate the probability of this infinite chain of states to occupy a particular state *in the limit*. The so-called *stationary distribution* π of a Markov chain transition matrix T is a probability distribution over its states S, i.e. it must satisfy

$$\pi \geq 0, \sum_{s \in S} \pi_s = 1 \ . \tag{3.22}$$

In mathematical terms, the stationary distribution has the property that performing another

iteration of the chain must map the distribution onto itself, i.e.

$$\pi = \pi \cdot T \ . \tag{3.23}$$

Based on these definitions, it is possible for a given Markov chain to have more than one stationary distribution. This is generally only the case when the state space consists of sub-partitions that cannot be reached from each other, as is the case when there is more than one absorbing state. The stationary distributions of the different systems whose input/mean-output mapping we visualised previously in Figure 3.7 are shown in Figure 3.8.

The stationary distributions confirm that the parameter α indeed works as intended: when $\alpha/2 < 1$, the chains spend most of their time in the extreme states corresponding to categorical usage of either of the two competing variants. When $\alpha/2 > 1$, on the other hand, the chains mostly consist of learners who mix the variants evenly. The behaviour with intermediate values $\alpha/2 \approx 1$ falls in between, with the exact distribution also depending on the type of learners.

The MAP learner, not considered in the original Reali and Griffiths papers, deserves special attention: as already hinted at above, only this learning strategy looks like a proper *regulariser* in the sense that an input proportion will, *on average*, result in an output proportion that is in fact more regular than the input. It is also the only learning strategy which, for any $\alpha/2 \leq 1$, does not introduce variation when there isn't any in the input, i.e. learners who receive homogeneous input will never spontaneously introduce variation into their output. Figure 3.8c shows that, as a consequence, chains of such learners will end up in either of two *absorbing states* corresponding to categorical usage of a variant, and remain there indefinitely.

3.4.2 Neutral evolution and s-shaped curves

So far our analysis of the stationary distribution limits us to describing the expected state of a model, but abstracted away from time. One particular claim of Reali and Griffiths (2010) concerning temporal dynamics is that even the neutral evolution model described by Bayesian regularisers will produce s-shaped curves. While we would not expect completely symmetric replication such as produced by neutral evolution to produce particularly directed transitions, they argue that this depends on which data is considered. In particular, since historical linguists only (or at least primarily) describe changes which have gone to completion, our assessment of whether a model produces s-shaped curves should equally be limited to data of this kind. They consequently go on to analyse only those chains that start off in a state where the first generation uses one of the competing variants categorically, while the last ends up in the opposite state where its productions contain only the other variant.

In order to get a better understanding of the underlying dynamics of our Markov model, we will therefore need to switch to an analysis that allows us to condition the Markov chains to be in specific states at specific points in time. One tool to do exactly this are *Hidden Markov Models* (HMMs). As the name suggests, HMMs are closely related to the Markov models described above. While in 'normal' Markov chains the state sequence is directly visible to the observer, Hidden Markov Models allow us to specify a certain level of uncertainty over the model's state at any given point in time. Of particular importance to is that, instead of just
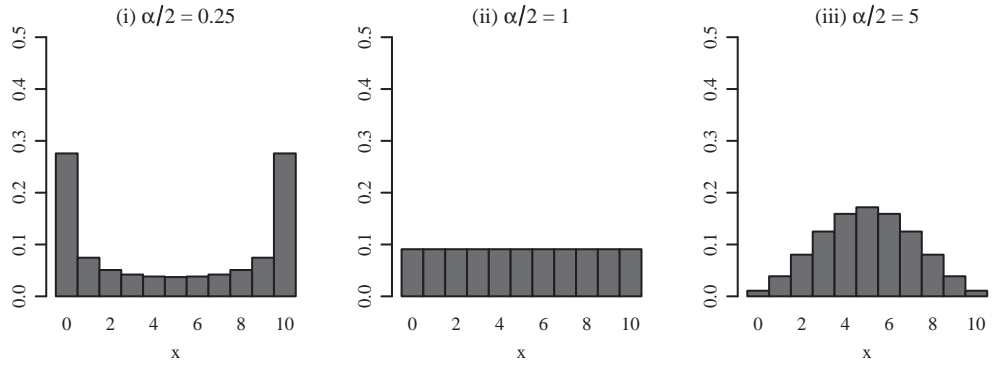
(a) Stationary distribution for chains of learners who are sampling from the posterior.
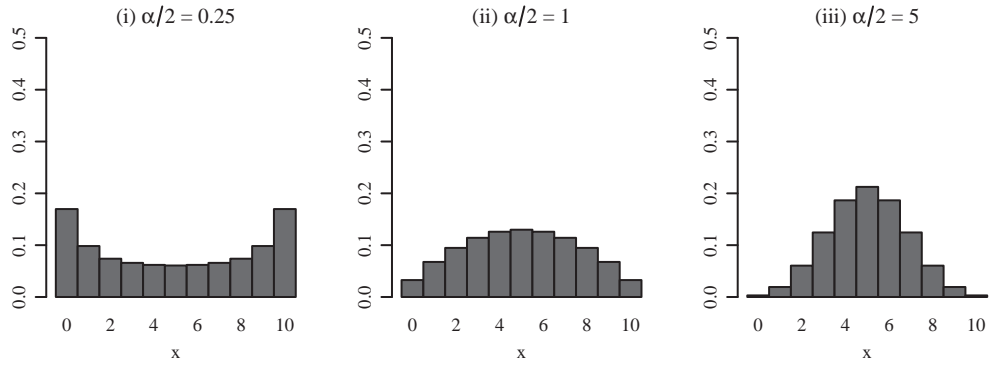
(b) average
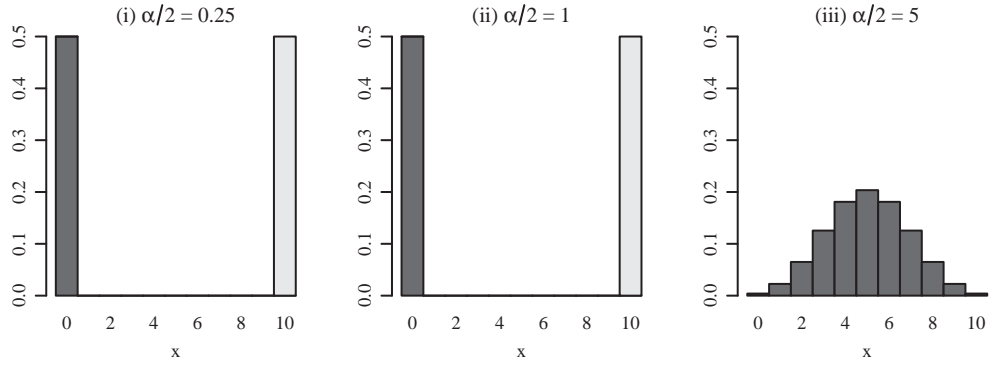
(c) Stationary distribution for *maximum a posteriori* (MAP) learners. The different colours indicate that for $\alpha/2 \leq 1$ the Markov chain has two absorbing states, corresponding to categorical usage of either of the variants.

Figure 3.8: Stationary distributions of the Markov chain transition matrices.

randomly generating state sequences, HMMs allow us to make probabilistic inferences about the most likely states or state sequences that our model is likely to be in.

In what follows, I used R's `HMM` package (Himmelmann, 2010) to both replicate and extend the results reported in Reali and Griffiths (2010). Firstly, Figure 3.9 shows a replication of the original analysis from their paper. All four subplots show the state probability distribution for Markov chains of length 50 where the input data presented to the first generation consisted of 50 instances of only one variant. The probability distribution is represented as a heat map where, for any specific generation, darker colors indicate a higher probability of being in a state at that time. The probabilities of all states per generation sum to 1. The particular probability distributions shown here were calculated for chains of learners which use the inferred mean $\hat{\theta}$ of the posterior distribution to sample data for the following generation, but results for learners sampling directly from the posterior are qualitatively similar.

Subfigures (i) on the left show the development of the chains when conditioning on this initial state only. These two plots, which differ only in their setting of α, neatly highlight the contrast between the two different regimes of the regularisation parameter *alpha*: in Figure 3.9a we set $\alpha = 0.5$, corresponding to *regularisation*. In this setting, chains of learners are drawn to produce either of the two variants (near-)categorically. Note that, even though the system starts off with only one variant as its input, the chance introduction of tokens of the competing variant leads some chains to eventually regularise in the 'other direction': whenever tokens of the other variant accumulate through random sampling, the chains start to be equally drawn towards the other fully regular state, i.e. categorical usage of the formerly unattested variant.

While even after 50 generations the majority of chains is still at or near the usage frequency that was presented to the first participant, increasingly chains will start to 'bunch up' against the top-most state corresponding to categorical usage of the other variant. Indeed, in the limit we should expect the the right-most 'slice' of Figure 3.9a(i) to become completely symmetric around the halfway-mark, as it approaches the Markov chain's stationary distribution shown above in Figure 3.8b(i).

In contrast, the left panel of Figure 3.9b with $\alpha = 10$ represents the *de-regularisation* regime, where individuals prefer to use both variants equally. This is borne out by the fact that chains of such learners are quickly drawn towards the middle states, indicating mixed usage.

Subfigures (ii) on the right-hand side show the expected distribution of states when conditioning on both the initial and final states of the chain, where the last individual only produces tokens of the competing variant that was not attested in the first generation's input data. While the probability distribution over possible states at most intermediate generations is extremely wide, Reali and Griffiths point to the *average trajectory* (shown in white) that is calculated by computing the *average state* of all chains at any given generation. They point out that, intriguingly, the shape of this average trajectory is dependent on the regularisation parameter α. In particular, the model produces s-shaped trajectories exactly when chains are geared towards regularising input, which experimental evidence suggests is in fact a feature of human language learning (Hudson Kam and Newport, 2005; Reali and Griffiths, 2009; Smith and Wonnacott, 2010).

It is crucial to point out here is that this *average* of all transitions is not necessarily rep-
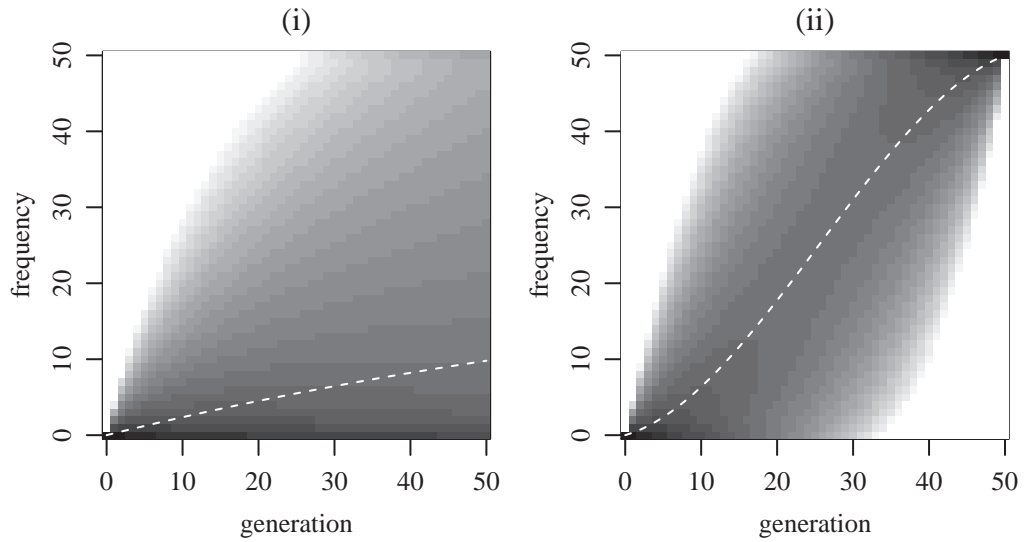
(a) Results with learners accepting the *mean* of the posterior as their hypothesis for θ with $\alpha = 0.5$.
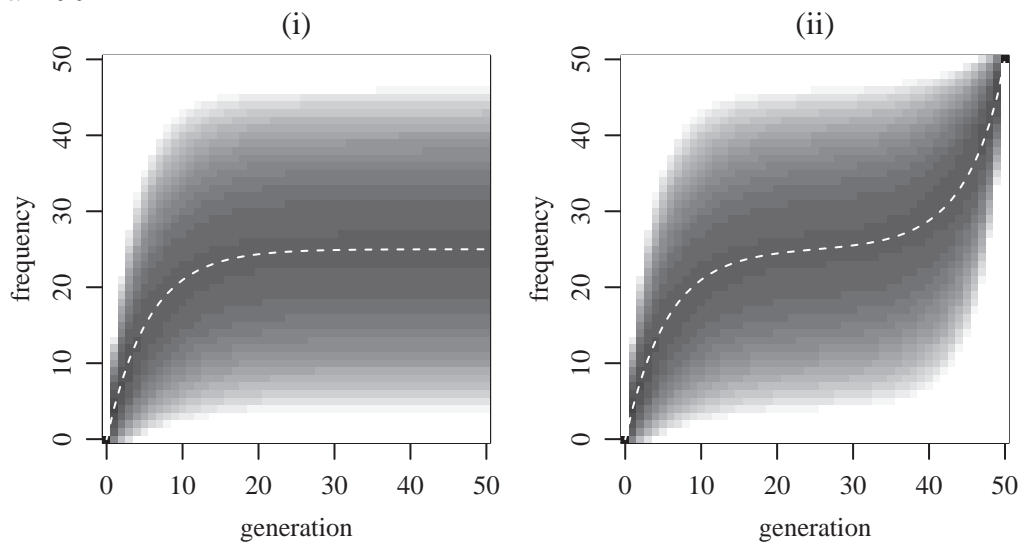
(b) Results with learners accepting the *mean* of the posterior as their hypothesis for θ with $\alpha = 10$.

Figure 3.9: State probability distribution for all Markov chains of length 50 where the input to the first generation consists of tokens of only one variant. The dashed white line indicates the trajectory through the 'average' states that the chain is in at any given point in time. *(i)* conditioning on the first generation's input only *(ii)* conditioning on both the first and final generations' data.

resentative of the model's *typical* transitions (Blythe, 2012a). In order to get an idea of what
individual trajectories of Iterated Learning chains actually look like, we can simply generate
state sequences of the underlying Markov model randomly and filter them according to the
start and end conditions (see Appendix B for the code).

Figure 3.10 shows three randomly generated chains that fulfill both the start and end condi-
tion specified above. The trajectories were generated using exactly the same parameter setting
as the one underlying the s-shaped average trajectory shown in Figure 3.9. Already here we
can see that individual trajectories are much more noisy, less directed and s-shaped than the
numerically computed 'average transition' above suggests.
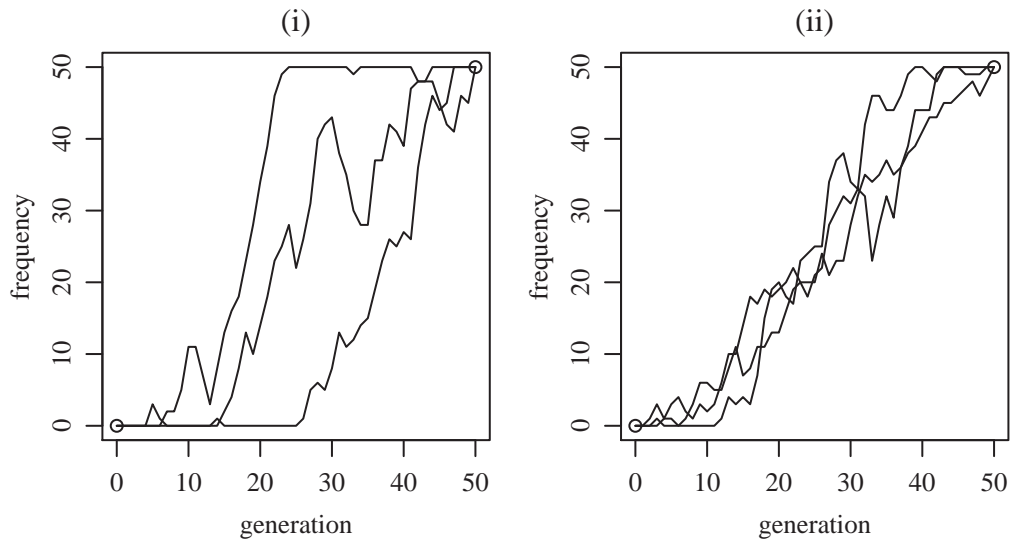
Figure 3.10: Three randomly generated Markov chains initiated at $0/N$ and terminating at
N/N after 50 iterations. *(i)* learners sampling from the posterior distribution $p(\theta|x)$. *(ii)* learn-
ers accepting the *mean* of the posterior as their hypothesis for θ.

What is also evident is that not all of the 'transitions' are actually of the length that we
specified: many chains either remain at the initial state for some time, or otherwise converge
on categorical usage of the other variant early and remain there until the remaining generations
have passed. This points to another more general problem, namely that termination after
exactly 50 generations is not actually well-motivated. To understand the dynamics of this model
even better we should therefore take a closer look at the expected duration of transitions.

Expected number of generations for a transition to complete

In order to get a more accurate picture of the typical trajectory exhibited by regularising
Iterated Leaners, we first need to know the likelihood of a transition completing in a given
number of generations. Figures 3.11 shows both the per-iteration probability as well as the
cumulative probability of a chain of Iterated Learners reaching categorical usage of the initially
non-existent, incoming variant over time.

For the *averaging* learner with parameters $\frac{\alpha}{2} = 0.25$ and $N = 50$ as above, the chain is most
likely to first reach categorical usage of the incoming variant at the distribution's mode after
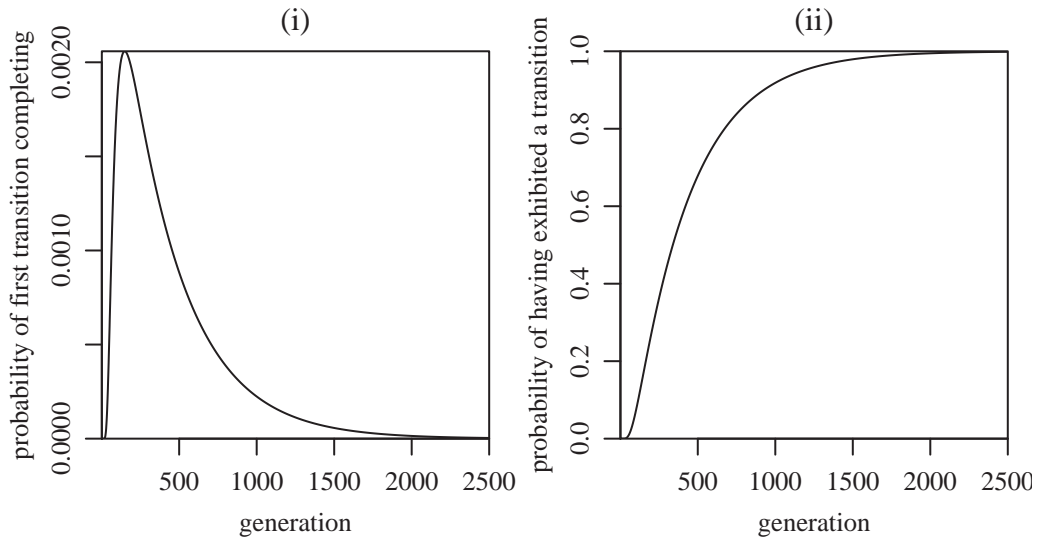
Figure 3.11: Probability of transitions from categorical usage of one to categorical usage of the other variant, for learners accepting the *mean* of the posterior as their hypothesis for θ. *(i)* probability of completing first transition after the given number of generations *(ii)* cumulative probability of having completed at least one transition.

149 generations, while on average the first transition takes 444 iterations to complete.

The distribution of the expected duration of a transition by a chain of learners sampling directly from the posterior distribution is qualitatively similar. Using the same parameter settings as above, the most likely and mean duration until completion of the first transition are 87 and 310 respectively).

Average trajectory of transitions that have the exact same duration

As pointed out above, the number of generations until a new variant has fixated isn't actually representative of the *duration* of a transition. Since chains might remain at their initial state for a few iterations before picking up, or also return back to the initial state before picking up again. If we are interested in the length of the actual transition (i.e. we only start to measure the duration of a transition when the new variant is first innovated) the distribution of transition durations looks quite different, as shown in Figure 3.12.

An immediately obvious difference between this and the earlier distribution of transition durations in Figure 3.11 is that the cumulative probability in subfigure (i) never reaches 1. Under Reali and Griffiths (2010)'s original condition on the final state only, which allowed all possible intermediate trajectories, all chains would eventually reach the target state at some point.

Not so when conditioning on transitions which have to last an exact number of generations: Figure 3.12 only considers transitions that, from their moment of actuation, actually reach the target state without ever 'failing' (i.e. returning back to the categorical initial state) in between. For a chain of learners who take the mean of their posterior distribution as their hypothesised underlying frequency θ, only about 1.46% of initial introductions of a new, competing variant
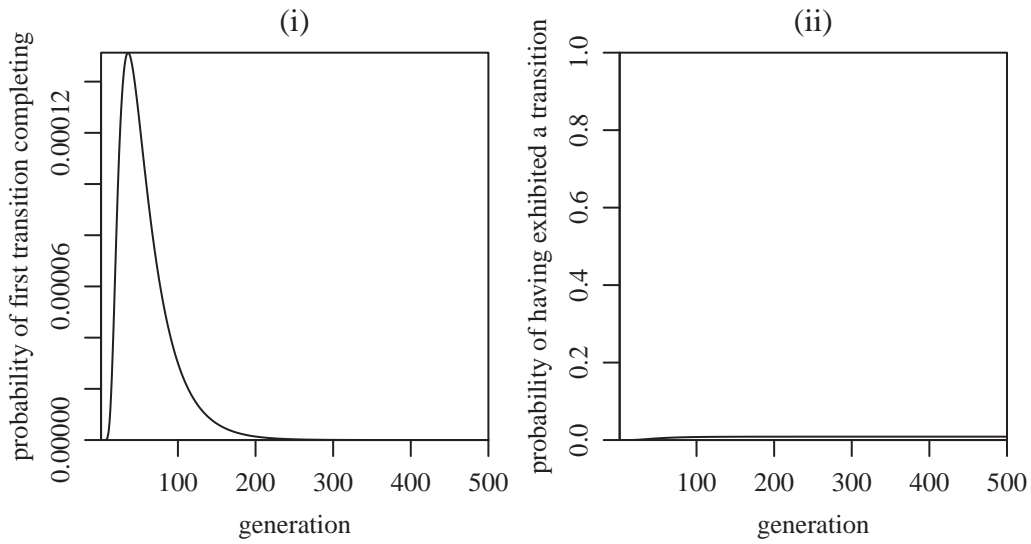
(a) Results with learners sampling from the posterior distribution $p(\theta|x)$.
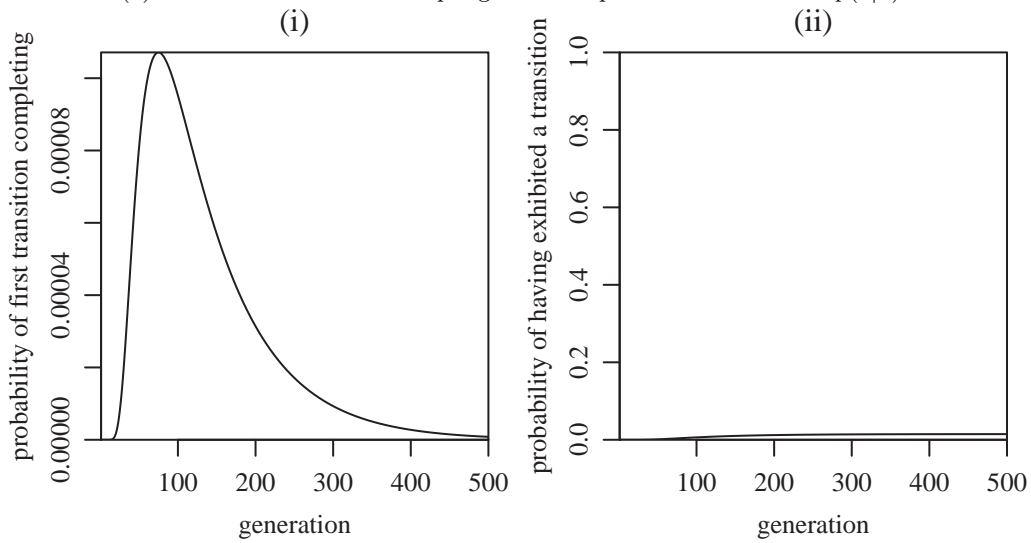
(b) Results with learners accepting the *mean* of the posterior as their hypothesis for θ.

Figure 3.12: Probability of having completed a transition in exactly the number of generations without ever reverting back to the initial state.

actually lead to successful transitions without any interruptions.

In terms of the distribution of durations of those transitions which *are* successful, the number of generations until completion are expectedly much lower than in Figure 3.11 above. For the *averaging* learner, the most like exact duration of a successful transition is much lower at 75 generations, with the mean duration at around 135 generations. For the *sampler* the values are even lower (mode 36, mean 58).

Figure 3.13 shows the state probability distribution as well as average trajectory of the Markov chains which are conditioned on introducing the initially unattested variant in the very first generation, as well as on only tokens of that variant at the maximum number of generations (and no earlier), without ever returning to the initial state. Results are shown for both *sampling* (Figure 3.13a) as well as *averaging* (Figure 3.13b) learners for two different representative durations, the most likely duration of a transition (the mode of the distributions in Figure 3.12) and the (higher) *mean* duration. The Figure shows that the average of all transitions, again indicated by the dashed white line, is actually more like an *S* bent in the 'wrong' direction. In other words, unlike what we find in empirical data on language changes, some of the slowest rates of growth occur at the mid-point of the change, similar to the average transition of chains of *de-regularising* learners shown in Figure 3.9b.

What Figure 3.13 also shows up, however, is that even using this arguably more accurate conditioning on exact start and end points of the transitions as well as on a more realistic time scale, the average trajectory is still not an accurate representation of a typical trajectory. Marked by the white dots is one of the *most likely* individual trajectories that the Markov chain passes through on its way from the initial to the final state. This trajectory is determined using the *Viterbi algorithm* (Jurafsky and Martin, 2008), a dynamic programming algorithm for Hidden Markov Models that allows one to infer the most likely sequence of states given a sequence of observations which only reveal partial information about the likely underlying states. The algorithm can be used for our purposes by providing it with a sequence of observations that indicate categorical usage of one variant at the start and categorical usage of the other at the end of the sequence, with a fixed number of observations representing an unspecified degree of mixed usage in between (the source code as well as a more detailed description of the approach can be found in Appendix B). The sequence of underlying state transitions which has the highest overall likelihood of all possible paths given these observations is one that rapidly crosses the mixed-usage area in 10-15 generations, and remains hovering at near-categorical usage of either variant for the rest of the time. (It should be noted that the exact position of this fast transition along the time axis is irrelevant, in fact all transitions *parallel* to the one indicated by the dots, i.e. ones with the same shape but actuating at earlier or later generations, have the exact same probability of occurring.)

To finish our study of the individual transitions generation by this model, we randomly generate a final set of transitions, limiting ourselves to only those that first complete after *exactly* the specific number of generations, i.e. we exclude ones that reach a frequency of 50 of the incoming variant early and stay there. Three such example transitions can be seen in Figure 3.14.
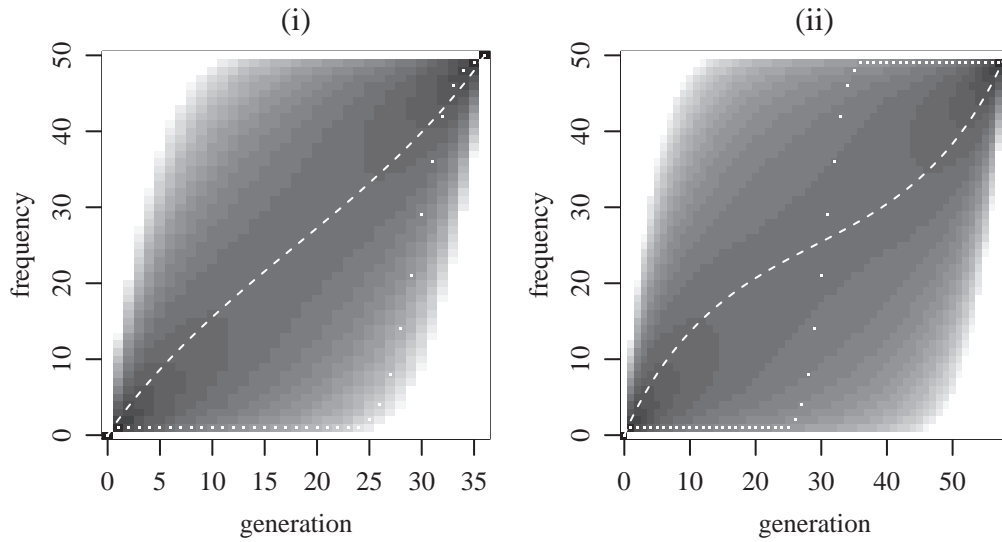
(a) Results with learners sampling from the posterior distribution $p(\theta|x)$, most likely duration and average duration of a transition are 36 and 58 generations, respectively.
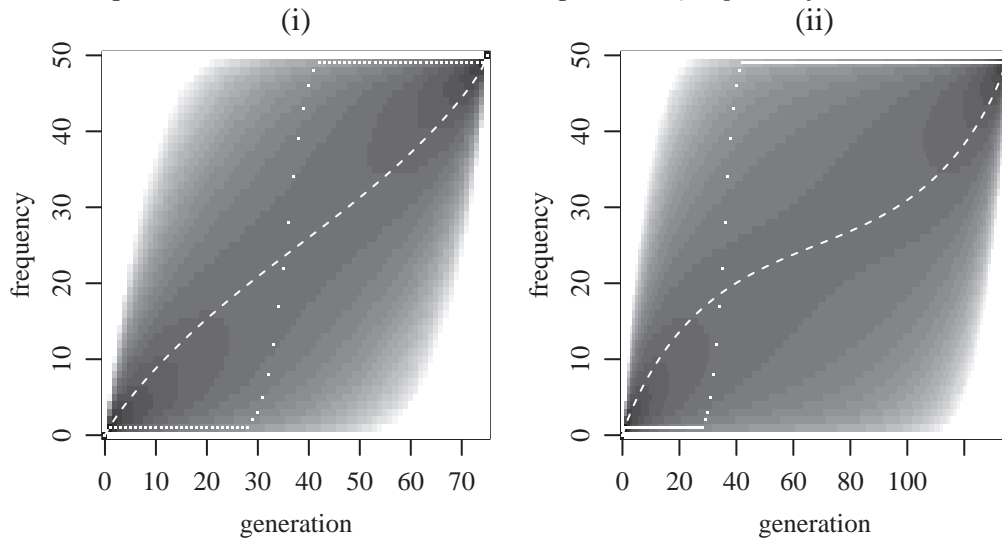
(b) Results with learners accepting the *mean* of the posterior as their hypothesis for θ, most likely duration and average duration of a transition are 75 and 135 generations, respectively.

Figure 3.13: State probability distribution for all Markov chains exhibiting a transition with the exact same duration. The dashed white line shows the *average* trajectory, while the white dots indicate one of the *most likely* transition paths. The duration is set to be equal to *(i)* the most likely duration of a transition and *(ii)* the average duration of all completed transitions respectively, as computed for parameters $N = 50, \alpha = 0.5$.
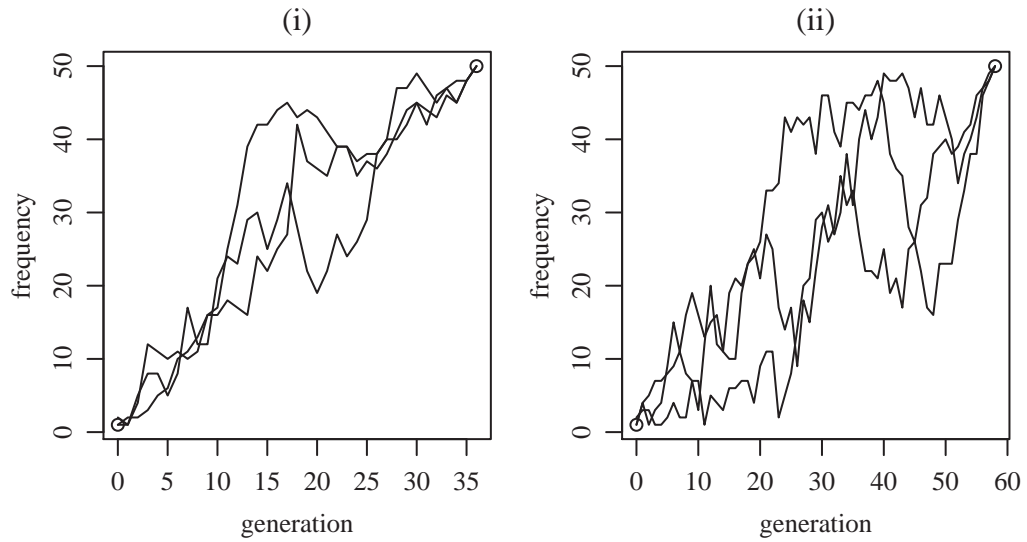
(a) Results with learners sampling from the posterior distribution $p(\theta|x)$.
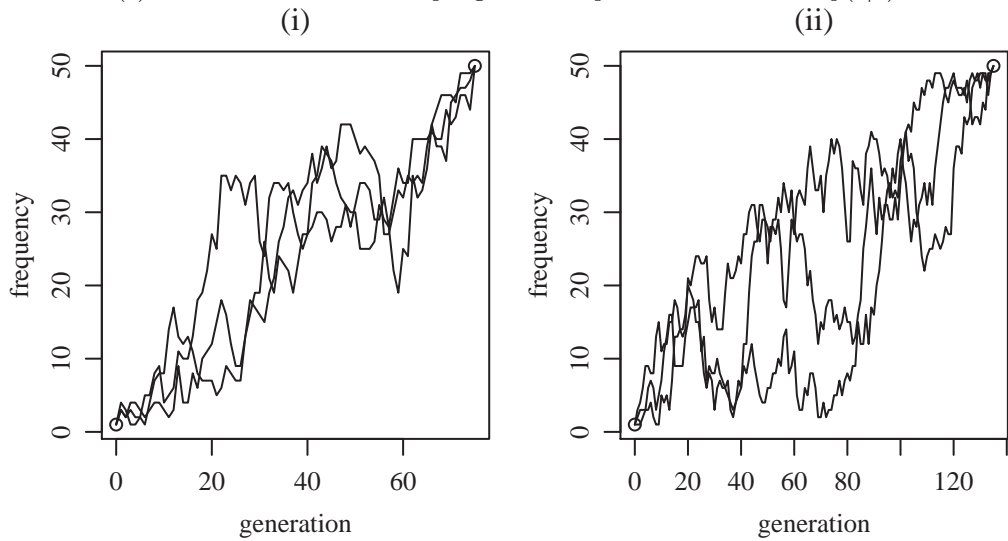
(b) Results with learners accepting the *mean* of the posterior as their hypothesis for θ.

Figure 3.14: Three randomly generated transitions which first exhibit categorical usage of the new variant exactly after the average number of generations it takes a chain to complete a transition. The duration of transitions is equal to *(i)* the most likely and *(ii)* the *average* duration of a transition given the parameter settings ($N = 50, \alpha = 0.5$).

3.4.3   Effect of sample size on the duration of transitions

No matter what the shape of the average trajectory might be, for the sake of cross-validating the general results of the neutral evolution models as implemented here as well as by the Utterance Selection Model, we can compare the two models' predictions regarding how the expected duration of transitions develops as a function of the 'population size'.

While in the Utterance Selection Model the 'population size' refers explicitly to the size of the *speech community* (i.e. it is a measure of the number of interacting individuals), Reali and Griffiths are more implicit about the precise meaning of their model parameter N. In Reali and Griffiths (2010) they show that a chain of learners employing a specific sample size N that accepts the average of the posterior as their hypothesis for the underlying frequency θ is identical to the Wright-Fisher model of neutral evolution with symmetric mutation rates, an equivalence that will be discussed more in-depth in Section 6.3.1. Taking the equivalence of these two models literally would mean that the parameter N in the present model corresponded to the population size of a group of Bayesian learners, each of which uses either of the variants categorically, with the probability of adopting either variant given by $\hat{\theta}$.

Another way to construe the meaning of parameter N corresponds to how it is mapped onto an Iterated Learning experiment on humans in Reali and Griffiths (2009). Here, the model is fit to a chain of single individuals, each of which first receives and then produces a sample of N tokens. While not exactly specifying a feature of the individual, the function that N fulfills in this context is to control the resolution at which the data is presented to and produced by individual participants in the chain. In this sense, the parameter fulfills a function very similar to the T parameter of the Utterance Selection Model described above.

On the other hand, the fact that the model does not allow for continuous updating of the internal representations once they are acquired, but is instead based on a one-time learning event of sample size N, means that the set of possible posterior distributions $p(\theta|x)$, as well as the resolution of possible values of θ for strategies that adopt one value deterministically, is completely constrained by N. As a consequence, the parameter inadvertently acts as something like a *memory capacity* of the individual which, unlike the USM's sample resolution T, also limits the individual agents' representational resolution of the frequency distribution they are trying to acquire.

Whichever way the parameter is to be construed, its setting does not just affect the likelihood of transitions occurring, but also the transitions' duration and shape. The parameter's effect on the average as well as most likely *duration* of completed transitions in chains of learners is shown in Figure 3.15. In all cases, N shows a linear relationship with the time until fixation for all measures with varying slopes, a result that is in line with findings for expected diffusion times obtained from other general models of neutral evolution (Kimura and Ohta, 1969).

3.4.4   Summary

To complement the study of various different replication regimes implement in the USM framework earlier, I presented a replication of Reali and Griffiths's Markov model of neutral evolution with symmetric innovation, a model that has been used to make concrete claims about the possi-
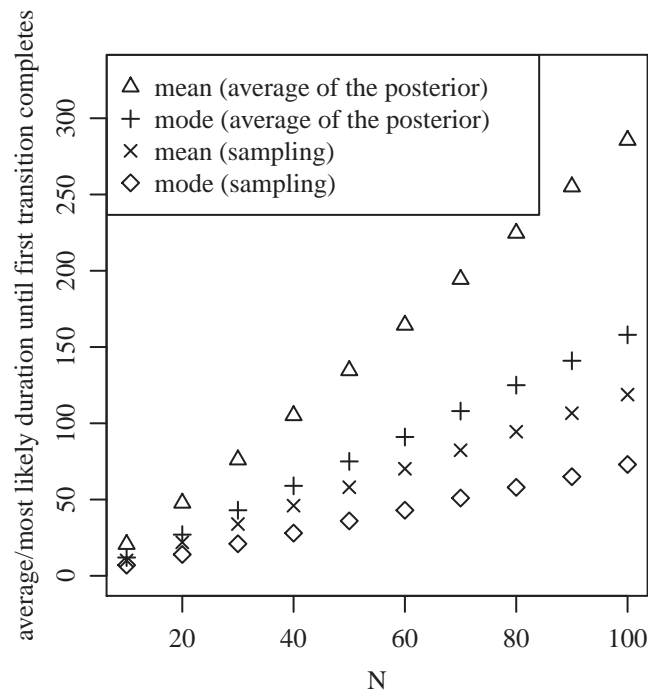
Figure 3.15: Mean and mode of the duration of transitions as a function of the parameter N, with $\alpha = 0.5$.

bility of s-shaped transitions in the absence of asymmetry between variants (Reali and Griffiths, 2010). However, neither the dynamics of individual transitions, nor a closer investigation of the *average* trajectories under different conditioning assumptions suggests that this model of neutral evolution based on regularising Bayesian learners exhibits curves that are particularly directed, instead producing noisy transitions with frequent reversals and restarts. Also, in agreement with other models of neutral evolution, the expected duration of a transition from categorical use of one variant to categorical use of another increases linearly with the population/memory size parameter. Another important conclusion regarding modelling more generally is that, when one is interested in the *temporal dynamics* of a system it is indispensable to look not only at the end states or average dynamics as a shortcut, but that a more exhaustive analysis of the actual dynamics and *typical* transitions is required.

3.5 Trend-amplification and momentum-based selection

The literature summarised above in Chapter 2 as well as the pressures investigated in this chapter cover the bulk of the established accounts and theories about language change. While the direct comparison of different pressures in the Utterance Selection Model just recapitulated indicates that an asymmetry between variants is necessary to account for the directed trajectories found in language change, a survey of the literature shows that there is no universal agreement on where exactly those asymmetries should be found. Adaptive pressures, most of which can be characterised as being language-internal, provide good explanations of the macro-level pat-

terns of language change found cross-linguistically, but they are subject to a methodological flaw known as the *actuation problem*. While general, universal pressures can be invoked to account for universal properties of human languages, they fail to explain why some linguistic features are only selected for occasionally in specific languages, thus leading to the actuation of particular language changes. In other words, none of the selection mechanisms investigated so far offer a theory of how asymmetries could emerge spontaneously and sporadically, as appears to be the case in language change. The remainder of this thesis is dedicated to the exploration of a relatively novel selection pressure that was briefly alluded to previously, namely that of *trend amplification* by the individual. In the spirit of an exhaustive model comparison, the next chapter will investigate the dynamics of such a mechanism, implemented as a *momentum-based* selection pressure within the Utterance Selection Model.

Chapter 4

Modelling momentum-based selection

The contents of this chapter are published in:

4.1 Introduction

Human languages are a prime example of a culturally evolving trait: they are made up of socially learned conventions which are constantly being replicated, and exhibit great diversity across the globe (Evans, 2009). Important aspects of the dynamics of language change are well-understood. Firstly, language change is *sporadic* (de Saussure, 1959; Labov, 2001). Of all the conventions that make up a single language, at any given point most of them are not undergoing change, but are replicated faithfully, from basic word order patterns down to the pronunciation details of individual words. Languages are transmitted robustly over many generations, a necessary requirement for their use as a tool for communication (Lewis and Laland, 2012). Secondly, when a convention *does* change, individuals will gradually replace an established variant with a new variant. This gradual replacement exhibits directed transitions in the form of *s-shaped curves* such as the one shown in Fig. 4.1, akin to the patterns of logistic growth found in biological evolution (Bailey, 1973; Altmann et al., 1983; Kroch, 1989b; Denison, 2003; Blythe and Croft, 2012)[1]. This similarity to the signature of adaptive selection in biology is puzzling (Labov, 2001, ch.1). Linguistic conventions are *arbitrary*, which means we should not expect an inherent advantage in particular linguistic variants, such as which basic word order is used by a language, or how exactly a distinctive phonemic segment is pronounced (as long as it maintains its contrastive function). How and why would an entire population of speakers go about replacing an existing convention with a different one "to say the same thing"?

4.1.1 Language-internal accounts

In order to explain *why* languages change, many studies have attempted to pin down the causes of individual changes by systematically comparing the states of the languages prior to and after a change (Hockett, 1965; McMahon, 1994). While many of the earliest such studies would attribute change to the gradual accumulation of performance and transmission errors alone (e.g. Jespersen, 1922; Hockett, 1958), the generativist paradigm with its focus on the language acquisition device shifted the attention to child-based language change. Studies of language change in the generative tradition trace changes back to the re-ordering or simplification of a language's grammatical rules (Kiparsky, 1968; Wang, 1969; Bailey, 1973; Lass, 1980; Vennemann, 1983),

---

[1]While the notion of 's-shaped curves' is notoriously ill-defined, for the purposes of this paper it will suffice to use Blythe and Croft's definition as any directed trajectory that does not feature "large fluctuations and a tendency for an upward or downward trend to reverse one or more times before an innovative variant goes extinct or wins out" (2012, p.285).
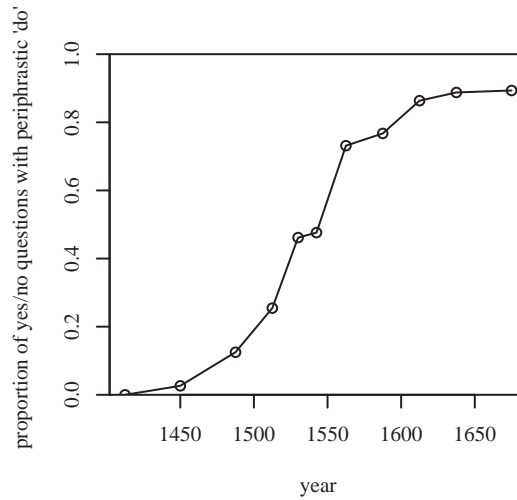
Figure 4.1: Competition between two syntactic patterns of *yes/no questions*, as observed in a corpus of Middle English writing (Ellegård, 1953). The established question syntax (e.g. "Went he?") was gradually replaced by its modern variant (e.g. "Did he go?") along an s-shaped trajectory.

typically assumed to be due to children's reanalysis of linguistic parameters based on their limited linguistic input (see Kroch (2001) and Foulkes and Vihman (2013) for reviews concerning syntactic and phonological change, respectively). Rather than characterising change as the result of imperfect transmission, a more recent strand of research regards language as a *complex adaptive system* which evolves to fulfill the communicative needs of its speakers, while at the same time adapting to the constraints imposed by their learning mechanisms (Kirby, 1999; Steels, 2000; Griffiths and Kalish, 2007; Beckner et al., 2009).

What unites these *language-internal* accounts is that they all rely on a qualitative difference between the language states prior to and after the change. This difference can be based on a variety of factors, such as the languages' expressivity, processing efficiency, or simply their stability with respect to error-prone language acquisition. Within historical and variationist linguistics such explanations of language change have long been criticised on the basis that they *overpredict* change (de Saussure, 1959; Greenberg, 1959; Weinreich et al., 1968; Lass, 1980; Ohala, 1989; Croft, 2000; Labov, 2001; Winter-Froemel, 2008). In their seminal paper, Weinreich et al. succinctly summarised the issue and coined it the *actuation problem*: "Why do changes in a structural feature take place in a particular language at a given time, but not in other languages with the same feature, or in the same language at other times?" (Weinreich et al., 1968, p.102).

In other words, language-internal pressures by themselves do not account for the *sporadicity* of language change: many non-adaptive or suboptimal structures that are claimed to have been selected against in one language will happily persist in other languages – and when they finally do change, language-internal accounts often offer no explanation of what triggered the *actuation* of the change (de Saussure, 1959; Postal, 1968; Ohala, 1993). While language-internal factors

offer good predictions of *which* changes are more likely to occur than others (Jaeger and Tily, 2010; Wedel et al., 2013), they do not explain *when* or *why* the stable transmission of language suddenly caves under functional pressures when it does. To account for the sporadic nature of language change, many have argued that it is not enough to rely on intra-linguistic factors alone.

4.1.2 Social accounts

Sociolinguistic research of the past five decades has shown that innovations do not spread uniformly across a given speech community, but that the progression of change is stratified based on factors such as a speaker's age, ethnicity, or socio-economic status (Foulkes and Docherty, 2006; Tagliamonte, 2012). *Social accounts* hold that social features of linguistic variants, rather than their inherent linguistic character, are responsible for driving language change (Sturtevant, 1947; Croft, 2000; Labov, 2001; Croft, 2006). Social accounts of language change are *evolutionary* in nature: they decouple the generation of *variation* from the process of *selection* which leads to the diffusion of variants through a speech community. The underlying mechanisms, however, are very different from biological evolution. While the generation of new variants is assumed to be driven by linguistic or functional factors, social accounts attribute the ultimate *selection* of variants to extra-linguistic social factors (Ohala, 1989; Croft, 2000; Labov, 2001; Stevens and Harrington, 2013). The 'division of labour' between language-internal and social pressures in this approach can simultaneously account for the arbitrary adoption of one linguistic convention from the pool of variants over another, while at the same time explaining the crosslinguistic distribution of linguistic features which reflect functional pressures.

Recent work on a mathematical model of language change suggests that only the presence of a bias which favours the replication of a newly incoming variant can reliably reproduce the s-shaped transitions observed in language change (Blythe and Croft, 2012). While this mechanism, known as *replicator selection*, is in principle also compatible with language-internal biases, the authors eschew this conclusion. In line with social accounts of language change they conclude instead that it is the *social prestige* of a new variant that is responsible for its preferential replication. Importantly, the sociolinguistic use of the term *prestige* actually refers to a *content bias*: rather than preferentially copying variants used by prestigious individuals, *prestige* is simply another name for a bias that, while social in origin, is actually inherent to the linguistic variant (Sturtevant, 1947; Labov, 2001). Crucially, social accounts do not solve the underlying logical problem of how a population would agree on the selection of a new variant if there is no objective advantage to that variant. The choice of the population to attach preferential prestige to some variant is as arbitrary and requires just as much explanation as the population's increased use of one linguistic variant over another. Because variant prestige is not accounted for within the theory (Meillet, 1926; Labov, 2001) and can only be attributed post-hoc (Sankoff, 1988; Trudgill, 2004; Maegaard et al., 2013), social accounts are typically unable to make a priori predictions about whether particular changes are likely to happen or not. If we saw competing variants as completely identical in terms of both their linguistic *and* social value, how could directed transitions come about? To address this question, it is useful

to consider ideas from the wider domain of cultural evolution.

4.1.3 Replicator-neutral accounts

The evolutionary approach to language variation and change outlined above has also been adopted widely to study processes of cultural change more generally (Boyd and Richerson, 1985; Mesoudi, 2011). Interestingly, even though replicator-neutral accounts – where individuals have no inherent preference for any of the competing variants – have been studied extensively in the context of cultural evolution (Bentley et al., 2004, 2007), such models have received relatively little attention in the study of linguistic change (e.g. Trudgill, 2008; Baxter et al., 2009).

One of the attempts to build a bridge between general models of cultural evolution and the dynamics of language change is Reali and Griffiths (2010). Starting from a model of pure neutral evolution by random copying – where individuals replicate the different variants proportionally to their current frequency – the authors add what they frame as an inferential bias for *regularisation*. They show that the trajectories produced by this 'regularising' neutral model exhibit s-shaped growth, as long as only those trajectories initiated at 0% use of a novel variant and terminating at 100% use are considered. Crucially, however, their mathematical model captures all possible trajectories between those two points, and their result holds only for the *average of all possible trajectories*. This idealised trajectory is highly unlike the 'typical' transitions produced by their neutral evolution model, which are characterised by a noisy trajectory, often with many reversals. The strict symmetry of their Markov model also predicts that there should be as many completed language changes as there are actuated changes that went to the 50% mark before being interrupted, a situation that does not seem to be the case. These considerations call into question whether neutral evolution by random copying can provide an adequate model of the dynamics of language change (Blythe, 2012a).

While in pure neutral evolution models the likelihood of replicating a variant is assumed to be dependent on that variant's current frequency alone, another class of replicator-neutral models that has received increased attention recently considers the effects of *temporal information* and *memory* on the diffusion of cultural traits. Labov (2001) for example suggested that the systematic incrementation of sound changes across generations could be explained by the notion of *age vectors*. He hypothesises that, following an initial stage where learners acquire the average community usage of linguistic variants, adolescents advance their productions in line with the age stratification of variable usage that can be observed in the population – in other words, it presumes that youngsters have a bias against sounding *outdated*. This idea was taken up by Mitchener (2011), who framed it in terms of *prediction-driven instability*: in his mathematical model, individuals are able to observe the usage levels of a categorical sociolinguistic variable among the 'older' and 'younger' individuals in the population. New individuals entering the population then adopt a usage rate according to the predicted future use of the variants, by extrapolating from the usage levels of the two groups along an idealised logistic curve. While the model exhibits spontaneous transitions between the two (or more) competing language states, it produces trajectories that exhibit rapid growth from the onset of the change, unlike the gradual uptake observed in empirical data such as shown in Fig. 4.1.

The individuals' usage rates also remain fixed after they are initally acquired, leaving open the question of whether the same mechanism could also give rise to directed changes when individuals adjust their usage rates throughout their lifetime, as has been observed in linguistic changes (Sankoff and Blondeau, 2007).

Another general model of cultural evolution based on a similar principle is *momentum-based selection* (Gureckis and Goldstone, 2009), which we will study more closely in the remainder of the current analysis. In this model, an individual's choice between competing cultural variants is influenced by the variants' *momentum*, i.e. by *changes to the variants' frequency of use* in the recent past. Individuals are assumed to be biased towards variants which have recently seen an increase in their usage rate, and conversely biased against variants that have been adopted relatively less frequently in the recent past.

Gureckis and Goldstone test their model on a dataset of the frequency of names given to children in the US over 127 years. Their prediction for the popularity of a name in a given year, which is based on the name's long-term popularity modulated by its short-term momentum, leads to a significantly better fit of the empirical data than the prediction made by pure random copying accounts which do not incorporate momentum. Importantly, their model was primarily intended to be fit to empirical data, but not meant as a generative model of individual behaviour. The authors rule this out, noting that "if rising names are preferred, which in turn causes them to rise, then a momentum bias might quickly lead to convergence on a single token" (p.668). They regard this as a negative property of the model, as they are interested in mechanisms that exhibit *cycles* in the popularity of traits, such as found in the realm of fashion (Kroeber, 1919; Berger and Le Mens, 2009; Acerbi et al., 2012). In language, on the other hand, convergence on a single convention is the rule rather than the exception, suggesting that momentum-based selection may be an appropriate model for language change.

4.2   Momentum-based selection

Our main contribution in this work is to investigate the dynamics of momentum-based selection by integrating it into an existing framework of language change, and evaluating it with respect to the characteristics of language change we identified above: the *sporadic* nature of changes which, once actuated, proceed in an orderly, *directed* manner. We begin by reviewing the original formulation of momentum-based selection in Gureckis and Goldstone (2009). The model is built around tracking exponentially weighted moving averages (EWMAs) of the relative frequencies of competing cultural traits over time. Given a time series of relative frequencies $\vec{n} = \langle n_1, n_2, n_3, \dots \rangle$, the weight of each data point towards the moving average, which we denote $\hat{n}_\alpha(t)$, decreases exponentially over time (hence the name). Given a new datum n_t received at time t, the moving average is updated iteratively using

$$\hat{n}_\alpha(t) = \alpha \cdot n_t + (1 - \alpha) \cdot \hat{n}_\alpha(t-1) \,. \tag{4.1}$$

The parameter $\alpha \in [0, 1]$ is a smoothing factor that determines both the weight given to the newest data point, as well as how quickly the data points' weight decreases over time. At

time t, the relative weight of datum n_{t-i} to the current average is $\alpha \cdot (1 - \alpha)^i$. The higher α, the more weight is given to more recent data points. Based on this, the momentum of a variant at time t, $m(t)$, is determined by calculating two EWMAs $\hat{n}_\alpha(t), \hat{n}_\gamma(t)$ of the variant's attested frequencies $\langle n_1 \cdots n_t \rangle$ with two distinct smoothing factors $\gamma > \alpha$, and taking their difference,

$$m(t) = \hat{n}_\gamma(t) - \hat{n}_\alpha(t). \tag{4.2}$$

Because the higher γ gives more weight to recent data points, the moving average $\hat{n}_\gamma(t)$ corresponds to the more recent popularity of a trait while $\hat{n}_\alpha(t)$ captures its long-running popularity. The momentum term $m(t)$ will consequently be positive if a variant has been more popular in the recent past compared to its long-term popularity, and negative if the variant has been adopted relatively less frequently in the recent past.

4.2.1 Mathematical properties of momentum

To understand just what is captured by the momentum term $m(t)$, we can investigate the general dynamics of the difference between two EWMAs $\hat{n}_\alpha(t), \hat{n}_\gamma(t)$ based on their parameters $\gamma > \alpha$. The strongest possible trend in changes to relative variant frequency can be achieved by initialising both EWMAs so that they indicate categorical usage of, say, the outgoing variant (i.e. $\hat{n}_\alpha(0) = \hat{n}_\gamma(0) = 0$), and then continuously updating both EWMAs with input data suggesting that, actually, everyone is using the novel, incoming variant categorically (i.e. $\vec{n} = \langle 1, 1, 1, \ldots \rangle$). Even in this simple case, the dynamics of the momentum term are complex, as can be seen in Fig. 4.2.

For the underlying EWMAs themselves, the higher the smoothing factor, the faster they approach the input values (Fig. 4.2a.i), and the more quickly they reflect changes in the distribution too (Fig. 4.2a.ii). The corresponding momentum terms that are derived by subtracting an EWMA with a high parameter γ from a more slowly changing one with a lower parameter α are shown directly underneath (Fig. 4.2b). What is of interest to us are the different *shapes* of these momentum curves: a parameter combination which exhibits a rapidly rising curve will cause an individual to posit a trend based on just a few suggestive input data points, while a curve that slopes off slowly means that a momentum bias will persist for a longer time after the initial detection of a trend.

Both the number of data points it takes to reach their maximum value as well as the amplitude of this highest possible momentum value depend on both smoothing factors in complex ways. The short-term memory parameter γ is of particular importance, as it controls the time depth at which the momentum term is most sensitive to underlying trends in the data: a high γ causes the momentum term to immediately reflect short-term variation in the input, while settings of γ closer to α lead to more conservative trend estimates which smooth over the noise present in individual input data points.

The sudden change in trend after 60 data points shown in Fig. 4.2b.ii illustrates this point: a momentum term based on high $\gamma = 0.15$ (dotted line), while very quick to reflect sudden changes in the input, is very unstable. After receiving only five data points of the new input value $n_t = 0$, the previous sustained upward trend is 'forgotten', with the momentum term

first quickly returning to 0, then going negative to reflect the new, short-term downwards trend from the series of 1s back to 0s.

Generally, assuming an abrupt change in the input values such as above, the number of iterations that both EWMAs have to be updated with the same constant input value before the maximum possible difference between the two is reached is

$$t_{\mathrm{mmax}}(\alpha, \gamma) = \frac{\ln \frac{\alpha}{\gamma}}{\alpha - \gamma} \ . \tag{4.3}$$

The maximum possible amplitude of the momentum term at that point is

$$m_{\max}(\alpha, \gamma) = e^{-\gamma t_{\mathrm{mmax}}(\alpha, \gamma)} - e^{-\alpha t_{\mathrm{mmax}}(\alpha, \gamma)} \ . \tag{4.4}$$

Knowing the mathematical boundaries of the momentum term $m(t)$ we can now go on to incorporate the momentum bias into a model of language change.

4.2.2 The Utterance Selection Model of language change

To investigate the dynamics of momentum-based selection as a model of individual behaviour, we implemented the momentum-based selection bias in the *utterance selection model* of language change (USM, Baxter et al., 2006; Blythe and Croft, 2012). Derived from Croft's evolutionary theory of language change (2000), the USM provides a well-studied multi-agent framework to study the dynamics of the competition and diffusion of *discrete* linguistic replicators, be they lexical items, constructions, or different categorical variants of a speech sound[2].

Two fundamental principles underlie the design of the USM: firstly, the individual agents use the competing variants *proportionally*, rather than categorically. In the minimal case with only two competing variants studied here, an agent's usage rates can be fully described by a single number, call it x, in the range $[0, 1]$. While this value can be interpreted as reflecting some cognitive state of the speaker, it also has a more direct behavioural correspondent: when an agent is selected to participate in an interaction, their probability of producing the novel variant is equal to x, while the probability of producing the competing variant is $1 - x$. This aspect of the USM is in line with linguistic evidence which shows that human language use is inherently variable and probabilistic (Kroch, 1994; Labov, 1994; Bybee, 2007; Nardy et al., 2013).

Secondly, agents continuously tune their own variable usage rate towards the production rates they observe in interactions with other agents, thus mimicking the human tendency to *align* linguistic behaviour with that of interlocutors (Giles et al., 1991; Branigan et al., 2000; Jaeger and Snider, 2013). This aspect of the USM is in line with the finding that many aspects of linguistic behaviour do not remain fixed, instead remaining malleable across an individual's life span (Kerswill, 1996; Sankoff and Blondeau, 2007; Beckner et al., 2009; Bowie and Yaeger-Dror, 2013; Stanford, 2014). According to the formal definition of the USM (Baxter et al., 2006), an agent's current proportion of use of a variant $x_{\alpha}(t)$, is simply an exponentially weighted moving

---

[2]For an account of how age vectors can drive change in a continuous dimension such as vowel productions, see Swarup and McCarthy (2012).
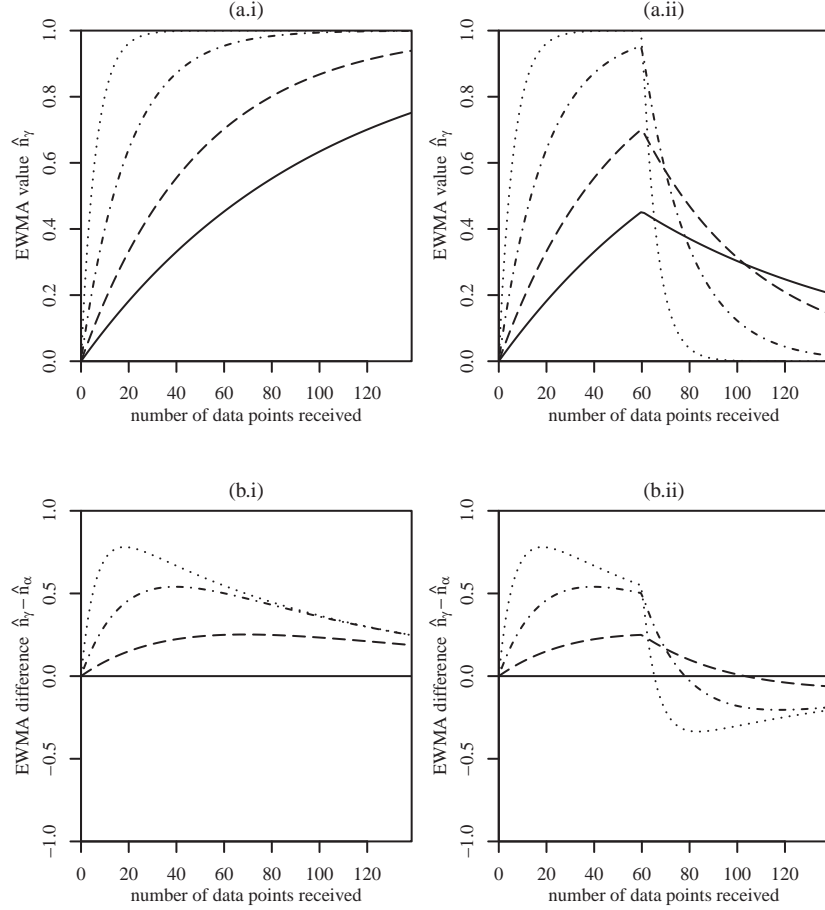
Figure 4.2: Exponentially weighted moving averages (EWMAs) of the same input data but with different smoothing factors, as well as their corresponding momentum terms. *(a.i)* Four EWMAs with smoothing factors $\gamma = 0.15, 0.05, 0.02, 0.01$ (from top to bottom) are initialised at $\hat{n}_\gamma(0) = 0$ and repeatedly updated using the same constant input data series $\vec{n} = \langle 1, 1, 1 \ldots \rangle$. *(a.ii)* same as (a.i), but with the input data series $\vec{n}$ switching from all 1s to all 0s after 60 data points. *(b)* Corresponding momentum terms $m(t) = \hat{n}_\gamma(t) - \hat{n}_\alpha(t)$ derived from the trajectories above, by taking each EWMA and subtracting the value of the EWMA with the lowest smoothing factor from above ($\alpha = 0.01$). Line styles correspond to those in (a).

average (EWMA) of the frequencies of the incoming variant that the agent has observed in their input over time[3]. The rate of alignment is controlled by the smoothing factor α of this EWMA, which can be understood as a *learning rate*. This learning rate is typically held small (in the range of 0.01): there is alignment, but the individual frequency adjustments after an interaction are very small and it takes many interactions for an agent to change their preferred variant.

On top of this basic update rule, a USM agent's alignment behaviour can be altered by applying biases to their input data before it gets incorporated into the EWMA. This is where momentum-based selection comes into play.

4.2.3 Momentum-based selection in the USM

We now explain how to minimally incorporate momentum-based selection into the USM. Assuming an agent using learning rate α has just engaged in its t-th interaction and observed another agent use the incoming variant with a relative frequency of y, then their own frequency of use x_α is updated to be

$$x_\alpha(t) = \alpha \cdot f(y) + (1-\alpha) \cdot x_\alpha(t-1) \ , \tag{4.5}$$

where $f(y)$ is a function from $[0,1]$ to $[0,1]$ which transforms the *objective* observed frequency y of the variant into a subjective *perceived frequency* which the agent then aligns to. Similar to Gureckis and Goldstone (2009) we can now simply define the perceived frequency $f(y)$ of an agent in the momentum-based USM as the objective frequency y of a variant observed in an interaction offset by that variant's momentum,

$$f(y) = y + b \cdot m'(t) \tag{4.6}$$

with the exception of

$$f(0) = 0 \quad \text{and} \quad f(1) = 1 \ . \tag{4.7}$$

We impose the latter since our focus lies on modelling the diffusion of existing linguistic variants, independent of how those variants were introduced into the population to begin with. It simply stops our momentum-biased selection function $f(y)$ from introducing novel, unattested variants, a constraint that is typical of models of selection generally (see e.g. Boyd and Richerson, 1985). The positive bias parameter b in equation 4.6 controls the strength with which the normalised momentum term $m'(t)$ as defined below in Equation 4.8 influences the perceived frequency. Should the momentum bias cause $f(y)$ to go below 0 or above 1, it is simply truncated at 0 and 1, respectively[4]. Crucially, because the momentum term can be positive or negative (depending on the direction of the trend), this perceived frequency function is *symmetric*, which makes it *replicator-neutral*: no matter which bias strength b is used, the function does not a priori favour one of the variants over the other.

---

[3]For simplicity of notation we will henceforth omit theˆabove the variables denoting EWMAs.

[4]The exact form of the bias function $f(x)$ matters much less than its monotonicity and the fact that $f(x) > x$ when the momentum term is positive (i.e. when the agent perceives an upward trend) and $f(x) < x$ when it is negative (indicating a downward trend).

Since the effect of different strengths of this bias parameter b on the model dynamics is relevant to our analysis, we have to make sure that its settings are comparable across settings of the other parameters. This isn't as straightforward as it might seem, because the range of values that the original momentum term definition $m(t)$ in Equation 4.2 can take on depends on both smoothing factors α and γ, as could be seen in Fig. 4.2. The absolute amplitude of the momentum curves is of little interest to us; on the contrary, the differences in maximum possible amplitude distort the effect of the bias parameter b which is supposed to control the strength with which momentum is applied. To counteract this, we normalise the momentum term $m(t)$ based on the α, γ used in a given simulation condition. For any given pair of smoothing factors α, γ, we can scale the momentum term to the $[-1, 1]$ range by defining the normalised momentum

$$m'(t) = \frac{x_\gamma(t) - x_\alpha(t)}{m_{\max}(\alpha, \gamma)} \ . \tag{4.8}$$

To calculate the momentum component in the numerator, the difference between two EW-MAs, we simply re-use the agent's own usage frequency, which according to the USM definition is also an EWMA. To augment the basic USM with momentum-based selection, every agent simply has to keep track of another x_γ on top of the long-term estimate x_α it already maintains.

4.3 Results

4.3.1 Analytical approximation

Before proceeding to a full population-based simulation we can establish the general dynamics of the model by investigating the behaviour of an individual agent set in an idealised, deterministic production-perception loop (Wedel, 2006). We initialise a single agent to use the incoming variant at some low level and repeatedly update their two EWMAs $x_\alpha(t), x_\gamma(t)$ by having them align to their own proportion of use $x_\alpha(t)$ for 100 iterations. As can be seen in Fig. 4.3, nothing happens: an agent aligning to their own usage rate simply remains at that proportion and, in the absence of any changes in the input sequence, the momentum term stays 0. To test how the model reacts to fluctuations in the input we alter the agent's input by fabricating a data point which suggests that their interlocutors are actually categorically using the incoming variant (see Fig. 4.3a). When the agent aligns to this input it leads to a small punctual increase in their variant use, but the sudden change in the input data also makes the momentum term take on a positive value (dashed grey line). Following the fabricated data point, the agent again receives their own samples as input data. But the bias exerted by the momentum term, which makes the agent's *perceived* usage rate higher than their actual usage rate, causes further increases in their use of the incoming variant. However, the lack of further perturbations causes the momentum to decay back towards 0, and the agent becomes stationary again at a usage level not far from their initial setting. If we introduce a second fabricated data point shortly after the first one, the model's behaviour changes dramatically: the system enters a regime where the momentum bias generated by the two fabricated data points affects the perceived frequency of the agent's input so much that it causes the momentum term to increase even further, leading
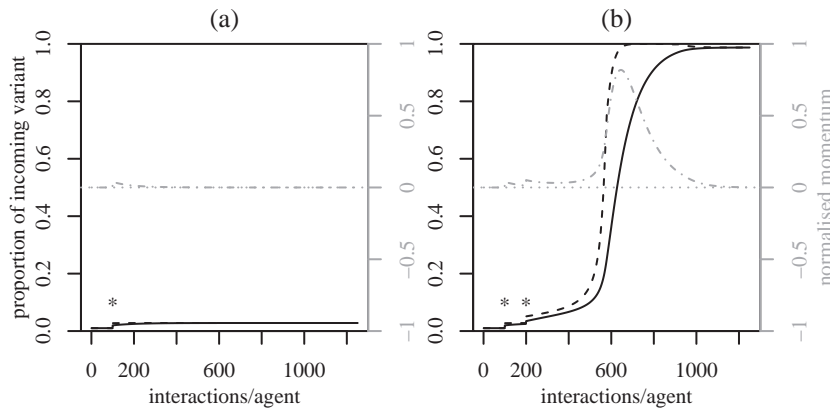
Figure 4.3: Momentum-based selection dynamics of a single agent's variable usage rate in a deterministic production-perception loop, with learning rates $\alpha = 0.01, \gamma = 0.02$ and momentum bias $b = 2$. At every time step the agent updates their own usage rate (solid black line) by aligning to their own average momentum-biased production with a sample resolution of $T = 5$ (indicated by the dashed black line). This stable loop is perturbed by administering fabricated input data suggesting 100% usage of the incoming variant at the time points marked by asterisks, demonstrating the two regimes of momentum-based selection: *(a)* stability: a single fabricated data point after 100 interactions causes a sudden increase in the agent's usage rate (solid black line) as well as the momentum term (dot-dashed grey line, right axis), but the feedback loop stabilises again. *(b)* directed transitions: adding another fabricated data point after 200 interactions raises the momentum term high enough to trigger self-reinforcing runaway change, giving rise to an s-shaped transition.

to self-reinforcing runaway change (Fig. 4.3b).

This preliminary analysis shows that the momentum-based selection model exhibits two different regimes, accounting for both periods of stability and of directed change. Capturing the dynamics of the transition between the two regimes is however not trivial: particularly the switch from stability to a directed transition depends crucially on both the strength of the momentum bias as well as random fluctuations in the agents' input as they sample data from their interlocutors. We therefore turn to numerical simulations, where the data production and agent interactions will be driven by stochastic processes.

4.3.2 Numerical simulation

In order to get a fuller picture of the momentum-based selection dynamics we ran simulations with a total of 2,520 parameter combinations[5]. The six parameters of the momentum-based USM are summarised below. Only one, the learning rate α, was held constant across all simulation runs, the other five parameters were varied at the levels given in parentheses:

- α: the agents' learning rate (0.01)

- γ: the agents' short-term memory smoothing factor $(0.015, 0.02, 0.025, 0.03, 0.35, 0.4)$

---

[5]The source code for running the simulations as well as the analytical approximation are available at `http://github.com/kevinstadler/momentum`

- T: the Binomial sample size determining the resolution at which agents can observe each other's relative usage frequencies $(2, 3, 4, 5)$

- b: the bias strength with which agents apply the normalised momentum to yield their *perceived* frequency of usage $(0.5, 1.0, 1.5, 2.0, 2.5)$

- N: number of agents in the population $(2, 5, 10, 20, 30, 50, 100)$

- x_0: initial proportion of the incoming variant used by all agents $(0.01, 0.02, 0.03)$

Combining all these possible parameter combinations and running the 2,520 conditions for 48 trials each resulted in a total of 120,960 simulation runs. On top of the conditions listed above, we also produced simulation runs where we set the bias strength $b = 0$, which is equivalent to pure neutral evolution. 24,192 runs from this additional condition provide a baseline that the dynamics of our momentum-based selection model can be compared against. Every simulation run proceeds as follows:

Firstly, initialise N agents, setting both their x_α and x_γ to x_0. Then, carry out interactions between agents by repeating the following steps:

1. randomly select two agents i, j from the pool of N agents – we assume that all pairs of agents have the same probability of interacting with each other.

2. let both agents produce T tokens of the variable by taking a random sample n_i, n_j for each agent from the Binomial distribution $B(T, x_\alpha)$, using the two agents' respective usage rates x_α at the time of the interaction.

3. calculate the perceived frequencies that the agents will align to, using equation 4.6. For agent i, who will align to j's productions, calculate $f(\frac{n_j}{T})$ using agent i's current normalised momentum term $m'(t)$; for agent j, calculate $f(\frac{n_i}{T})$ using j's $m'(t)$.

4. update both agents' x_α as well as x_γ by incorporating their perceived frequency according to equation 4.5.

The simulations were run until every individual in the population had converged to within a ten-thousandth of a percent of using only one of the two competing variants, or for a maximum of 200,000 interactions per agent[6].

4.3.3 Simulation results

For the sake of our analysis we use a simple definition of what a 'transition' is. Taking a fixed threshold (say 5%), we can define the two extreme areas where the mean population usage level of the minority variant is below this threshold as the two regions of 'near-categorical use' of either variant. A transition, then, is the period in which the mean usage levels of the population crosses from near-categorical use of one to near-categorical use of the other variant. A first striking finding when analysing the simulation results is that changes are rare: of the

---

[6]More than 99% of simulation runs had terminated before this time limit was reached.
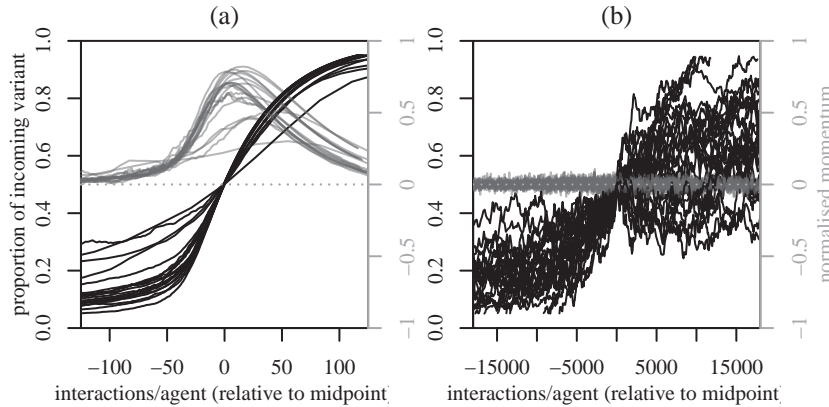
Figure 4.4: Successful transitions generated by simulation runs in conditions with and without momentum-based selection. The graphs show the development of the average proportion of use of the incoming variant across the population (black line, left axis) from the point where it crosses the 5% mark until it reaches 95%, alongside the average momentum term during that period (grey line, right axis). Transitions are aligned at the point where the trajectory first crosses the 50% mark of incoming variant usage. *(a)* 20 trajectories randomly drawn from the 21,909 successful transitions generated by momentum-based selection with momentum bias $b \geq 1$, population sizes $N \geq 5$ and various settings of γ, T, x_0. *(b)* all 28 transitions generated in 17,280 simulation runs with $b = 0$, equivalent to neutral evolution, with various settings of γ, T, x_0 and population sizes $N \geq 5$. Note the different time scales. The momentum term, ineffective when $b = 0$, is shown for reference.

120,960 simulation runs using the momentum bias, only 18,040 (around 15%) ever exhibit a transition, while the majority of runs simply converge on categorical use of the majority variant. This result is in line with the observation that the actuation of language change is *sporadic*: even when a novel variant is known to the entire population, this alone is not likely to lead to a community-wide language change.

When we investigate the distribution of transitions across the different parameter settings, we find that the bias strength b carves the space into two regions with distinct dynamics: while simulation runs with $b \geq 1$ exhibit directed transitions at comparable time scales, the neutral evolution condition with $b = 0$ as well as the weak momentum bias setting at $b = 0.5$ yield both fewer and temporally less consistent transitions, as shown in Fig. 4.4. The difference between those two regimes is exacerbated as population sizes become larger, making transitions in the neutral evolution conditions even rarer and slower.

Beyond this qualitative difference in successful transitions, our earlier prediction regarding the general directedness of trajectories in the neutral evolution condition are also borne out by the results: of all simulation runs where the incoming variant ever reaches the half-way mark (i.e. average 50% usage of both variants across the population), only 55% of trajectories in conditions with $b \leq 0.5$ actually result in the diffusion of the incoming variant. The remaining half-completed transitions are interrupted and revert back to majority usage of the established variant. In contrast, in conditions with $b \geq 1.0$, 97% of the trajectories that reach the half-way mark also lead to the population-wide adoption of the incoming variant.

In contrast to the low-bias conditions which exhibit the dynamics of neutral evolution, conditions with a sufficiently high momentum bias b reliably produce s-shaped transitions between the two regions of near-categorical use at irregular intervals. before eventually converging on categorical use of either of the variants. The dynamics are robust under many different parameter settings which give rise to highly similar transition dynamics (see Fig. 4.4; the parameters' much greater influence on the likelihood of transitions occurring is beyond the scope of this paper). While similar transitions are also found in models driven by replicator selection, an important difference is that our model has no a priori preference for any of the variants built in. Instead of having a constant bias applied from outwith the model, the momentum term provides the opportunity for a bias to emerge dynamically from within the system, as can be seen from the temporal development of the momentum term in Figs. 4.5. Crucially, rather than relying on an external trigger, the s-shaped transitions are *self-actuating*: agents constantly read weak trends into the random fluctuations in their input but these temporary individual biases will vary across the population, and more often than not cancel each other out. There is, however, always the possibility for these weak biases to overlap, which could cause a subset of agents to slowly shift their variant use in parallel. When this shift is detected by other agents they will themselves start to amplify it, leading to a self-reinforcing feedback loop. The directed transitions in a momentum-based model of language change are triggered *spontaneously* and, while it is the most likely outcome, changes are not guaranteed to succeed either: even if a change is actuated, its propagation is not completely inevitable, as can be seen in interrupted changes such as the one shown in Fig. 4.5b. The dynamics of momentum-based selection provide an intriguing account of the unpredictability of the actuation of linguistic changes without the need for an external bias or trigger.

The trajectories shown in Figs. 4.5 are exemplary of the dynamics of momentum-based selection across the full range of parameter settings we explored. Only for settings of the momentum bias b close to 0 as well as for short-term smoothing factors γ very close to the learning rate α do the momentum-based selection dynamics break down, and the model reverts to pure neutral evolution-like behaviour. In comparison to the prediction-driven model of Mitchener (2011), the momentum-based selection model shows that it is not necessary for learners to engage in active prediction of the population's *future* state along a particular trajectory. Rather, having a simple bias based on variant history is sufficient to drive orderly directed changes, and the transitions generated by our model appear to exhibit a more gradual uptake than the trajectories reported by Mitchener. We also find that having a bias for *regularisation* is not absolutely necessary to guarantee an orderly progression of the changes. In a population of agents who are continuously updating their usage rates, the momentum bias presented here is robust enough to drive changes to near-completion.

4.4 Discussion

We have shown that the momentum-based selection model fulfills two defining requirements of a model of language change: the spontaneous, sporadic actuation of changes, and their progression in the form of a directed, s-shaped curve. However, other accounts of language change which
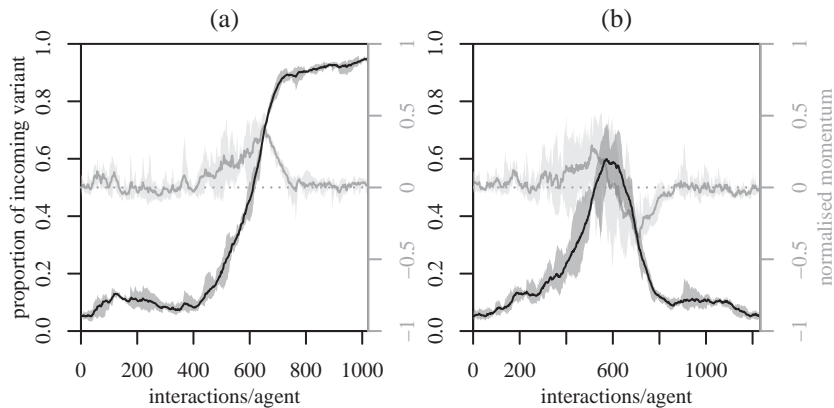
Figure 4.5: Transitions generated by two simulation runs using identical parameter settings ($N = 5, b = 2.0, T = 2, \alpha = .01, \gamma = .04$). The graphs show the development of the average proportion of use of the incoming variant across the population (black line, left axis) as well as the average momentum term influencing the agents' perception (grey line, right axis). Shaded intervals indicate the range (minimum and maximum values) attested in the population. *(a)* A successful, s-shaped transition typical of momentum-based selection: an initially noisy momentum value rises high enough to trigger self-reinforcement of the momentum bias (at around 450 interactions) until it saturates and tails off again *(b)* Example of a rare, interrupted transition: despite the onset of a directed shift, the wide range of momentum biases across the population destabilises the feedback loop, causing the average momentum to break down and invert, returning the usage frequency of the incoming variant back towards its initial low level.

posit a selection bias in favour of the incoming variant also predict s-shaped trajectories, so how can we know which account best describes the empirical data? While the progression of every instance of language change will be influenced by several factors concurrently or at different times (see e.g. Ghanbarnejad et al., 2014; Stanford, 2014), it is still interesting to investigate which (if any) of the mechanisms of language change discussed in the introduction can be identified as the main driving force behind language change. Here, we want to highlight some of the more subtle differences in the predictions made by different accounts of language change which would allow us to tease apart the momentum-based, language-internal and social accounts of language change based on cross-linguistic data.

4.4.1   The two rates of linguistic change

An interesting (and to our knowledge novel) way to evaluate competing theories of language change is to look at the predictions they make regarding the *rates* of linguistic change. It is important to note that 'rate' can refer to two different things in the context of language change: one interpretation of 'rate' refers to the probability of a particular change occurring, such as when talking about different English past tense forms becoming regularised over time (Lieberman et al., 2007) or the rate of lexical replacement more generally (Sankoff, 1970; Wichmann and Holman, 2009; Greenhill et al., 2010; Monaghan, 2014). Rather than referring to the time frame within which a specific change takes place, this really describes the *likelihood of a (type*

of) change, or an *actuation probability* (Cowgill, 1963). The other use of 'rate' refers to the *speed* of the transition of one particular change, i.e. it is a measure of the time span from the introduction of a new variant to its completely replacing an established one. Under the assumption that language change follows an s-shaped pattern, this second rate of change is often taken to be the growth rate parameter of the logistic function (Pintzuk, 2003), and it is this 'rate' that is referred to by the 'Constant Rate Effect' observed in syntactic change (Kroch, 1989b).

What is interesting about these two rates of change is that different accounts of language change make implicit predictions regarding the relationship between them, in particular whether the likelihood of a change occurring is correlated with the rate at which the change proceeds once it has been actuated. Under the assumption that the same pressures that lead to the introduction of more functional or 'adaptive' variants are also responsible for their preferred selection once they have been innovated, language-internal accounts would predict that changes which occur more often cross-linguistically should also be selected for more strongly in individual languages. This would translate into faster changes so that, controlling for other factors such as frequency and size of the speech community, the two rates of change should be positively correlated. This differs from the prediction made by the momentum-based account: while the probability of a new variant appearing, and consequently its random actuation from the pool of variants, is dependent on linguistic factors, these factors are not what drives the diffusion of the variant. Assuming that individuals apply similar momentum biases to all linguistic variables, a momentum-based account would therefore predict the speed of individual transitions and the changes' actuation probability to be uncorrelated.

The situation with social accounts is trickier: the fact that many different social factors have been posited to influence the selection of linguistic variants, both positively and negatively, makes it difficult to derive a general prediction regarding the speed of individual changes. What determines the probability of actuation is an equally open question: it has been proposed that the occurrence of changes might be driven by the need to create distinct social identities within a community (Labov, 2002; Matthews et al., 2012; Roberts, 2013), implying that we should not expect actuation probabilities to be constant cross-linguistically.

While it is difficult to derive specific predictions regarding the correlation between the two rates of change from social accounts of language change, many insights into the respective roles of the different pressures could be gleaned from studying cross-linguistic datasets of changes (see also Bickel, 2015). The crucial issue is that the three qualitatively very different accounts discussed here might predict quantitatively similar selection pressures for particular language changes, making it impossible to distinguish the contribution of the different types of pressures on a *per-change* basis. Our understanding of the issue could therefore profit immensely from investigating the empirical distribution of *both* rates of change as well as their relationship based on cross-linguistic data.

4.4.2 Momentum-sensitivity in the individual

While momentum-based selection successfully reproduces the macro-level s-shaped curves that are characteristic of linguistic change, this raises the question of whether the model makes valid

assumptions about individuals' micro-level behaviour (Mesoudi and Lycett, 2009). Firstly, it is clear that both linguistic knowledge and performance are embedded in diachrony – language users are sensitive to changes in the frequencies of variants (Jaeger and Snider, 2013) and well aware of diachronic connotations (Labov, 2001; Guy, 2003; Tagliamonte, 2012), both types of information that could drive momentum-based selection. In the general cultural evolution literature it is well-established that frequency-dependent biases are a natural strategy for social learning tasks, since frequency can be an indicator of the *social value* of a variant (Boyd and Richerson, 1985). Similarly, *changes* in frequency can be a good indicator of the *future* social value of a cultural variant (Gureckis and Goldstone, 2009). Laboratory experiments on cultural evolution in humans have provided empirical evidence for the self-perpetuating nature of trends, where people will amplify trends even against their own personal preferences (Salganik and Watts, 2008; Willer et al., 2009), suggesting that individuals might also have an incentive to use metalinguistic information about the history of linguistic variants. While there is plenty of qualitative and anecdotal evidence on speakers' explicit evaluation of language changes (see e.g. Trudgill 1972; Labov 2001; Guy 2003; Tagliamonte 2012), quantitative research on the extent of people's explicit or implicit knowledge about the direction of ongoing changes is just starting. Experimental evidence shows that listeners employ their implicit knowledge about ongoing sound changes during speech perception (Hay et al., 2006; Drager, 2011), and there is evidence of explicit knowledge both in the area of phonetic (Carrera-Sabaté, 2014) and syntactic change (Chapter 5).

4.5   Conclusion

In this paper we investigated the *momentum-based selection* model and studied its evolutionary dynamics. Our analysis shows that this model, where individuals are biased towards variants which have recently seen an increase in their frequency of use, exhibits two features characteristic of language change: the spontaneous, sporadic actuation of changes, and their progression in the form of directed, s-shaped curves.

   Crucially, the momentum-based selection mechanism demonstrates that the apparent selection of a particular cultural variant in a population is not sufficient evidence for any inherent asymmetry between the variants in competition. Instead, selection biases can be an emergent property of the system, particularly in the case of social learning where individuals possess meta-level knowledge about the variants. The human capacity to acquire meta-linguistic knowledge about ongoing language changes, for example by tracking changes in the variants' frequencies of use over time, therefore deserves further study.

   Finally, we highlighted the importance of collecting and studying cross-linguistic data sets of comparable historical changes to test the general predictions made by different accounts of language change. This strand of research in particular needs to be expanded further in order to help us gain deeper insights into the respective roles of the myriad pressures involved in language change.

Chapter 5

Probing momentum-awareness in the individual

5.1   Introduction

While the degree to which the historical development of languages is inferred and used by language learners has long been of interest to sociolinguists (Labov, 1989), empirically this question has only been tackled relatively recently as part of a general effort to study the acquisition of sociolinguistic knowledge by individuals (Labov, 2014; Foulkes and Hay, 2015). Of particular relevance to this thesis is the question of how specific linguistic variants and their relative usage levels can come to be associated with specific age groups. The concept of 'age vectors' captures the age-based stratification of variable usage, and the idea that individual language users possess knowledge of the age vectors of their community has been invoked as an explanation of how language changes are transmitted and increment across generations (Labov, 2001). The model presented in the previous chapter demonstrated how a similar mechanism, based on tracking changes to frequency distributions of discrete variants in *real time*, can equally account for spontaneous directed transitions of change in a speech community.

While there is an increasing body of empirical evidence on individuals' knowledge of ongoing changes which I reviewed in Section 2.2.5, the experimental data presented there was limited to continuous phonetic changes. Although the fact that this sub-domain of language change still encompasses the largest part of sociolinguistic research might in part be attributed to the generativist sovereignty over morphosyntactic research which did not leave much space for an empiricist-variationist methodology, it is worth noting that the type of sociolinguistic variable (i.e. whether different variants differ continuously or categorically) influences not just how the linguist might describe or represent variation, but also how that variation is acquired by individuals.

In particular, the type of variable impacts on the amount of information on inter-individual differences that can be extracted from individual realisations of a variable that is observed in an interaction. While continuous phonetic tokens are potentially very noisy and it might therefore help to have access to a speaker's full distribution of realisations to get a complete picture of their variable usage, given sufficient stratification of variable realisations along a continuous dimension even a single token can potentially contain enough information to place the speaker along a cline from 'more outdated' to 'more modern' in their variant usage. Not only that, continuous variants allow speakers to sound '*even more* novel' by extending the change along that cline and producing variants that 'overshoot' even the most advanced novel variants, productions which can nevertheless be immediately understood as instantiations of the same innovation.

The same is not true in the case of categorical variants. Firstly, in order to learn about an individual speaker's variable usage, one truly need to learn about the overall distribution of realisations, i.e. the relative frequencies with which the different variants are mixed by the speaker. Except when different variants are strong social markers which are exclusively used by non-overlapping speaker groups, very little information can be extracted from individual productions. Instead, several realisations of a variable by one and the same speaker are necessary to make strong inferences about a speaker. In combination with the fact that morphosyntactic variables can only be observed much less frequently than most phonetic and phonological ones, it

is not obvious that people would be as good at acquiring or making inferences about categorical variables as they are for continuous ones, like the vowel realisations tested by Drager (2011).

In one of the rare studies investigating age effects for categorical rather than continuous (phonetic) traits, Walker and Hay (2011) showed the influence of congruence between 'word age' and 'voice age' in facilitating lexical access: listeners of all ages exhibited a speed-up in processing words produced by a voice indicative of an age group, exactly when the word was more likely to be used by speakers of that age group. Although this experiment speaks to the influence of perceived age directly, it does not involve a sociolinguistic variable, since the different stimuli belong to different semantic domains, rather than being different ways of 'saying the same thing'.

The goal of the present work is to extend the body of research on 'age vectors', as they are perceived and used by the individual, to the domain of syntactic change. Since this an understudied area of research, we will also present a novel questionnaire methodology designed to help quantify people's explicit knowledge about ongoing language changes, in particular their impressions of the changes' direction.

Here, a disclaimer is in order regarding the term *awareness* which, in contemporary linguistic research, is typically used to refer to an individual's explicit, *meta-linguistic* knowledge about their own and their community's language use (Preston and Robinson, 2005). This knowledge is consequently described as being "above the level of conscious awareness" (Baranowski, 2013, p.283) of the individual. While the terminology used in this chapter reflects the fact that the present methodology is based on participants' *explicit* statements about their meta-linguistic knowledge, it should be noted that conscious awareness is just one possible proxy that allows one to test individuals' ability to detect (and potentially amplify) linguistic trends. Although this chapter is dedicated exclusively to the study of linguistic *awareness* of changes, the mechanism of momentum-based language change does not strictly rely on awareness per se, but would most likely be driven by implicit knowledge which could be tested through more sophisticated experimental methods such as the ones used by Drager (2005, 2011).

5.2 Quantifying the awareness of syntactic changes

In this work we investigate the human capacity for tracking changes in syntactic variables by probing speakers' awareness of three instances of the loss of verb movement in the variety of Scots spoken in Shetland. Shetland is a group of islands approximately 200km North of Great Britain with around 23,000 inhabitants across 15 inhabited islands (Shetland Islands Council, 2014, see also Figures 5.1 and 5.2). While Shetland forms part of the United Kingdom, it was only passed from Denmark to the Crown of Scotland in the late 1400s, and the islands' linguistic history is correspondingly diverse. Although virtually all toponyms on the island can be traced back to Viking origins, the Scots settlers who emigrated to the islands following the annexation to the Scottish Crown brought their own West-Germanic vernaculars with them. These vernaculars gradually replaced the local *Norn* language, a North Germanic variety most closely related to Old Norwegian, which however continued to be spoken on the isles until at least the 18th century (Knooihuizen, 2009). Today, the primary native vernacular of Shetland
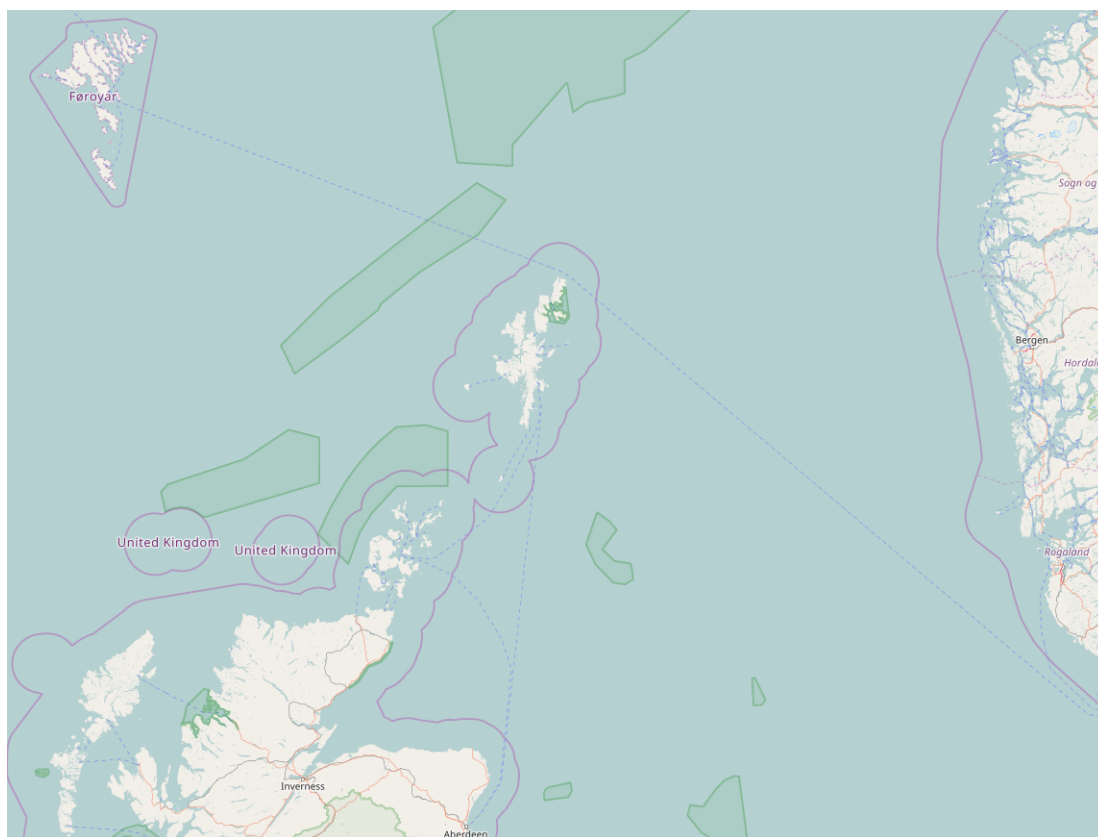
Figure 5.1: Shetland's location in the North Sea, showing the territorial waters of the United Kingdom, Norway and the Faroe Islands as well as ferry links and natural conservation areas. Projection: Web Mercator. © OpenStreetMap contributors, licensed under the Creative Commons Attribution-ShareAlike 2.0 license (CC BY-SA).

can be characterised as a variety of Scots, which is itself a continuum of language varieties spoken throughout the Scottish Lowlands that has developed largely in parallel to (rather than being derived from) the more well known English varieties that spread from England to many other parts of the globe (Millar, 2007, p.15).

Due to its insular location and sustained contact with North Germanic languages (Jamieson, 2016), Shetland Scots has retained many linguistic features typical of Germanic languages that the varieties of English and Scots on the British mainland have long lost. Some examples which will be evident in the questionnaire examples below are the relatively rich verbal inflection, as well as the maintenance of a number distinction in the second person pronoun *du* (cognate with Middle English *thou*). The features under investigation in this study are also of the historically conservative kind, namely the inversion of the verb position in several syntactic contexts, all of which occur alongside the emergence of periphrastic *do* (or 'do-support'). While for most areas of England the change away from the historically original verb-initial constructions (see Jamieson, 2015, for a more detailed analysis) can be dated to the period from 1500-1700 (Ellegård, 1953; Kroch, 1989a), the homologous change in Shetland Scots has only been unfolding over the course of the 20th century (Jonas, 2002). The three related changes currently ongoing
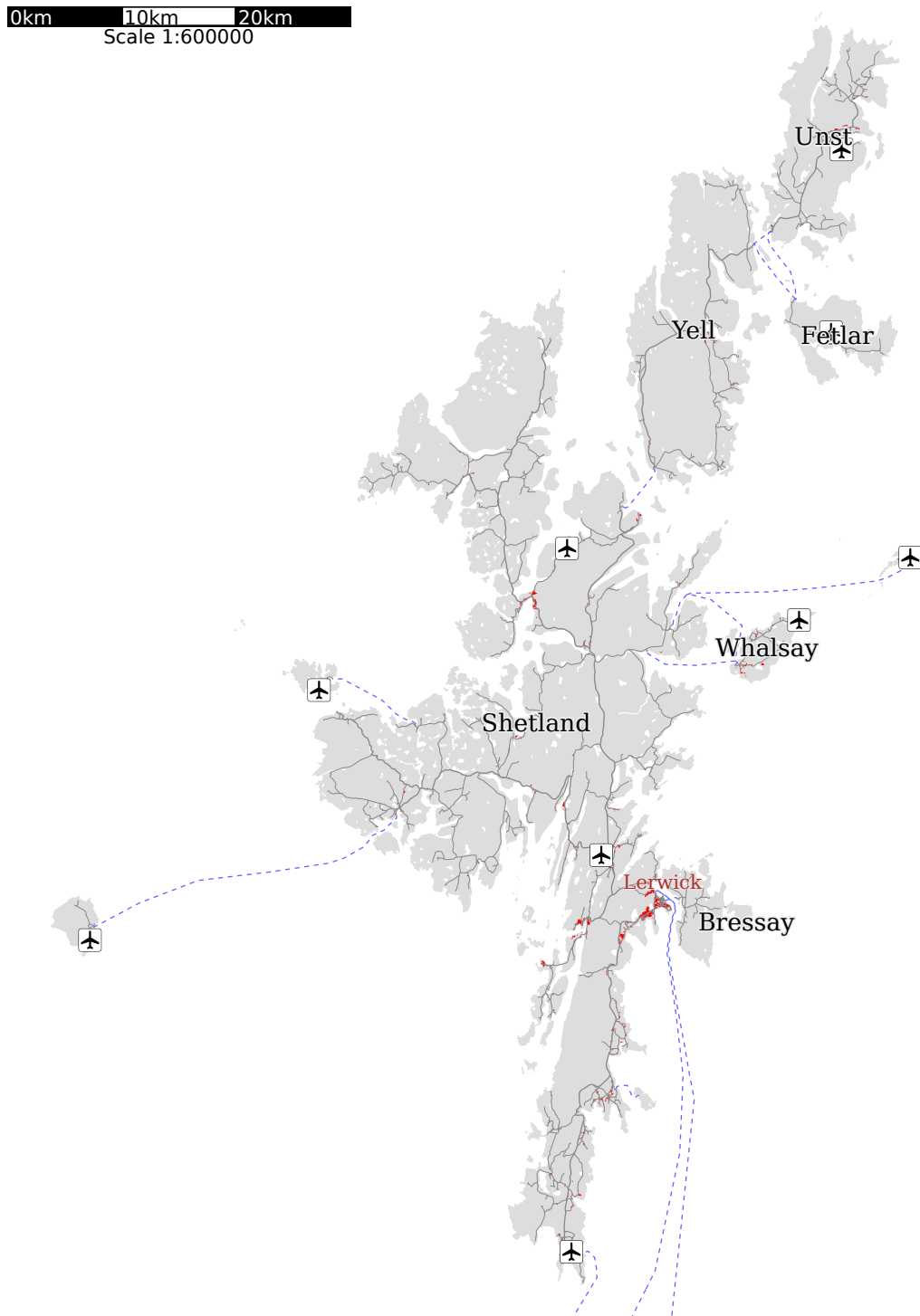
Figure 5.2: Detail of the Shetland archipelago showing paved roads, ferry routes, airports and areas of human residence (in red). Ferry connections to the Scottish mainland leave from the capital town of Lerwick, also the largest settlement. Regular flights to several Scottish airports as well as Bergen (Norway) depart from Sumburgh Airport at the Southern tip of the main island, with many more airfields for local planes scattered across the archipelago. Projection: orthographic, centered on 60.345°N and 1.4°W. Map data © OpenStreetMap contributors.

in Shetland which we investigate in the current work are as follows:

- verb positioning in imperatives, which is changing from a raised verb (surface realisation `VSO`) to Standard English `SVO` structure. An example in Shetland Scots would be "Mak du dy ain denner!" vs. "Du mak dy ain denner!", with the latter (incoming) variant akin to Standard English syntax, i.e. 'You (sg.) make your (sg.) own dinner!'

- yes/no question syntax: change from a main verb-initial to a 'periphrastic do' structure, e.g. from "Kens du Sarah?" to "Does du ken Sarah?", with 'ken' being the Scots lexeme for 'to know'.

- wh question syntax: change from plain `WhVSO` with a fronted main verb to a 'periphrastic do' structure, i.e. `Wh-`do'-SVO`. An example of the two constructions would be old "Whit gae du him?" to "Whit did du gie him?", with 'gie'/'gae' the Scots equivalents of Standard English 'give'/'gave'.

In all three cases, usage of the incoming variants is already common. The two question variables are more advanced, with younger speakers almost categorically using the incoming variants (with the exception of a few lexically specific items, see Jamieson 2015). So while the changes are in a sense nearing completion, all members of the speech community are still exposed to both outgoing and incoming variants due to their being used frequently by older speakers.

5.2.1   Methodology

To quantify people's explicit knowledge about ongoing language changes we adapted a self-evaluation method originally used to investigate the perception of phonetic changes by Labov (1966b) and Trudgill (1972), who asked speakers to report their relative usage of several phonetic variables. We refined the methodology, so that every sociolinguistic variable under investigation was covered by a one page questionnaire eliciting speakers' estimates of their own usage, as well as that of other social groups, alongside other (folk-)linguistic beliefs about the linguistic variants themselves[1]. At the top of each questionnaire page, the two competing syntactic variants were introduced in the following way:

> You are probably familiar with these two ways of asking somebody to do something:
>   "*Mak du dy ain denner!*"   "*Du mak dy ain denner!*"

The order of the two variants was randomised between individuals, in the above example the outgoing variant is on the left, the incoming one (akin to Standard English "You make your own dinner!") on the right. The dialectal spelling of the example sentences is quasi-standardised on Shetland, and their mixing with the Standard English formulations of the questionnaire is

---

[1]The complete materials of the paper-based questionnaire can be found in Appendix C.

not unusual. The actual questionnaire consisted of the following five questions (referred to as *Q1* through *Q5* throughout the text) which were intended to probe different aspects of people's explicit knowledge about the changes in question:

Question 1: "How much do you use either of these variants?"
This explicit question regarding speakers' own frequency of use could be answered on a 5-point scale, with the options labelled 'I use only (variant 1)', 'I use mostly (variant 1)', 'I use both equally', 'I use mostly (variant 2)' and 'I use only (variant 2)', with the order of the two variants matching those of their initial presentation at the top of the page.

Question 2: "How much do you think are people around you using either of the variants?"
This question could again be answered on a 5-point scale with options 'People use only (variant 1/2)/mostly (variant 1/2)/both equally'. This question does not just provide information on speakers' perception of their average interlocutors' frequency of use, but the *relative difference* between the answers to questions 1 and 2 can potentially provide information on whether speakers think of themselves as being 'ahead' or 'behind' the curve of a particular change relative to their speech community.

Question 3: "Which of the two variants do you think is *older*?"
This (intentionally vague) question is intended to get at speakers' beliefs or connotations regarding the 'age' of the competing variants, without drawing explicit attention to the fact that the variable is in fact changing. The three possible answers were '(variant 1) is older', '(variant 2) is older' and 'People have always used both', with the order of the two variants randomised.

Questions 4+5: "How much do you think *younger/older speakers* use either of the variants?"
The final two questions tap into speakers' awareness of the apparent time development of a change, with the same 5-point options as above: 'Younger/older speakers use only (variant 1/2)/mostly (variant 1/2)/both equally'. The order of the two questions was randomised between individuals.

Data collection proceeded in three stages: first, to pilot the methodology, 8 participants were asked to complete the paper version of the questionnaire on site in Shetland in January 2015. The pilot questionnaire consisted of just two sociolinguistic variables with the following example sentences:

1. verb positioning in imperatives: *Mak du dy ain denner!* vs. *Du mak dy ain denner!*, with the latter (incoming) variant akin to Standard English syntax, i.e. 'You (sg.) make your (sg.) own dinner!'

2. negation marking: *He didna go* vs. *He didnoo go* – this sociolinguistic variable is not undergoing change and was added as a control, with 'didna' being the negation variant used categorically in most of Shetland, set against the 'didnoo' variant which is categorical only on the island of Whalsay to the East of Shetland's main island (see Figure 5.2). With a population of around 1,000 and close links to the mainland, we expected the

local 'didnoo' variant (as well as its geographical patterning) to be known to all Shetland locals, an assumption that was borne out by several explicit references to the the variants' distribution by locals during data collection.

Following the successful pilot, 16 more participants were asked to complete an extended 4-page version of the questionnaire which covered two further variables:

3. yes/no question syntax: *Kens du Sarah?* vs. *Does du ken Sarah?*, i.e. "Do you (sg.) know Sarah?", with 'ken' being the Scots lexeme for 'to know'

4. wh question syntax: *Whit gae du him?* vs. *Whit did du gie him?*, the latter again akin to Modern English syntax but with the Scots lexical items for Standard English 'what' and 'give'.

These first 24 participants were part of a balanced sample matched for binary gender, age, and geographic location within Shetland. All participants grew up in and were currently living in rural locations in Shetland (i.e. outside the island's main town, Lerwick). Participants had on average spent 3.7 years living away from Shetland, typically for higher education, work or training purposes. In all cases, the questionnaire was administered as an exit-questionnaire following a ≈ 40 minute task which involved providing grammaticality judgments for a large number of examples of the changing variables in question as well as fillers, which was carried out in pairs (see Jamieson 2015 for a full description of the methodology and analysis of results).

Finally, we created an identical online version of the 4-variable questionnaire which was advertised via email and social networks. The online questionnaire was self-contained (i.e. not preceded by the grammaticality judgment task) and provided us with a convenience sample of another 53 participants from all over Shetland, who completed the questionnaire in April 2015[2].

5.2.2   Hypotheses & predictions

Based on previous research on both the specific syntactic changes under investigation as well as sociolinguistic knowledge more generally, we can derive several predictions about what effects we might expect so see *a priori*.

Regarding differences between the sociolinguistic variables, prior work on verb inversion in Shetland as well as the results derived from the acceptability judgments reported in Jamieson (2015) suggests that the two question variables should pattern differently from the imperatives. The change in question syntax is more advanced, with younger speaker using the incoming question variants almost exclusively. Across the population, we would therefore expect generally higher incoming variant usage for those two variables, as opposed to still relatively mixed usage of the two imperative variants. This general pattern should be observable to different degrees across all of our questions related to usage rates, i.e. all questions but Q3.

If our informants are indeed aware of the ongoing changes and their directionality, this should be evident in their responses to questions 3 through 5. Here, all three changing variables under

---

[2]The online version of the questionnaire is still available for reference at `http://spellout.net/ibexexps/kstadler/shetland/experiment.html`.

investigation should pattern differently from the stable negation variable which acts as our control: since this variable is not undergoing change, it provides us with a baseline for responses to explicit questions about variant age when variable usage is really patterned geographically rather than temporally.

In terms of between-participant differences, we can test a number of hypotheses that are implicit in much contemporary sociolinguistic research: because many studies of sound changes in progress have found there to be a significant effect of gender (typically with females leading a change), it has been argued that, primarily due to their position in Western societies, women might be more sensitive to linguistic cues (Trudgill 1972, but see Eckert 1989; Labov 1990). While the studies of automatic implicit sociophonetic knowledge discussed above have not revealed any effect of gender, Carrera-Sabaté (2014) showed that young females were more explicitly aware of an ongoing sound change in the Lleidatà dialect of Catalan. Although no gender differences have been reported for the current syntactic changes in Shetland, neither in production nor in terms of grammatical acceptability (Jamieson, 2015), we can still assess the claim that female speakers might be more sensitive or aware of ongoing changes.

Another interesting question regarding between-speaker differences pertains to how the sociolinguistic knowledge of age patterning might differ between participants of different ages. For example, in her experiment Drager (2011) found an effect of listener *age*, where only *older* participants' perception of vowels that were currently undergoing a chain shift were affected by the perceived age of the speaker whose tokens they were asked to classify. In other words, older speakers were actively compensating more strongly for the manipulated age difference, with younger speakers exhibiting less sociophonetic sensitivity, at least in the sense that they were not actively employing their knowledge of ongoing changes in the classification task. Although we should consider the possibility that the age of our participants will affect their sociolinguistic knowledge of the variables under investigation in this work, it is not possible to derive a straightforward prediction regarding the presence or direction of an effect from the literature. While age effects are also attested in the large body of empirical research on language *attitudes* (Giles and Billings, 2004), it is not immediately obvious how and whether the *qualitative* evaluation of innovations (often assumed to be primarily negative, see e.g. Labov 2001; Tagliamonte 2012) corresponds to the *quantitative* evaluation and perception of changes, with currently no conclusive results regarding the effect of age on the latter.

5.3 Results

Pooling together the data from the paper-based and online questionnaires, the total number of responses was $N = 77$ for the imperative (`imp`) and negation (`neg`) variables, and $N = 69$ for the yes/no question (`ynq`) as well as wh question (`whq`) syntax. Both the locally collected and online samples had a similar age distribution, with participants ranging from 18 to 73 years, with a median age of 32.

In terms of the geographical location of our participants there was a bigger difference between the two samples, as can be seen in Figure 5.3. The balanced sample explicitly excluded speakers originally from Shetland's capital Lerwick, home to 7,500 of the islands' total population of
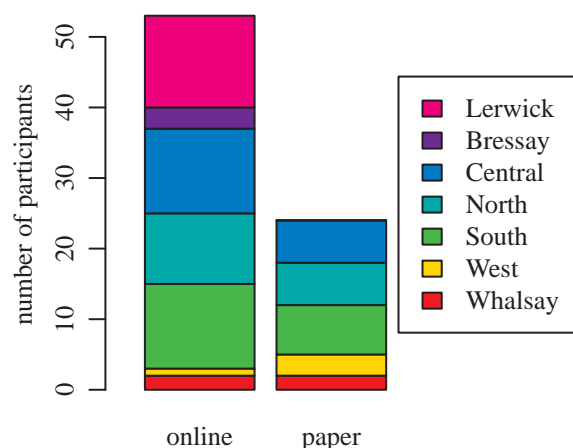
Figure 5.3: Origin of questionnaire participants within Shetland, by condition (online convenience sample vs. paper-based controlled sample)

23,000. The Scots vernacular of Lerwick is undergoing a more rapid change towards Standard Scottish English (SSE) forms than rural variants (Sundkvist, 2011), a development that can be attributed to the larger influx of speakers of other English varieties due to the capital's role as a hub for offshore oil drilling in the surrounding sea. The convenience sample on the other hand naturally includes a large proportion of Lerwick respondents. However, when it comes to their questionnaire responses, we did not find the Lerwick participants to pattern differently from the rest of the population for any of the questions.

Also, despite the fact that we might have expected the on-site participants to be more aware or sensitive to the questionnaire based on the preceding 40 minute acceptability judgment task on related syntactic variables, the type of data collection (paper-based on site vs. online) did not come out as a significant predictor in any of the statistical models reported below.

5.3.1 Assessing the reliability of subjective usage judgments

Before we turn to the actual data analysis, we have to address a central issue of our methodology, namely the type of data collected and its reliability. While questionnaires are still one of the standard tools employed in dialectological research, explicit questions about language use have fallen into disfavour in the quantitative sociolinguistic tradition. One reason for this is that lay-people's subjective evaluation of linguistic forms is traditionally not regarded as a reliable indicator of usage, as overt evaluations are often assumed to reflect the participants' qualitative sociolinguistic attitudes rather than people's actual quantitative usage (Labov, Trudgill, inter alia). In combination with the increased availability of speech and text data analysis technology over the past few decades, the relevance of subjective data on language use outside dedicated areas of research such as language attitudes has all but disappeared. There has, however, been a recent resurgence in interest in the beliefs that non-linguists have about language variation, bridging the two fields with its own set of research methodologies often referred to as *perceptual dialectology* (Montgomery and Beal, 2011). Rather than completely discarding the opinions of laypersons on the topic of language, this approach raises a number of own research questions

regarding how naive speakers' 'folk beliefs' about language are related to language use as studied by linguists (Preston, 1996).

It is in this domain that broadly sociolinguistic approaches come closer to the methodologies still most frequently used to study syntax and syntactic variation, in particular by means of *grammaticality judgments* which have over the years been replaced by more gradual *acceptability judgments* provided by naive native speakers rather than linguistic researchers themselves (Cornips and Corrigan, 2005). Based on this continuum of related research methods based on explicit linguistic opinions expressed by speakers, there is also an increasing amount of literature on the question of actual usage is reflected in acceptability judgments and processing preferences (Sorace and Keller, 2005; Featherston, 2005) as well as attitudinal data (see e.g. Maegaard et al., 2013; Durham, 2014). In order to better understand the nature of the estimated usage levels obtained through our present methodology it is therefore insightful to cross-validate the results with other measures. While we have no quantitative production data available for the three changing syntactic variables in question which are all very low in frequency, we can, however, compare the relative usage rates against the grammatical acceptability judgments which were collected independently for the 24 participants during the first, on site phase of data collection (see Jamieson 2015). If the novel methodology presented here is indeed reliable, we should expect good correlation between the two measures.

Despite the fact that both grammaticality judgments and usage rate estimates were gathered through explicit elicitation, there are two big differences between the two types of data: firstly, the usage rate estimates draw explicit attention to the type of speaker that it is envisaged to be representative of, i.e. the speaker themselves, the 'average' interlocutor in their community, or a 'younger' or 'older' speaker specifically. This framing focusses explicitly on the variants' use in a specific context, while the acceptability judgment draws the informants' attention primarily to the linguistic variant itself. In this way, a quantitative acceptability judgment does not distinguish in a principled manner between utterances that the informant would use in their own production, and what they would accept (or expect) from some of their interlocutors, but never actually produce themselves.

The second difference is that the usage rates as collected here capture the *relative* usage of the two competing variants of a sociolinguistic variable by directly contrasting the two equivalent forms of a single example sentence. The acceptability judgments on the other hand express the *absolute* acceptability of an individual example sentence on a 7-point Likert scale, without speaking directly to the *relative* usage of the two competing variants.

In order to assess the reliability of our usage estimates by correlating it with the acceptability judgment data, we first need to transform the two measures to comparable scales, keeping in mind those differences. The basic idea here is to convert the absolute judgments per variant into relative ratings per variable, by comparing the per-variant ratings of the incoming and outgoing realisations of the same example sentences. Transforming the acceptability judgments to relative scores can be done in several different ways and boils down to three decisions. While none of the choices turned out to have a strong effect on the results, it is worth discussing them to get an idea of how a methodology based around acceptability judgments can be related to the present questionnaire methodology.

The first question regards exactly *which* acceptability judgments should be correlated with the usage estimates. The acceptability judgment data collected by Jamieson (2015) is more abundant: each of the 24 informants provided judgments for a total of 49 example sentences across the three changing variables, covering 17 different verb types, and some of the verbs were chosen because they are known to exhibit strong lexical effects that affect the choice of syntactic construction. The two principled ways to limit the lexical effects on the acceptability measures are, on one hand, to only correlate the judgments for those sentences where the verbs match those used in the respective example sentences from the questionnaire or, alternatively, to wash out lexical effects by taking the average acceptability score over all verbs used in the acceptability judgment task. Both approaches turn out to result in almost identical correlation coefficients, and even the inclusion of all individual lexical items leads to a highly significant (if lower) correlation coefficient.

The second decision relates to how the two acceptability judgments for the two competing variants should be converted into one measure capturing their relative acceptability, akin to the direct juxtaposition of the "which variant do you use more" measure employed by the questionnaire. The two most straightforward ways to combine them into one measure is by taking the difference between the ratings for the incoming and the outgoing variant, either by subtraction (absolute difference in acceptability) or division (relative difference in acceptability). Both approaches produce a numeric scale with a neutral centre point occupied by pairs of judgments where the incoming and outgoing variants were rated to be equally acceptable (0 for the absolute difference by subtraction, and 1 for the relative difference by division). In terms of the relative ranking of pairs with differing judgments the two scales only differ marginally, as can be seen in Figure 5.4 below, where it is relatively easy to identify the matching pairs of datapoints between the two graphs based on their y-axis position.[3]

Having transformed the absolute acceptability judgments to a relative acceptability scale, there is still a third decision to be made, namely which of the usage estimate ratings it should be correlated with. The separate questions in the questionnaire gathered data regarding different speaker groups, including the informants themselves as well as several idealised interlocutor groups. Intuitively, we would expect people's acceptability judgments to reflect their own probability of producing the respective variants, and it is indeed only their self-usage estimates that yield a significant correlation with the derived relative acceptability scores.

The relationship between the reported relative self-usage rates and both the absolute and relative difference in acceptability reported by the 24 participants in the paper condition are shown in Figure 5.4. Due to the large number of ties along the 5-point ordinal scale of the questionnaire we chose Kendall's τ_B to calculate the correlation between the two measures. We found the strongest correlation between informants' self-usage estimates and the relative difference of their acceptability judgments by division, with $\tau_B = 0.22207$. This correlation is significant at $p < .01$, as determined using the `pvrank` R package which provides p-value

---

[3]The main difference between the two different ways to convert the acceptability judgments concerns the resolution of the resulting scale: pairwise subtraction of the 7 possible ranks of the ordinal acceptability judgments yields a total of 13 possible relative ratings, while the division method results in up to 35 theoretically possible values, and consequently fewer ties. The impact of choosing either of the two approaches on the resulting correlation coefficients is still only marginal, and our results are not highly sensitive to either choice, as can be seen below.

calculation for rank order correlation tests accounting for ties (Amerise et al., 2016).
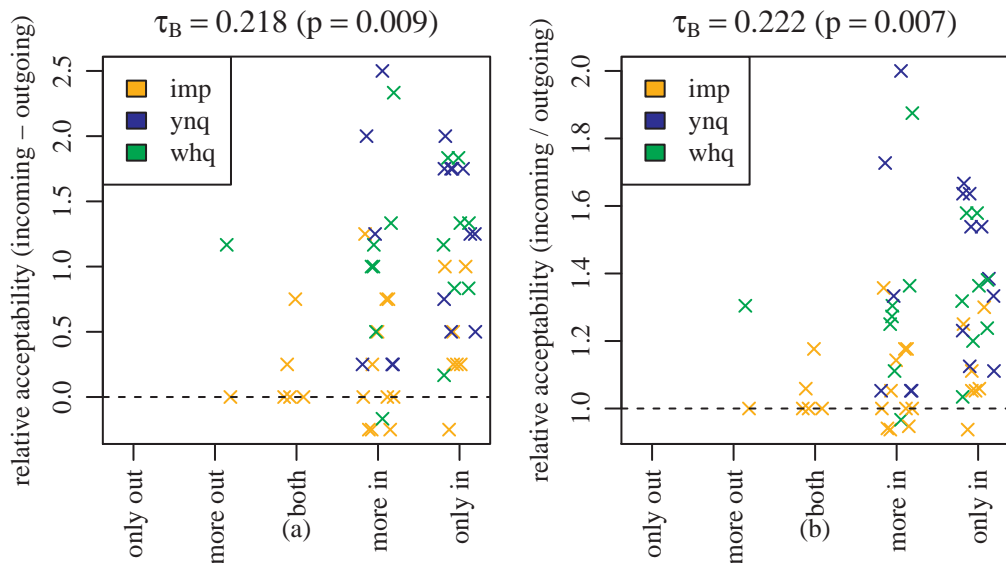
Figure 5.4: Correlation between the 24 informants' reported self-usage rate of the two variants for the three changing variables (x-axis, jittered) and the relative acceptability derived from the average acceptability judgments (y-axis), for two ways of calculating the relative preference of the variants: (a) *absolute* difference in acceptability, calculated by subtracting the rating of the outgoing variant from the incoming one (b) *relative* difference of the incoming vs. outgoing ratings, calculated by division. The mid-points of the two relative acceptability scales (where two competing variants are judged as equally acceptable) are indicated by dashed lines.

Having established that the subjective responses that make up our data set pattern closely with an independent measure of use in the form of acceptability judgments, we can now turn to analysing our participants' responses, and the patterns found therein.

5.3.2 Self-estimates of own usage

The first measure elicited from our participants was an estimate of their own usage levels for each of the linguistic variables in question. The overall data, split by sociolinguistic variable and informant age, is shown in Figure 5.5. The first impression is that the response distribution is highly uneven, with the majority of responses falling onto the three categories of our 5-point-scale that indicate at least 50% usage of the incoming variants.[4]

As expected, the self-reports on the stable negation variable are patterned by the informants' geographical location, with the only four informants indicating categorical usage of the more localised variant all originating from the isle of Whalsay. In contrast, responses for the three changing variables appear to reflect differences in usage patterns in *apparent time*, the familiar phenomenon where the language use of younger speakers is found to be more *advanced*, i.e. they exhibit higher usage rates of the new, incoming variants (Wagner, 2012). While the mean age

---

[4]Since the control variable exhibits stable variation, there is no 'incoming' or 'outgoing' variant in this sense. Instead, we have coded the more frequently used (i.e. geographically more widespread) variant as the 'incoming' one.
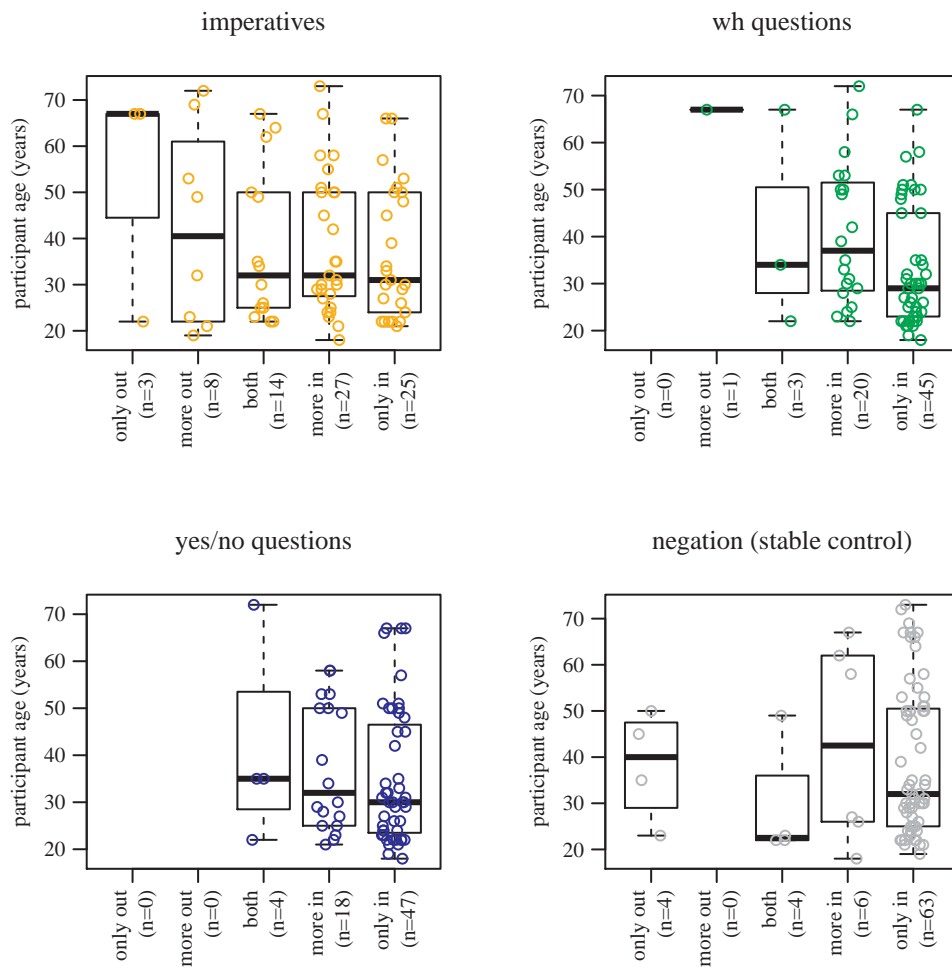
Figure 5.5: Distribution of informant age per reported level of own usage for the three changing variables as well as the stable, geographically conditioned control. The five response levels correspond to the five possible responses described in Section 5.2.1, ranging from the leftmost 'I only use the incoming variant' to 'I only use the outgoing variant' at the very right.

of respondents tends to decrease for higher reported usage of the incoming variants for the three changing variables, the same is not true for the stable control variable.

To test our hypotheses we used ordered logistic regression, an extension of logistic regression that allows for more than two (ordered) response categories. We fit a number of models of increasing complexity using R's `ordinal` package (Christensen, 2015), with participant as a random effect. The results from these models are shown in Table 5.1: the four coefficients at the very bottom of the table, present for all models, are equivalent to the *intercept* in (generalised) linear models, only that ordinal regression requires $n - 1$ intercepts to capture the baseline distribution of responses across the n response levels. The coefficients above it capture the inferred effect of the various predictor variables on the outcome distribution of self-evaluation responses along the 5-point ordinal scale. Ordered logistic models can be read like any regression model, except that the intercepts are given as *log odds ratios*, and the coefficients β_i as *difference in log odds* per unit change in the predictor variable. In other words, with every unit change in the predictor, the relative odds ratio of responding in a lower vs. a higher category as given by the intercepts is multiplied by $\exp(\beta_i)$.

Looking at the succession of models in Table 5.1 as well as the pairwise model comparison between them in Table 5.2, we first find a strong effect of the type of sociolinguistic variable (model 1): in comparison to the stable negation variant for which the majority of speakers reported using only the 'incoming' variant, the imperatives (`varimp`) exhibit significantly more responses in the center of the 5-point scale, with the two question types (`varynq`, `varwhq`) falling somewhere in between. While we find no evidence for an effect of informant age across all variables (model 2) there appears to be an effect of gender, with females being more likely to report increased use of the majority variants (model 3). When taking into account differences between the different sociolinguistic variables, we do find evidence for an effect of age for the changing variables only. Rather than using the `var` term, a 4-level factor that distinguishes all sociolinguistic variables covered by the questionnaire, model 4 adds an interaction between the age coefficient and the *type* of sociolinguistic variable, with the binary `stable` factor opposing the three changing variables against the stable negation variable. Even though the model *comparison* between models 3 and 4 in Table 5.2 is not significant, the model coefficient $\beta_{age} = -0.028$ means that the relative probability of responding in a higher category is multiplied by $\exp(-0.028) = 0.972$ for every year that a participant is *older*. This is equivalent to their probability of reporting a relatively higher usage level of the incoming variant *decreasing* slightly (by about 2.8%). The coefficient of the `age:stable` interaction term, which is of a similar amplitude but in the opposite direction ($\beta_{age:stable} = 0.039$), implies that this effect of age is effectively cancelled out for the stable control variable.

Finally, fitting a model with separate interaction terms for each of the four variables (by substituting the `age:stable` term with `age:var`, model 5) does not improve the model fit significantly, indicating that the presence of the age effect is mostly explained by whether the sociolinguistic variable is undergoing change or not.

To help visualise the effect of age and aid in interpreting the log odd ratio coefficients, the empirical responses by age as well as the corresponding predictions made by the ordered logistic models are shown in Figure 5.6. For this purpose we split the responses we collected into two

	(1)	(2)	(3)	(4)	(5)
age		−0.019	−0.020	−0.028*	−0.019
		(0.011)	(0.011)	(0.012)	(0.016)
stable				1.031	
				(0.901)	
gendermale			−0.788*	−0.817*	−0.838*
			(0.359)	(0.365)	(0.370)
age:varynq					−0.005
					(0.023)
age:varwhq					−0.033
					(0.023)
age:varneg					0.030
					(0.024)
varynq	1.909***	1.874***	1.864***	1.865***	2.052*
	(0.372)	(0.372)	(0.372)	(0.374)	(0.960)
varwhq	1.770***	1.736***	1.732***	1.733***	2.997**
	(0.365)	(0.365)	(0.366)	(0.367)	(0.985)
varneg	2.558***	2.575***	2.567***		1.403
	(0.414)	(0.414)	(0.415)		(0.979)
age:stable				0.039	
				(0.021)	
only out|more out	−3.204***	−3.965***	−4.261***	−4.648***	−4.290***
	(0.450)	(0.651)	(0.665)	(0.711)	(0.801)
more out|both	−2.282***	−3.036***	−3.336***	−3.713***	−3.366***
	(0.348)	(0.579)	(0.595)	(0.642)	(0.738)
both|more in	−0.996***	−1.747**	−2.051***	−2.407***	−2.065**
	(0.280)	(0.532)	(0.547)	(0.590)	(0.692)
more in|only in	0.884**	0.132	−0.172	−0.493	−0.135
	(0.274)	(0.512)	(0.522)	(0.558)	(0.669)

Note: *p<0.05; **p<0.01; ***p<0.001

Table 5.1: Ordered logistic regression model (coefficients and standard errors) of participants' own usage estimates.

	Model	Res. df	-2LL	Test	df	LR	P(>Chi)
(0)	1	287	606.566				
(1)	var	284	551.283	0 vs 1	3	55.282	<.001
(2)	age + var	283	548.353	1 vs 2	1	2.930	.087
(3)	age + gender + var	282	543.518	2 vs 3	1	4.835	.028
(4)	age:stable + gender + var	281	540.054	3 vs 4	1	3.464	.063
(5)	age:var + gender	279	537.920	4 vs 5	2	2.134	.344

Table 5.2: Pairwise comparison of the models in Table 5.1.
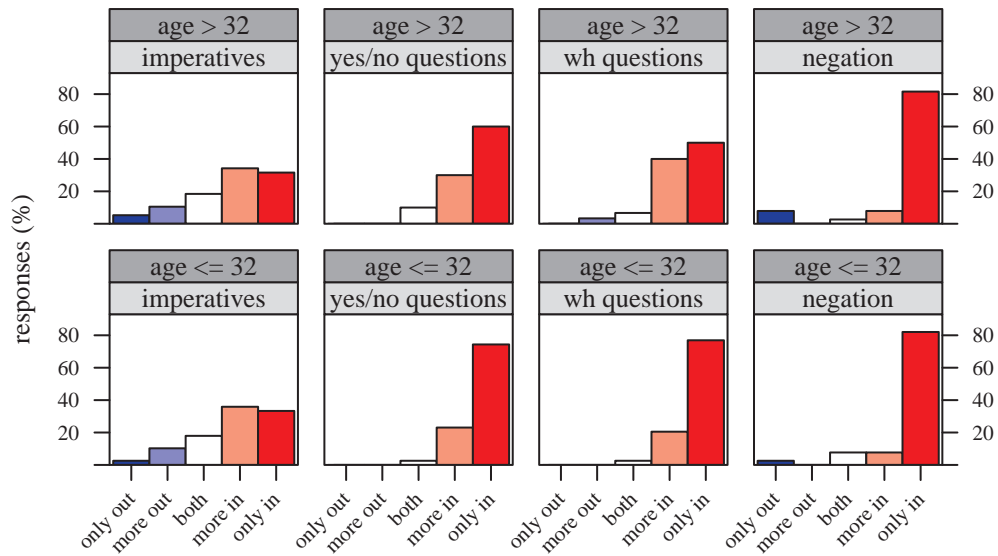
(a) Empirical distribution of the self-reported usage levels for the three changing variables, with the participants split into two age groups (≤ 32 years, $N = 39$; > 32 years, $N = 38$). The younger the speaker, the more likely they are to report higher usage of the incoming variant.
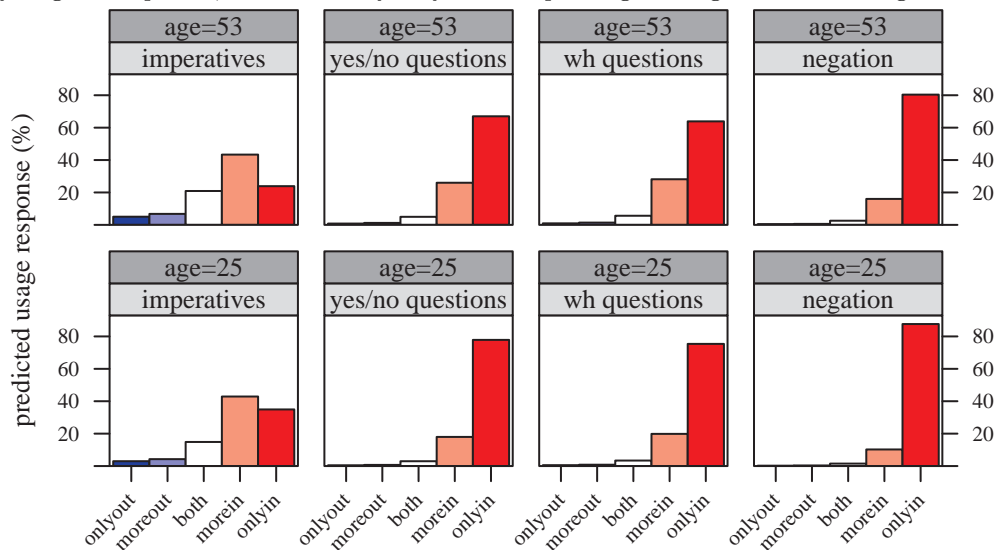
(b) Prediction of response distributions of ordered logistic model (4) from Table 5.1 for the mean ages of the two age groups plotted above.

Figure 5.6: Comparison of empirical own usage reports and corresponding model prediction.

evenly sized age groups, and the distribution of responses per age group and sociolinguistic variable is shown in Figure 5.6a: despite differing baseline distributions (with imperatives generally exhibiting a flatter distribution), we can see the general trend of increasing incoming variant usage for younger speakers. The predictions made by model 4 for the different variables and a typical member of each of the two age groups are plotted underneath in Figure 5.6b, showing good agreement with the empirical data.

5.3.3 'Other people' usage estimates

When it comes to reporting on the linguistic usage levels of other individuals in their speech community, the overall pattern of responses is similar to the self-usage estimates, but with an added central tendency or edge-avoiding effect in the responses, as can be seen in Figure 5.7. This presumably stems from the fact that, when imagining an 'average' individual, the informants will model this on the *population average*, which, given any amount of within- or between-individual variation, is almost necessarily non-categorical.

Again, we performed ordered logistic regression models with participant as a random effect, reported in Tables 5.3 and 5.4. In contrast to the self-usage reports, we find no effect of age for any of the variables (models 1+2). Instead, the models reveal that the only significant predictors of the reported usage levels are the type of sociolinguistic variable (model 3) as well as the participant's gender, with females again estimating slightly *higher* usage of the incoming variant in the community (model 4). Importantly we find no significant interaction between the type of variable and any of the other predictors, i.e. the effect of gender again pertains to all four variables, and not just the changing ones (model 5).

	(1)	(2)	(3)	(4)	(5)
age	0.005	−0.002			
	(0.011)	(0.012)			
stable		−0.560			
		(0.771)			
age:stable		0.023			
		(0.019)			
varynq			1.750***	1.749***	1.751***
			(0.404)	(0.404)	(0.404)
varwhq			1.888***	1.883***	1.884***
			(0.410)	(0.410)	(0.410)
varneg			1.491***	1.487***	1.222**
			(0.387)	(0.388)	(0.450)
gendermale				−0.832*	−1.019*
				(0.368)	(0.406)
gendermale:stable					0.743
					(0.654)
only out|more out	−5.821***	−6.045***	−5.294***	−5.615***	−5.717***
	(1.092)	(1.120)	(1.037)	(1.053)	(1.059)
more out|both	−3.465***	−3.686***	−2.829***	−3.141***	−3.232***
	(0.537)	(0.591)	(0.408)	(0.440)	(0.451)
both|more in	−1.659***	−1.873***	−0.835**	−1.136***	−1.215***
	(0.446)	(0.509)	(0.289)	(0.322)	(0.332)
more in|only in	2.471***	2.300***	3.825***	3.519***	3.464***
	(0.469)	(0.516)	(0.419)	(0.421)	(0.424)

Note: *p<0.05; **p<0.01; ***p<0.001

Table 5.3: Ordered logistic regression model (coefficients and standard errors) predicting participants' answers to the question "How much do you think are people around you using either of the variants?"
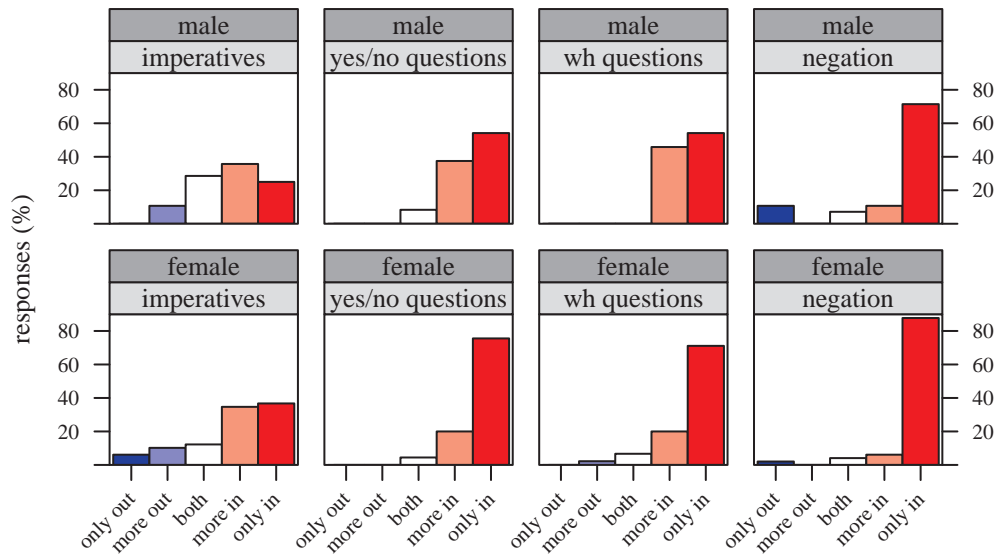
(a) Overall distribution of reported *own* usage, per sociolinguistic variable.
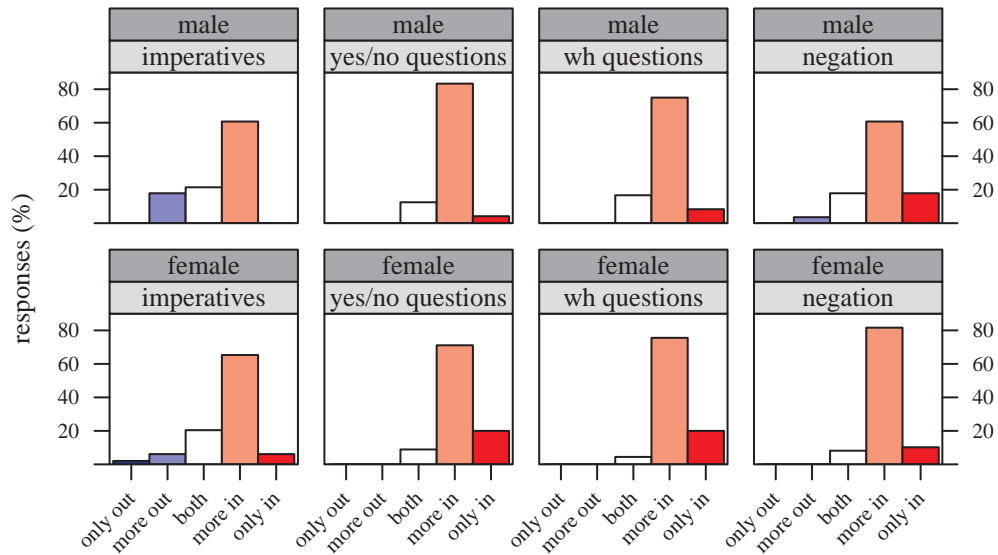
(b) Responses to the question "How much do you think are people around you using either of the variants?". While the distribution of responses to the imperative question is again consistently flatter than for the other responses, in comparison to the self-usage estimates shown in (a) there is a clear shift away from both of the extreme response options, indicating that individuals perceive the population average to be variable rather than categorical.

Figure 5.7: Comparison of own vs. community-level usage estimates by binary gender.

Table 5.4: Pairwise comparison of the models in Table 5.3.

	Model	Res. df	-2LL	Test	df	LR	P(>Chi)
(0)	1	287	507.936				
(1)	age	286	507.702	0 vs 1	1	0.234	.628
(2)	age:stable	284	505.271	1 vs 2	2	2.431	.297
(3)	var	284	478.018	1 vs 3	2	29.684	<.001
(4)	var + gender	283	472.928	3 vs 4	1	5.089	.024
(5)	var + gender:stable	282	471.625	4 vs 5	1	1.303	.254

The data from the first two questions implicitly contains another piece of information, namely where the speakers regard their own variable usage to be relative to the community-level. We measure this by looking at the number of ordinal categories along the 5-point scale that the self-usage response *differs* from the reported community-level usage, and subtracting their ordinal ranks. In this representation of the relative difference, positive numbers indicate that a speaker reported a relatively higher usage of the incoming variant for themselves than for the community, and vice versa.

If we look at the distribution of this measure in Figure 5.8, we see that the participants' age is a good predictor of this difference between themselves and the community. Ordinal logistic regression models paralleling those fit to the self-usage reports in Table 5.2 above confirm this picture: as can be seen in Tables 5.5 and 5.6, the most parsimonious model predicts the difference between individuals' reported own and community usage based on the type of sociolinguistic variable as well as the informant's age. Although the effect of age is again greater for the changing variables (compare models 2 and 3), the models indicate no significant difference between the stable as opposed to any of the changing variables (models 3+4).
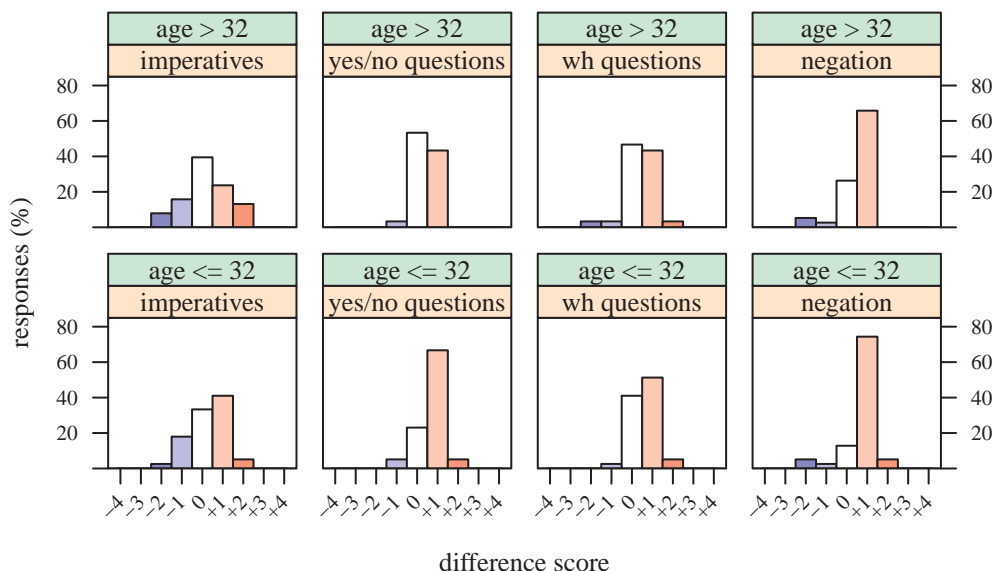
Figure 5.8: Relative difference in number of ordinal categories between participants' reported *own* vs community usage levels of the competing variants, by sociolinguistic variable and age group. Positive numbers indicate that a speaker reported a higher usage of the incoming variant for themselves than for the community, and vice versa. While the 5-point response scale allows for differences of up to ±4, no speaker indicated their own usage to be more than two ordinal categories away from the community level. Younger speakers are more likely to perceive themselves to be ahead of the community level usage (more red), while older speakers are most likely to report their usage to be level with the community (white) or behind (blue).

5.3.4 Beliefs about the age of competing linguistic variants

The third question of the questionnaire aimed at eliciting the speakers' beliefs about the variants by explicitly asking which they thought was the 'older' of the two, with 'people have always

	(1)	(2)	(3)	(4)
age		−0.022*	−0.027*	−0.023
		(0.010)	(0.011)	(0.016)
stable			0.633	
			(0.745)	
varynq	0.883**	0.850*	0.841*	1.481
	(0.337)	(0.336)	(0.337)	(0.879)
varwhq	0.677*	0.636	0.624	0.494
	(0.336)	(0.336)	(0.337)	(0.881)
varneg	1.305***	1.322***		0.796
	(0.343)	(0.343)		(0.858)
age:stable			0.018	
			(0.017)	
age:varynq				−0.017
				(0.022)
age:varwhq				0.004
				(0.022)
age:varneg				0.013
				(0.020)
-2|-1	−3.312***	−4.156***	−4.377***	−4.223***
	(0.422)	(0.583)	(0.627)	(0.759)
-1|0	−1.934***	−2.788***	−3.000***	−2.847***
	(0.307)	(0.504)	(0.549)	(0.692)
0|+1	0.432	−0.426	−0.618	−0.461
	(0.269)	(0.462)	(0.500)	(0.652)
+1|+2	4.180***	3.324***	3.129***	3.302***
	(0.414)	(0.535)	(0.568)	(0.715)

Note: *p<0.05; **p<0.01; ***p<0.001

Table 5.5: Ordered logistic regression model (coefficients and standard errors) of the relative difference between participants' reported own and community-level usage.

	Model	Res. df	-2LL	Test	df	LR	P(>Chi)
(0)	1	287	657.692				
(1)	var	284	642.109	0 vs 1	3	15.583	.001
(2)	age + var	283	637.137	1 vs 2	1	4.972	.026
(3)	age:stable + var	282	636.056	2 vs 3	1	1.081	.298
(4)	age:var	280	635.065	3 vs 4	2	0.990	.609

Table 5.6: Pairwise comparison of the models in Table 5.5.

used both' given as a neutral third option. As can be seen in Figure 5.9, people reliably identify the outgoing variant as the 'older' one for the three changing variables. For the stable negation variable results are more mixed, but many also report the less widespread 'didnoo' variant as being older.

The ordered logistic regression models reported in Tables 5.7 and 5.8 show that, while participants generally tend to report the less frequent variants to be older, they are more likely to do so for the three changing variables (model 3). There is also again a significant interaction between the type of variable and the age of the informant: older informants are more likely to identify the outgoing variant as the 'older' one, but again only for the sociolinguistic variables that are changing (models 2-5). Beyond the split between the stable and changing variables, there was no significant difference between the individual variables (models 3-6), and we also found no evidence for any effect of gender.
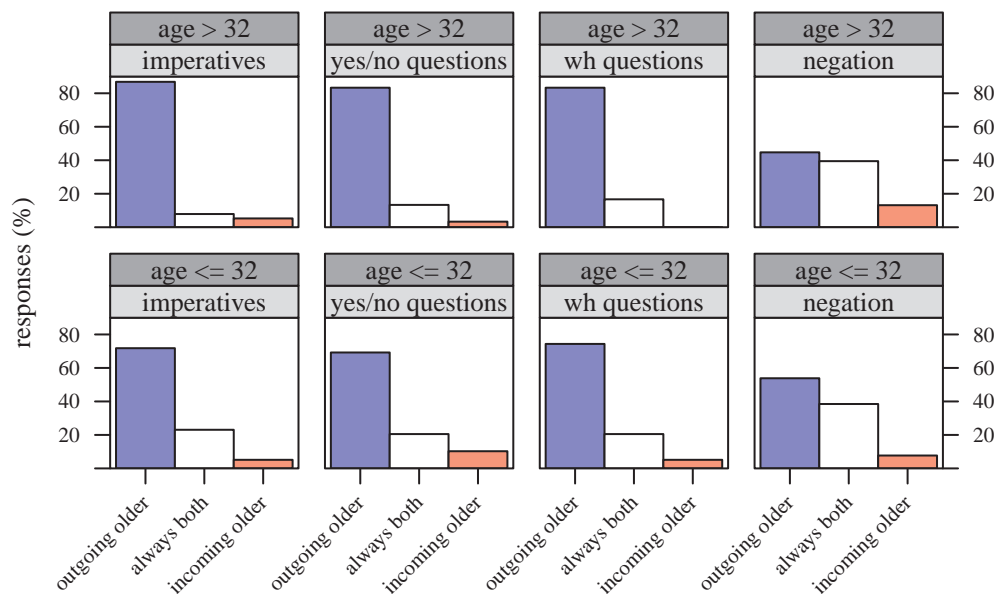
Figure 5.9: Distribution of responses to the question "Which of the two variants do you think is *older*?", split by sociolinguistic variable and age of the informant. In comparison to the stable negation variant, where about half of the people considered the geographically more limited (minority) variant 'didnoo' to be older, the response distribution was more extreme for the three changing variables. Here, around 80% of respondents report the outgoing forms (which are already the minority forms) to be older, with even fewer answers falling on the majority variant as well as the "people have always used both" option.

5.3.5 Perception of apparent time differences

The final pair of questions, which ask for the participants' impression of the relative usage level of the two competing variants by 'younger' and 'older' speakers was aimed at helping us to determine whether individuals can perceive and report apparent time differences in categorical variables. As can be seen in Figure 5.10, speakers consistently report higher usage of the majority variants among younger speakers but, surprisingly, this effect is present across all

	(1)	(2)	(3)	(4)	(5)	(6)
gendermale	0.494					
	(0.373)					
age		0.007			−0.016	−0.016
		(0.012)			(0.015)	(0.015)
stable			1.404***		−0.751	−0.617
			(0.308)		(0.789)	(0.843)
varynq				0.373		0.320
				(0.435)		(0.438)
varwhq				0.080		0.028
				(0.440)		(0.444)
varneg				1.557***		
				(0.401)		
age:stable					0.058**	0.057**
					(0.020)	(0.020)
outgoing older|always both	1.232***	1.315**	1.531***	1.685***	0.963	1.103
	(0.249)	(0.490)	(0.248)	(0.359)	(0.604)	(0.678)
always both|incoming older	3.251***	3.332***	3.727***	3.891***	3.254***	3.403***
	(0.355)	(0.556)	(0.384)	(0.472)	(0.674)	(0.747)

Note: *p<0.05; **p<0.01; ***p<0.001

Table 5.7: Ordered logistic regression model (coefficients and standard errors) of responses to the question "Which of the two variants do you think is *older*?".

	Model	Res. df	-2LL	Test	df	LR	P(>Chi)
(0)	1	288	434.312				
(1)	gender	287	432.539	0 vs 1	1	1.773	.183
(2)	age	287	433.952	0 vs 2	1	0.360	.549
(3)	stable	287	412.654	0 vs 3	1	21.658	<.001
(4)	var	285	411.839	3 vs 4	2	0.815	.665
(5)	age:stable	285	403.374	3 vs 5	2	9.279	.01
(6)	age:stable + var	283	402.723	5 vs 6	2	0.652	.722

Table 5.8: Pairwise comparison of the models in Table 5.7.

sociolinguistic variables, including the stable control. While there appears to be a general trend to attribute the usage of minority variants (whether outgoing or just geographically limited) to older speakers, the ordinal logistic regression model 2 in Tables 5.9 and 5.10 shows that the effect is significantly stronger for the three changing variables, but with no significant differences between them (compare models 3+4). In contrast to the responses to question 3, the age of our informants does not appear to have a systematic effect on how they perceive apparent time differences for any of the variables under investigation (models 1+4).
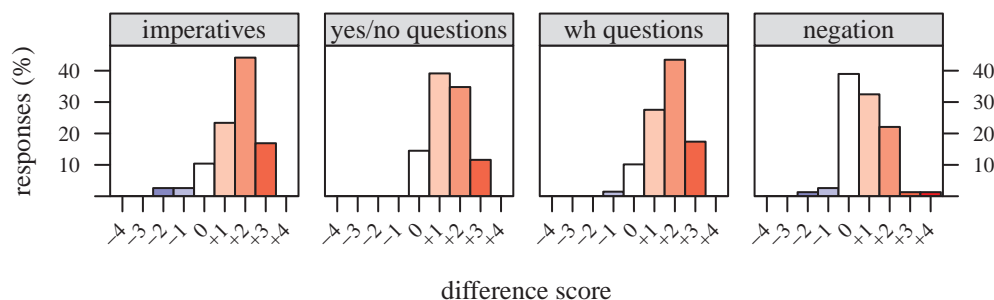
Figure 5.10: Relative difference between the reported usage of the two variants between 'older' and 'younger' speakers, for each of the four variables. The relative difference indicates the number of ordinal categories that separates an individuals' responses for the two age groups along the 5-point scale, where positive numbers (in shades of red) correspond to reporting higher usage of the majority variant among younger speakers, and vice versa. The most extreme points of the scale at ±4 correspond to reporting the 'older' and 'younger' speakers to be categorical users of opposing variants, the mid-point at 0 to reporting the same usage levels for both.

While this might suggest that individuals can accurately perceive and report on apparent age differences in changing categorical variables, we cannot straightforwardly jump to this conclusion. Model 5 shows that our participants' responses to question 3 (i.e. their reported belief about which if any of the competing variants is 'older') is a significant predictor of how far they reported younger speakers to be 'ahead' or 'behind' in their relative usage of the incoming variant. So while all results presented so far show that individuals are clearly able to (correctly) report on the directionality of the changes under investigation, the exact *source* of this metalinguistic knowledge is still in question. We turn to this issue in our discussion.

5.4   Discussion

Having presented the questionnaire data as well as statistical models of the individual responses, we now turn our attention to the particular research questions and predicted effects discussed in Section 5.2.2.

5.4.1   Identifying the source of awareness of ongoing changes

The main goal of our research was to quantify if and to what degree individuals are aware of and able to report on the direction of ongoing changes in their community, with a particular eye on whether individuals are sensitive to differences in 'apparent time', i.e. the traditionally more

	(1)	(2)	(3)	(4)	(5)
age	−0.011			−0.005	
	(0.008)			(0.010)	
stable		−1.535***		−0.804	−1.215***
		(0.266)		(0.653)	(0.274)
varynq			−0.432		
			(0.310)		
varwhq			0.052		
			(0.312)		
varneg			−1.674***		
			(0.322)		
age:stable				−0.020	
				(0.016)	
oldervaralways both					−1.325***
					(0.317)
oldervarincoming older					−2.316***
					(0.575)
-2|-1	−5.109***	−5.445***	−5.590***	−5.697***	−6.176***
	(0.673)	(0.620)	(0.648)	(0.730)	(0.661)
-1|0	−4.108***	−4.435***	−4.581***	−4.680***	−5.121***
	(0.494)	(0.417)	(0.457)	(0.567)	(0.466)
0|+1	−1.801***	−1.970***	−2.114***	−2.171***	−2.454***
	(0.354)	(0.215)	(0.284)	(0.432)	(0.254)
+1|+2	−0.328	−0.301	−0.432	−0.490	−0.636***
	(0.327)	(0.164)	(0.242)	(0.402)	(0.189)
+2|+3	1.690***	1.914***	1.810***	1.726***	1.730***
	(0.354)	(0.220)	(0.275)	(0.424)	(0.230)
+3|+4	5.397***	5.689***	5.591***	5.500***	5.572***
	(1.049)	(1.013)	(1.026)	(1.077)	(1.016)

Note: $^{*}p<0.05;$ $^{**}p<0.01;$ $^{***}p<0.001$

Table 5.9: Ordered logistic regression model (coefficients and standard errors) of the relative difference in individuals' estimates for 'younger' and 'older' speaker groups in number of ordinal categories, a proxy for perceived apparent time differences. Raw data is shown in Figure 5.10.

	Model	Res. df	-2LL	Test	df	LR	P(>Chi)
(0)	1	284	829.962				
(1)	age	283	828.119	0 vs 1	1	1.842	.175
(2)	stable	283	794.396	0 vs 2	1	35.565	<.001
(3)	var	281	791.493	2 vs 3	2	2.904	.234
(4)	age:stable	281	791.456	2 vs 4	2	2.941	.23
(5)	stable + oldervar	281	764.235	2 vs 5	2	30.161	<.001

Table 5.10: Pairwise comparison of the models in Table 5.9.

advanced usage of incoming variants among younger speakers. While the results presented in Sections 5.3.4 and 5.3.5 indicate that people have accurate knowledge about ongoing changes, our data does not allow us to give a definite answer regarding the *source* of this knowledge. In principle, beliefs could be based on any or all of (a) meta-linguistic commentary or other connotations of variants being archaic, (b) impressions of the change in *real time*, i.e. awareness of changing speech patterns that accumulate throughout a speaker's lifetime, or (c) apparent time differences in variable usage across age groups. The questionnaire was designed with these different sources of knowledge in mind. In particular, while questions 4+5 tapped into speakers knowledge about apparent time differences between speakers, the responses to question 3 offer a more general window into individuals' folk-linguistic beliefs about the 'age' or novelty of the competing variants.

As already indicated above, the two measures turned out to be highly correlated: model (5) in Table 5.9 showed the folk-linguistic belief about variant age is itself a good predictor of individuals' perceived apparent time differences, as derived from questions 4+5. Importantly, the converse is also true: Tables 5.11 and 5.12 report expanded models of individuals' responses to question 3, showing that the apparent time differences gathered from the responses to questions 4+5 are a good predictor of the responses to question 3. The number of ordinal categories between the reported 'older' and 'younger' usage levels (coded as a continous variable `apparentdiffN`) is in fact a better predictor of the participants' folk-linguistic belief than whether the variable is actually changing or not (compare models 1 and 2). What the models do not suggest, however, is that folk-linguistic beliefs are based on perceived apparent time differences alone (model 3). Other predictors remain significant, in particular the interaction between informant age and the type of sociolinguistic variable reported earlier in Section 5.3.4. Even when the participants' impression of apparent age differences is taken into account, there is still a significant effect of informant age on their responses, with older participants less likely to report the less frequent variant as older for the stable control variable only (compare models 4 and 5).

Based on these results it remains an open question whether people might have inferred their answer to question 3 independently from a perceived apparent time difference, or if participants felt led to answer questions 4 and 5 in a way that would post-hoc justify their earlier response, which was itself based on other (socio-indexical) knowledge. In order to understand the different sources of socio-indexical knowledge that might be at play here, it is insightful to take a closer look at the responses to our control variable.

5.4.2 Variant age and perceived apparent time differences: evidence from the control variable

Some more evidence regarding the primary source of inividuals' knowledge can be glanced from the responses to our control variable. As the distribution of negation variants on Shetland is stable, determined by geography rather than by generational change, we would not expect any information on the history of the two variants to be available to our speakers. Regarding their answers to the first two questions, the control patterned as expected. Own usage reports were

	(1)	(2)	(3)	(4)	(5)
age				0.002	−0.020
				(0.013)	(0.016)
stable	1.404***		0.816*	0.815*	−1.229
	(0.308)		(0.332)	(0.332)	(0.841)
apparentdiffN		−1.124***	−1.002***	−1.001***	−0.984***
		(0.185)	(0.190)	(0.190)	(0.190)
age:stable					0.055**
					(0.021)
outgoing older|always both	1.531***	−0.234	0.176	0.264	−0.504
	(0.248)	(0.280)	(0.332)	(0.618)	(0.684)
always both|incoming older	3.727***	2.199***	2.675***	2.764***	2.075**
	(0.384)	(0.348)	(0.415)	(0.678)	(0.726)

Note: *p<0.05; **p<0.01; ***p<0.001

Table 5.11: Extension of the ordered logistic regression model in Table 5.7, adding individuals' perceived apparent time difference as an additional predictor for their responses to the question "Which of the two variants do you think is older?".

	Model	Res. df	-2LL	Test	df	LR	P(>Chi)
(0)	1	288	434.312				
(1)	stable	287	412.654	0 vs 1	1	21.658	<.001
(2)	apparentdiffN	287	384.260	0 vs 2	1	50.053	<.001
(3)	stable + apparentdiffN	286	378.250	2 vs 3	1	6.009	.014
(4)	age + stable + apparentdiffN	285	378.222	3 vs 4	1	0.029	.866
(5)	age:stable + apparentdiffN	284	370.925	4 vs 5	1	7.297	.007

Table 5.12: Pairwise comparison of the models in Table 5.11.

predicted perfectly by geographical location, with all four participants indicating categorical usage of the 'noo' variant (coded as 'outgoing' in the Figures) coming from the isle of Whalsay. The estimates of the community-wide usage consequently fell chiefly in the 'more incoming' category, with no clear internal patterning of responses along any of the participant variables (see Figure 5.7).

When it came to beliefs about the relative age of the two negation variants in question 3 we found that, despite the apparent absence of any evidence for it, almost half of the participants stated that they believed that the geographically limited negation variant was 'older' than its more widespread counterpart, as could be seen in Figure 5.9. This result is puzzling: the geographical distribution of the two variants is well-known to Shetland inhabitants, an assumption that was confirmed by the fact that 17 out of 77 participants used the optional comment field to point out that the variable realisation of negation patterned geographically and not by time, the dimension investigated by the questionnaire.

The distribution of perceived apparent time differences for our control revealed a similarly intriguing result: as the negation variable isn't changing we would have expected the distribution of difference scores to be centered around the zero mark. While Figure 5.10 showed that identical estimates for the 'older' and 'younger' speaker groups was indeed the most frequently reported result across all our participants, the overall distribution of responses is skewed towards reporting higher usage of the more widespread variant for younger speakers, with almost no apparent time differences reported in the opposite direction.

This raises the question of why speakers would extrapolate from a geographical pattern to history, or at least to the historical origin of variants. While there is currently no evidence that the relative usage of the more localised negation variant is decreasing (Jamieson, p.c.), it is imaginable that it shares a few meta-linguistic features with the outgoing variants of the changing variables. As the more 'local' variant, it could have gathered connotations of 'dialectality' and 'authenticity' that are typically associated with older speakers. This (possibly imagined) prevalent usage among older speakers could in turn give rise to both effects observed in the responses: firstly, it could be regarded to speak to the variant's historical primacy (Bailey, 2002), and secondly, the usage pattern by authentic 'older' speakers stands in natural opposition to younger speakers, explaining the reported differences in apparent time, an effect that was possibly amplified by the ordering of questions in the survey.

As it stands, it is not possible for us to identify the exact pathway of these socio-indexical connotations. Does 'authenticity' speak to both speaker group differences *and* variant history, or does one affect the other? While the responses to the changing variables provide evidence that apparent time differences are perceived independently, we can again not rule out that the responses to questions 4+5 were influenced so as to provide a post-hoc justification of the earlier answer regarding the variants' relative age. Conversely, the presence of seemingly non-substantiated beliefs about variant age for the control variable might indicate that, when forced to identify one variant as 'older', our participants might attribute historical primacy based on imagined age stratification even when they know that, synchronically, usage is determined by geography alone.

To avoid the same problem and aid in identifying the influence of either type of knowledge,

an improved methodology should therefore randomise the order of questions 3 vs. 4+5 between individuals.The two randomisations could then serve as separate conditions: measuring the effect of first explicitly drawing attention to either linguistic history *or* apparent time differences between participants would allow us to quantify what effect either meta-linguistic dimension has on the other.

5.4.3 Between-participant differences: age

Our results revealed several significant effects of the informants' age on their responses: firstly, age is a good predictor for participants' reported self-use of the changing variables, as reported in Section 5.3.2. In combination with the fact that age did not seem to affect participants' estimates of the community-level usage, age is consequently a good predictor for how much the respondents think they are 'ahead' or 'behind' their estimated community level usage, as measured by the relative difference between people's estimated self and community usage levels (Section 5.3.3). This suggests that individuals can correctly identify where, during an ongoing change, they are positioned relative to the rest of the community.

While Drager (2011) found that older speakers were more sensitive in their automatic compensation for ongoing phonetic changes, we find no such effect for the explicit reports on syntactic variables. For the changing variables, there is no significant effect of informant age on either the perceived 'age' of the variants or the perceived differences in apparent time (Section 5.3.5).

5.4.4 Between-participant differences: gender

The only significant effects of binary gender were found in the first two questions, i.e. the reported 'self' as well as 'other speakers' usage levels: females tended to report *higher* levels of the incoming variant both for themselves and the population as a whole. This tendency was present across all four sociolinguistic variables, whether changing or not. The raw data of the responses split by variable and gender is shown in Figure 5.11.

This gender effect does *not* translate to a higher difference between perceived self-usage and perceived community usage, i.e. females do *not* perceive themselves to be further 'ahead' the community than males do, the answers to both questions appear to be shifted in unison. Also, despite an often purported increased sensitivity to linguistic changes among females, neither the identification of the 'older' variants (Section 5.3.4) nor the strength of the perceived apparent time differences (Section 5.3.5) showed significant effects of participant gender.

5.5 Summary & conclusion

In this work, we presented a new methodology to assess and measure quantitatively whether people can report the direction of currently ongoing changes based on their awareness of the variable usage levels of categorical sociolinguistic variables. Results from 77 speakers show that people are reliably capable of reporting apparent time differences within their speech community, as well as provide evaluations of the age of the different linguistic variants that are compatible with the direction of the changes. Complementing earlier experimental results on
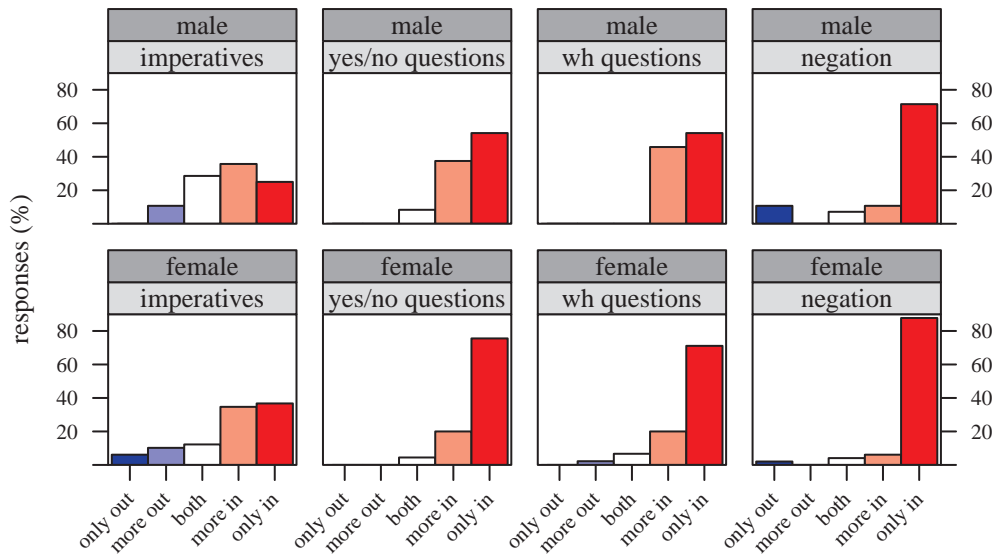
Figure 5.11: Raw data for responses regarding self-usage by sociolinguistic variable and binary gender. For the changing variables, females generally report higher level usage of the incoming majority variants.

individual speakers' use of their implicit knowledge of continuous sound changes, we obtained quantitative evidence for the explicit awareness of changes in three related morphosyntactic variables which are only attested at very low frequencies in everyday speech. Even though their low frequency means that we have no quantitative production data to compare to, the absolute usage levels reported by our informants reflect what is known about the relative progress of the three changes, indicating that naive individuals can learn about variable usage rates even from very limited data. While it is possible that explicit socio-indexical commentary on the variants could be underlying the shared knowledge about the direction of changes, the quantitative difference in responses between variables as well as the age-stratified distribution of self-usage assessments within variables suggests that the degree of knowledge about the variants cannot be straightforwardly reduced to qualitative connotations of specific variants as being 'older', 'more dialectal' etc.

Connecting the present work to related approaches, I briefly discussed the continuum of methodologies that are based on eliciting explicit judgments on language and language use from laypeople. Our extension of a simple and fast survey method originally employed by Labov (1966b) and Trudgill (1972) falls on a spectrum somewhere between current quantitative methods employed in syntactic research and the qualitative approaches to variation pervasive in the study of language attitudes. Our work is best understood as a contribution to the re-emerging field of *perceptual dialectology* dedicated to the study of naive language users' impression of language variation and change. This branch of research embraces the assumption that subjective evaluations, whether qualitative or quantitative, should not be discarded a priori but that they can complement and support evidence collected by other means in important ways. In Section 5.3.1 in particular we showed that subjective reports of relative quantitative usage correlate well with measures derived from grammatical acceptability judgments.

In the context of this thesis, the present work represents an empirical contribution to the study of individuals' capacity to detect and monitor ongoing changes in their community. As argued in Chapter 2, the *metalinguistic awareness* of innovations and changes in the individual provides a rich source for social biases which could underlie selection mechanisms at play in language change, such as the momentum-based selection bias presented in the previous chapter. While we complemented and extended existing work to morphosyntactic, categorical variables of low frequency, the study of individuals' sociolinguistic knowledge is still a new topic that will require much further work, particularly in respect to innovations and changes.

Chapter 6

Symmetric selection of asymmetric innovation

6.1   Connecting innovation and selection

In its opening chapters, this thesis provided a very general overview of the myriad 'pressures' and 'biases' that have been proposed to influence language change. Chapter 4 investigated a new selection pressure based on the idea of trend-amplification, with Chapter 5 contributing to the existing empirical evidence that such a momentum-based approach is a psychologically viable candidate mechanism for language change. Having honed in on one particular mechanism, this final chapter is intended to 'zoom out' again to connect the idea of momentum-based selection, a *symmetric* mechanism which does not predict any of the competing linguistic variants to be more prevalent than any other, with other pressures which might be responsible for the asymmetry underlying the uneven distributions and unidirectional paths that we find in real language changes. By combining a model of asymmetric innovation of variants with a symmetric selection mechanism we arrive at a more complete, if more complex, picture of the dynamics of language change.

Section 6.2 starts off with a recapitulation of the evolutionary framework of language change and provides theoretical motivations for why momentum could (and should) be relevant for language change research. In order to get a more systematic grip on different *types* of pressures, Section 6.3 will introduce the *Wright-Fisher model* of biological evolution and investigate it using the Markov model toolkit that was already used in Section 3.4.1. The main purpose of this is to allow us to study a toy model of the dynamics and interaction of the two basic evolutionary pressures, namely the *innovation* of variants as opposed to their *selection*. Having covered the basics of the Wright-Fisher model as well as its predictions regarding the interaction of the different pressures, Section 6.4 will provide an extension that implements a momentum-based selection bias. The point of the model is to show how the interaction of asymmetric innovation and a neutral selection bias like momentum can provide an explanation not just of the dynamics of language change, but also of the skewed synchronic distribution of linguistic variants that we observe cross-linguistically.

6.2   Language change, arbitrariness, replicator-neutrality

A core feature of momentum as it was presented in Chapter 4 was that it created directed selection pressures while being *replicator-neutral*, i.e. without favouring either of the competing variants a priori. Why is this particularly relevant to language change? When going back to the roots of modern-day linguistics, we can uncover a strong theoretical reason for why language changes should be replicator-neutral. One of the central tenets of Ferdinand de Saussure's *Course in general linguistics* (1916; 1959) is the principle of the *arbitrariness of the sign*. The principle states that linguistic signs, i.e. the association of a specific signifier (typically a sequence of speech sounds) to a specific signified (an object or concept referred to by using the signifier) is not 'motivated' by any (physical) semblance between the two. The linguistic sign is *arbitrary* in the sense that the association between form and meaning is primarily established on *social* grounds. According to this definition linguistics is chiefly the study of *conventional* as opposed to 'natural' associations and, while Saussure was explicitly referring to lexical con-

ventions, modern approaches to grammar which challenge the categorical distinction between lexicon and morphosyntax (e.g. Construction Grammar, Goldberg, 1995) suggest that his principle can be extrapolated to morphosyntactic constructions (such as syntactic constituent order) as well.

Even though the principle of arbitrariness has been challenged based on work related to iconicity and sound-symbolism (see Perniss et al., 2010, for a review), it has maintained its place as a cornerstone of linguistics to this day (Joseph, 2000). But how is the principle of arbitrariness relevant to language change? While Saussure's distinction of the 'synchronic' versus the 'diachronic' study of language was also the origin of the schism between research on the two topics that characterised much of 20th century linguistics discussed in Chapter 2, the arbitrariness of the sign also speaks to diachronic processes. As made explicit by evolutionary frameworks, the linguistic conventions we find today are there because they have come about and spread through historical processes, in particular their replication and selection. So, if we want to maintain that the conventions we find synchronically are in fact *arbitrary*, we have to believe that their diachronic development, and in particular any *selection* of specific conventions over their competitors, occurred on arbitrary grounds as well (but see Joseph, 2015, p.93 for an argument why iconicity and arbitrariness are not contradictory to begin with).

Here we find, once again, the diverging priorities of researchers working on language change on the *macro* and *micro* level, as well as the potential disconnect between the two approaches. While skewed macro-level patterns and distributions across languages will be taken to suggest some degree of non-arbitrariness or motivatedness in the selection of specific linguistic traits, researchers working on an individual change (which might or might not conform to the larger scale patterns) will have very different, particular criteria to determine the presence (or cause) of selection in their micro-level data. This disconnect between the two explanatory levels of language change is acknowledged very explicitly by Ohala when he states that

> As for the immediate triggers of sound change in a particular language at a particular time, I will have little to say about them except to suggest that these things are bound to happen and that it is not so interesting to try to identify them. (Ohala, 1989, p.174)

This dismissal of the (study of) factors responsible for triggering individual changes could be taken to indicate a certain degree of frustration about the struggle to unify the reliable and frequently attested large-scale patterns across languages with the idiosyncracy of individual changes.

Unlike Ohala, I suggest that the shortcomings in the prediction of individual changes can also be framed in constructive terms, as a test for competing accounts of language change. What the theoretical arguments presented above as well as the empirical evidence suggest is that a good theory of language change should predict *that* changes happen, *which* changes are more likely to happen than others, but all the while accounting for the sporadic nature of changes, i.e. not making overspecific claims about being able to accurately predict which changes are going to happen where. What this theory should be able to predict is not just language changes themselves, but also *the degree of their unpredictability*, i.e. it should have a

theory of and explanation for its own limitations.

This is exactly the promise made by the evolutionary, replicator-based stance taken by Labov and Croft which decouples the two processes of *innovation* of linguistic variants as opposed to their *selection.* While the function-driven repeated innovation of similar variants across languages can account for the distribution or likelihood of different types of changes overall, as measured by their cross-linguistic 'actuation probability', the actuation of specific changes is explained by arbitrary (social) pressures that *explicitly underspecify* predictions about whether a specific change is about to happen in a language or not.

While the advantages of this 'division of labour' have long been argued for theoretically, the approach has not seen wide adoption in the empirical and modelling literature, as I discussed in Section 2.5. Arguably, this is largely due to the absence of concrete candidate mechanisms for the second, arbitrary selection step. As was argued in Section 2.2.5, the default account of socially driven change invokes the concept of sociolinguistic *prestige*, which is based on asymmetric selection biases that can only be posited *post-hoc* on an individual basis. The subsequent lack of testable predictions made by social accounts that would generalise across changes means that, on the macro-level, asymmetric social pressures have largely been disregarded at the expense of equally replicator selection-based functional accounts of language change, which do however make a priori predictions.

The momentum-based selection account presented in this thesis constitutes a more mechanistic explanation of socially-driven changes beyond 'prestige' accounts. With this new mechanism in mind, the remainder of this chapter is dedicated to demonstrating based on a toy model how the interaction between innovation and selection biases more generally, as well as between asymmetric innovation and momentum-based selection in particular, can satisfy the requirements of a theory of language change outlined above.

6.3   Modelling the interaction of different pressures

In this section I will present a simple model to investigate how a symmetric momentum bias, a selection bias favouring trending conventions whether they are beneficial or deterious, interacts with asymmetric *generation* of variants. The model is an extension of the Markov model of Bayesian Iterated Learning that was analysed in-depth in Section 3.4.

To recapitulate, Reali and Griffiths (2009) proposed a model of regularisation by Iterated Learning. In their model, a Bayesian learner infers the underlying relative frequencies of several competing variants based on a sample of productions they observe from a teacher. By iteratively passing one learner's output on to the next and analysing the *stationary distribution* of this Markov chain, they showed that by setting the parameter α to values that lead individual learners to slightly increase the variability of their own productions, the chains end up *regularising* the input distribution: over time, the population ends up mostly producing input distributions where one of the competing variables is used categorically.

Moreover, in Reali and Griffiths (2010) the authors showed that under some circumstances this model of Bayesian Iterated Learning is equivalent to the Wright-Fisher model of biological evolution with mutation (see e.g. Hartl and Clark, 2007). The Wright-Fisher model is a general

tool from population genetics which is used to predict the expected change in the frequency of competing alleles in a population in the presence or absence of different pressures such as *mutation* and *selection*.

6.3.1 The Wright-Fisher model

In its very simplest form the Wright-Fisher model (for its original formulation see Wright, 1931) describes the dynamics of competition between two variants in a finite population of size N. Call the two variants 0 and 1 and their respective absolute frequencies n_0 and n_1. The entire population of N individuals is replaced at discrete time steps so that it is always the case that $n_0 + n_1 = N$. To simplify notation and assuming a constant population size N, we can again drop the indices referring to individual variants: we will henceforth write n_t to mean n_1 at time t, from which the frequency of the competing variant can be trivially computed.

The state of the population can be described simply by the relative frequency x of that variant, i.e.

$$x_t = \frac{n_t}{N} \ . \tag{6.1}$$

The Wright-Fisher model assumes that the individual generations of the population as it evolves over time are *non-overlapping*. In other words, the generation makeup at the following timestep, n_{t+1}, is determined by creating a new population of N tokens, each generated by replicating a randomly selected 'ancestor' from the previous generation. The probability distribution over the likely frequencies n_{t+1} in the next generation is consequently distributed according to a Binomial distribution

$$n_{t+1} \sim Bin(N, f(x_t)) \ . \tag{6.2}$$

In the absence of any other pressures, the probability of replicating an instance of a particular variant is simply equivalent to the relative frequency of that variant in the previous generation, i.e.

$$f(x) = x \ . \tag{6.3}$$

The behaviour of this simplest form of the Wright-Fisher model describes the dynamics of completely neutral drift which was not only fundamental in informing the *neutral theory* of genetic evolution (Kimura, 1968, 1983), but still forms an important baseline for evaluating the presence or absence of evolutionary pressures in empirical data sets to this day.

Without any mechanisms that influence the selection of existing variants or mutations that would introduce new ones, this is a very simple model of the *diffusion* of traits through replication. Before we turn to more complex versions of the Wright-Fisher model which incorporate the influence of mutation and selection pressures, it is insightful to have a closer look at the similarity between diffusion models from biology and the Utterance Selection Model as described in Section 3.2.

Relationship to the Utterance Selection Model (USM)

The similarity between Equation 6.2 and the data production function of the USM in Equation 3.16 is not just superficial: in its most general formulation the Wright-Fisher model is in fact identical to the trivial case of a single USM agent with a learning rate $\alpha = 1$ and USM sample resolution $T = N$ that is engaged in a production-perception loop.

More fully-fledged versions of the USM with populations of speakers also find parallels in more complex models of biological evolution. Blythe (2007) showed how the learning and alignment dynamics of the USM are in fact identical to Wright's *island* model, an extension of the simple Wright-Fisher model above that is used to study the diffusion of variants between subdivided populations with limited migration between them (Wright, 1943).

While models of pure diffusion in complex population structures are interesting in themselves and many results concerning fixation times and fixation probabilities can be derived from them analytically (see e.g. Imhof and Nowak, 2006; Baxter et al., 2008; Blythe, 2012b; Michaud, 2016), we shall return to the simpler model of just one population with fixed turnover, which greatly simplifies the analysis of different innovation and selection pressures that we are particularly interested in here.

The Wright-Fisher model with mutation

Under the Wright-Fisher model with *mutation* (henceforth referred to more generally as *innovation*), the probability of producing an instance of variant 0 at the next generation is not completely equivalent to its current relative frequency x. Instead, with probability μ_1, any of the n_0 type 0 variants present in the population can spontaneously mutate into an instance of variant 1 during its replication. Conversely, there is a probability of μ_0 that any of the instances of the other variant (of which there are n_1) will spontaneously mutate into one of variant 0. Under this assumption, the relative probability of producing an instant of variant 1 at the next generation is

$$f(x) = x \cdot (1 - \mu_0) + (1 - x) \cdot \mu_1 . \tag{6.4}$$

It is this version of the Wright-Fisher model that is mathematically equivalent to Reali and Griffiths's model of Bayesian inference, in particular to the version of the model where the learners derive a hypothesis from their input sample of size N by deterministically adopting the mean $\hat{\theta}$ of the posterior distribution $p(\theta|x)$ as their production probability. In this case, the learner's production behaviour is identical to that of a population of N individuals who are replicating with innovation rates set to

$$\mu_0 = \mu_1 = \frac{\alpha}{2 \cdot (\alpha + N)} . \tag{6.5}$$

Reframed in terms of biological evolution, the values of the Bayesian model parameter that are associated with regularisation behaviour, $\alpha < 1$, all satisfy the limit of low mutation, $\mu \ll 1/N$ (Tarnita and Taylor, 2014).

The Wright-Fisher model with innovation allows for pressures in favour of specific variants, by setting the innovation rates $\mu_1 \neq \mu_0$. But it does not support *selection* of variants, as would

be needed for a replicator or momentum-based selection bias.

The Wright-Fisher model with innovation and selection

To allow selection on top of innovation, we have to move on to the Wright-Fisher model with innovation and selection, which takes the form

$$f(x) = \frac{x \cdot (1+s) \cdot (1-\mu_0) + (1-x) \cdot \mu_1}{x \cdot (1+s) + 1 - x} \ . \tag{6.6}$$

The parameter $s \geq 0$ in this equation represents a *selection coefficient* which causes the n_1 tokens of variant 1 that are present in the population to be preferentially selected, i.e. their likelihood of replication is increased at the expense of the competing variant. In contrast to the innovation pressures μ, the effectiveness of the selection coefficient depends on the current prevalence of the selected variant in the population. In particular, the selection pressure in favour of an advantageous variant is completely ineffective as long as no tokens of that variant are present in the population. In other words, the selection coefficient cannot introduce new variants. This can be appreciated best when we consider the force of selection in isolation, by setting the probability of randomly generating new variants $\mu_0 = \mu_1 = 0$. In this case we arrive at the Wright-Fisher model with selection only,

$$f(x) = \frac{x + s \cdot x}{1 + s \cdot x} \ . \tag{6.7}$$

6.3.2 Wright-Fisher model dynamics for infinite population size

At this point we can already get a glimpse of how these two different types of pressures differ in their dynamics as well as how they interact. Assuming an infinite population size, random sampling effects are completely washed out and we recover an idealised image of the dynamics of the Wright-Fisher model.

The dynamics of innovation

Figure 6.1 shows the relative impact of innovation and selection under the assumption of an infinitely large population size. The graphs plot the relative change Δx to the frequency of one of the variants as a function of its current prevalence x in the population. As can be seen from panels (i) and (ii) in Figure 6.1a the impact of innovation, as measured by how much it affects the current frequency of variants, is stronger when there are fewer tokens of the innovated variant in the population. More than anything else, innovation is a pressure *away* from homogeneous population states.

The dynamics of selection

The dynamics of selection look very different: from Figure 6.1a(iii) we can see that selection is ineffective as long as there are no instances of the selected-for variant present, and it acts strongest when there is *most* variation – in the case of two competing variants this is when both variants are equally represented in the population.

The panels underneath in Figure 6.1b demonstrate how these pressures affect the distribution of variants over time in the limiting case of an infinitely large population, an idealised condition in which the replication dynamics are unaffected by the noise of random sampling. Asymmetric innovation pressures lead to rapid spread of the frequently innovated variant initially, but the growth then tails off towards some asymptotic frequency, which corresponds to the points in the upper panels where the line indicating the relative change to the variant frequency crosses the y-axis, i.e. where $\Delta x = 0$.

In contrast to this *r-shaped* growth, for the temporal dynamics of the selection pressure shown in Figure 6.1b(iii) we recover the logistic growth pattern that we also found for the replicator selection regime of the Utterance Selection Model in Section 3.2. Also, unlike innovation pressures, the function indicating the relative impact of selection in Figure 6.1a(iii) crosses the $\Delta x = 0$ line twice, meaning that the selection regime possesses two stable, asymptotic states, at $x = 1$ as well as $x = 0$. Unlike the innovation pressure, the s-shaped growth pattern cannot escape the variationless state at $x = 0$ but requires some low level of variation in order to 'kick in'.
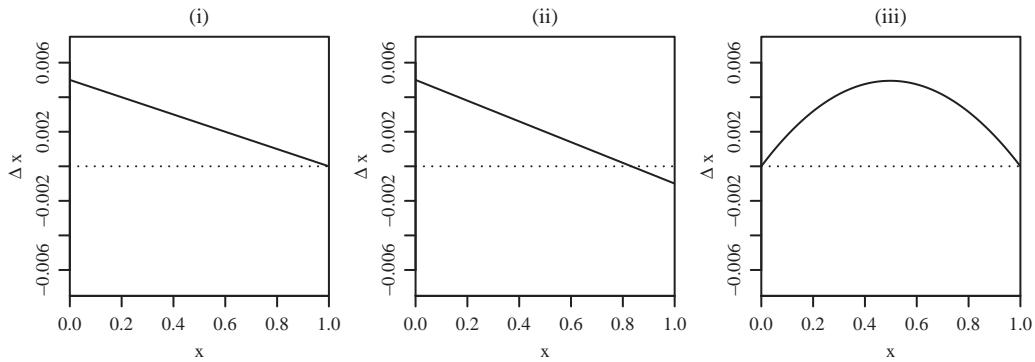
(a) Influence of innovation and selection pressures as a function of the current population makeup. *(i)* innovation of one variant only, $\mu_1 = 0.005$. *(ii)* innovation of both variants at different rates, $\mu_1 = 0.005, \mu_0 = 0.001$. *(iii)* selection only, $s = 0.02$.
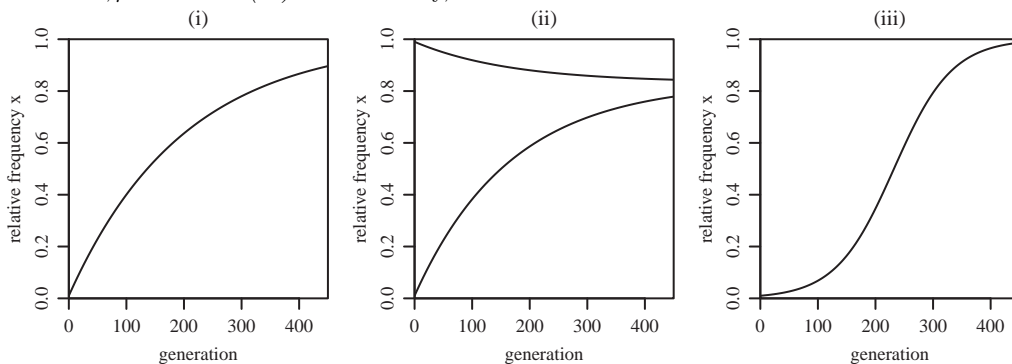
(b) Temporal development of the population given the corresponding pressures from the panels directly above, assuming initial states $x = 0.01$ and 0.99.

Figure 6.1: Dynamics of innovation and selection in the Wright-Fisher model.

The interaction of innovation and selection

Now that we have an idea of when the different types of pressures of innovation and selection affect the evolution of a system most strongly, we can go on and ask how the two interact, i.e. what happens when both pressures apply simultaneously?

Assuming the same constant selection pressure as above, there are still three different scenarios of interaction to consider. Figure 6.2 shows the infinite population size dynamics assuming (i) selection of a variant that is also preferentially innovated, (ii) selection on top of *unbiased* (symmetric) innovation, and (iii) selection *against* a preferentially innovated variant.

In all cases we can see that, while innovation of the selected for variant helps it spread more quickly open first introduction, innovation of the competing variant actually stops the selection pressure from taking over the entire population. Only as the innovation rates μ diminish towards zero do we recover the logistic growth dynamics of selection that we saw above, which lead to complete dominance by the preferred variant in the limit.

While this simple analysis gives us an idea of when innovation and selection are expected to impact most strongly on an evolving systems dynamics', understanding the *interaction* of these different types of pressures in finite populations requires more in-depth study for which we will again turn to the Markov model framework.

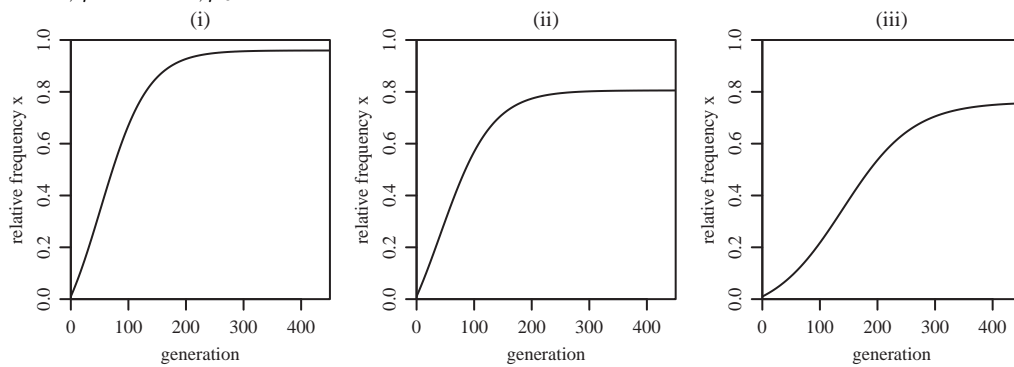## 6.3.3 Innovation and selection in finite populations

While the assumption of an infinite population size allows for an idealised study of the effects of innovation and selection as well as an idea of when they balance each other out, the instances of biological and cultural evolution that we can observe empirically all play out in finite populations, where changes to the distribution of variants over time are subject to random sampling effects. In order to disentangle the relative roles of the two pressures of innovation and selection in finite populations, we will make use of some of the same analytical tools that were already used to study the model of Reali and Griffiths in Section 3.4.1.

Following the same order of pressures as above, we will be investigating their dynamics using two different tools: firstly, in order to get an idea of the relative frequencies at which the competing variants are present across time we will again be looking at the *stationary distributions* of the Markov chains that correspond to the Wright-Fisher models with the respective innovation and selection parameters.

Like before, we will focus on parameter combinations which lead the models to preferrably occupy states of (near-)categorical usage of either of the variants, as is the case for low mutation rates. Since we are again also interested in the nature of the *transitions* between those two extreme states, we will have to move beyond merely looking at the stationary distributions, since this type of analysis generalises over the temporal aspect of the model dynamics. To get an idea of the nature of the transitions we will primarily look at the *completion probabilities* of actuated changes, as measured by the probability of the incoming variant *diffusing* through the entire population. To explain the concept, we will start with the simplest possible model of pure *diffusion* in which both of the those pressures are completely absent.
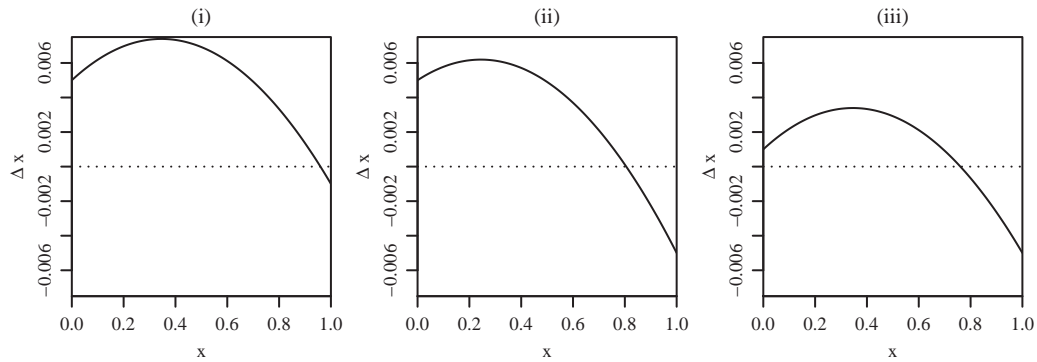
(a) Influence of innovation and selection pressures as a function of the current population makeup. All panels assume selection of the incoming variant at $s = 0.02$. *(i)* preferred innovation of the selected for variant, $\mu_1 = 0.005, \mu_0 = 0.001$ *(ii)* symmetric innovation of both variants, $\mu_0 = \mu_1 = 0.005$ *(iii)* antagonistic pressures, where selection competes against the preferentially innovated variant, $\mu_1 = 0.001, \mu_0 = 0.005$.

(b) Temporal development of the population given the corresponding pressures from the panels directly above, starting at initial state $x = 0$.

Figure 6.2: Dynamics of the interaction of innovation and selection pressures in the Wright-Fisher model.

Diffusion without innovation or selection

In the absence of any innovation of new variants, the two extreme states that correspond to categorical usage of either variant are *absorbing* states. Starting off with a population that exhibits variation, repeated replication of a finite population will eventually lead one variant to *diffuse* through the entire population, with all of the other competing variants being eliminated. The exact probability of either variant diffusing across the entire population depends on the initial state of the population. Figure 6.3a shows the probability distribution across all model states, based on three different initial conditions. This distribution, computed through numerical simulation, shows that the probability of a variant diffusing is equal to its initial frequency, a finding that is in agreement with analytical results (Clifford and Sudbury, 1973).

A more efficient way to look at this is by simply considering the *diffusion probability* of the incoming variant given a range of initial states, all in one plot. The left panels in Figure 6.3b show the diffusion probability for pure neutral drift given different starting states of the model for increasing population sizes. The probabilities lie on the diagonal of the chart in each case, showing that the result that the probability of diffusion is identical to the variant's initial relative frequency is independent of population size.

The dynamics of selection

In contrast to purely neutral diffusion, the presence of a selection bias for individual variants (a *replicator bias*) leads to a systematic increase of the selected variant as soon as it is introduced. How a selection pressure favouring the variant alters the dynamics of diffusion can be seen in Figure 6.3b. In contrast to the pure diffusion case ($s = 0$), a positive selection coefficient causes the relative probability of the selected for variant to diffuse to the entire population to *increase* above its baseline probability which is indicated by the dotted line. While a higher selection coefficient s increases the diffusion probability of the variant, the reliability of the selection bias in terms of guaranteeing the variant to win out depends crucially on the size of the population: the smaller the population, the more the dynamics are influenced by random sampling which, in the absence of mutation, increases the chance of driving even a positively selected for variant to extinction. Particularly for larger population sizes a sufficiently high selection coefficient can almost guarantee the diffusion of the variant, as long as it manages to avoid extinction during the fragile initial low frequency region.

Symmetric innovation as (de-)regularisation

Before considering the case of interacting pressures, we should also investigate the dynamics of the spontaneous innovation of variants acting by itself. The case of symmetric innovation, where the probabilities of spontaneously introducing either of the competing variants are equal ($\mu_0 = \mu_1$), was shown to be equivalent to Bayesian Iterated Learning chains of *averaging* learners by Reali and Griffiths and an in-depth quantitative analysis of the dynamics under such settings was provided in Section 3.4. To briefly recapitulate, under moderately low innovation rates the stationary distribution is similar to the one of the pure diffusion model, only that the non-zero probability of randomly producing an unattested variant means that variation is never fully
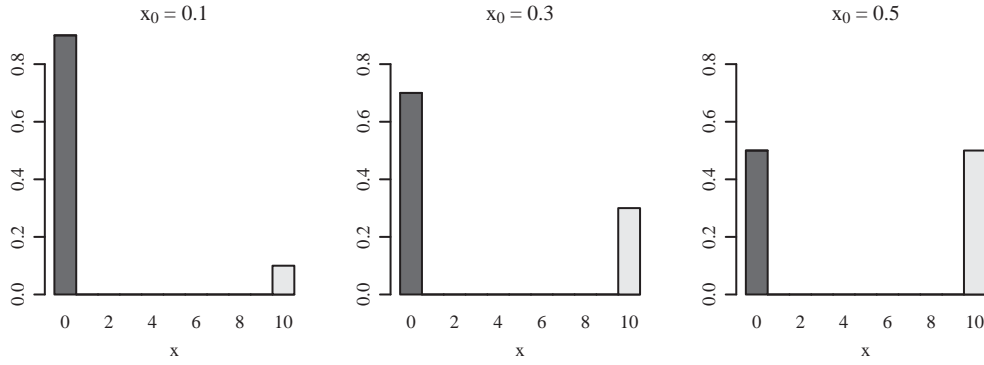
(a) Stationary distribution of diffusion without selection given different initial states x_0.
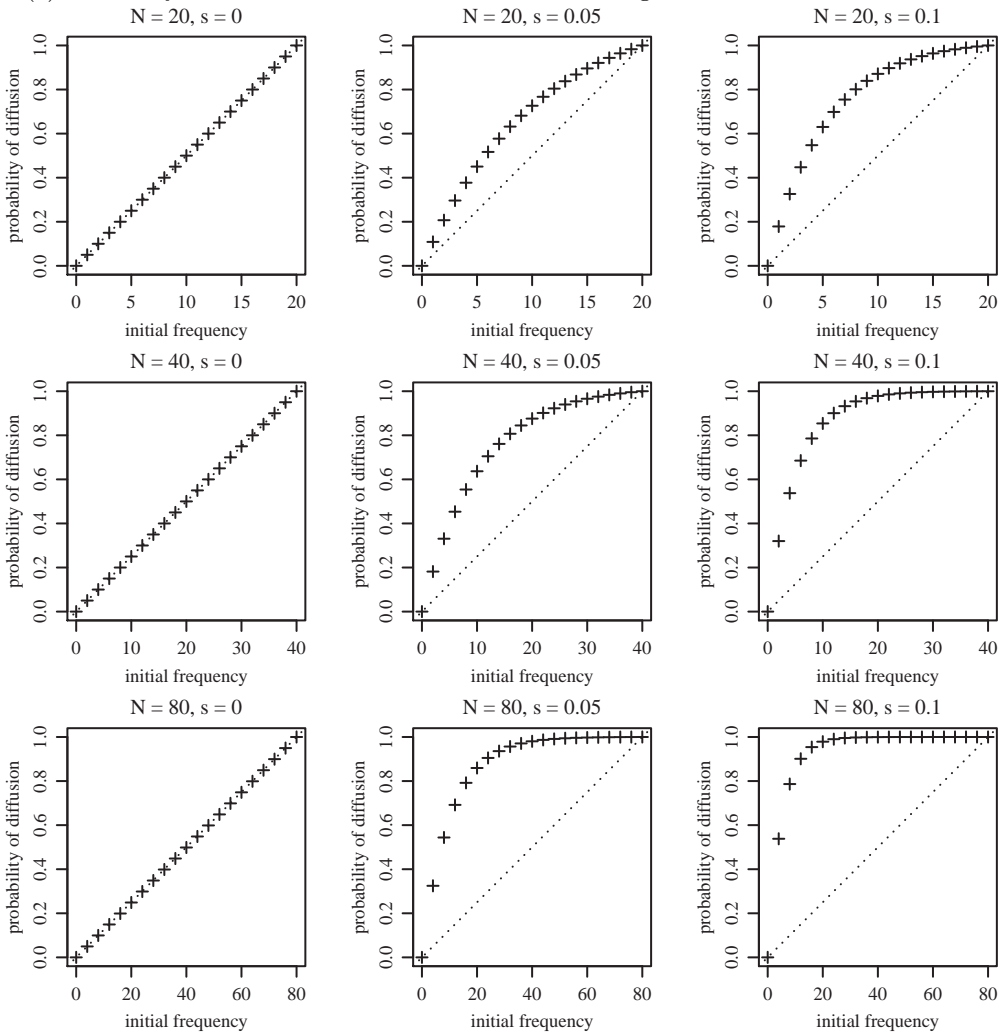
(b) Diffusion probabilities for different strenghts of the selection coefficient s and population sizes N. The baseline diffusion probabilities given purely neutral drift are indicated by the dotted line.

Figure 6.3: Diffusion outcomes for different population sizes N and selection coefficients s.

eliminated from the system. High innovation rates on the other hand will lead the populations to mostly consist of an even mix of all possible variants, with the exact cutoff point between the behaviours depending both on the innovation rate μ and the population size N as well as the relationship of the two (see Section 6.3.1).

A summary of the different regimes of the stationary distribution is shown in Figure 6.4a. It should be noted that, given the same innovation rate μ, the shape of the stationary distribution depends on the population size (or, in Reali and Griffiths's framing, the size of the sample provided to the learner). Calculating the innovation rate based on a regularisation parameter α which scales with the population size according to Equation 6.5 offers a more reliable way of picking an innovation rate that leads to the regime of stationary distributions that are of interest to us, i.e. ones that primarily occupy states of categorical usage of either competing variant. For the remainder of the analysis we will therefore be specifying innovation rates using this parameter α.

In order to calculate either variant's probability of diffusion in a model with innovation, we have to adjust the definition of *diffusion* slightly. Whenever there is a non-zero probability to spontaneously innovate either of the variants, neither of the model states corresponding to categorical usage of a variant are absorbing states, since even a variant that's been eliminated from the population can always be innovated anew and spread to completion. To nevertheless capture the influence of initial states, we define the diffusion probability of the incoming variant simply as the probability of first reaching a state of categorical usage of that variant, as opposed to first reaching a state where the variant is completely unattested, and we will be using this definition for the remainder of the analysis.

The relative probability of first reaching either state of categorical usage given low symmetric innovation rates is shown in Figure 6.4b. As already indicated in Section 3.4, Reali and Griffiths's model exhibits unusual behaviour for a model of regularisation in the sense that the probability of diffusing to the currently *less* attested variant is actually slightly *raised* above the baseline probability given by neutral diffusion, a baseline which is approached as $\alpha \to 0$.

Asymmetric innovation as a model of the accumulation of errors

While symmetric innovation probabilities can be conceived of as a bias for regularisation, the probabilities can also be set unequal to introduce a bias that favours one of the variants over the other. This is an alternative way to introduce asymmetry between variants, apart from the selection bias discussed above. This configuration of the Wright-Fisher model parallels one of the earliest and simplest theories of language change discussed in Chapter 2: the assumption that a linguistic variant is accidentally 'misproduced' as another some of the time, for example due to coarticulation effects, maps neatly onto a scenario with moderately low mutation rates, where the innovation probability in one direction vastly outweighs the one in the other.

Such changes through the *gradual accumulation of errors* are more typically considered for phonetic changes in a *continuous* dimension, but the same process could also apply to categorical variables, such as is the case for syntactic patterns that are in competition. While models of (syntactic) grammar competition typically consider more complex (external) triggers for actuation, we can nevertheless study the dynamics of the spread of variants assuming that
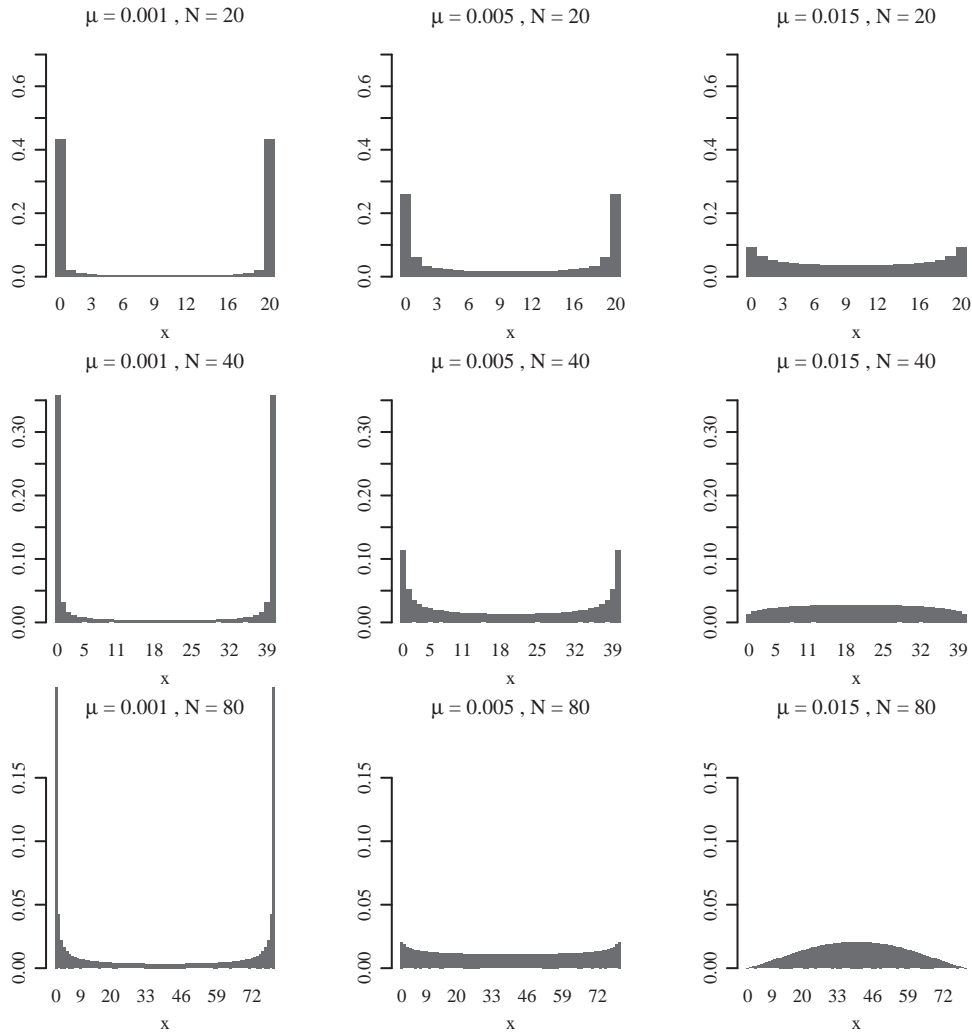
(a) Stationary distributions of the Wright-Fisher model under different symmetric mutation rates $\mu_0 = \mu_1 > 0$ and population sizes N.
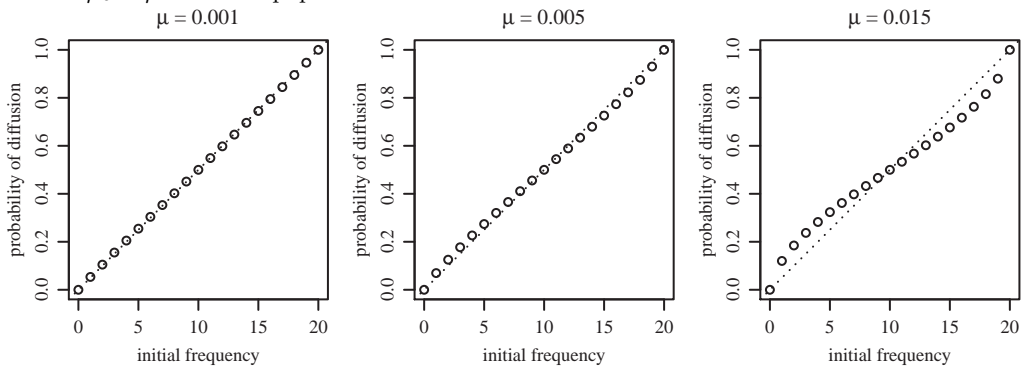
(b) Probability of the incoming variant diffusing through the entire population as a function of the current frequency of the variant.

Figure 6.4: The dynamics of symmetric innovation in the Wright-Fisher model.

asymmetric innovation probabilities are in place in which the incoming variant is preferrably innovated, i.e. $\alpha_1 > \alpha_0$.

In order to avoid raising the preferred innovation rate to unrealistically high levels, all comparisons between different degrees of innovation asymmetries will be based on holding the preferred innovation probability α_1 constant and setting the opposite rate α_0 to a diminishing fraction of the baseline rate. In this way, the degree of asymmetry can simply be expressed by the fraction α_1/α_0, with higher numbers corresponding to greater asymmetries.

That even small asymmetries in the innovation probabilities can have a strong effect on the expected synchronic distribution of variants can be seen from the stationary distributions in Figure 6.5a, showing domination of the preferrably innovated variant to become increasingly more likely as its relative likelihood of innovation over competitor variants increases. While this suggests that asymmetries in innovation could be just as influential as an asymmetric selection bias, we also find evidence against the strongly directed nature of individual transitions predicted under the infinite population size assumption in Section 6.3.1. Under this simplifying assumption it was demonstrated that changes in one direction actuate instantly and are not s-shaped, but rather exhibit *r-shaped* or *logarithmic growth* that starts off rapidly but slows down as it approaches saturation.

However, a look at the diffusion probability of the incoming variant in Figure 6.5b shows that the directedness of growth derived from the infinite population size approximation above is not actually representative of the dynamics of innovation in finite populations. In contrast to asymmetries through selection, diffusion probabilities do not stray far from the neutral evolution baseline, suggesting that typical transitions are very similar to those of pure diffusion, i.e. noisy and far from strongly directed. While the stationary distribution shows that this model is much more likely to remain in the categorical preferred variant state for a long time, the diffusion probabilities show that it typically takes a large number of initiated transitions before one actually succeeds in spreading to the entire population, even when the variant has already managed to spread to a large part of it. At least as long as the innovation rates are relatively low in absolute terms, even an increase in asymmetry does not significantly raise the probability of the favoured variant diffusing without interruption.

A comparison of the dynamics of selection and innovation in Figures 6.3 and 6.5 highlights that, while the expected stationary distributions of these two modes of asymmetry are very similar in that they predict a strong *dominance* of the preferred variant, the typical trajectories of each might differ wildly. The latter point is of particular interest since, as discussed above, the population size can interact with the effectiveness with which different asymmetric biases express themselves in the evolutionary dynamics. As a next step we can therefore analyse the effect that a *bottleneck* as implemented by a constrained population or sample size has on the two types of pressures as well as the interaction between the two.

6.3.4 Innovation, selection and the bottleneck in Iterated Learning

As could be seen above, the *size* of a population plays a crucial rule in how effective different kinds of pressures are. In the same way that completely neutral *diffusion* of a variant is much

(a) Stationary distribution

(b) Probability of the incoming variant diffusing through the entire population as a function of the current frequency of the variant.
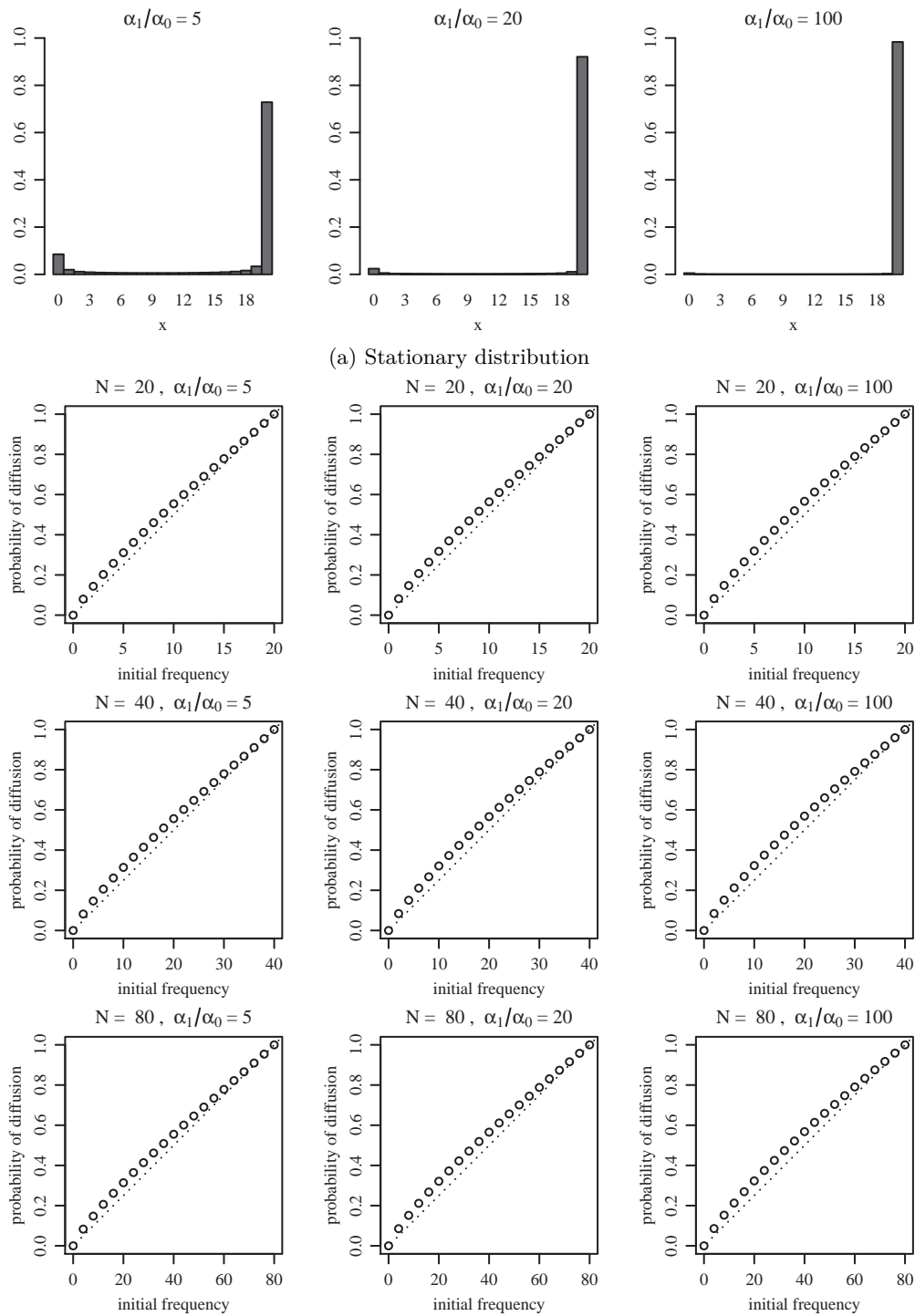
Figure 6.5: Dynamics of the Wright-Fisher model with asymmetric innovation rates. The innovation probability of the biased for variants is held fixed at $\alpha_1 = 0.2$.

more likely in smaller populations, the relative power of one and the same selection bias declines the smaller the population gets. This characterisation casts new light on the workings of the Iterated Learning Model (ILM) that was discussed in Section 2.2.3, particularly on the role of the *bottleneck*. One of the core ideas of the ILM is that, by imposing a limit on the amount of information or learning data that is transmitted between generations of learners, biases become exacerbated and express themselves more quickly in the linguistic systems produced by iterated learning chains.

The purpose of this section is to attempt to shed light on the bottleneck result by teasing apart the relative impact of innovation and selection pressures under the effect of drift. The motivation for this is that, while several results showing the mathematical equivalence between Bayesian inference and specific models of biological evolution exist (Reali and Griffiths, 2010; Harper, 2009), there are still important conceptual differences between the two types of models. In the Bayesian approach any and all types of bias or pressure are amalgamated in the prior distribution $p(h)$, as well as to some degree in the production probabilities $p(d|h)$. This formulation does not allow us to distinguish between biases for innovation and selection in a principled way, as can be done in more biologically-minded frameworks such as the Wright-Fisher model.

Based on Reali and Griffiths's demonstration of the equivalence between their iterated learning chain of *averaging* learners and the Wright-Fisher model, we want to investigate the respective role that innovation and selection pressures have as a bottleneck is imposed. To do so we will compare the *relative effectiveness* of different types of asymmetric biases given different values of the population/sample size parameter N. We will generally limit ourselves to instances of the model with low innovation rates which exhibit regularisation behaviour, i.e. $\alpha \ll 1$. As could be seen from the various stationary distributions above, asymmetric pressures in combination with diffusion through a finite population typically lead to situations where either one or the other variant prevails in the population. Based on these stationary distributions we can calculate the average relative frequency that the preferred variant has over the other in terms of its expected synchronic distribution,

$$\hat{x} = \sum_{i=0}^{N} \frac{i}{N} \cdot \pi_{x=i} \ , \tag{6.8}$$

where π again refers to the stationary distribution as defined earlier in Equation 3.23. Since we are only looking at cases with low innovation rates α corresponding to regularising behaviour, the stationary distributions are bimodal, with the populations mostly remaining in regions of near-categorical use of either of the competing variants. This means that intermediate values of $\hat{x}$ that we compute do not indicate the frequency at which the selected for variant is *typically* present in the population, but rather *how much of the time* that variant is used near-categorically, as opposed to its competing variant. If the replication of the two variants was completely unbiased we would expect the average frequency to be at the 0.5 mark meaning that, across time, we would find both variants to be equally frequent. The further away the average frequency moves from this neutral mark, the more *effective* that bias is at expressing itself in the expected synchronic distribution of variants. Based on this measure we can now compare

how the two different types of asymmetric biases are affected by different population/sample sizes N.

The effectiveness of selection

The first type of asymmetric bias considered here is that of *selection*, which is implemented by setting the selection coefficient $s > 0$. This asymmetric pressure applies on top of some low *symmetric* innovation probability corresponding to $\alpha/2 = 0.1$, which we impose to stop both variants from simply diffusing to the entire population. As can be seen in Figure 6.6(i), in smaller populations the impact of selection pressures is increasingly *reduced*, with the exact point at which the effect of the bias starts to falter depending on the magnitude of the selection coefficient. As the population size decreases towards its minimum at $N = 1$, the expected frequency of the selected for variant approaches 50%, indicating no preference for either of the competing variants.

This effect can be explained based on the dynamics of selection demonstrated above: when modelling discrete (quantitative) traits, selection pressures rely on *variation* in the population and act most strongly when variation is highest. As the population size decreases the impact of neutral *drift* through random sampling effects increases, diminishing the force of selection that is reliant on the relatively stable maintenance of variation.
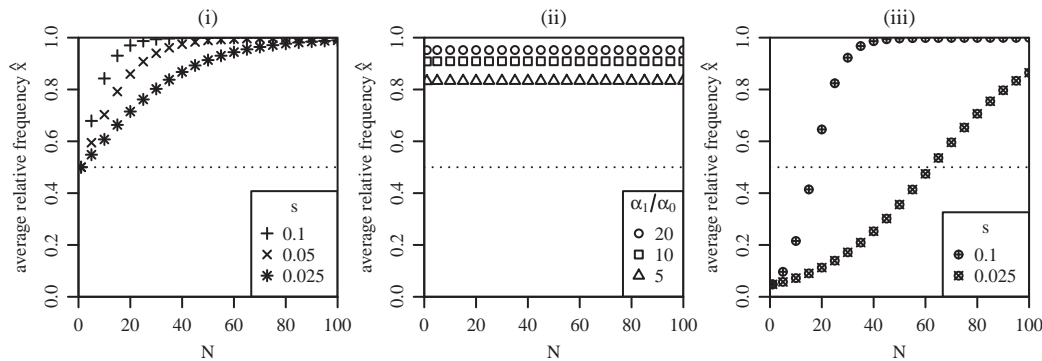
Figure 6.6: Relative frequency of variants as a function of population size given different regimes of innovation and selection pressures. *(i)* selection of symmetric innovation, $\alpha = 0.01$ *(ii)* asymmetric innovation only, $\alpha_1 = 0.01$ *(iii)* asymmetric innovation with selection against the preferentially innovated variant, $\alpha_0 = 0.01$, $\alpha_1 = 0.0005$.

The effectiveness of asymmetric innovation

It was shown above that asymmetric innovation rates which favour the spontaneous production of one of the variants can have a strong effect on their expected synchronic distribution. Figure 6.5 already indicated that the influence of innovation is robust to changes in population size, a result that is also borne out by the present effectiveness measure. Rather than being affected by smaller population sizes, Figure 6.6(ii) shows that systems tend to converge towards some consistent low usage level of the less frequently innovated variant, with that usage frequency

dependent on the two innovation rates, but not population size[1]. In other words, unlike with selection, asymmetric pressures due to biased *innovation* appear to be robust to bottlenecks.

The interaction of innovation and selection

While we have discovered a change in the effectiveness of selection pressures for reduced population sizes, neither of the pressures taken individually has seen an *increase* in how much it is expressed synchronically, so how can the bias-amplifying effect observed in iterated learning chains be explained? Figure 6.6(iii) shows the expected frequency of variants based on pressures of innovation and selection that work *antagonistically*: while the selection coefficient favours the incoming variant at the rates indicated, there is also an underlying asymmetry in the generation of variants that preferably innovates the competing variant. In this configuration we see an inversion of the frequency distribution at a population size that is again dependent on the strength of selection. Where for larger population sizes the selection pressure prevails, the innovation pressure is more robust and thus wins out whenever a sufficiently small bottleneck is imposed.

The present analysis of the relative force of innovation and selection pressures sheds the amplification of biases in iterated learning in a new light: adopting Reali and Griffiths's equivalence between the Wright-Fisher model and iterated learning chains, the imposition of a bottleneck corresponds to a reduction of population size in the biological sense. Based on our characterisation of the effectiveness of different asymmetric pressures in the Wright-Fisher model I conclude that, at least when it comes to the general interpretation of bottlenecks causing an amplification of biases, the biases that see themselves relatively amplified during iterated learning experiments are more likely of the *innovation*, rather than the *selection* type, which see a decrease in their impact.

Most studies in the Language as a Complex Adaptive System tradition are based on the assumption that biases which become exacerbated in simulation and experimental settings are of the *selection* type (e.g. Kirby 1999, ch.6, Chater and Christiansen 2010), without taking into account the effect that other evolutionary pressures, in particular innovation and drift, have on the systems under investigation (see Henrich et al., 2008, p.127-129).

Recall that the original finding of iterated learning models was that the bottleneck triggers *generalisation*, i.e. the emergence of new items that are consistent with other known elements of a production system. In situations where learners have to infer productive *systems*, such as reconstructing recursive language (Kirby, 2002) or function learning (Griffiths et al., 2013), the pressure that is amplified is one towards the systematic *innovation* of new signals. Especially when extrapolating the role of the bottleneck to the acquisition of single, holistic traits, our analysis predicts that the effect seen in experimental work is unlikely to be an amplification of selection, but rather the *relaxation* of other selective pressures that hold in real life language use, causing asymmetries that are due to preferential innovation biases to prevail.

---

[1]Identical results are obtained when using a set mutation rate μ rather than the population size-dependent regularisation rate α, assuming the same asymmetry between innovation of the two variants. While the same absolute level of μ leads to more temporal instability in smaller populations, the mean value of the stationary distributions only depends on the ratio between the two innovation rates.

At this point two more caveats regarding modelling are in order: firstly, it should be noted that this result regarding the ineffectiveness of selection pressures does not just hold for asymmetric selection as implemented by the Wright-Fisher model's selection coefficient s. In other words, the analysis does not rely on the unusual configuration of directly antagonistic innovation and selection pressures. *Symmetric* selection pressures, such as the frequency-dependent regularisation biases proposed for the USM in Section 3.2 as well general confirmity pressures, are equally affected by the imposition of a bottleneck. So while iterated learning can help reveal human biases in directed innovation, the methodology does not so far speak to how these innovations manage to overcome conformity pressures to help them spread through real language communities.

Secondly, the present analysis of the dynamics of selection is based on the discrete replicator view implied by evolutionary approaches such as the USM proposed by Croft (2000). It should be noted that the nature of selection in this framework, as a process that merely favours the replication of instances of a variant that are already present in the population, is not the only possible way to frame selection. In fact, the model predictions might look very different if we considered *continous* rather than *discrete* replicators (e.g. Wedel, 2006, akin to the modelling of *quantitative traits* in biological evolution), or if we considered the effect of *guided variation*, a cultural selection pressure proposed by Boyd and Richerson (1985, pp.136) that is independent of the amount of variation present in the population.

So while other analyses are possible, which of these theoretical models of selection comes closest to reality is a matter that can only be established empirically, a task for which an effort must be made to more systematically disentangle the mechanisms of innovation and selection in experimental work on humans. The approach in this chapter was taken both because it is based on a concrete model of iterated learning (Reali and Griffiths, 2009), but also because it offers a clear explanation of the exact nature of the asymmetric biases that are amplified by a bottleneck, whether that bottleneck is construed in terms of limited learning input sample or population size.

6.4   Momentum-based selection in the Wright-Fisher model

The analysis of the Wright-Fisher model with different asymmetric pressures above confirmed the theoretical criticisms presented earlier: while asymmetries due to either accumulation of error or selection express themselves in the synchronic distribution of traits as expected, the investigation of the diffusion probabilities indicated that, under most conditions, selection pressures would not allow dispreferred traits to spread, while innovation pressures by themselves would only ever exhibit noisy trajectories far from the directed transitions we normally see in language change. In other words, universal innovation and selection alone seem *too strong* a predictor to account for particular, as opposed to universal, language features, as I argued for in the the discussion of the actuation problem in Section 2.2.4.

The momentum-based selection model presented in Chapter 4 on the other hand promised to account for the sporadic nature of language changes, as well as their spontaneous actuation. The multi-agent model presented in the earlier chapter considered *selection* only, without taking

into account the effect that *innovation*, and in particular *asymmetries* in innovation probabilities, might have on the macro-level dynamics of the model, such as the expected synchronic distribution of competing traits. To flesh out the predictions of these interactions, I will therefore present a modification of the Wright-Fisher model that incorporates a simplified version of the momentum-based selection mechanism.

6.4.1 A Markov model state space for momentum

So far, we have used the Wright-Fisher model to explore the quantitative dynamics of the well-known pressures of innovation and selection, both of which have direct parallels in biological evolution. The logical next step is to investigate how these results compare to a trend-amplifying bias such as the one implemented by momentum-based selection. On the face of it, the idea of momentum and the assumptions of a Markov model discussed in Section 3.4.1 seem at odds: to reiterate, the Markov assumption states explicitly that the probability of transitioning into a particular state must only be influenced by a system's current state, not by any previous states or state trajectories.

The Markov model framework itself is oblivious to the structure of a model's state space and the 'meaning' of individual states in terms of how they are interpreted by the modeller. In order to represent a population of N individuals (or memory size of N tokens), Reali and Griffiths constructed a space of $N + 1$ states. Just looking at the level of connectivity between states, this state space might seem unstructured: all states had non-zero (if very small) transition probabilities to each other, forming one fully connected graph. But on top of this homogeneous structure there was a semantically meaningful organisation of states: each state corresponded to a certain memory state of an agent, with corresponding transition probabilities corresponding to the 'proximity' to the other states in terms of the usage rates they represented, as shown in Figure 6.7a.

Starting from this basic pattern of connectivity, we can construct a Markov model that lends a sense of time to the model state space. In order to augment the with a momentum bias, we simply multiply the number of states: for every state of the Reali and Griffiths model which represents a certain prevalence x of variant 0, we create two additional states, representing the same value of x, but with copies of the state that indicate positive and negative momentum terms m.

A schematic visualisation of the shape of this state space is shown in Figure 6.7b. While the number of states of this momentum model is almost threefold in comparison to the baseline model, the pattern of transitions is actually not much more complex. In particular, the model is not fully connected: every state has exactly N outward transitions, exactly one each to every level of $x = 0 \ldots N$. The semantics of these three parallel states determines which of them a given previous state will transition into: all transitions from states with a lower to a higher x go into the $m = 1$ state, transitions from higher to lower values of x go into the $m = -1$ state, and only transitions from identical values of x enter the state with $m = 0$, indicating stagnation and therefore the absence of a trend.

In order to affect the dynamics of the system, the probabilities of transitioning between

(a) State space of the Reali and Griffiths (2009) Markov model with $N = 5$.

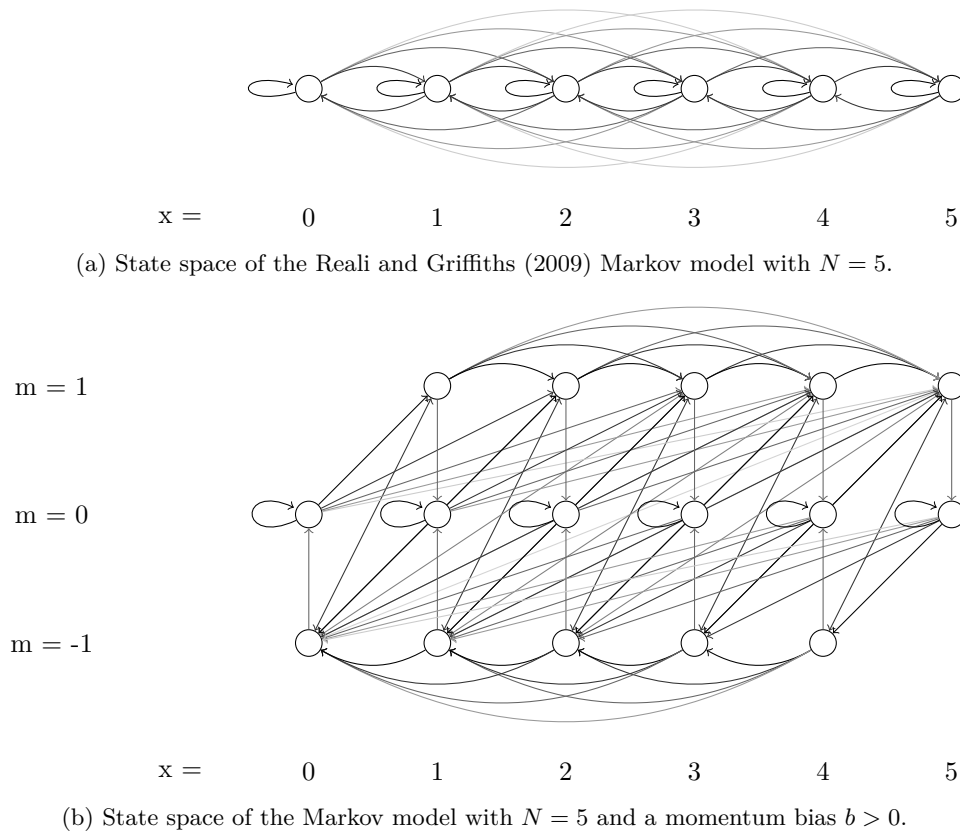(b) State space of the Markov model with $N = 5$ and a momentum bias $b > 0$.

Figure 6.7: Schematic visualisation of the Markov model state space. Colouring of the edges indicates relative probability of the transitions, with darker edges representing more likely transitions.

different levels of x are affected by the value of the momentum term: for the middle row with $m = 0$, the probabilities of producing a given x are equivalent to the Reali and Griffiths model, which means they correspond to the Wright-Fisher model with innovation only, as in Equation 6.4.

For the upper and lower rows in the diagram, corresponding to a positive (top) or negative (bottom) momentum term, the momentum affects the probabilities of producing a certain number of x tokens by exerting a selection pressure on the variant that is currently 'trending', i.e. whose frequency x has increased at the last time step. To this end the transition probabilities out of a state with $m = 1$ are calculated according to the Wright-Fisher model with innovation and selection as given in Equation 6.6, with the selection coefficient s in favour of the incoming variant 1 set to a fixed constant. Conversely, the transition probabilities for states with $m = -1$ are controlled by the same equation, only that the same selection coefficient is selecting variant 1. Note how, as long as the innovation probabilities are equal, this system is again symmetric, and thus replicator-neutral. Even though there are clear paths of directed selection, e.g. towards higher values of x along the top of the state space, these paths are mirrored exactly on the other side. A concrete example of a Markov chain transition matrix for such a momentum model is given in Table 6.1.

### 6.4.2   The interaction of momentum and innovation

So how do the dynamics of this momentum selection model differ from the original innovation-only version of the Wright-Fisher model? A direct comparison of the two models' stationary distributions for various settings of the regularisation parameter α and momentum bias strength b is shown in Figure 6.8. While a population size of $N = 80$ only allows for 81 different states in terms of the *frequency* distribution of the competing variants, the corresponding momentum model possesses 241 states. To aid interpretability, the stationary distributions are therefore grouped by the absolute frequency x of the variants, with the stationary probability of states with the same frequency but different momentum values stacked on top of each other. The three colours indicate the momentum: the red portions show the probability of being in a state with positive momentum, where productions are biased towards higher levels of x, while the blue sections express the same but for negative momentum. The white sections of the bars represent the probability of being in a state with a momentum of 0, resulting in sampling that is not biased towards either variant.

In order to confirm that this specific model of selection works as expected, we can set the momentum bias to $b = 0$, in which case we recover exactly the same stationary distribution as the Wright-Fisher model with innovation only, as shown in Figure 6.8a. Despite the selection bias being ineffective in this model, the colour indication of the momentum term is still informative: it shows how much of the time the model *remains* at a given proportion, resulting in a momentum term of 0 as indicated in white. As the impact of the momentum bias is increased, the probability of remaining in a state of mixed usage of variants decreases, as can be seen in Figure 6.8a.

So how does the model with a momentum bias differ from a simple model of regularisation

	x′ = 0−	x′ = 0=	x′ = 0+	x′ = 1−	x′ = 1=	x′ = 1+	x′ = 2−	x′ = 2=	x′ = 2+	x′ = 3−	x′ = 3=	x′ = 3+
x = 0−	0.0000	0.9524	0.0000	0.0000	0.0000	0.0468	0.0000	0.0000	0.0008	0.0000	0.0000	0.0000
x = 0=	0.0000	0.9524	0.0000	0.0000	0.0000	0.0468	0.0000	0.0000	0.0008	0.0000	0.0000	0.0000
x = 0+	0.0000	0.9524	0.0000	0.0000	0.0000	0.0468	0.0000	0.0000	0.0008	0.0000	0.0000	0.0000
x = 1−	0.3656	0.0000	0.0000	0.0000	0.4371	0.0000	0.0000	0.0000	0.1742	0.0000	0.0000	0.0231
x = 1=	0.2892	0.0000	0.0000	0.0000	0.4444	0.0000	0.0000	0.0000	0.2276	0.0000	0.0000	0.0389
x = 1+	0.2189	0.0000	0.0000	0.0000	0.4329	0.0000	0.0000	0.0000	0.2855	0.0000	0.0000	0.0627
x = 2−	0.0627	0.0000	0.0000	0.2855	0.0000	0.0000	0.0000	0.4329	0.0000	0.0000	0.0000	0.2189
x = 2=	0.0389	0.0000	0.0000	0.2276	0.0000	0.0000	0.0000	0.4444	0.0000	0.0000	0.0000	0.2892
x = 2+	0.0231	0.0000	0.0000	0.1742	0.0000	0.0000	0.0000	0.4371	0.0000	0.0000	0.0000	0.3656
x = 3−	0.0000	0.0000	0.0000	0.0008	0.0000	0.0000	0.0468	0.0000	0.0000	0.0000	0.9524	0.0000
x = 3=	0.0000	0.0000	0.0000	0.0008	0.0000	0.0000	0.0468	0.0000	0.0000	0.0000	0.9524	0.0000
x = 3+	0.0000	0.0000	0.0000	0.0008	0.0000	0.0000	0.0468	0.0000	0.0000	0.0000	0.9524	0.0000

Table 6.1: Markov chain transition matrix with momentum, $N = 80$, $\alpha = 0.1$, $b = 0.3$. The sign at the end of the name of the state indicates whether the state's momentum is positive (upward-trending), negative (downward-trending) or neutral (stagnant).
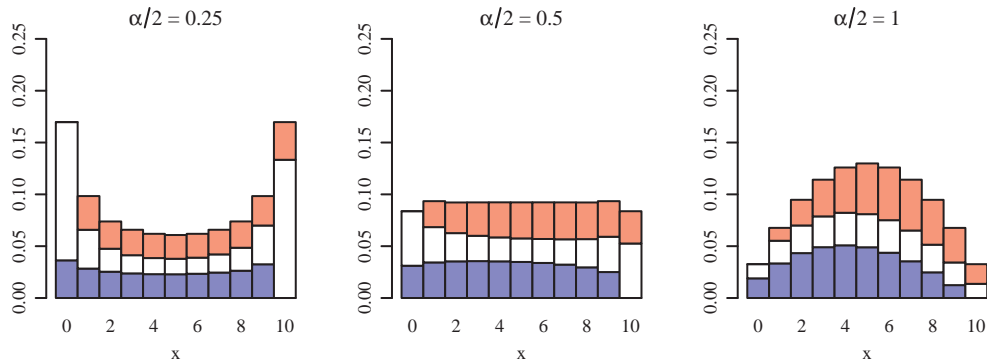
(a) Without a momentum bias ($b = 0$) the stationary distribution is identical to Reali and Griffiths (2009)'s *averaging* learner (compare Figure 3.8b). The differentiation between the triples of states representing the same variable frequency but different momentum values (colour-coded as red for positive, blue for negative, and white for no momentum) makes the temporal dynamics of the model more explicit, but the extra states do not otherwise change the dynamics of the original model.
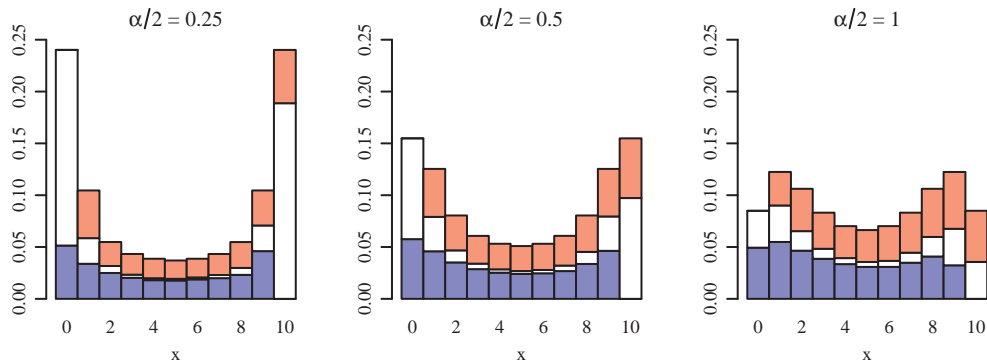
(b) With a momentum bias of $b = 3$, the model naturally avoids states corresponding to mixed usage of the competing variants, as it preferentially sweeps through this middle region in a directed fashion. The higher the momentum bias, the less time the model spends in regions of mixed usage of the two variants.

Figure 6.8: Stationary distributions of Markov chains with momentum in a population of $N = 10$ for various regularisation parameters α and momentum bias strengths b.

through symmetric selection of the more frequent variant? To appreciate the temporal dynamics of the momentum-based Wright-Fisher model we have to again move beyond simply looking at its stationary distributions and instead consider the likelihood of actuated transitions to diffuse to the entire population, as was done in the analyses above. In the absence of momentum-based selection shown in Figure 6.9(i) we again recover the same dynamics as for simple symmetric innovation (compare Figure 6.4b), where the diffusion probabilities do not stray far from the neutral evolution baseline. This changes drastically with the introduction of a momentum bias shown in Figure 6.9(ii) and (iii): the panels show the probability of diffusing based on the model states with the given frequency that correspond to a positive momentum term, i.e. trajectories that have come to occupy the relative frequency of that variant as part of a recent rise in frequency. Trajectories that start off with positive momentum have significantly higher completion probabilities, even at low initial frequencies of the variant.

Interestingly, this effect is slightly reversed for states corresponding to near completion. As was shown above, changes in variant frequency due to selection pressures slow down as the variant starts to prevail in the population. The increased likelihood of stagnation towards the end of transitions in combination with the simplistic ternary distinction of momentum based on the direction of change at the last time step in the present model means that, the initiation of a momentum-driven trend in the opposite direction becomes more likely as the changes slow down near completion. This artefact of the current toy model would therefore be greatly alleviated by a more gradual measure of trends that captures directedness on a greater time depth, such as the one implemented in Chapter 4.
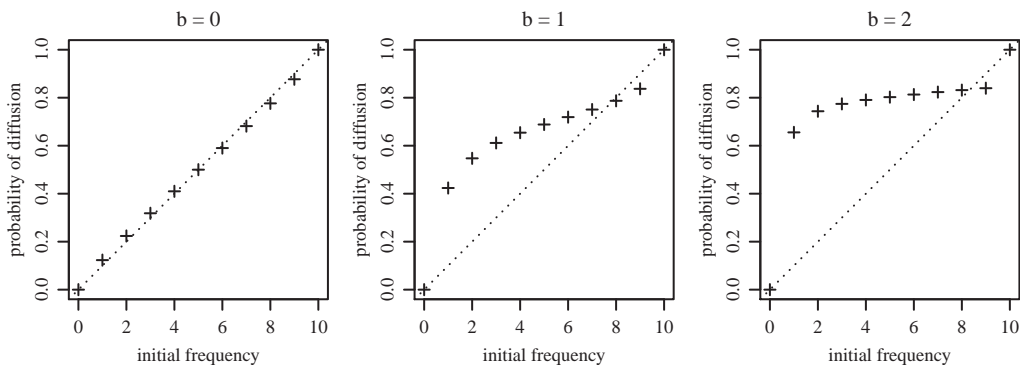

Figure 6.9: Probability of an actuated transition completing as a function of the initial frequency of the variant, assuming that the variant has positive (upwards) momentum. Innovation probabilities of both variants are equal, with $\alpha/2 = 0.1$.

A fact that is possibly not immediately appreciated about the present diffusion probability plot is that the increased average likelihood of completion actually holds for *both* of the competing variants. In the earlier analysis of the general Wright-Fisher model it was sufficient to only plot the completion probabilities of the incoming variant. The likelihood of the other variant succeeding was simply the complement of the probability for its competitor, meaning that an increased success rate of one variant was automatically associated with a decrease in the other. Under the present model on the other hand we are particularly interested in the completion

probabilities of directed transitions which have positive momentum. When taking this temporal dimension of our model into account while calculating the respective diffusion probabilities for either variant, we are therefore not actually computing those probabilities based on the same initial state, but based on the two corresponding states with positive and negative momentum respectively. The important consequences of this point will become even more apparent when we consider the case of momentum-based selection applying on top of asymmetric innovation rates.

6.4.3 Momentum and asymmetric innovation

In the simple Wright-Fisher model with asymmetric innovation rates discussed earlier, the preferentially innovated variant was shown to dominate the dynamics, with the model by remaining in a state of categorical usage of that variant most of the time. However, this bias towards one variant was not strongly evident in the dynamics of individual transitions, where even strong asymmetries resulted in only slight increases in the diffusion probability of the preferred variant at the expense of competing variants, indicative of rare and noisy transitions between the two extreme states.

While Figure 6.10a shows that having symmetric momentum-based selection on top of asymmetric innovation biases does not significantly alter the expected synchronic distribution of variants, the probabilities of diffusion out of states with positive momentum draw a picture of a very different dynamic. Figure 6.10b shows the probability of successful diffusion for both of the competing variants across different initial frequencies. While in the case of asymmetric innovation without a momentum bias ($b = 0$) an increase in one variant's diffusion probability entails a decrease in that of the other, the presence of a momentum bias actually raises the likelihood of successful transitions above the neutral evolution baseline for *both* of the variants, albeit at slightly different rates. That there is still a difference between the diffusion probabilities of the two variants throughout can be explained by the asymmetry in innovation, since the less frequently generated variant is still slightly more likely to have its actuated transition interrupted by the spontaneous generation of instances of its competitor variant.

6.4.4 Momentum, asymmetric innovation and the synchronic distribution of variants

After having considered several different pressures and combinations of pressures in this chapter we found that, based on the Wright-Fisher model as a simplified model of cultural evolution, only the presence of biases for the spontaneous *innovation* of new variants in combination with a symmetric selection bias such as momentum-based selection was able to capture both asymmetries in synchronic distributions while at the same time producing directed transitions towards the categorical usage of both preferred as well as less frequently attested variants.

Even though the present model of the interaction between asymmetric variation and momentum-based selection is highly simplified, it is interesting to get an idea of the degree to which the momentum-based selection bias projects asymmetries in innovation onto population-level distributions of variants. In other words, assuming that individual language changes are driven by
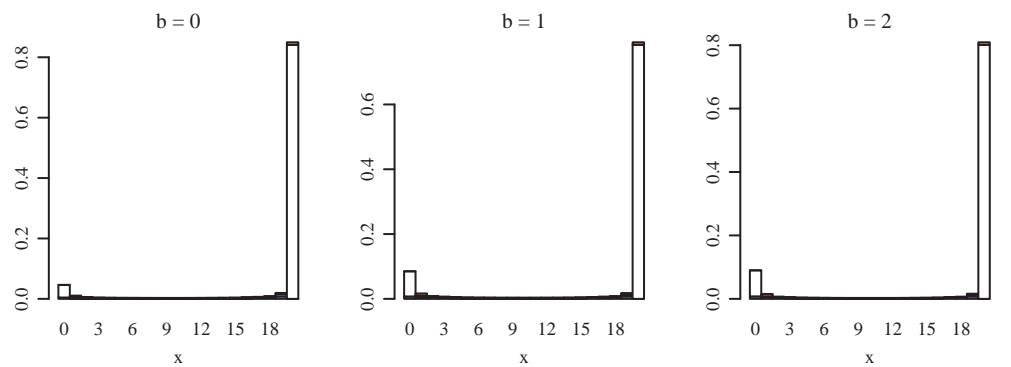
(a) Stationary distributions. The presence of a momentum bias increases the relative likelihood of synchronically observing the less frequently innovated variant.
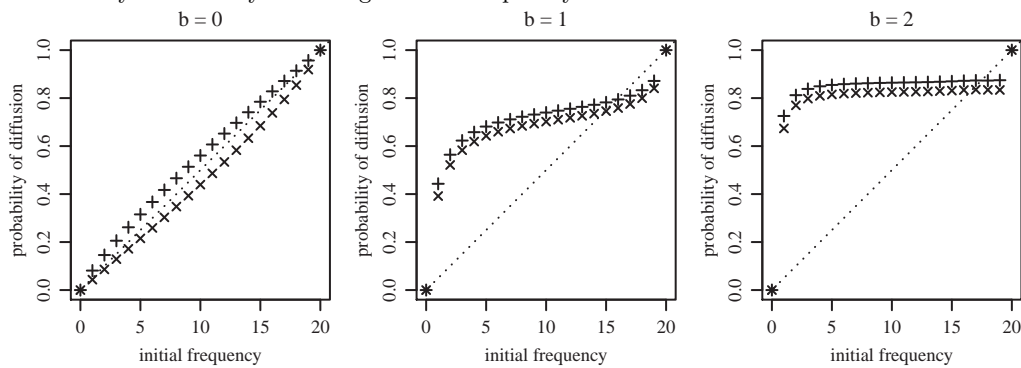
(b) Completion probabilities for transitions where the incoming variant is the one that is *more* likely to be introduced through innovation (plusses) as well as for transitions where the incoming variant is the one that is *less* likely to be introduced through innovation (crosses).

Figure 6.10: Dynamics of momentum-based selection with asymmetric innovation probabilities for different strengths of the momentum bias b, with $\alpha_1 = 0.2$, $\alpha_1/\alpha_0 = 10$.

replicator-neutral trend-amplification mechanisms, to what degree should we expect innovation biases that we find in individuals to be reflected in cross-linguistic distributions? To investigate this relationship, Figure 6.11b plots the relative synchronic frequency of the two competing variants as a function of an increasing difference in their innovation probabilities. Here, we return to the direct specification of the innovation rates μ, since we are actually interested in how one and the same probability of spontaneously producing a variant interacts with population size, rather than trying to model a specific regime of (de-)regularisation.

Figure 6.11b(i) shows the effect of increasing the parameter N, indicating population size in the biological framing of the Wright-Fisher model, or otherwise the size of the learning sample in Reali and Griffiths's model of iterated learning. In either case, the argument will correspond to a measure of how precisely variable usage of competing variants can be represented, either as distributed knowledge across the population or within the individual. While we find a linear mapping from innovation asymmetries to predicted synchronic frequency for all population sizes, the strength of this mapping *decreases* with increasing population size so that, assuming the same absolute level of innovation rates, we should find asymmetries in innovation to be relatively *less* expressed in larger populations.

Figure 6.11b(ii) shows how different baseline rates of innovation for both variants affect the degree to which any asymmetry should be reflected in synchronic data. The results indicate that, the *less* likely spontaenous innovations in a specific trait occur, the *more* should any asymmetries in those innovation probabilities be manifested in synchronic data.

While general trends can be derived from the present models, caution should be taken in attempting to derive empirical predictions about social factors such as community size from them directly. To be more precise, the predictions shown here are based on the assumption that innovation rates are constant across populations, which is not necessarily the case since innovation might itself be affected by external pressures. A momentum-based selection account thus offers another explanation of how environmental and sociocultural factors that differ between societies can affect the evolution of their linguistic systems, but through steering the *innovation* of novel variants and traits, rather than by causing the selection of specific variants directly.
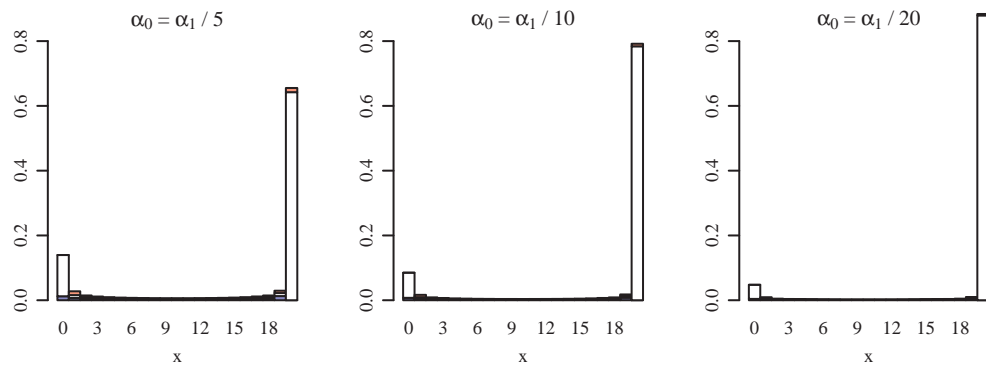
(a) Stationary distributions with asymmetric variation, where $\alpha_1 = 0.2$.
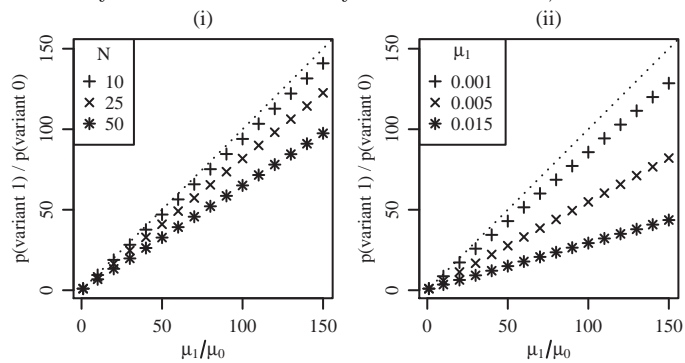
(b) Expected odds of synchronically observing one variant over the other as a function of the odds of the spontaneous innovation of one variant over the other. The degree to which the same asymmetry in innovation is expressed in synchronic distributions varies, depending on *(i)* the Wright-Fisher model parameter N, corresponding to population or sample size (data for $\mu_1 = 0.015$), and *(ii)* absolute innovation probabilities, given $N = 20$.

Figure 6.11: Expected synchronic distribution of variants given asymmetric innovation probabilities with momentum-based selection ($b = 1$).

Chapter 7

Summary, conclusion & outlook

In this thesis I have adopted an evolutionary approach to language change and argued for its importance in explaining and reconciling some seemingly contradictory results and scientific standpoints regarding the nature of and mechanisms behind language change. Following a broad overview of different explanations and accounts of language change as well as formal models of change in particular, I introduced and analysed the dynamics of a new model of the momentum-based selection of linguistic variants in Chapter 4. The fieldwork on individuals' explicit awareness of ongoing syntactic changes in Shetland presented in Chapter 5 contributed to existing evidence regarding the acquisition of sociolinguistic knowledge about language changes, knowledge which represents a fundamental ingredient of a mechanism based on detecting and amplifying trends in language use. Finally, in Chapter 6 I investigated a simplified model of momentum-based *selection* in combination with an independent pressure for the *innovation* of new variants. While the importance of distinguishing between pressures of innovation and selection entailed by an evolutionary approach has been emphasized by other researchers on language change (particularly Croft, 2000), I hope to have demonstrated practically how this separation of concerns can synthesize the explanation of both the *universal* patterns found in language changes as well as the *particular* nature of the sporadic and unpredictable actuation of individual changes.

The momentum-based trend amplification mechanism that formed the core of this thesis allows for the neutral but *directed* selection of variants from the pool of synchronic variation, a pool whose internal distribution is heavily affected by functional motivations (Ohala, 1989). This reconciliation of the functional origin of many language changes with their ultimately arbitrary selection and adoption by a community is summarised neatly by Labov (2001) when he states that his comprehensive studies of sound changes

> made it clear that linguistic change in progress is heavily constrained by the physical environment in which it takes place, and by structural factors that limit the course of change. At the same time, it is argued that the forces that move and motivate change, and are responsible for incrementation and transmission across generations, are largely social in nature. (p.498)

While this framework of thought is widely accepted within micro-level sociolinguistic work,

I have argued that similar evolutionary approaches have not been fully adopted by the wider language change research community. The relevance of distinguishing universal asymmetries in innovation from those in selection was diminished by the lack of a concrete neutral selection mechanism that could account for the arbitrary selection of linguistic innovations without having to refer to concepts like sociolinguistic 'prestige' which can only be attributed post-hoc. I hope to have filled this conceptual gap in the evolutionary, two-step model of language change with the momentum-based selection mechanism, a candidate mechanism which can sporadically trigger a population to adopt linguistic innovations. The dynamics of the model exhibit directed transitions that are indicative of the selection of the incoming variants, irrespective of whether the innovations are actually functionally superior, inferior, or simply neutral with respect to the existing variants.

While Labov speaks of "largely social" forces that move and motivate actual language changes, I have presented momentum-based selection as a highly *mechanistic* pressure in this thesis. This is not to say that the two types of mechanism are necessarily contradictory, or that they are even distinct. The apparent disconnect between the two should not be regarded as an unbridgable gap, but rather as an opportunity for research into whether much arbitrary-seeming social variation could in fact be reduced to a more mechanistic explanation such as the one presented in this thesis, and thus maybe even provide a mechanistic *grounding* for the concept of sociolinguistic prestige.

For example, the sociolinguistic literature has seen repeated claims about specific speaker groups (in particular females as well as the more 'socially mobile' middle classes, see e.g. Labov 2001, p.501) to be leading linguistic change, at least as far as the primarily available data from language changes in Western societies is concerned. While this pattern has already been proposed to reflect a relatively greater stake in the 'social marketplace' of linguistic conventions by those groups, the momentum-based selection mechanism raises the question of whether an increased sensitivity to changes could not just have to do with social status, but could also be explained due to those individuals' position in their social network which allows them a better overview of the state of language (as well as current linguistic 'trends') in their community. A comparison between the stratified diffusion of language changes through real social communities with the predictions made by momentum-based selection in structured social networks could yield further insights into the relevance of such mechanisms.

Rather than simply contribute another model of language change to aid the field take a step towards some well-defined goal post on its course to 'explaining' language change, I hope that my review of the diverse accounts and approaches to language change in Chapter 2 has succeeded in pointing out that there are in fact wildly differing opinions on the position of these very goal posts. I argued that different subfields of language change research are concerned with explaining very different aspects of change, from the emergence of language universals over probabilistically predictable socio-culturally influenced features to the unpredictability of particular, idiosyncratic changes. In trying to adjust and unify the position of the goal posts, by pointing out the complementary nature of the approaches when viewed in an evolutionary framework, I also hope to have raised awareness regarding the scientific end goal for a theory of language change as an area of research that straddles the social and historical sciences (Blute,

1997). In line with Weinreich et al.'s original formulation of the *actuation problem* (1968), I have argued that a complete theory of language change should not just be limited to correctly predicting the predictable aspects of the phenomenon, but also provide an explicit account or explanation of the *unpredictable* nature of language change (p.186).

While in this thesis I have chiefly focussed on computational modelling as a tool for scientific enquiry, I hope to have highlighted the need for more empirical, particularly *quantitative* research into how language changes actually unfold in communities, as well as how different individuals participate in language changes over time. Although I have brought forward theoretical arguments for why a replicator-neutral selection mechanism such as momentum-based selection can go a long way in explaining both the universal patterns of language changes as well as the constant diversification of languages, there is still much need for empirical evidence for such a mechanism. Particularly on the micro-level of the individual, the nature of sociolinguistic knowledge is just becoming a focus of research, and many more results on this matter can be expected in the coming years from work within the framework of perceptual dialectology as well as from experimental methods such as the one used by Drager (2011).

On the macro-level of historical changes and the cross-linguistic comparison of similar changes, I have pointed out a divergence in research goals as well as a relative lack of cumulative work within any unified framework. In Section 4.4.1 I argued that different researchers have approached the quantitative question of the *rate of language change* on two very different levels, where work in diachronic typology (such as Bickel, 2015) is interested in the probability of a certain type of change to occur on a macro-level, whereas traditional historical linguistic work hones in on the speed with which a particular change diffuses through a population. While there has generally been much talk of adaptation and selection in language change, quantitative evaluations of how much the selection of linguistic structures actually plays a role within particular historical changes has been minimal, as I argued in Section 2.4.2.

With respect to the cross-linguistic actuation probabilities of similar changes, the evolutionary approach to language change shows that the synchronic prevalence or preference of certain linguistic structures does not necessarily have to do with them being selected for based on their adaptive features. Instead, I hope to have demonstrated that many asymmetries in language change, including the unidirectional patterns found in domains ranging from sound change to grammaticalisation, might instead be due to the preferred innovation of specific variants. Separate from these asymmetries in the *direction* of language changes due to innovation, I showed how selection mechanisms based on social learning, such as the trend-amplification dynamics of momentum-based selection, can yield *directed* transitions that, while superficially indicative of an asymmetry between variants, are based on biases which are only temporary and emergent from the dynamics of social learning.

Bibliography

Acerbi, Alberto, Stefano Ghirlanda and Magnus Enquist, 2012. The logic of fashion cycles. *PloS one* 7(3): e32541. doi:10.1371/journal.pone.0032541.

Acerbi, Alberto, Stefano Ghirlanda and Magnus Enquist, 2014. Regulatory traits: cultural influences on cultural evolution. In Stefano Cagnoni, Marco Mirolli and Marco Villani (eds.), *Evolution, Complexity and Artificial Life*, pp. 135–147. Berlin, Heidelberg: Springer Berlin Heidelberg. ISBN 978-3-642-37576-7. doi:10.1007/978-3-642-37577-4.

Ahern, Christopher Andrew, Mitchell G. Newberry, Robin Clark and Joshua B. Plotkin, 2016. Evolutionary forces in language change. *arXiv* 1608.00938.

Altmann, Eduardo G., Zakary L. Whichard and Adilson E. Motter, 2013. Identifying trends in word frequency dynamics. *Journal of Statistical Physics* 151(1-2): 277–288. doi:10.1007/s10955-013-0699-7.

Altmann, Gabriel H., Haro von Buttlar, Walter Rott and Udo Strauß, 1983. A law of change in language. In Barron Brainerd (ed.), *Historical linguistics*, pp. 104–115. Bochum: Studienverlag Dr. N. Brockmeyer.

Amerise, Ilaria L., Marco Marozzi and Agostino Tarsitano, 2016. pvrank: Rank Correlations. `https://cran.r-project.org/package=pvrank`.

Andersen, Henning, 2008. Grammaticalization in a speaker-oriented theory of change. In Thórhallur Eythórsson (ed.), *Grammatical Change and Linguistic Theory: The Rosendal papers*, ch. 1, pp. 11–44. John Benjamins.

Arita, Takaya and Yuhji Koyama, 1998. Evolution of linguistic diversity in a simple communication system. *Artificial Life* 4(1): 109–124.

Atkinson, Mark, 2016. *Sociocultural determination of linguistic complexity*. Ph.D. thesis, The University of Edinburgh.

Atkinson, Mark, Simon Kirby and Kenny Smith, 2015. Speaker input variability does not explain why larger populations have simpler languages. *PLOS ONE* 10(6): e0129463. doi:10.1371/journal.pone.0129463.

Atkinson, Quentin, Geoff Nicholls, David Welch and Russell Gray, 2005. From words to dates: water into wine, mathemagic or phylogenetic inference? *Transactions of the Philological Society* 103(2): 193–219. doi:10.1111/j.1467-968X.2005.00151.x.

Atkinson, Quentin D. and Russell D. Gray, 2005. Curious parallels and curious connections - Phylogenetic thinking in biology and historical linguistics. *Systematic Biology* 54(4): 513–526.

Bailey, Charles-James N., 1970. Building rate into a dynamic theory of linguistic description. *Working Papers in Linguistics* 2(9): 161–233.

Bailey, Charles-James N., 1973. *Variation and linguistic theory.* Arlington, VA: Center for Applied Linguistics, `http://files.eric.ed.gov/fulltext/ED125238.pdf`.

Bailey, Guy, 2002. Real and apparent time. In J. K. Chambers, Peter Trudgill and Natalie Schilling (eds.), *The Handbook of Language Variation and Change*, ch. 12, pp. 312–332. Oxford: Blackwell.

Baker, Adam, 2008. Computational approaches to the study of language change. *Language and Linguistics Compass* 2(2): 289–307. doi:10.1111/j.1749-818X.2008.00054.x.

Baker, Adam, Diana Archangeli and Jeff Mielke, 2011. Variability in American English s-retraction suggests a solution to the actuation problem. *Language Variation and Change* 23(03): 347–374. doi:10.1017/S0954394511000135.

Baranowski, Maciej, 2013. On the role of social factors in the loss of phonemic distinctions. *English Language and Linguistics* 17(02): 271–295. doi:10.1017/S1360674313000038.

Baronchelli, Andrea, Maddalena Felici, Vittorio Loreto, Emanuele Caglioti and Luc Steels, 2006. Sharp transition towards shared vocabularies in multi-agent systems. *Journal of Statistical Mechanics: Theory and Experiment* p. P06014. doi:10.1088/1742-5468/2006/06/P06014.

Baronchelli, Andrea, Vittorio Loreto and Luc Steels, 2008. In-depth analysis of the Naming Game dynamics: the homogeneous mixing case. *International Journal of Modern Physics C* 19(05): 785–812. doi:10.1142/S0129183108012522.

Baxter, Gareth and William Croft, 2016. Modeling language change across the lifespan: Individual trajectories in community change. *Language Variation and Change* 28(02): 129–173. doi:10.1017/S0954394516000077.

Baxter, Gareth J., Richard A. Blythe, William Croft and Alan J. McKane, 2006. Utterance selection model of language change. *Physical Review E* 73(4): 046118. doi:10.1103/PhysRevE.73.046118.

Baxter, Gareth J., Richard A. Blythe, William Croft and Alan J. McKane, 2009. Modeling language change: An evaluation of Trudgill's theory of the emergence of New Zealand English. *Language Variation and Change* 21(02): 257. doi:10.1017/S095439450999010X.

Baxter, Gareth J., Richard A. Blythe and Alan J. McKane, 2008. Fixation and consensus times on a network: a unified approach. *Physical Review Letters* 101(25): 258701. doi:10.1103/Phys-RevLett.101.258701.

Beckner, Clay, Richard Blythe, Joan Bybee, Morten H. Christiansen, William Croft, Nick C. Ellis, John Holland, Jinyun Ke, Diane Larsen-Freeman and Tom Schoenemann, 2009. Language is a complex adaptive system: position paper. *Language Learning* 59(December): 1–26. doi:10.1111/j.1467-9922.2009.00533.x.

Bentley, R. Alexander, Matthew W. Hahn and Stephen J. Shennan, 2004. Random drift and culture change. *Proceedings of the Royal Society B: Biological Sciences* 271(1547): 1443–1450. doi:10.1098/rspb.2004.2746.

Bentley, R. Alexander, Carl P. Lipo, Harold A. Herzog and Matthew W. Hahn, 2007. Regular rates of popular culture change reflect random copying. *Evolution and Human Behavior* 28(3): 151–158. doi:10.1016/j.evolhumbehav.2006.10.002.

Berg, Thomas, 1998. *Linguistic structure and change: an explanation from language processing.* Oxford: Clarendon Press.

Berger, Jonah and Gaël Le Mens, 2009. How adoption speed affects the abandonment of cultural tastes. *Proceedings of the National Academy of Sciences of the United States of America* 106(20): 8146–8150.

Bickel, Balthasar, 2015. Distributional typology: statistical inquiries into the dynamics of linguistic diversity. In Bernd Heine and Heiko Narrog (eds.), *The Oxford handbook of linguistic analysis*, pp. 901–923. Oxford: Oxford University Press. doi:10.5167/uzh-109110.

Bikhchandani, Sushil, David Hirshleifer and Ivo Welch, 1992. A theory of fads, fashion, custom, and cultural change as informational cascades. *Journal of Political Economy* 100(5): 992–1026. doi:10.1086/261849.

Bikhchandani, Sushil, David Hirshleifer and Ivo Welch, 1998. Learning from the behavior of others: conformity, fads, and informational cascades. *Journal of Economic Perspectives* 12(3): 151–170. doi:10.1257/jep.12.3.151.

Bloomfield, Leonard, 1933. *Language.* Chicago: University of Chicago Press. ISBN 9780226060675.

Blute, Marion, 1997. History versus science: the evolutionary solution. *Canadian Journal of Sociology / Cahiers canadiens de sociologie* 22(3): 345–364. doi:10.2307/3341626.

Blythe, Richard A., 2007. The propagation of a cultural or biological trait by neutral genetic drift in a subdivided population. *Theoretical population biology* 71(4): 454–472. doi:10.1016/j.tpb.2007.01.006.

Blythe, Richard A., 2010. Ordering in voter models on networks: exact reduction to a single-coordinate diffusion. *Journal of Physics A: Mathematical and Theoretical* 43(38): 385003. doi:10.1088/1751-8113/43/38/385003.

Blythe, Richard A., 2012a. Neutral evolution: a null model for language dynamics. *Advances in Complex Systems* 15(03n04): 1150015. doi:10.1142/S0219525911003414.

Blythe, Richard A., 2012b. Random copying in space. *Advances in Complex Systems* 15(1 & 2): 1150012–1 – 1150012–26. doi:10.1142/S0219525911003396.

Blythe, Richard A., 2015. Hierarchy of scales in language dynamics. *The European Physical Journal B* 88: 295. doi:10.1140/epjb/e2015-60347-3.

Blythe, Richard A. and William Croft, 2012. S-curves and the mechanisms of propagation in language change. *Language* 88(2): 269–304. doi:10.1353/lan.2012.0027.

Blythe, Richard A. and William A. Croft, 2009. The speech community in evolutionary language dynamics. *Language Learning* 59(Supplement s1): 47–63. doi:10.1111/j.1467-9922.2009.00535.x.

de Boer, Bart, 2006. Computer modelling as a tool for understanding language evolution. In Nathalie Gontier, Jean Paul Van Bendegem and Diederik Aerts (eds.), *Evolutionary Epistemology, Language and Culture*, *Theory and Decision Library A: Rational Choice in Practical Philosophy and Philosophy of Science*, vol. 39, pp. 381–406. Springer. ISBN 978-1-4020-3395-7. doi:10.1007/1-4020-3395-8_17.

de Boer, Bart, 2012. Modelling and language evolution: beyond fact-free science. In Luke McCrohon (ed.), *Five Approaches to Language Evolution: Proceedings of the Workshops of the 9th International Conference on the Evolution of Language*, pp. 83–92. Tokyo: Evolang9 Organizing Committee.

Borgerhoff Mulder, Monique, 2001. Using phylogenetically based comparative methods in anthropology: More questions than answers. *Evolutionary Anthropology: Issues, News and Reviews* 10(3): 99–111. doi:10.1002/evan.1020.

Bourdieu, Pierre, 1977. The economics of linguistic exchanges. *Social Science Information* 16(6): 645–668. doi:10.1177/053901847701600601.

Bowie, David and Malcah Yaeger-Dror, 2013. Phonological change in real time. In Joseph Salmons and Patrick Honeybone (eds.), *The Oxford handbook of historical phonology*, ch. 34. Oxford: Oxford University Press. doi:10.1093/oxfordhb/9780199232819.013.004.

Boyd, Robert and Peter J. Richerson, 1985. *Culture and the evolutionary process.* Chicago: University of Chicago Press. ISBN 9780226069333, `https://books.google.nl/books?id=MBg4oBsCKU8C&printsec=frontcover`.

Branigan, Holly P., Martin J. Pickering and Alexandra A. Cleland, 2000. Syntactic coordination in dialogue. *Cognition* 75(2): B13–B25. doi:10.1016/S0010-0277(99)00081-5.

Brighton, Henry, 2002. Compositional syntax from cultural transmission. *Artificial Life* 8(1): 25–54. doi:10.1162/106454602753694756.

Bucholtz, Mary and Kira Hall, 2004. Language and Identity. In Alessandro Duranti (ed.), *A Companion to Linguistic Anthropology*, ch. 16, pp. 369–394. Oxford: Blackwell Publishing Ltd. ISBN 978-1405144308. doi:10.1002/9780470996522.ch16.

Buchstaller, Isabelle, 2015. Exploring linguistic malleability across the life span: Age-specific patterns in quotative use. *Language in Society* 44: 457–496. doi:10.1017/S0047404515000391.

Busemeyer, Jerome R. and Adele Diederich, 2010. *Cognitive Modeling.* Sage Publications. ISBN 9780761924500.

Bush, Robert R. and Frederick Mosteller, 1955. *Stochastic models for learning.* Oxford: John Wiley & Sons, Inc.

Bybee, Joan, 2007. *Frequency of use and the organization of language.* Oxford: Oxford University Press. ISBN 9780195301564.

Bynon, Theodora, 1977. *Historical Linguistics.* Cambridge: Cambridge University Press. ISBN 9781139165709. doi:10.1017/CBO9781139165709.

Campbell, Lyle, 2013. *Historical Linguistics: An Introduction.* Edinburgh: Edinburgh University Press, 3rd edn. ISBN 9780748645947.

Cangelosi, Angelo and Domenico Parisi, 2002. Computer simulation: a new scientific approach to the study of language evolution. In Angelo Cangelosi and Domenico Parisi (eds.), *Simulating the evolution of language*, ch. 1, pp. 3–28. London: Springer-Verlag. ISBN 978-1-85233-428-4, `http://cogprints.org/2014/`.

Carrera-Sabaté, Josefina, 2014. Does meta-linguistic awareness play any role at the beginning of an ongoing sound change? The case of some vowel-ended verbs in Catalan. *Sociolinguistic Studies* 8(2): 193–221. doi:10.1558/sols.v8i2.193.

Castellano, Claudio, Santo Fortunato and Vittorio Loreto, 2009. Statistical physics of social dynamics. *Reviews of Modern Physics* 81(2): 591–646. doi:10.1103/RevModPhys.81.591.

Cedergren, Henrietta J., 1987. The spread of language change: verifying inferences of linguistic diffusion. In *Language Spread and Language Policy: Issues, Implications, and Case Studies (Proceedings of GURT '87)*, pp. 45–59. Georgetown.

Chambers, J. K. and Peter Trudgill, 1998. *Dialectology.* Cambridge: Cambridge University Press. ISBN 9780511805103. doi:10.1017/CBO9780511805103.

Chater, Nick and Morten H. Christiansen, 2010. Language evolution as cultural evolution: how language is shaped by the brain. *Wiley Interdisciplinary Reviews: Cognitive Science* 1(5): 623–628. doi:10.1002/wcs.85.

Chen, M. Keith, 2013. The effect of language on economic behavior: evidence from savings rates, health behaviors, and retirement assets. *American Economic Review* 103(2): 690–731. doi:10.1257/aer.103.2.690.

Choudhury, Monojit, Anupam Basu and Sudeshna Sarkar, 2006. Multi-agent simulation of emergence of schwa deletion pattern in Hindi. *Journal of Artificial Societies and Social Simulation* 9(2): 2.

Choudhury, Monojit, Vaibhav Jalan, Sudeshna Sarkar and Anupam Basu, 2007. Evolution, optimization, and language change: the case of Bengali verb inflections. In *Proceedings of the Ninth Meeting of the ACL Special Interest Group in Computational Morphology and Phonology (SigMorPhon '07)*, pp. 65–74. Morristown, NJ, USA: Association for Computational Linguistics. doi:10.3115/1626516.1626525.

Christensen, Rune Haubo Bojesen, 2015. ordinal - Regression Models for Ordinal Data. `https://cran.r-project.org/package=ordinal`.

Clifford, Peter and Aidan Sudbury, 1973. A model for spatial conflict. *Biometrika* 60(3): 581–588. doi:10.1093/biomet/60.3.581.

Cornips, Leonie M. E. A. and Karen P. Corrigan (eds.), 2005. *Syntax and Variation.* Current Issues in Linguistic Theory. John Benjamins Publishing Company. ISBN 978 90 272 4779 7. doi:10.1075/cilt.265.

Cornish, Hannah, Mónica Tamariz and Simon Kirby, 2009. Complex adaptive systems and the origins of adaptive structure: what experiments can tell us. *Language Learning* 59:Suppl.1(December): 187–205. doi:10.1111/j.1467-9922.2009.00540.x.

Cowgill, Warren, 1963. A search for universals in Indo-European diachronic morphology. In Joseph H. Greenberg (ed.), *Universals of language*, ch. 6, pp. 91–113. Cambridge, Massachusetts: MIT Press, `https://archive.org/details/universalsoflang00unse`.

Croft, William, 2000. *Explaining language change: an evolutionary approach.* Harlow: Pearson Education Limited.

Croft, William, 2008. Evolutionary Linguistics. *Annual review of anthropology* 37(1): 219–234. doi:10.1146/annurev.anthro.37.081407.085156.

Croft, William A., 2006. The relevance of an evolutionary model to historical linguistics. In Ole Nedergaard Thomsen (ed.), *Competing models of linguistic change: evolution and beyond, Current Issues in Linguistic Theory*, vol. 279, pp. 91–132. Amsterdam: John Benjamins Publishing Company. doi:10.1075/cilt.279.08cro.

Cuskley, Christine F., Martina Pugliese, Claudio Castellano, Francesca Colaiori, Vittorio Loreto and Francesca Tria, 2014. Internal and external dynamics in language: evidence from verb regularity in a historical corpus of english. *PLoS ONE* 9(8): e102882. doi:10.1371/journal.pone.0102882.

Darwin, Charles, 1871. *The descent of man, and selection in relation to sex.* London: John Murray.

De Oliveira, Viviane M., Paulo R. A. Campos, Marcelo A. F. Gomes and Ing Ren Tsang, 2005. Bounded fitness landscapes and the evolution of the linguistic diversity. *Physica A: Statistical Mechanics and its Applications* 368(1): 10.

De Vylder, Bart and Karl Tuyls, 2006. How to reach linguistic consensus: a proof of convergence for the naming game. *Journal of Theoretical Biology* 242(4): 818–831.

Dediu, Dan and D. Robert Ladd, 2007. Linguistic tone is related to the population frequency of the adaptive haplogroups of two brain size genes, ASPM and Microcephalin. *Proceedings of the National Academy of Sciences* 104(26): 10944–10949. doi:10.1073/pnas.0610848104.

Denison, David, 2003. Log(ist)ic and simplistic S-curves. In Raymond Hickey (ed.), *Motives for language change*, pp. 54–70. Cambridge: Cambridge University Press. ISBN 978-0-521-79303-2, `https://www.escholar.manchester.ac.uk/uk-ac-man-scw:3b1441`.

Dixon, Robert M. W., 1997. *The rise and fall of languages.* Cambridge: Cambridge University Press. ISBN 9780511612060. doi:10.1017/CBO9780511612060.

Drager, Katie, 2005. From bad to bed: The relationship between perceived age and vowel perception in New Zealand English. *Te Reo* 48: 55–68.

Drager, Katie, 2011. Speaker age and vowel perception. *Language and Speech* 54(1): 99–121. doi:10.1177/0023830910388017.

Driscoll, Anna and Emma Lape, 2014. Reversal of the Northern Cities Shift in Syracuse, New York. In *43rd New Ways of Analyzing Variation conference.* Chicago, Illinois: University of Illinois.

Durham, Mercedes, 2014. 30 years later: real-time change and stability in attitudes towards the dialect in Shetland. In Robert Lawson (ed.), *Sociolinguistics in Scotland*, pp. 296–318. Basingstoke: Palgrave MacMillan, `http://orca.cf.ac.uk/id/eprint/45627`.

Eckert, Penelope, 1989. The whole woman: Sex and gender differences in variation. *Language Variation and Change* 1(03): 245–267. doi:10.1017/S095439450000017X.

Ellegård, Alvar, 1953. *The auxiliary do: the establishment and regulation of its use in English.* Gothenburg studies in English. Stockholm: Almqvist & Wiksell, `http://hdl.handle.net/2077/10782`.

Enke, Dankmar, Roland Mühlenbernd and Igor Yanovich, 2016. The emergence of the progressive to imperfective diachronic cycle in reinforcement-learning agents. In Sean G. Roberts, Christine F. Cuskley, Luke McCrohon, Lluis Barceló-Coblijn, Olga Fehér and Tessa Verhoef (eds.), *The Evolution of Language: Proceedings of the 11th International Conference (EVOLANG11).* New Orleans.

Ettlinger, Marc, 2007a. An exemplar-based model of chain shifts. In *Proceedings of the 16th International Congress of the Phonetic Sciences*, pp. 685–688.

Ettlinger, Marc, 2007b. Shifting categories: an exemplar-based computational model of chain shifts. In *Proceedings of the 29th annual conference of the Cognitive Science Society*, pp. 239–244. Lawrence Erlbaum.

Evans, Bethwyn, 2009. Beyond pronouns: further evidence for South Bougainville. In Bethwyn Evans (ed.), *Discovering history through language: Papers in honour of Malcolm Ross*, ch. 5, pp. 73–101. Canberra: Pacific Linguistics. ISBN 978-0-85883-605-1, `http://hdl.handle.net/11858/00-001M-0000-000F-F9C3-D`.

Fagyal, Zsuzsanna, Samarth Swarup, Anna María Escobar, Les Gasser and Kiran Lakkaraju, 2010. Centers and peripheries: Network roles in language change. *Lingua* 120(8): 2061–2079. doi:10.1016/j.lingua.2010.02.001.

Featherston, Sam, 2005. Magnitude estimation and what it can do for your syntax: some wh-constraints in German. *Lingua* 115(11): 1525–1550. doi:10.1016/j.lingua.2004.07.003.

Fedzechkina, Maryia, 2014. *Communicative efficiency, language learning, and language universals*. Ph.D. thesis, University of Rochester. `http://hdl.handle.net/1802/28852`.

Fedzechkina, Maryia, Becky Chu, T. Florian Jaeger and John Trueswell, 2016. Processing preferences shape language change. In *The Evolution of Language: Proceedings of the 11th International Conference (EVOLANG11)*. New Orleans.

Ferdinand, Vanessa, 2015. *Inductive evolution: cognition, culture, and regularity in language*. Ph.D. thesis, The University of Edinburgh. `http://hdl.handle.net/1842/11741`.

Fortescue, Michael, 2006. The non-linear nature of diachronic change. In Ole Nedergaard Thomsen (ed.), *Competing models of linguistic change: evolution and beyond*, *Current Issues in Linguistic Theory*, vol. 279, pp. 17–31. Amsterdam: John Benjamins Publishing Company. doi:10.1075/cilt.279.03for.

Foulkes, Paul and Gerard Docherty, 2006. The social life of phonetics and phonology. *Journal of Phonetics* 34(4): 409–438.

Foulkes, Paul and Jennifer B. Hay, 2015. The emergence of sociophonetic structure. In Brian MacWhinney and William O'Grady (eds.), *The Handbook of Language Emergence*, ch. 13, pp. 292–313. Hoboken: John Wiley & Sons, Inc. doi:10.1002/9781118346136.ch13.

Foulkes, Paul and Marilyn Vihman, 2013. First language acquisition and phonological change. In Patrick Honeybone and Joseph Salmons (eds.), *The Oxford handbook of historical phonology*, ch. 18. Oxford: Oxford University Press. doi:10.1093/oxfordhb/9780199232819.013.001.

Fries, Charles C. and Kenneth L. Pike, 1949. Coexistent phonemic systems. *Language* 25(1): 29–50. doi:10.2307/409907.

Fruehwald, Josef, 2013. *Phonological involvement in phonetic change*. Ph.D. thesis, University of Pennsylvania. `http://jofrhwld.github.io/dissertation/fruehwald_dissertation.html`.

Fruehwald, Josef, Jonathan Gress-Wright and Joel Wallenberg, 2009. Phonological rule change: the constant rate effect. In *40th annual meeting of the North East Linguistic Society*. Cambridge, Massachusetts: Massachusetts Institute of Technology.

Futrell, Richard, Kyle Mahowald and Edward Gibson, 2015. Large-scale evidence of dependency length minimization in 37 languages. *Proceedings of the National Academy of Sciences* 112(33): 10336–10341. doi:10.1073/pnas.1502134112.

Garrett, Andrew and Keith Johnson, 2013. Phonetic bias in sound change. In Alan C. L. Yu (ed.), *Origins of Sound Change: Approaches to Phonologization*, ch. 3, pp. 51–97. Oxford: Oxford University Press. ISBN 9780199573745.

Gerlee, Philip, 2015. Directional variation in evolution: consequences for the fitness landscape metaphor. *bioRxiv* 015529. doi:10.1101/015529.

Ghanbarnejad, Fakhteh, Martin Gerlach, José M. Miotto and Eduardo G. Altmann, 2014. Extracting information from S-curves of language change. *Journal of the Royal Society, Interface* 11(101): 20141044. doi:10.1098/rsif.2014.1044.

Gibson, Edward and Kenneth Wexler, 1994. Triggers. *Linguistic Inquiry* 25(3): 407–454.

Giles, Howard and Andrew C. Billings, 2004. Assessing language attitudes: speaker evaluation studies. In Alan Davies and Catherine Elder (eds.), *The Handbook of Applied Linguistics*, ch. 7, pp. 187–209. Oxford: Blackwell Publishing Ltd. doi:10.1002/9780470757000.ch7.

Giles, Howard, Justine Coupland and Nikolas Coupland (eds.), 1991. *Contexts of accommodation*. Studies in Emotion & Social Interaction. Cambridge: Cambridge University Press. ISBN 9780521361514.

Goldberg, Adele E., 1995. *Constructions: a construction grammar approach to argument structure*. Chicago: University of Chicago Press.

Goldstone, Robert L., Michael E. Roberts and Todd M. Gureckis, 2008. Emergent processes in group behavior. *Current Directions in Psychological Science* 17(1): 10–15. doi:10.1111/j.1467-8721.2008.00539.x.

Gong, Tao, Lan Shuai, Mónica Tamariz and Gerhard Jäger, 2012. Studying language change using Price equation and Pólya-urn dynamics. *PLoS ONE* 7(3): e33171. doi:10.1371/journal.pone.0033171.

Granovetter, Mark S., 1973. The strength of weak ties. *American Journal of Sociology* 78(6): 1360–1380. doi:10.1086/225469.

Graudina, L.N., 1964. Razvitie nulevoj formy roditelnogo mnozestvennogo u sescestvitelnych - jedinic izemerenija. In Iosif Pavlovich Muchnik and Mikhail Viktorovich Panov (eds.), *Razvitie grammatiki i leksiki sovremennogo russkogo jazyka*. Moscow: Nauka.

Greenberg, Joseph H., 1959. Language and evolution. In Betty J. Meggers (ed.), *Evolution and anthropology: a centennial appraisal.* Washington DC: The Anthropological Society of Washington.

Greenberg, Joseph H., 1963. Some universals of grammar with particular reference to the order of meaningful elements. In Joseph H. Greenberg (ed.), *Universals of language.* Cambridge, Massachusetts: MIT Press, https://archive.org/details/universalsoflang00unse.

Greenberg, Joseph H., Charles Egerton Osgood and Sol Saporta, 1954. Language change. In Charles E. Osgood and Thomas A. Sebeok (eds.), *Psycholinguistics: a survey of theory and research problems*, ch. 6.3, pp. 146–163. Baltimore: Waverly Press, Inc.

Greenhill, Simon J., Quentin D. Atkinson, Andrew Meade and Russell D. Gray, 2010. The shape and tempo of language evolution. *Proceedings of the Royal Society B: Biological Sciences* 277(1693): 2443–2450. doi:10.1098/rspb.2010.0051.

Griffiths, Thomas L. and Michael L. Kalish, 2007. Language evolution by Iterated Learning with Bayesian agents. *Cognitive Science* 31(3): 441–480. doi:10.1080/15326900701326576.

Griffiths, Thomas L., Stephan Lewandowsky and Michael L. Kalish, 2013. The effects of cultural transmission are modulated by the amount of information transmitted. *Cognitive Science* 37(5): 953–967. doi:10.1111/cogs.12045.

Grimm, Jacob, 1848. *Geschichte der deutschen Sprache*, vol. 1. Weidmannsche Buchhandlung, https://archive.org/details/geschichtederde14grimgoog.

Gureckis, Todd M. and Robert L. Goldstone, 2009. How you named your child: understanding the relationship between individual decision making and collective outcomes. *Topics in Cognitive Science* 1(4): 651–674. doi:10.1111/j.1756-8765.2009.01046.x.

Guy, Gregory R., 2003. Variationist approaches to phonological change. In Brian D. Joseph and Richard D. Janda (eds.), *The Handbook of Historical Linguistics*, ch. 8, pp. 369–400. Oxford, UK: Blackwell Publishing Ltd. ISBN 9780470756393. doi:10.1002/9780470756393.

Harper, Marc, 2009. The replicator equation as an inference dynamic. *arXiv* 0911.1763.

Hartl, Daniel L. and Andrew G. Clark, 2007. *Principles of population genetics.* Sunderland, Massachusetts: Sinauer Associates, Inc., fourth edn.

Haspelmath, Martin, 1999. Optimality and diachronic adaptation. *Zeitschrift für Sprachwissenschaft* 18(2): 180–205. doi:10.1515/zfsw.1999.18.2.180.

Haspelmath, Martin, 2004. On directionality in language change with particular reference to grammaticalization. In Olga Fischer, Muriel Norde and Harry Perridon (eds.), *Up and down the cline – the nature of grammaticalization, Typological Studies in Language*, vol. 59, pp. 17–44. Amsterdam: John Benjamins Publishing Company. ISBN 9789027229694. doi:10.1075/tsl.59.03has.

Haspelmath, Martin, 2006. Against markedness (and what to replace it with). *Journal of Linguistics* 42. doi:10.1017/S0022226705003683.

Haspelmath, Martin, 2008. Creating economical morphosyntactic patterns in language change. In Jeff Good (ed.), *Linguistic Universals and Language Change*, ch. 8, pp. 185–214. Oxford: Oxford University Press. ISBN 9780199298495. doi:10.1093/acprof:oso/9780199298495.001.0001.

Haspelmath, Martin, 2010. Comparative concepts and descriptive categories in crosslinguistic studies. *Language* 86(3): 663–687. doi:10.1353/lan.2010.0021.

Haspelmath, Martin, 2016. The evolution (or diachrony) of "language evolution". `http://dlc.hypotheses.org/894`.

Hay, Jennifer and Laurie Bauer, 2007. Phoneme inventory size and population size. *Language* 83(2): 388–400. doi:10.1353/lan.2007.0071.

Hay, Jennifer, Paul Warren and Katie Drager, 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.

Henrich, Joseph, Robert Boyd and Peter J. Richerson, 2008. Five misunderstandings about cultural evolution. *Human Nature* 19(2): 119–137. doi:10.1007/s12110-008-9037-1.

Herold, Ruth, 1997. Solving the actuation problem: Merger and immigration in eastern Pennsylvania. *Language Variation and Change* 9(02): 165. doi:10.1017/S0954394500001861.

Hickey, Raymond, 2012. Internally- and externally-motivated language change. In Juan Manuel Hernandez-Campoy and Juan Camilo Conde-Silvestre (eds.), *The Handbook of Historical Sociolinguistics*, ch. 21, p. 387. Blackwell.

Himmelmann, Lin, 2010. HMM - Hidden Markov Models. `https://cran.r-project.org/package=HMM`.

Hirshleifer, David, 1995. The blind leading the blind: social influence, fads and informational cascades. In Mariano Tommasi and Kathryn Ierulli (eds.), *The New Economics of Human Behaviour*, ch. 12, pp. 188–215. Cambridge University Press. ISBN 9780521479493, `http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1278625`.

Hockett, Charles F., 1958. *A course in modern linguistics*. New York: The Macmillan Company.

Hockett, Charles F., 1965. Sound change. *Language* 41(2): 185–204.

Hodge, Carleton E., 1970. The linguistic cycle. *Language Sciences* 13: 1–7.

Hopper, Paul J. and Elizabeth Closs Traugott, 1993. *Grammaticalization*. Cambridge: Cambridge University Press.

Hruschka, Daniel J, Morten H. Christiansen, Richard A. Blythe, William Croft, Paul Heggarty, Salikoko S. Mufwene, Janet B. Pierrehumbert and Shana Poplack, 2009. Building social cognitive models of language change. *Trends in Cognitive Sciences* 13(11): 464–469. doi:10.1016/j.tics.2009.08.008.

Hudson Kam, Carla L. and Elissa L. Newport, 2005. Regularizing unpredictable variation: the roles of adult and child learners in language formation and change. *Language Learning and Development* 1(2): 151–195. doi:10.1080/15475441.2005.9684215.

Hull, David L., 1988. *Science as a Process.* University of Chicago Press. ISBN 9780226360515.

Imhof, Lorens A. and Martin A. Nowak, 2006. Evolutionary game dynamics in a Wright-Fisher process. *Journal of Mathematical Biology* 52(5): 667–681. doi:10.1007/s00285-005-0369-8.

Jaeger, Herbert, Luc Steels, Andrea Baronchelli, Ted Briscoe, Morten H. Christiansen, Thomas L. Griffiths, Gerhard Jäger, Simon Kirby, Natalia L. Komarova, Peter J. Richerson and Jochen Triesch, 2009. What can mathematical, computational and robotic models tell us about the origins of syntax? In Derek Bickerton and Eörs Szathmáry (eds.), *Biological Foundations and Origin of Syntax*, ch. 18, pp. 385–410. Cambridge, Massachusetts: MIT Press. ISBN 978-0-262-01356-7.

Jaeger, T. Florian and Neal E. Snider, 2013. Alignment as a consequence of expectation adaptation: syntactic priming is affected by the prime's prediction error given both prior and recent experience. *Cognition* 127(1): 57–83. doi:10.1016/j.cognition.2012.10.013.

Jaeger, T. Florian and Harry Tily, 2010. On language 'utility': processing complexity and communicative efficiency. *Wiley Interdisciplinary Reviews: Cognitive Science* 2(3): 323–335. doi:10.1002/wcs.126.

Jäger, Gerhard, 2008. Language evolution and George Price's "General theory of selection". In Robin Cooper and Ruth Kempson (eds.), *Language in flux: Dialogue coordination, language variation, change and evolution*, pp. 53–80. London: College Publications, `http://www.sfs.uni-tuebingen.de/~gjaeger/publications/leverhulme.pdf`.

Jamieson, E, 2015. *An investigation of verb raising in the Shetland dialect of Scots.* MSc Dissertation, The University of Edinburgh. `http://hdl.handle.net/1842/16546`.

Jamieson, E, 2016. Imperative commands in the Shetland dialect of Scots: Nordic origins? In Ian Giles, Laura Chapot, Christian Cooijmans, Ryan Foster and Barbara Tesio (eds.), *Beyond Borealism: New Perspectives on the North*, pp. 98–116. London: Norvik Press Ltd. ISBN 9781909408333.

Jespersen, Otto, 1922. *Language: its nature, development and origin.* London: George Allen & Unwin Ltd., `https://archive.org/details/languageitsnatur00jespiala`.

Jespersen, Otto, 1949. *Efficiency in linguistic change.* Det Kgl. Danske videnskabernes selskab. Historisk-filologiske meddelelser. København: Ejnar Munksgaard, 2nd edn.

Jonas, Dianne, 2002. Residual V-to-I. In David W. Lightfoot (ed.), *Syntactic effects of morphological change*, ch. 14, pp. 251–270. Oxford: Oxford University Press, `https://books.google.com/books?id=C-MXVYAg2H8C&pg=PA251`.

Jones, Sir William, 1799. The third anniversary discourse, delivered 2 February, 1786: "On the Hindús". In *The Works of Sir William Jones*, vol. 1, pp. 19–34. London: Robinson, `http://gretil.sub.uni-goettingen.de/gretil_elib/Jon786__Jones_3rdAnniversaryDiscourse1799.pdf`.

Joseph, Brian D., 2013. Demystifying drift: a variationist account. In *Shared Grammaticalization: With special focus on the Transeurasian languages*, Studies in Language Companion Series 132, ch. 3, pp. 43–66. Amsterdam: John Benjamins Publishing Company. doi:10.1075/slcs.132.08jos.

Joseph, John E., 2000. *Limiting the Arbitrary: linguistic naturalism and its opposites in Plato's Cratylus and modern theories of language*, *Studies in the History of the Language Sciences*, vol. 96. Amsterdam: John Benjamins Publishing Company. ISBN 978 90 272 4585 4. doi:10.1075/sihols.96.

Joseph, John E., 2015. Iconicity in Saussure's linguistic work, and why it does not contradict the arbitrariness of the sign. *Historiographia Linguistica* 42(1): 85–105. doi:10.1075/hl.42.1.05jos.

Jurafsky, Dan and James H. Martin, 2008. *Speech and Language Processing*. Prentice Hall, 2nd edn. ISBN 978-0131873216.

Kanwal, Jasmeen, Kenny Smith, Jennifer Culbertson and Simon Kirby, 2017. Language users optimise form-meaning mappings for efficient communication. *Cognition* (in press).

Kaplan, Jonathan, 2008. The end of the adaptive landscape metaphor? *Biology & Philosophy* 23(5): 625–638. doi:10.1007/s10539-008-9116-z.

Kataoka, Nobuhide and Kunihiko Kaneko, 2000. Natural language from function dynamics. *Bio Systems* 57(1): 1–11.

Kauhanen, Henri, 2017. Neutral change. *Journal of Linguistics* 53(2): 327–358. doi:10.1017/S0022226716000141.

Kauhanen, Henri and George Walkden, 2015. A population dynamic mechanism for the constant rate effect (and beyond). In *22nd International Conference on Historical Linguistics*. Naples.

Kerswill, Paul, 1996. Children, adolescents, and language change. *Language Variation and Change* 8(2): 177–202. doi:10.1017/S0954394500001137.

Kerswill, Paul and Ann Williams, 1994. A new dialect in a new city Childrens' and adults' speech in Milton Keynes. Tech. rep., Economic and Social Research Council (ESRC), Swindon, UK. `http://hdl.handle.net/10068/452815`.

Kimura, Motoo, 1968. Evolutionary rate at the molecular level. *Nature* 217: 624–626.

Kimura, Motoo, 1983. *The neutral theory of molecular evolution*. Cambridge: Cambridge University Press. ISBN 0-521-23109-4.

Kimura, Motoo and Tomoko Ohta, 1969. The average number of generations until fixation of a mutant gene in a finite population. *Genetics* 61(3): 763–771.

Kingsland, Sharon, 1982. The refractory model: the logistic curve and the history of population ecology. *The Quarterly Review of Biology* 57(1): 29. doi:10.1086/412574.

Kiparsky, Paul, 1968. Linguistic universals and linguistic change. In Emmon Bach and Robert T. Harms (eds.), *Universals in linguistic theory*, pp. 170–202. New York: Holt, Rinehart, and Winston. ISBN 0039100774.

Kiparsky, Paul, 2014. New perspectives in historical linguistics. In Claire Bowern and Bethwyn Evans (eds.), *The Routledge Handbook of Historical Linguistics*. London: Routledge.

Kirby, James and Morgan Sonderegger, 2013. A model of population dynamics applied to phonetic change. In *Proceedings of the 35th Annual Meeting of the Cognitive Science Society*, pp. 776–781.

Kirby, James P, 2013. The role of probabilistic enhancement in phonologization. In Alan C L Yu (ed.), *Origins of Sound Change: Approaches to Phonologization*, ch. 11. Oxford: Oxford University Press. ISBN 978-0-19-957374-5.

Kirby, Simon, 1999. *Function, selection, and innateness*. Oxford: Oxford University Press. ISBN 978-0-19-823811-9.

Kirby, Simon, 2000. Syntax without natural selection: how compositionality emerges from vocabulary in a population of learners. In Chris Knight, M. Studdert-Kennedy and James R. Hurford (eds.), *The evolutionary emergence of language: social function and the origins of linguistic form*, pp. 303–323. Cambridge: Cambridge University Press. doi:10.1.1.52.8080.

Kirby, Simon, 2002. Learning, bottlenecks and the evolution of recursive syntax. In Edward J. Briscoe (ed.), *Linguistic evolution through language acquisition. Formal and computational models*. Cambridge: Cambridge University Press.

Kirby, Simon, Hannah Cornish and Kenny Smith, 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 of the United States of America* 105(31): 10681–10686.

Kirby, Simon, Mike Dowman and Thomas L. Griffiths, 2007. Innateness and culture in the evolution of language. *Proceedings of the National Academy of Sciences of the United States of America* 104(12): 5241–5245. doi:10.1073/pnas.0608222104.

Kirby, Simon, Kenny Smith and Henry Brighton, 2004. From UG to universals: linguistic adaptation through iterated learning. *Studies in Language* 28(3): 587–607. doi:10.1075/sl.28.3.09kir.

Kirby, Simon, Monica Tamariz, Hannah Cornish and Kenny Smith, 2015. Compression and communication in the cultural evolution of linguistic structure. *Cognition* 141: 87–102. doi:10.1016/j.cognition.2015.03.016.

Knooihuizen, Remco, 2009. Shetland Scots as a new dialect: phonetic and phonological considerations. *English Language and Linguistics* 13(03): 483. doi:10.1017/S1360674309990207.

Kroch, Anthony S., 1989a. Function and grammar in the history of English: periphrastic do. In Ralph Fasold and Deborah Schiffrin (eds.), *Language Change and Variation*, Current issues in linguistic theory 52, pp. 133–172. Amsterdam/Philadelphia: John Benjamins Publishing Company, `ftp://babel.ling.upenn.edu/papers/faculty/tony_kroch/papers/function-grammar-do.pdf`.

Kroch, Anthony S., 1989b. Reflexes of grammar in patterns of language change. *Language Variation and Change* 1(03): 199–244. doi:10.1017/S0954394500000168.

Kroch, Anthony S., 1994. Morphosyntactic variation. In Katherine Beals (ed.), *Papers from the 30th Regional Meeting of the Chicago Linguistics Society: Parasession on Variation and Linguistic Theory*.

Kroch, Anthony S., 2001. Syntactic change. In Mark Baltin and Chris Collins (eds.), *The Handbook of Contemporary Syntactic Theory*, Blackwell handbooks in linguistics, ch. 22, pp. 698–729. Malden, Massachusetts: Blackwell Publishers. ISBN 978-0-631-20507-4, `http://web.stanford.edu/~kiparsky/Papers/kroch01.pdf`.

Kroch, Anthony S., Susan Pintzuk and John Myhill, 1982. Understanding "do". In *Papers from the 18th Regional Meeting of the Chicago Linguistics Society*, pp. 282–294. Chicago: Chicago Linguistics Society.

Kroeber, Alfred Louis, 1919. On the principle of order in civilization as exemplified by changes of fashion. *American Anthropologist* 21(3): 235–263. doi:10.1525/aa.1919.21.3.02a00010.

Labov, William, 1966a. Hypercorrection by the lower middle class as a factor in linguistic change. In William Bright (ed.), *Sociolinguistics*. The Hague: Mouton.

Labov, William, 1966b. *The social stratification of English in New York City*. Washington DC: Center for Applied Linguistics.

Labov, William, 1972. *Sociolinguistic patterns*. Philadelphia: University of Pennsylvania Press.

Labov, William, 1989. The child as linguistic historian. *Language Variation and Change* 1(1): 85–97. doi:10.1017/S0954394500000120.

Labov, William, 1990. The intersection of sex and social class in the course of linguistic change. *Language Variation and Change* 2(2): 205–254. doi:10.1017/S0954394500000338.

Labov, William, 1994. *Principles of linguistic change: Internal factors*, *Language in Society 20*, vol. 1. Malden, Massachusetts: Wiley-Blackwell.

Labov, William, 2001. *Principles of linguistic change: Social factors*, *Language in Society 29*, vol. 2. Malden, Massachusetts: Wiley-Blackwell.

Labov, William, 2002. Driving forces in linguistic change. In *2002 International Conference on Korean Linguistics*. Seoul: Seoul National University.

Labov, William, 2010. *Principles of linguistic change: Cognitive and cultural factors*, *Language in Society 39*, vol. 3. Malden, Massachusetts: Wiley-Blackwell. ISBN 978-1-4051-1214-7.

Labov, William, 2012. What is to be learned: The community as the focus of social cognition. *Review of Cognitive Linguistics* 10(2): 265–293. doi:10.1075/rcl.10.2.02lab.

Labov, William, 2014. The sociophonetic orientation of the language learner. In Chiara Celata and Silvia Calamai (eds.), *Advances in Sociophonetics*, Studies in Language Variation 15, ch. 1, pp. 17–29. John Benjamins Publishing Company. ISBN 9789027234957. doi:10.1075/silv.15.01lab.

Ladd, D. Robert, Seán G. Roberts and Dan Dediu, 2015. Correlational studies in typological and historical linguistics. *Annual Review of Linguistics* 1(1): 221–241. doi:10.1146/annurev-linguist-030514-124819.

Larsen-Freeman, Diane, 1997. Chaos/Complexity Science and Second Language Acquisition. *Applied Linguistics* 18(2): 141–165. doi:10.1093/applin/18.2.141.

Lass, Roger, 1980. *On explaining language change.* Cambridge: Cambridge University Press. ISBN 0521228360.

Lass, Roger, 1990. How to do things with junk: exaptation in language evolution. *Journal of Linguistics* 26(1): 79–102. doi:10.1017/S0022226700014432.

Lau, Tyler, 2016. Modeling the gender system mergers from Latin to Romance. In *90th Annual Meeting of the Linguistic Society of America (LSA)*. Washington DC.

Le Page, Robert B. and Andrée Tabouret-Keller, 1985. *Acts of identity: creole-based approaches to language and ethnicity.* Cambridge: Cambridge University Press. ISBN 9780521316040.

Lewis, Hannah M. and Kevin N. Laland, 2012. Transmission fidelity is the key to the build-up of cumulative culture. *Philosophical Transactions of the Royal Society B: Biological Sciences* 367(1599): 2171–2180. doi:10.1098/rstb.2012.0119.

Lieberman, Erez, Jean-Baptiste Michel, Joe Jackson, Tina Tang and Martin A. Nowak, 2007. Quantifying the evolutionary dynamics of language. *Nature* 449(7163): 713–716. doi:10.1038/nature06137.

Lightfoot, David, 1991. *How to set parameters: Arguments from language change.* Cambridge, Massachusetts: MIT Press.

Lightfoot, David, 2010. Language acquisition and language change. *Wiley Interdisciplinary Reviews: Cognitive Science* 1(5): 677–684. doi:10.1002/wcs.39.

Lindgren, Kaj Brynolf, 1953. *Die Apokope des Mhd. -e in seinen verschiedenen Funktionen*, *Annales Academiae scientiarum fennicae*, vol. 78. Helsinki: Suomalaisen Tiedeakatemian Toimituksia.

Livingstone, Daniel, 2000. Neutral evolution and linguistic diversity. `http://cis.paisley.ac.uk/research/reports/`.

Lupyan, Gary and Rick Dale, 2010. Language structure is partly determined by social structure. *PloS one* 5(1): e8559. doi:10.1371/journal.pone.0008559.

Luraghi, Silvia, 2010. Causes of language change. In Silvia Luraghi and Vit Bubenik (eds.), *Continuum Companion to Historical Linguistics*, pp. 358–370. London: Continuum International Publishing Group.

MacWhinney, Brian, Andrej Malchukov and Edith Moravcsik (eds.), 2014. *Competing motivations in grammar and usage.* Oxford: Oxford University Press. ISBN 9780198709848. doi:10.1093/acprof:oso/9780198709848.001.0001.

Maegaard, Marie, Torben Juel Jensen, Tore Kristiansen and Jens Normann Jørgensen, 2013. Diffusion of language change: Accommodation to a moving target. *Journal of Sociolinguistics* 17(1): 3–36. doi:10.1111/josl.12002.

Mahowald, Kyle, Evelina Fedorenko, Steven T. Piantadosi and Edward Gibson, 2013. Info/information theory: Speakers choose shorter words in predictive contexts. *Cognition* 126(2): 313–318. doi:10.1016/j.cognition.2012.09.010.

Martinet, André, 1955. *Économie des changements phonétiques: traité de phonologie diachronique.* Bern: A. Francke.

Matthews, Cristina, Gareth Roberts and Christine A. Caldwell, 2012. Opportunity to assimilate and pressure to discriminate can generate cultural divergence in the laboratory. *Evolution and Human Behavior* 33(6): 759–770. doi:10.1016/j.evolhumbehav.2012.06.004.

McMahon, April and Robert McMahon, 2005. *Language classification by numbers.* Oxford University Press. ISBN 9780199279012, `https://books.google.com/books?id=Hl3HPyT4WcQC&printsec=frontcover`.

McMahon, April M. S., 1994. *Understanding language change.* Cambridge: Cambridge University Press. ISBN 9781139166591. doi:10.1017/CBO9781139166591.

Meillet, Antoine, 1926. *Linguistique historique et linguistique générale.* Paris: Librairie Ancienne Honore Champion, 2nd edn., `https://archive.org/details/linguistiquehist00meil`.

Mesoudi, Alex, 2011. *Cultural Evolution.* Chicago: University of Chicago Press. ISBN 9780226520445, `http://www.press.uchicago.edu/ucp/books/book/chicago/C/bo8787504.html`.

Mesoudi, Alex and Stephen J. Lycett, 2009. Random copying, frequency-dependent copying and culture change. *Evolution and Human Behavior* 30(1): 41–48. doi:10.1016/j.evolhumbehav.2008.07.005.

Michaud, Jérôme, 2016. Linking the processes of language evolution and language change: a five-level hierarchy. In *The Evolution of Language: Proceedings of the 11th International Conference (EVOLANG11)*. New Orleans.

Michaud, Jérôme, 2017. Continuous time limits of the utterance selection model. *Physical Review E* 95(2): 022308. doi:10.1103/PhysRevE.95.022308.

Millar, Robert McColl, 2007. *Northern and Insular Scots*. Dialects of English. Edinburgh: Edinburgh University Press.

Milroy, James and Lesley Milroy, 1985. Linguistic change, social network and speaker innovation. *Journal of Linguistics* 21(2): 339–384. doi:10.1017/S0022226700010306.

Mitchener, W. Garrett, 2011. A mathematical model of prediction-driven instability: how social structure can drive language change. *Journal of Logic, Language and Information* 20: 385–396. doi:10.1007/s10849-011-9136-y.

Monaghan, Padraic, 2014. Age of acquisition predicts rate of lexical evolution. *Cognition* 133(3): 530–534. doi:10.1016/j.cognition.2014.08.007.

Montgomery, Chris and Joan Beal, 2011. Perceptual dialectology. In Warren Maguire and April McMahon (eds.), *Analysing Variation in English*, ch. 6, pp. 121–148. Cambridge: Cambridge University Press. ISBN 9780511976360. doi:10.1017/CBO9780511976360.007.

Nardy, Aurélie, Jean-Pierre Chevrot and Stéphanie Barbu, 2013. The acquisition of sociolinguistic variation: Looking back and thinking ahead. *Linguistics* 51(2): 255–284. doi:10.1515/ling-2013-0011.

Nedergaard Thomsen, Ole (ed.), 2006. *Competing models of linguistic change: evolution and beyond*, *Current Issues in Linguistic Theory*, vol. 279. Amsterdam: John Benjamins Publishing Company. ISBN 978 90 272 4794 0. doi:10.1075/cilt.279.

Nettle, Daniel, 1999. *Linguistic diversity*. Oxford: Oxford University Press. ISBN 9780198238577.

Nettle, Daniel, 2012. Social scale and structural complexity in human languages. *Philosophical Transactions of the Royal Society B: Biological Sciences* 367(1597): 1829–1836. doi:10.1098/rstb.2011.0216.

Nevalainen, Terttu, Helena Raumolin-Brunberg and Heikki Mannila, 2011. The diffusion of language change in real time: Progressive and conservative individuals and the time depth of change. *Language Variation and Change* 23(1): 1–43. doi:10.1017/S0954394510000207.

Nevalainen, Terttu, Tanja Säily and Turo Vartiainen, 2014. Upcoming resource: An online Language Change Database. In *International Society for the Linguistics of English (ISLE 3)*. Zurich, Switzerland.

Nevalainen, Terttu, Turo Vartiainen, Tanja Säily, Joonas Kesäniemi, Agata Dominowska and Emily Öhman, 2016. Language Change Database: A new online resource. *ICAME Journal* forthc.

Newmeyer, Frederick J., 1998. *Language form and language function*. Cambridge, Massachusetts: MIT Press. ISBN 9780262640442.

Newmeyer, Frederick J., 2014. Where do motivations compete? In Brian MacWhinney, Andrej Malchukov and Edith Moravcsik (eds.), *Competing motivations in grammar and usage*, ch. 18, pp. 299–314. Oxford: Oxford University Press. doi:10.1093/acprof:oso/9780198709848.003.0018.

Niyogi, Partha, 2006. *The computational nature of language learning and evolution*. Current Studies in Linguistics. Cambridge, Massachusetts: The MIT Press. ISBN 978-0262513395.

Niyogi, Partha and Robert C. Berwick, 1995. The logical problem of language change. Tech. rep., MIT AI Laboratory and Center for Biological and Computational Learning: A.I. Memo No. 1516/CBCL Paper No. 115, Cambridge, Massachusetts. `ftp://publications.ai.mit.edu/ai-publications/pdf/AIM-1516.pdf`.

Niyogi, Partha and Robert C. Berwick, 1997. Evolutionary consequences of language learning. *Linguistics and Philosophy* 20(6): 697–719. doi:10.1023/A:1005319718167.

Nowak, Martin A., Natalia L. Komarova and Partha Niyogi, 2001. Evolution of universal grammar. *Science* 291(5501): 114–118. doi:10.1126/science.291.5501.114.

Ohala, John J., 1989. Sound change is drawn from a pool of synchronic variation. In Leiv Egil Breivik and Ernst Håkon Jahr (eds.), *Language change: contributions to the study of its causes*, pp. 173–198. Berlin: Mouton de Gruyter.

Ohala, John J., 1993. The phonetics of sound change. In Charles Jones (ed.), *Historical linguistics: problems and perspectives*, pp. 237–278. London: Longman.

Otero-Espinar, M. Victoria, Luís F. Seoane, Juan J. Nieto and Jorge Mira, 2013. An analytic solution of a model of language competition with bilingualism and interlinguistic similarity. *Physica D: Nonlinear Phenomena* 264: 17–26. doi:10.1016/j.physd.2013.08.011.

Paul, Hermann, 1880. *Prinzipien der Sprachgeschichte*. Halle: Max Niemeyer.

Pearl, Lisa and Amy Weinberg, 2007. Input filtering in syntactic acquisition: answers from language change modeling. *Language Learning and Development* 3(1): 43–72. doi:10.1080/15475440709337000.

Perniss, Pamela, Robin L. Thompson and Gabriella Vigliocco, 2010. Iconicity as a general property of language: evidence from spoken and signed languages. *Frontiers in Psychology* 1(227). doi:10.3389/fpsyg.2010.00227.

Pierrehumbert, Janet B., Forrest Stonedahl and Robert Daland, 2014. A model of grassroots changes in linguistic systems. *arXiv* 1408.1985.

Pintzuk, Susan, 2003. Variationist approaches to syntactic change. In Brian D Joseph and Richard D Janda (eds.), *The Handbook of Historical Linguistics*, ch. 15, pp. 509–528. Oxford: Blackwell Publishing Ltd. doi:10.1002/9780470756393.ch15.

Piotrovskaja, A. A. and R. G. Piotrovskij, 1974. Matematiceskie Modeli Diachronii i Tekstoobrazovanija. In *Statistika Reci i Avtomaticeskij Analiz Teksta*, pp. 361–400. Leningrad: Nauka.

Postal, Paul Martin, 1968. *Aspects of phonological theory.* New York: Harper & Row.

Preston, Dennis R., 1996. Whaddayaknow?: The modes of folk linguistic awareness. *Language Awareness* 5(1): 40–74. doi:10.1080/09658416.1996.9959890.

Preston, Dennis R. and Gregory C. Robinson, 2005. Dialect perception and attitudes to variation. In Martin J. Ball (ed.), *Clinical Sociolinguistics*, ch. 11, pp. 133–152. Malden, Massachusetts: Blackwell Publishing Ltd. doi:10.1002/9780470754856.ch11.

Reali, Florencia and Thomas L. Griffiths, 2009. The evolution of frequency distributions: Relating regularization to inductive biases through iterated learning. *Cognition* 111(3): 317–328. doi:10.1016/j.cognition.2009.02.012.

Reali, Florencia and Thomas L. Griffiths, 2010. Words as alleles: connecting language evolution with Bayesian learners to models of genetic drift. *Proceedings of the Royal Society B: Biological Sciences* 277(1680): 429–436. doi:10.1098/rspb.2009.1513.

Renton, Jessie, 2016. *The effect of population structure on spontaneous dialect formation.* MPhys dissertation, The University of Edinburgh.

Ritt, Nikolaus, 2004. *Selfish Sounds and Linguistic Evolution: A Darwinian Approach to Language Change.* Cambridge: Cambridge University Press. ISBN 9780521120630, `https://books.google.com/books?id=jGGAaOZxA_gC&printsec=frontcover`.

Ritt, Nikolaus and Andreas Baumann, 2012. Transferring mathematics to English studies: prospects and problems. In Sabine Coelsch-Foisner, Manfred Markus and Herbert Schendl (eds.), *Transfer in English studies*, Austrian Studies in English. Vienna: Braumüller.

Roberts, Gareth, 2013. Perspectives on language as a source of social markers. *Language and Linguistics Compass* 7(12): 619–632. doi:10.1111/lnc3.12052.

Roberts, Seán and James Winters, 2013. Linguistic diversity and traffic accidents: lessons from statistical studies of cultural traits. *PLoS ONE* 8(8): e70902. doi:10.1371/journal.pone.0070902.

Roberts, Seán G. and James Winters, 2012. Constructing knowledge: nomothetic approaches to language evolution. In Luke McCrohon, T. Fujimura, K. Fujita, R. Martin, K. Okanoya, R. Suzuki and N. Yusa (eds.), *Five Approaches to Language Evolution: Proceedings of the Workshops of the 9th International Conference on the Evolution of Language*, pp. 148–157.

Rogers, Everett M., 1962. *Diffusion of innovations.* New York: Free Press of Glencoe. ISBN 0-612-62843-4.

Ross, Malcolm and Mark Durie, 1996. Introduction. In Mark Durie and Malcolm Ross (eds.), *The Comparative Method reviewed: regularity and irregularity in language change*, ch. 1, pp. 3–38. Oxford: Oxford University Press. ISBN 0-19-506607-3.

Salganik, Matthew J. and Duncan J. Watts, 2008. Leading the herd astray: An experimental study of self-fulfilling prophecies in an artificial cultural market. *Social Psychology Quarterly* 71(4): 338–355. doi:10.1177/019027250807100404.

Salmons, Joseph and Patrick Honeybone, 2013. Structuralist historical phonology: systems in sound change. In Joseph Salmons and Patrick Honeybone (eds.), *The Oxford handbook of historical phonology*, ch. 3. Oxford: Oxford University Press, `http://www.lel.ed.ac.uk/homes/patrick/hphhndbk-structuralism.pdf`.

Sankoff, David, 1970. On the rate of replacement of word-meaning relationships. *Language* 46(3): 564–569. doi:10.2307/412307.

Sankoff, David, 1988. Sociolinguistics and syntactic variation. In Frederick J. Newmeyer (ed.), *Linguistics: The Cambridge Survey*, ch. 8, pp. 140–161. Cambridge: Cambridge University Press. ISBN 9780511620577. doi:10.1017/CBO9780511620577.

Sankoff, Gillian, 2013. Language change across the lifespan. In *Plenary address to the Linguistic Society of America Annual Meeting*, p. 81. Boston, Massachusetts: Linguistic Society of America.

Sankoff, Gillian and Hélèn Blondeau, 2007. Language change across the lifespan: /r/ in Montreal French. *Language* 83(3): 560–588. doi:10.1353/lan.2007.0106.

Sankoff, Gillian and Suzanne Laberge, 1973. On the acquisition of native speakers by a language. *Kivung* 6: 32–47.

Sapir, Edward, 1921. *Language: An Introduction to the Study of Speech.* Harcourt, `http://www.bartleby.com/186/`.

Sapir, Edward, 1929. The status of linguistics as a science. *Language* 5(4): 207–214. doi:10.2307/409588.

de Saussure, Ferdinand, 1916. *Cours de linguistique générale.* New Haven: Yale University Press.

de Saussure, Ferdinand, 1959. *Course in general linguistics.* New York: The Philosophical Library, Inc., `https://archive.org/details/courseingeneral00saus`.

Scott-Phillips, Thomas C. and Simon Kirby, 2010. Language evolution in the laboratory. *Trends in Cognitive Sciences* 14(9): 411–417.

Shetland Islands Council, 2014. *Shetland in Statistics*, vol. 41. Lerwick, UK: Shetland Islands Council, Economic Development Unit, Solarhus, 3 North Ness Business Park, Lerwick, Shetland, ZE1 0LZ. ISBN 978 0 905924 71 7, `http://www.shetland.gov.uk/economic_development/documents/ShetlandinStatisticsforwebsite.pdf`.

Silva, Eduardo J. S. and Viviane M. de Oliveira, 2008. Evolution of the linguistic diversity on correlated landscapes. *Physica A: Statistical Mechanics and its Applications* 387(22): 5597–5601. doi:10.1016/j.physa.2008.06.005.

Situngkir, Hokky, 2004. On selfish memes: culture as complex adaptive system. *Journal of social complexity* 2(1): 20–32.

Skyrms, Brian, 2010. *Signals. Evolution, learning, and information.* Oxford: Oxford University Press. ISBN 9780199582945.

Smith, Andrew D. M., 2016. Dynamic models of language evolution: the linguistic perspective. In *The Palgrave Handbook of Economics and Language*, ch. 2, pp. 61–100. Basingstoke: Palgrave, `http://hdl.handle.net/1893/22890`.

Smith, Kenneth, 2003. *The transmission of language: models of biological and cultural evolution.* Ph.D. thesis, Dept of Linguistics, University of Edinburgh. `http://www.lel.ed.ac.uk/~kenny/thesis.html`.

Smith, Kenny, Mónica Tamariz and Simon Kirby, 2013. Linguistic structure is an evolutionary trade-off between simplicity and expressivity. In Markus Knauff, Michael Pauen, Natalie Sebanz and Ipke Wachsmuth (eds.), *Proceedings of the 35th Annual Meeting of the Cognitive Science Society*, pp. 1348–1353. Austin, Texas: Cognitive Science Society.

Smith, Kenny and Elizabeth Wonnacott, 2010. Eliminating unpredictable variation through iterated learning. *Cognition* 116(3): 444–449. doi:10.1016/j.cognition.2010.06.004.

Sonderegger, Morgan and Partha Niyogi, 2010. Combining data and mathematical models of language change. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, pp. 1019–1029.

Sorace, Antonella and Frank Keller, 2005. Gradience in linguistic data. *Lingua* 115(11): 1497–1524. doi:10.1016/j.lingua.2004.07.002.

Sóskuthy, Márton, 2013. *Phonetic biases and systemic effects in the actuation of sound change.* Ph.D. thesis, The University of Edinburgh. `https://www.era.lib.ed.ac.uk/handle/1842/8946`.

Spike, Matthew, 2016. *Minimal requirements for the cultural evolution of language.* Ph.D. thesis, The University of Edinburgh.

Spike, Matthew, Kevin Stadler, Simon Kirby and Kenny Smith, 2016. Minimal requirements for the emergence of learned signalling. *Cognitive Science* in press. doi:10.1111/cogs.12351.

Stadler, Kevin, 2009. *Cultural transmission and inductive biases in populations of Bayesian learners*. MSc Dissertation, The University of Edinburgh. `http://hdl.handle.net/1842/3617`.

Stadler, Kevin, Richard A. Blythe, Kenny Smith and Simon Kirby, 2016. Momentum in language change: a model of self-actuating s-shaped curves. *Language Dynamics and Change* 6(2): 171–198. doi:10.1163/22105832-00602005.

Stanford, James N., 2014. Language acquisition and language change. In Claire Bowern and Bethwyn Evans (eds.), *The Routledge Handbook of Historical Linguistics*, ch. 21, pp. 466–483. London: Routledge.

Stanford, James N., Nathan A. Severance and Kenneth P. Baclawski, 2014. Multiple vectors of unidirectional dialect change in eastern New England. *Language Variation and Change* 26(01): 103–140. doi:10.1017/S0954394513000227.

Steels, Luc, 1995. A self-organizing spatial vocabulary. *Artificial Life* 2(3): 319–332. doi:10.1162/artl.1995.2.3.319.

Steels, Luc, 1998. The origins of ontologies and communication conventions in multi-agent systems. *Autonomous Agents and Multi-Agent Systems* 1(2): 169–194. doi:10.1023/A:1010002801935.

Steels, Luc, 2000. Language as a Complex Adaptive System. In Marc Schoenauer, Kalyanmoy Deb, Günter Rudolph, Xin Yao, Evelyne Lutton, Juan Julian Merelo and Hans-Paul Schwefel (eds.), *Parallel Problem Solving from Nature PPSN VI*, pp. 17–26. Springer. doi:10.1007/3-540-45356-3_2.

Steels, Luc and Angus McIntyre, 1998. Spatially distributed naming games. *Advances in Complex Systems* 1(4): 301–323. doi:10.1142/S021952599800020X.

Stevens, Mary and Jonathan Harrington, 2013. The individual and the actuation of sound change. *Loquens* 1(1): e003. doi:10.3989/loquens.2014.003.

Sturtevant, Edgar Howard, 1947. *An introduction to linguistic science*. New Haven: Yale University Press, `https://archive.org/details/introductiontoli032816mbp`.

Sundgren, Eva, 2001. Men and women in language change. A Swedish case study. *NORA - Nordic Journal of Feminist and Gender Research* 9: 113–123. doi:10.1080/080387401753355344.

Sundkvist, Peter, 2011. The Shetland Islands: globalisation and the changing status of Standard English. *English Today* 27(04): 19–25. doi:10.1017/S0266078411000496.

Swarup, Samarth and Corrine McCarthy, 2012. Representational momentum may explain aspects of vowel shifts. In *Artificial Life 13*, pp. 267–274. MIT Press. ISBN 9780262310505. doi:10.7551/978-0-262-31050-5-ch036.

Tagliamonte, Sali A., 2012. *Variationist sociolinguistics. Change, observation, interpretation.* Language in Society. Wiley-Blackwell.

Tagliamonte, Sali A., 2015. *Making waves: the story of variationist sociolinguistics.* Wiley-Blackwell. ISBN 978-1-118-45516-6.

Tagliamonte, Sali A. and Alexandra D'Arcy, 2009. Peaks beyond phonology: adolescence, incrementation, and language change. *Language* 85(1): 58–108. doi:10.1353/lan.0.0084.

Tarde, Gabriel, 1903. *The laws of imitation.* New York: Henry Holt and Company, `https://archive.org/details/lawsofimitation00tard`.

Tarnita, Corina E. and Peter D. Taylor, 2014. Measures of relative fitness of social behaviors in finite structured population models. *The American Naturalist* 184(4): 477–488. doi:10.1086/677924.

Traugott, Elizabeth Closs, 2011. Grammaticalization and mechanisms of change. In Bernd Heine and Heiko Narrog (eds.), *The Oxford Handbook of Grammaticalization*, ch. 2, pp. 19–30. Oxford: Oxford University Press. ISBN 9780199586783. doi:10.1093/oxfordhb/9780199586783.013.0002.

van Trijp, Remi, 2013. Linguistic assessment criteria for explaining language change: a case study on syncretism in German definite articles. *Language Dynamics and Change* 3: 1–28. doi:10.1163/22105832-13030106.

Troutman, Celina, Brady Clark and Matthew Goldrick, 2008. Social networks and intraspeaker variation during periods of language change. *University of Pennsylvania Working Papers in Linguistics* 14(1): 325–338.

Trudgill, Peter, 1972. Sex, covert prestige and linguistic change in the urban British English of Norwich. *Language in Society* 1(02): 179–195. doi:10.1017/S0047404500000488.

Trudgill, Peter, 2004. *New-dialect formation: the inevitability of Colonial Englishes.* Edinburgh: Edinburgh University Press.

Trudgill, Peter, 2008. Colonial dialect contact in the history of European languages: On the irrelevance of identity to new-dialect formation. *Language in Society* 37(02): 241–254. doi:10.1017/S0047404508080287.

Vachek, Josef, 1962. On the interplay of external and internal factors in the development of language. *Lingua* 11: 433–448. doi:10.1016/0024-3841(62)90053-0.

Vennemann, Theo, 1983. Causality in language change: theories of linguistic preferences as a basis for linguistic explanations. *Folia Linguistica Historica* VI(1): 5–26. doi:10.1515/flih.1983.4.1.5.

Vennemann, Theo, 1993. Language change as language improvement. In Charles Jones (ed.), *Historical linguistics: problems and perspectives*, ch. 11, pp. 319–344. London: Longman, `https://books.google.com/books?id=gODKAgAAQBAJ&pg=PA319`.

Vogt, Paul and Elena Lieven, 2010. Verifying theories of language acquisition using computer models of language evolution. *Adaptive Behavior* 18(1): 21–35. doi:10.1177/1059712309350970.

Vulanović, Relja, 2007. Fitting periphrastic do in affirmative declaratives. *Journal of Quantitative Linguistics* 14(2-3): 111–126. doi:10.1080/09296170701379369.

Vulanović, Relja and Harald Baayen, 2007. Fitting the development of periphrastic do in all sentence types. In Peter Grzybek and Reinhard Köhler (eds.), *Exact methods in the study of language and text: dedicated to Gabriel Altmann on the occasion of his 75th birthday*, Quantitative linguistics 62, pp. 679–688. Berlin: Mouton de Gruyter. ISBN 9783110193541, `http://hdl.handle.net/2066/42139`.

Wagner, Suzanne Evans, 2012. Age grading in sociolinguistic theory. *Language and Linguistics Compass* 6: 371–382. doi:10.1002/lnc3.343.

Walker, Abby and Jennifer Hay, 2011. Congruence between 'word age' and 'voice age' facilitates lexical access. *Laboratory Phonology* 2(1): 219–237. doi:10.1515/labphon.2011.007.

WALS Online, 2013. The World Atlas of Language Structures Online. `http://wals.info`.

Wang, William Shi-Yuan, 1969. Competing changes as a cause of residue. *Language* 45(1): 9–25. doi:10.2307/411748.

Wang, William Shi-Yuan, Jinyun Ke and James W. Minett, 2004. Computational studies of language evolution. In Chu-Ren Huang and Winfried Lenders (eds.), *Computational linguistics and beyond: perspectives at the beginning of the 21st century*, Frontiers in Linguistics 1, pp. 65–106. Taiwan: Academia Sinica. doi:10.1.1.535.3186.

Wedel, Andrew, 2004. Category competition drives contrast maintenance within an exemplar-based production/perception loop. In *Proceedings of the 7th Meeting of the ACL Special Interest Group in Computational Phonology*, pp. 1–10. Barcelona: Association for Computational Linguistics. doi:10.3115/1622153.1622154.

Wedel, Andrew, 2009. Resolving pattern conflict: variation and selection in phonology and morphology. In James P. Blevins and Juliette Blevins (eds.), *Analogy in Grammar: Form and Acquisition*, ch. 4. Oxford: Oxford University Press. ISBN 9780199547548. doi:10.1093/acprof:oso/9780199547548.003.0004.

Wedel, Andrew, Abby Kaplan and Scott Jackson, 2013. High functional load inhibits phonological contrast loss: a corpus study. *Cognition* 128(2): 179–186. doi:10.1016/j.cognition.2013.03.002.

Wedel, Andrew B., 2006. Exemplar models, evolution and language change. *The Linguistic Review* 23(3): 247–274. doi:10.1515/TLR.2006.010.

Wedel, Andrew B., 2007. Feedback and regularity in the lexicon. *Phonology* 24(01): 147–185. doi:10.1017/S0952675707001145.

Weinreich, Uriel, William Labov and Marvin Herzog, 1968. Empirical foundations for a theory of language change. In Winfred P. Lehmann and Yakov Malkiel (eds.), *Directions for Historical Linguistics: A Symposium*, pp. 95–188. University of Texas Press, `https://liberalarts.utexas.edu/lrc/resources/books/directions/5-weinreich.php`.

Wellens, Pieter, 2012. *Adaptive strategies in the emergence of lexical systems.* Ph.D. thesis, Vrije Universiteit Brussel.

Wells, John C., 1982. *Accents of English.* Cambridge: Cambridge University Press. ISBN 9780511611759. doi:10.1017/CBO9780511611759.

Wichmann, Søren, 2015. Diachronic stability and typology. In Claire Bowern and Bethwyn Evans (eds.), *The Routledge Handbook of Historical Linguistics*, ch. 8, pp. 212–224. London: Routledge.

Wichmann, Søren and Eric W. Holman, 2009. Population size and rates of language change. *Human Biology* 81(2): 259–274.

Willer, Robb, Ko Kuwabara and Michael W. Macy, 2009. The false enforcement of unpopular norms. *American Journal of Sociology* 115(2): 451–490. doi:10.1086/599250.

Williams, Ann and Paul Kerswill, 1999. Dialect levelling: change and continuity in Milton Keynes, Reading and Hull. In Paul Foulkes and Gerard Docherty (eds.), *Urban Voices: Accent Studies in the British Isles*, ch. 8, pp. 141–162. Routledge. ISBN 9780340706084.

Winter, Bodo and Martijn Wieling, 2016. How to analyze linguistic change using mixed models, Growth Curve Analysis and Generalized Additive Modeling. *Journal of Language Evolution* 1(1): 7–18. doi:10.1093/jole/lzv003.

Winter-Froemel, Esme, 2008. Towards a comprehensive view of language change: Three recent evolutionary approaches. In Ulrich Detges and Richard Waltereit (eds.), *The Paradox of Grammatical Change: Perspectives from Romance*, *Current Issues in Linguistic Theory*, vol. 293, pp. 215–250. Amsterdam: John Benjamins Publishing Company. doi:10.1075/cilt.293.10win.

Winters, James, Simon Kirby and Kenny Smith, 2015. Languages adapt to their contextual niche. *Language and Cognition* 7(3): 415–449. doi:10.1017/langcog.2014.35.

Wray, Alison and George W. Grace, 2007. The consequences of talking to strangers: Evolutionary corollaries of socio-cultural influences on linguistic form. *Lingua* 117(3): 543–578. doi:10.1016/j.lingua.2005.05.005.

Wright, Sewall, 1931. Evolution in Mendelian populations. *Genetics* 16: 97–159. doi:10.1016/S0092-8240(05)80011-4.

Wright, Sewall, 1943. Isolation by distance. *Genetics* 28(2): 114–138.

Yang, Charles D, 2002. *Knowledge and learning in natural language.* Oxford: Oxford University Press. ISBN 0-19-925414-1.

Zipf, George Kingsley, 1935. *The psycho-biology of language: an introduction to dynamic philology.* Boston: Houghton Mifflin Co.

Zipf, George Kingsley, 1949. *Human behaviour and the principle of least effort.* Addison-Wesley.

Index

Appendix A

Derivation of replicator selection trajectories in the Utterance Selection Model

The mathematical appendix to Blythe and Croft (2012) lays out how one can analytically derive the average trajectory resulting from the differential replication of variants according to some biasing function $f(.)$ as a function of some of the USM's other parameters, in particular the learning rate λ. These analyses rely on a number of simplifying assumptions, most importantly the *homogeneity principle*, which "requires two speakers to have the same probability of using an innovation at any point in time (after an initial relaxation)" (p.2). Although this principle has only been proven to hold for the case of random copying (Blythe, 2010) some of our assumptions, such as the mutually strong weighting of the agents' productions through a single high accommodation term h as well as the densely connected network specified by the interaction matrix G, go a long way towards satisfying the criterion, which essentially states that there should be no *systematic* differences between the different agents' usage levels across the population.

In combination with the *mean-field assumption* that, "at any given time, a speaker's usage frequency is well-represented by its mean value" (p.3), in other words disregarding the influence of random fluctuations due to sampling effects, the dynamics of the population can then be described as a deterministic function of time[1].

In particular, given a bias function of the general form $f(u) = u + \lambda \cdot g(u)$ the mean change to the average usage frequency x of the population in one interaction is:

$$\langle (x' - x) \rangle = \frac{\lambda^2}{N} \cdot \langle g(\frac{n}{T}) \rangle + O(\lambda^3)$$

where $\langle g(\frac{n}{T}) \rangle$ represents the weighted mean value of $g(.)$ over all possible sample productions $n \sim Bin(T, x)$. To abstract away from the number of interactions to a time scale that generalises across different learning rates and population sizes we take one interaction to last λ^2/N time units (see Baxter et al., 2006, 2009), which yields the deterministic mean-field equation

$$\frac{d}{dt}x(t) = 2 \cdot \langle g(\frac{n}{T}) \rangle \ .$$

In other words, the average trajectory is simply dependent on the average of the function $g(u)$ over the distribution of token productions for a given current population mean x. Assuming that tokens are produced independently, this is a binomial distribution with $n = T$ and $p = x$, which is the probability distribution $P(.)$ referred to in all equations below. Filling in the two different replicator biases (original multiplicative and additive) and solving the differential equations for them we can thus deduce the average trajectories of the biases for different T.

A.1   Multiplicative replicator selection

To recapitulate,

$$f(u) = u \cdot (1 + b) = u + u \cdot b$$

---

[1]For an alternative approach that uses *stability analysis* to determine the influence of random fluctuations, see Renton (2016).

which means that

$$g(u) = \frac{b \cdot u}{\lambda} \, .$$

Assuming that b is sufficiently small so that $g(\frac{n}{T}) = \frac{b}{\lambda}\frac{n}{T}$ for all $n < T$ (i.e. that the biased value $(1 + b) \cdot \frac{T-1}{T}$ never exceeds the maximum of 1 which is true as long as $b < \frac{1}{T-1}$), we can write the average of $g(u)$ as:

$$\langle g(\frac{n}{T}) \rangle = \frac{b}{\lambda} \sum_{n=0}^{T-1} P(n)\frac{n}{T} = \frac{b}{\lambda}(\langle \frac{n}{T} \rangle - P(T))$$

where the final transformation is simply a reformulation of the weighted mean of the Binomial distribution for all $n < 0 \ldots T - 1$. Invoking the deterministic mean-field assumption we can substitute the average population usage x for $\langle \frac{n}{T} \rangle$ to find

$$\langle g(\frac{n}{T}) \rangle = \frac{b}{\lambda}x(1 - x^{T-1})$$

and thus

$$\frac{d}{dt}x(t) = \frac{2b}{\lambda}x(1 - x^{T-1})$$

and

$$\frac{1}{x(1 - x^{T-1})}dx(t) = \frac{2b}{\lambda}dt \, .$$

Integrating

$$\int (\frac{1}{x} + \frac{x^{T-2}}{1 - x^{T-1}}) = \frac{2bt}{\lambda}$$

we find

$$\ln|x| - \frac{\ln|1 - x^{T-1}|}{T - 1} = \frac{2bt}{\lambda} + c \, .$$

Choosing c so that $x(0) = x_0$,

$$c = \ln|x_0| - \frac{\ln|1 - x_0^{T-1}|}{T - 1}$$

$$C = \frac{x_0}{(1 - x_0^{T-1})^{\frac{1}{T-1}}} \, .$$

Returning to the original equation and solving for x

$$\frac{|x|}{|1 - x^{T-1}|^{\frac{1}{T-1}}} = C\exp^{2bt/\lambda}$$

$$x = C\exp^{2bt/\lambda}|1 - x^{T-1}|^{\frac{1}{T-1}}$$

$$x^{T-1} = C^{T-1} \exp^{(2bt/\lambda)(T-1)}(1 - x^{T-1})$$

$$x^{T-1}(1 + C^{T-1}\exp^{(2bt/\lambda)(T-1)}) = C^{T-1}\exp^{(2bt/\lambda)(T-1)}$$

$$x^{T-1} = \frac{C^{T-1}\exp^{(2bt/\lambda)(T-1)}}{1 + C^{T-1}\exp^{(2bt/\lambda)(T-1)}} = \frac{C^{T-1}}{1 - \exp^{-(2bt/\lambda)(T-1)}} \ .$$

Substituting $C^{T-1} = \frac{x_0^{T-1}}{1-x_0^{T-1}}$ from above we arrive at

$$x^{T-1} = \frac{\frac{x_0^{T-1}}{1-x_0^{T-1}}\exp^{(2bt/\lambda)(T-1)}}{1 + \frac{x_0^{T-1}}{1-x_0^{T-1}}\exp^{(2bt/\lambda)(T-1)}}$$

$$x^{T-1} = \frac{x_0^{T-1}}{(1 - x_0^{T-1})\exp^{(-2bt/\lambda)(T-1)} + x_0^{T-1}}$$

$$x(t) = \frac{x_0}{(x_0^{T-1} + (1 - x_0)^{T-1}\exp^{-2(T-1)bt/\lambda})^{\frac{1}{T-1}}} \ .$$

For $T = 2$ this reduces to

$$x(t) = \frac{x_0}{x_0 + (1 - x_0)\exp^{-2bt/\lambda}}$$

which is equivalent to logistic growth with $r = 2b/\lambda$.

A.2   Additive replicator selection

Performing the same steps for

$$f(u) = u + b$$

we obtain

$$g(u) = \frac{b}{\lambda} \ .$$

Under the same assumption that $\frac{T-1}{T} + b$ never exceeds 1 we find

$$\langle g(\frac{n}{T})\rangle = \frac{b}{\lambda}\sum_{n=1}^{T-1} P(n) = \frac{b}{\lambda}[1 - P(T) - P(0)] = \frac{b}{\lambda}[1 - x^T - (1 - x)^T]$$

which yields the differential equation

$$\frac{d}{dt}x(t) = \frac{2b}{\lambda}(1 - x^T - (1 - x)^T)$$

$$\frac{1}{1 - x^T - (1 - x)^T}dx = \frac{2b}{\lambda}dt \ .$$

For $T = 2$ and $T = 3$ *only* the integral can be solved as

$$\frac{1}{T}(\log(x) - \log(1-x)) = \frac{2bt}{\lambda} + c$$

$$\frac{x}{1-x} = \exp^{cT} \exp^{2btT/\lambda} \ .$$

Substituting

$$C = \exp^{cT} = \frac{x_0}{1-x_0}$$

in

$$x = \frac{C \exp^{2btT/\lambda}}{C \exp^{2btT/\lambda} + 1}$$

we derive

$$x(t) = \frac{x_0}{x_0 + (1-x_0) \exp^{-2btT/\lambda}} \ .$$

For these two sampling rates, the additive replicator selection dynamic is thus equivalent to logistic growth with growth rate $r = \frac{2bT}{\lambda}$.

Appendix B

Markov model code

```r
1  suppressMessages(library(markovchain))
2  library(magrittr)
3
4  newchain <- function(unnormalisedmatrix, name, states=ifelse(sapply(colnames
       (unnormalisedmatrix), length, USE.NAMES=FALSE), colnames(
       unnormalisedmatrix), rownames(unnormalisedmatrix)))
5    new("markovchain", name=name, states=states, transitionMatrix=unname(
       unnormalisedmatrix/rowSums(unnormalisedmatrix)))
6
7  statenames <- function(N)
8    as.character(0:N)
9
10 # ps = binomial production probability in that state
11 binomialsampling.markov.matrix <- function(ps)
12   new("markovchain", name="Binomial sampling", states=statenames(length(ps)
       -1), transitionMatrix=t(sapply(ps, function(p) dbinom(0:(length(ps)-1),
       length(ps)-1, p))))
13 # transition probabilities out of a state are organised in rows:
14 # x[i,j] := P(i -> j) where i, j in [1, length(ps)]
15 # also by definition rowSums(binomialsampling.markov.matrix(...)) == 1
16
17 # assume (deterministic) selection of the mean of the posterior distribution
18 bilm.transition.matrix.average <- function(N, alpha)
19   new("markovchain", name=paste("BILM by averaging with alpha", alpha, sep="
       ="), states=statenames(N), transitionMatrix=t(sapply(0:N, function(x)
       dbinom(0:N, N, (x + alpha/2) / (N + alpha)))))
20
21 # assume (deterministic) maximum a posteriori calculation of theta
22 bilm.transition.matrix.map <- function(N, alpha)
23   # catch abnormal modes: when x=0 (or x=N) then alpha<1 (or beta<1) and the
```

```
24    # mode is simply 0 (or 1)
25    new("markovchain", name=paste("BILM by averaging with alpha", alpha, sep="
        ="), states=statenames(N), transitionMatrix=t(sapply(0:N, function(x)
        dbinom(0:N, N, pmax(0, pmin(1, (x + alpha/2 - 1) / (N + alpha - 2)))))))))

27  # assume sampling from the posterior Reali & Griffiths 2009 (p.321)
28  bilm.transition.matrix.sample <- function(N, alpha) {
29    mx <- sapply(0:N, function(target)choose(N, target)*beta(0:N+target+alpha/
        2, 2*N-0:N-target+alpha/2) / beta(0:N + alpha/2, N - 0:N + alpha/2))
30    new("markovchain", name=paste("BILM by sampling from the posterior with
        alpha", alpha, sep="="), state=statenames(N), transitionMatrix=mx/
        rowSums(mx))
31  }

33  # calculate the Wright-Fisher model mutation rate equivalent to the BILM's N
34  # and alpha. the bracketing in the paper itself is garbled and there's a '/'
35  # missing somewhere, the correct version of the transformation can be found
        on
36  # the top of page 5 of the supplementary material of Reali & Griffiths 2010
37  N.alpha.to.u <- function(N, alpha)
38    alpha / (2 * (alpha + N))

40  # m0 and m1 are mutation probabilities of spontaneously generating variants.
41  # b is a replicator bias (fitness advantage of variant 1 over 0) in (-inf,
        inf)
42  repl.mut.eq <- function(N, alpha0=0, alpha1=alpha0, m0=N.alpha.to.u(N,
        alpha0), m1=N.alpha.to.u(N, alpha1), b=0, ks=0:N) {
43    # apply selection (discrete replicator equation)
44    # Wright-Fisher http://www.stats.ox.ac.uk/~etheridg/orsay/selection.pdf
45    if (b < 0) {
46      biasedks <- (N-ks) * (1 + abs(b))
47      1 - (biasedks*(1-m0) + ks*m1 ) / (biasedks + ks)
48    } else {
49      biasedks <- ks * (1+b)
50      (biasedks*(1-m0) + (N-ks)*m1 ) / (biasedks + N - ks)
51    }
52    # alternatively: Fermi-style selection coefficient: b in (-inf,inf)
53    # (see also http://web.evolbio.mpg.de/~traulsen/paper/05.pdf)
54  #   biasedks <- ks*exp(b)
55    # limiting to [0,1] necessary because exp() causes overflow with negative
        b
56  #   pmax(0, pmin(1, ( biasedks*(1-m0) + (N-ks)*m1 ) / (biasedks + N - ks)))
57  }
58  #repl.mut.eq(5)
59  #repl.mut.eq(5, 0.1)
60  # a higher m0 means more pressure for x -> 0
```

```r
61  #repl.mut.eq(5, 0.1, 0)
62
63  repl.mut.matrix <- function (...)
64    binomialsampling.markov.matrix(repl.mut.eq(...))
65
66  plotstationary <- function(markovchain, st=steadyStates(markovchain),
          absorbingstates=ifelse(missing(markovchain), 1, dim(st)[1]), xlab="x",
          names.arg=0:(dim(st)[2]-1), ...)
67    barplot(st/absorbingstates, xlab=xlab, names.arg=names.arg, space=0, ...)
68
69  averagefrequency <- function(markovchain, st=steadyStates(markovchain)[1,],
          N=length(st)-1)
70    sum(0:N/N * st)
71
72  # advantage of variant 1 over 0, as derived from the stationary distribution
73  advantage <- function(m, nstates=1, maxmomentum=1, nmomentumstates=nstates*
          (1+2*maxmomentum), stationary=steadyStates(m)) {
74    v1freq <- averagefrequency(st=colSums(matrix(stationary, nrow=
          nmomentumstates)))
75    v1freq / (1 - v1freq)
76  }
77  #   sum(stationary[1,(ncol(stationary)-nmomentumstates):ncol(stationary)]) -
          sum(stationary[1,1:nmomentumstates])
78
79
80  # transition matrix fiddling
81  makestickytop <- function(m, N=dim(m)[1]-1)
82    rbind(m[-(N+1),], c(rep(0, N), 1))
83
84  makestickybottom <- function(m, N=dim(m)[1]-1)
85    rbind(c(1, rep(0, N)), m[-1,])
86
87  # compute the development of the Markov chain probability distribution for
          the
88  # specified number of iterations. initstate is a vector of length N and
89  # transitionmatrix an NxN matrix (with rows summing to 1)
90  markov.chain <- function(transitionmatrix, generations=2500, initstate=c(1,
          rep(0, dim(transitionmatrix)-1)), exactduration=FALSE) {
91    if (exactduration) { # use actuation as initstate, see below
92      return(markov.chain.exactduration(transitionmatrix, generations))
93    }
94    out <- matrix(nrow=generations+1, ncol=length(initstate))
95    out[1,] <- initstate
96    for (i in 2:(generations+1))
97      out[i,] <- out[i-1,] * transitionmatrix
98    invisible(out)
```

```r
99  }
100
101 # construct markov chain probability matrix
102 markov.chain.exactduration <- function(transitionmatrix, generations) {
103   # fixate once chain reaches N/N or reverts back to 0/N
104   onetimetransitions <- transitionmatrix[] %>% makestickytop %>%
        makestickybottom %>% newchain
105   # determine generation 2 start state (just after the initial pickup)
106   transitionjustpickedup <- c(0, transitionmatrix[1,-1])
107   data <- markov.chain(onetimetransitions, generations-1, initstate=
        transitionjustpickedup)
108   # clear out all the chains that reverted back to the 0/N state
109   data <- cbind(rep(0, generations), data[,-1])
110   # prepend generation 0 initstate
111   rbind(c(sum(transitionjustpickedup), rep(0, dim(transitionmatrix)-1)),
        data)
112   # rowSums(data) isn't actually equal (and neither is
113   # rowSums(apply.conditioning(data)), meaning the sticktop/bottom approach
        is
114   # actually leeching? would actually *cutting* the state space by two
        states
115   # be a more valid approach?
116 }
117
118 # return per-generation completion probabilities for the given markov chain
119 completionprobabilities <- function(transitionmatrix, exactduration=FALSE,
        ...) {
120   data <- transitionmatrix %>%
121   makestickytop %>% # only count first transitions
122   newchain %>%
123   # exactduration=TRUE causes markov.chain to use actuation as a start state
124   markov.chain(exactduration=exactduration, ...)
125
126   diff(data[, dim(transitionmatrix)[1]])
127 }
128
129 # return the mode and average duration of completions of this markov chain
130 completionstats <- function(transitionmatrix, ...) {
131   data <- completionprobabilities(transitionmatrix, ...)
132   c(mode=which.max(data), mean=weighted.mean(0:(length(data)-1), data))
133 }
134 #completionstats(bilm.transition.matrix.average(20, .05), 100)
135 #completionstats(bilm.transition.matrix.average(20, .05), 100, exactduration
        =TRUE)
136
137 plotcompletionprobabilities <- function(transitionmatrix, ylim=NULL, ...) {
```

```
138    ps <- completionprobabilities(transitionmatrix, ...)
139    tightmargin(pty="s", mfrow=c(1, 2))
140    plot(ps, type="l", xaxs="i", yaxs="i", xlab="generation", ylab="
         probability of first transition completing", main="(i)", ylim=ylim)
141    plot(cumsum(ps), type="l", xaxs="i", yaxs="i", ylim=0:1, xlab="generation"
         , ylab="probability of having exhibited a transition", main="(ii)")
142    invisible(ps) # for postprocessing
143  }
144
145  # numerically compute the probability of a chain succeeding from the given
146  # initial state (specified as an index of the transition/state matrix)
147  successprobability <- function(transitionmatrix, initstate, precision=.99,
         stepsize=500) {
148    transitionmatrix %<>% makestickytop %>% makestickybottom %>% newchain
149    population <- rep(0, dim(transitionmatrix))
150    population[initstate] <- 1
151    while (population[1] + population[length(population)] < precision)
152      population <- markov.chain(transitionmatrix, generations=stepsize,
         initstate=population)[stepsize,]
153    population[dim(transitionmatrix)]
154  }
155
156  # merge the absorbingstates of the transition matrix together into one
         absorbing state that only transitions to itself
157  mergeabsorbingstates <- function(m, absorbingstates) {
158    if (length(absorbingstates) == 1)
159      return(m)
160    # take subset of transition matrix, only leave one state for all states
161    ma <- m[-absorbingstates[-1], -absorbingstates[-1]]
162    # merge (sum) the final absorbingrows columns together for every row
163    ma[,absorbingstates[1]] <- rowSums(m[-absorbingstates[-1], absorbingstates
         ])
164    # make new merged absorbing state only transition to itself
165    ma[absorbingstates[1],] <- 0
166    ma[absorbingstates[1], absorbingstates[1]] <- 1
167    return(ma)
168  }
169
170  plotcompletionprobabilitiesperstart <- function(transitionmatrix,
         maxmomentum=0, nstatestomerge=1 + 2*maxmomentum, N=dim(transitionmatrix)
         /nstatestomerge-1, freqs=0:N, add=FALSE, pch=ifelse(add, 4, 3), ...) { #
          4 x, 3 +, 1 o
171    # squash top/bottom states together into two absorbing states
172    print(N)
173    m <- mergeabsorbingstates(transitionmatrix[], 1:nstatestomerge)
174    transitionmatrix <- newchain(mergeabsorbingstates(m, (nrow(m)-
```

```
            nstatestomerge+1):nrow(m)))
175    if (maxmomentum == 0)
176      startindices <- 1+freqs
177    else
178      # indices of states which have positive momentum
179      startindices <- c(1, 1+freqs[-c(1, length(freqs))]*3, dim(
       transitionmatrix))
180    ps <- sapply(startindices, function(i) successprobability(transitionmatrix
       , i))
181    if (add)
182      points(freqs, ps, pch=pch, ...)
183    else {
184      plot(freqs, ps, pch=pch, xlab="initial frequency", ylab="probability of
       diffusion", ylim=0:1, ...)
185      abline(a=0, b=1/N, lty=3)
186    }
187  }
188
189  #graylevels=round(0.75*length(hmmargs$hmm$States)) / round(0.75*ncol(data))
190  plotchain <- function(data, xlab="generation", ylab="frequency", graylevels
       =24, ...)
191    image(data, x=0:(nrow(data)-1), y=0:(ncol(data)-1), xlab=xlab, ylab=ylab,
       col=gray(graylevels:0/graylevels), breaks=c(0, 1.5^(-graylevels:0))*max(
       data), ...)
192
193
194  # typesetting
195  formatalpha <- function(alpha)
196    bquote(alpha/2 ~ "=" ~ .(alpha/2))
197
198  latextable <- function(m, caption=NULL, floating.environment="table", ...) {
199    rownames(m) <- paste("x", rownames(m), sep="=")
200    colnames(m) <- paste("x'", colnames(m), sep="=")
201    if (!is.null(caption) & length(caption) == 1) {
202      sep <- regexpr("[\\.,:]", caption)[1]
203      if (sep != -1)
204        caption <- c(caption, substr(caption, 1, sep-1))
205    }
206    print(xtable(m, caption=caption, ..., digits=4), floating.environment=
       floating.environment)
207  }
```

R/hmm.R

```
1  library(HMM)
2
```

```r
# transitionmatrix [] has to be organised row-wise
newhmm <- function(transitionmatrix, emissionProbs, startProbs=c(1, rep(0,
    dim(transitionmatrix)-1)))
  initHMM(States=states(transitionmatrix), Symbols=1:ncol(emissionProbs),
    startProbs=startProbs, transProbs=transitionmatrix[], emissionProbs=
    emissionProbs)

chain.mean <- function(markovchain)
  apply(markovchain, 1, function(row) weighted.mean(0:(length(row)-1), row))

# draw a b/w heatmap based on a matrix of positive numbers, with higher
    color resolution closer to 0
# the first argument is a list with elements 'hmm' and 'observation'

plotposterior <- function(hmmargs, addmean=TRUE, addmostlikely=FALSE, ...) {
  data <- t(do.call(posterior, hmmargs))
  plotchain(data, ...)
  if (addmean)
    lines(0:(nrow(data)-1), chain.mean(data), lty=2, col="white")
  if (addmostlikely)
    points(0:(nrow(data)-1), as.numeric(do.call(viterbi, hmmargs)), pch=".",
     col="white") # 4 for cross, 20 for small bullet
}

# all states emit the same symbol - posterior is same as running markov.
    chain()
noconditioning <- function(transitionmatrix, duration)
  list(hmm=newhmm(transitionmatrix, matrix(1, nrow=dim(transitionmatrix))),
    observation=rep(1, 1+duration))
#plotposterior(noconditioning(bilm.transition.matrix.average(50, .5), 100),
    addmostlikely=FALSE)

# all actuated trajectories without interruptions
#actuationconditioning <- function(transitionmatrix, duration)
#  list(hmm=newhmm(transitionmatrix, cbind(c(1, rep(0, dim(transitionmatrix)
    -1)), c(0, rep(1, dim(transitionmatrix)-1)))), observation=c(1, rep(2,
    duration)))

# gotta fiddle a little bit with this one to stop the model from avoiding
    the
# final state until the very last generation: every state has a chance of
# emitting one of two symbols (.99 vs .01), but the lower probability one of
# all the non-categorical states is never actually emitted
naiveconditioning <- function(transitionmatrix, duration)
  list(hmm=newhmm(transitionmatrix, cbind(rep(.99, dim(transitionmatrix)), c
    (rep(0, dim(transitionmatrix)-1), .01), c(rep(.01, dim(transitionmatrix)
```

```
        -1), 0))), observation=c(rep(1, duration), 2))
37 #plotposterior(naiveconditioning(bilm.transition.matrix.average(10, .01),
        30))

38
39 # exact initiation doesn't matter, chain might stay at 0 for some time
40 completionconditioning <- function(transitionmatrix, duration)
41   list(hmm=newhmm(transitionmatrix, cbind(c(rep(1, dim(transitionmatrix)-1),
        0), c(rep(0, dim(transitionmatrix)-1), 1))), observation=c(rep(1,
        duration), 2))
42 #plotposterior(completionconditioning(bilm.transition.matrix.average(30,
        .01), 50))

43
44 # exact initiaton and termination
45 exactconditioning <- function(transitionmatrix, duration)
46   list(hmm=newhmm(transitionmatrix, cbind(c(1, rep(0, dim(transitionmatrix)
        -1)), c(0, rep(1, dim(transitionmatrix)-2), 0), c(rep(0, dim(
        transitionmatrix)-1), 1))), observation=c(1, rep(2, duration-1), 3))
47 #plotposterior(exactconditioning(bilm.transition.matrix.average(50, .01),
        100))
```

R/randomtransitions.R

```
1 # numerical simulation

2
3 # stochastically generate chains that start off in state 0/n and are in
4 # state n/n after chainlength generations
5 generate.transitioning.chains <- function(transitionmatrix, chainlength,
        numchains=1, initstate=0) {
6   N <- dim(transitionmatrix)-1
7   p <- markov.chain(transitionmatrix, chainlength)[chainlength+1,N+1]
8   message("probability of being in state N/N after exactly ", chainlength, "
         generations is ", p)
9   message("this means a chain will be found roughly every ", round(1/p), "
        attempts")
10   sapply(1:numchains, function(chain) {
11     attempts <- 0
12     while (TRUE) {
13       attempts <- attempts+1
14       pop <- vector("numeric", chainlength+1)
15       pop[1] <- initstate
16       for (i in 2:(chainlength+1))
17         pop[i] <- sample(0:N, 1, prob=transitionmatrix[1+pop[i-1],])
18       if (pop[chainlength+1] == N) {
19         message("found transition ", chain, " after ", attempts, " attempts"
      )
20         return(pop)
```

```r
21        }
22      }
23    })
24  }
25
26  # stochastically generate chains that start off in state 0/n and that first
27  # reach state n/n after EXACTLY chainlength generations
28  generate.transitioning.chains.exact <- function(transitionmatrix,
        chainlength, numchains=1, initstate=NULL) {
29    N <- dim(transitionmatrix)-1
30    p <- markov.chain.exactduration(transitionmatrix, chainlength)
31    p <- p[chainlength+1,N+1] - p[chainlength,N+1]
32    message("probability of first arriving in state N/N after exactly ",
      chainlength, " generations is at most ", p)
33    message("this means a chain will be found roughly every ", round(1/p), "
      attempts")
34    sapply(1:numchains, function(chain) {
35      attempts <- 0
36      while (TRUE) {
37        attempts <- attempts+1
38        pop <- vector("numeric", chainlength+1)
39        pop[1] <- ifelse(is.null(initstate), sample(1:N, 1, prob=
      transitionmatrix[1,-1]), initstate)
40        for (i in 2:(chainlength+1)) {
41          if (pop[i-1] == N || pop[i-1] == 0)
42            break
43          pop[i] <- sample(0:N, 1, prob=transitionmatrix[1+pop[i-1],])
44        }
45        if (pop[chainlength+1] == N) {
46          message("found transition ", chain, " after ", attempts, " attempts"
      )
47          return(pop)
48        }
49      }
50    })
51  }
52
53  plotchains <- function(data, ...) {
54    # plot circles for start+end conditioning
55    plot(c(0, nrow(data)-1), range(data), xlab="generation", ylab="frequency",
        ...)
56  #   paste("frequency of variant 1 (out of ", max(data), ")", sep="")
57    for (i in 1:ncol(data))
58      lines(0:(nrow(data)-1), data[,i])
59  }
```

Appendix C

Questionnaire materials

This appendix contains the various materials used for the questionnaire data collection and analysis: Section C.1 provides the four questionnaire sheets described in Chapter 5 that were handed to participants on site. Only one randomisation is given. The online version of the questionnaire is still available at `http://spellout.net/ibexexps/kstadler/shetland/experiment.html`, and the source code for the generation of the randomised questionnaire sheets is provided in Section C.2. Section C.3 provides the R code used to load and arrange the questionnaire and acceptability judgment data that form the basis of the statistical analyses in Chapter 5.

C.1   Paper questionnaires

Momentum exit-questionnaires

A questionnaire for testing people's awareness of (the directionality of) changes, with questions tapping into their explicit knowledge of:

1. their own frequency of use
2. their interlocutors' frequency of use
3. the 'age' of a variant (a naive way to get at their impression of the real-time nature of a change)
4. the apparent time development of a change (asking about younger/older speakers' frequency of use)

Now with 4 pages per individual: questions about a changing variable (imperatives, p.1), a stable one (negation, p.2), and two more changing and almost completed ones (yes/no questions and wh questions, p.3+4).

The order of presentation of the two variants, their order in the 'which is older' question, and the order of the younger/older speakers question are all randomised, so there's 8 different versions of the same 4-page questionnaire.

Open questions/things to try out/change away from too much linguistics lingo:

1. replace 'use this variant' with 'say'?
2. do people understand 'negating a sentence'?
3. change from "people around you" to "people in Shetland" or "people you normally talk to"
4. change from "Which variant do you think is *older*" to "Which variant do you think has been around for longer?"

You are probably familiar with these two ways of asking somebody to do something:

"*Mak du dy ain denner!*"   "*Du mak dy ain denner!*"

How much do you use either of these variants?

☐   ☐   ☐   ☐   ☐
I use only   I use more   I use both   I use more   I use only
'Mak du..'   'Mak du..'   equally   'Du mak..'   'Du mak..'

How much do you think are people around you using either of the variants?

☐   ☐   ☐   ☐   ☐
People use only People use more People use both People use more People use only
'Mak du..'   'Mak du..'   equally   'Du mak..'   'Du mak..'

Which of the two variants do you think is *older*?

☐   ☐   ☐
'Mak du..' is   'Du mak..' is   People have
older   older   always used
                                  both

How much do you think *younger speakers* use either of the variants?

☐   ☐   ☐   ☐   ☐
younger   younger   younger   younger   younger
speakers use   speakers use   speakers use   speakers use   speakers use
only 'Mak du..' more 'Mak du..' both equally   more 'Du mak..' only 'Du mak..'

How much do you think *older speakers* use either of the variants?

☐   ☐   ☐   ☐   ☐
older speakers   older speakers   older speakers   older speakers   older speakers
use only   use more   use both equally use more   use only
'Mak du..'   'Mak du..'   'Du mak..'   'Du mak..'

1

You are probably familiar with these two ways of negating a sentence:

"*He didna go*"   "*He didnoo go*"

How much do you use either of these variants?

☐	☐	☐	☐	☐
I use only 'didna'	I use more 'didna'	I use both equally	I use more 'didnoo'	I use only 'didnoo'

How much do you think are people around you using either of the variants?

☐	☐	☐	☐	☐
People use only 'didna'	People use more 'didna'	People use both equally	People use more 'didnoo'	People use only 'didnoo'

Which of the two variants do you think is *older*?

☐	☐	☐
'didna' is older	'didnoo' is older	People have always used both

How much do you think *younger speakers* use either of the variants?

☐	☐	☐	☐	☐
younger speakers use only 'didna'	younger speakers use more 'didna'	younger speakers use both equally	younger speakers use more 'didnoo'	younger speakers use only 'didnoo'

How much do you think *older speakers* use either of the variants?

☐	☐	☐	☐	☐
older speakers use only 'didna'	older speakers use more 'didna'	older speakers use both equally	older speakers use more 'didnoo'	older speakers use only 'didnoo'

You are probably familiar with these two ways of asking somebody a question:

"*Kens du Sarah?*"   "*Does du ken Sarah?*"

How much do you use either of these variants?

☐	☐	☐	☐	☐
I use only 'Kens du..?'	I use more 'Kens du..?'	I use both equally	I use more 'Does du ken..?'	I use only 'Does du ken..?'

How much do you think are people around you using either of the variants?

☐	☐	☐	☐	☐
People use only 'Kens du..?'	People use more 'Kens du..?'	People use both equally	People use more 'Does du ken..?'	People use only 'Does du ken..?'

Which of the two variants do you think is *older*?

☐	☐	☐
'Kens du..?' is older	'Does du ken..?' is older	People have always used both

How much do you think *younger speakers* use either of the variants?

☐	☐	☐	☐	☐
younger speakers use only 'Kens du..?'	younger speakers use more 'Kens du..?'	younger speakers use both equally	younger speakers use more 'Does du ken..?'	younger speakers use only 'Does du ken..?'

How much do you think *older speakers* use either of the variants?

☐	☐	☐	☐	☐
older speakers use only 'Kens du..?'	older speakers use more 'Kens du..?'	older speakers use both equally	older speakers use more 'Does du ken..?'	older speakers use only 'Does du ken..?'

3

You are probably familiar with these two ways of asking somebody a question:

"*Whit gae du him?*"   "*Whit did du gie him?*"

How much do you use either of these variants?

☐	☐	☐	☐	☐
I use only 'Whit gae du..?'	I use more 'Whit gae du..?'	I use both equally	I use more 'Whit did du gie..?'	I use only 'Whit did du gie..?'

How much do you think are people around you using either of the variants?

☐	☐	☐	☐	☐
People use only 'Whit gae du..?'	People use more 'Whit gae du..?'	People use both equally	People use more 'Whit did du gie..?'	People use only 'Whit did du gie..?'

Which of the two variants do you think is *older*?

☐	☐	☐
'Whit gae du..?' is older	'Whit did du gie..?' is older	People have always used both

How much do you think *younger speakers* use either of the variants?

☐	☐	☐	☐	☐
younger speakers use only 'Whit gae du..?'	younger speakers use more 'Whit gae du..?'	younger speakers use both equally	younger speakers use more 'Whit did du gie..?'	younger speakers use only 'Whit did du gie..?'

How much do you think *older speakers* use either of the variants?

☐	☐	☐	☐	☐
older speakers use only 'Whit gae du..?'	older speakers use more 'Whit gae du..?'	older speakers use both equally	older speakers use more 'Whit did du gie..?'	older speakers use only 'Whit did du gie..?'

C.2 Questionnaire source code

questionnaire/data/questionnaire.Rmd

```
1  ---
2  title: Momentum exit-questionnaires
3  output: pdf_document
4  classoption: a4paper,12pt
5  header-includes:
6    - \usepackage{fullpage}
7    - \renewcommand{\familydefault}{\sfdefault}
8    - \usepackage{array}
9    - \newcolumntype{x}{>{\centering\arraybackslash}p{2.75cm}}
10 ---
11
12 \newenvironment{likert}
13  {\parskip=0cm\par\nopagebreak\centering\vspace{0.4cm}}
14  {\vspace{1.45cm}\par\noindent\ignorespacesafterend}
15
16 \thispagestyle{empty}
17
18 A questionnaire for testing people's awareness of (the directionality of)
      changes, with questions tapping into their explicit knowledge of:
19
20 1. their own frequency of use
21 2. their interlocutors' frequency of use
22 3. the `age' of a variant (a naive way to get at their impression of the
      real-time nature of a change)
23 4. the apparent time development of a change (asking about younger/older
      speakers' frequency of use)
24
25 Now with 4 pages per individual: questions about a changing variable (
      imperatives, p.1), a stable one (negation, p.2), and two more changing
      and almost completed ones (yes/no questions and wh questions, p.3+4).
26
27 The order of presentation of the two variants, their order in the 'which is
      older' question, and the order of the younger/older speakers question
      are all randomised, so there's 8 different versions of the same 4-page
      questionnaire.
28
29 Open questions/things to try out/change away from too much linguistics lingo
      :
30
31 1. replace '*use* this variant' with '*say*'?
32 2. do people understand 'negating a sentence'?
```

```
33  3. change from "people around you" to "people in Shetland" or "people you
       normally talk to"
34  4. change from "Which variant do you think is *older*" to "Which variant do
       you think has been around for longer?"
35
36  <!-- \newpage \begin{center} This page semi-intentionally left blank \end{
       center} -->
37
38  ```{r echo=FALSE, results="asis"}
39
40  likert <- function(labels, question="") {
41    cat(question, "\\begin{likert}\\begin{tabular}{", rep("x", length(labels))
       , "}",
42    paste0(rep("$\\square$", length(labels)), collapse=" & "), "\\\\",
43    paste0(labels, collapse=" & "), "\\end{tabular}\\end{likert}")
44  }
45
46  usage <- function(prefix, left, right) {
47    c(paste(prefix, c("only", "more"), left), paste(prefix, "both equally"),
       paste(prefix, c("more", "only"), right))
48  }
49
50  questionnaire <- function(left, right, variabledescription) {
51    cat("\\newpage You are probably familiar with these two ways of",
       variabledescription)
52    cat("\\begin{likert}", left$presentation, "\\hspace{3cm}",
       right$presentation, "\\end{likert}")
53
54    # How much of the time are you using the variants? If you are strictly
       using one of the variants circle in that variant.
55    likert(usage("I use", left$use, right$use), "How much do you use either of
        these variants?")
56
57    likert(usage("People use", left$use, right$use), "How much do you think
       are people around you using either of the variants?")
58
59    # Which of these variants do you think is the older one/has been used for
       longer? (i.e. tapping into either real time or apparent time knowledge)
       # (I.e. which one has been used for a longer period of time?) #"They
       have been around for the same time"
60    likert(c(paste(c(left$realtime, right$realtime), "is older"), "People have
        always used both"), "Which of the two variants do you think is *older
       *?")
61    # people used to say more "..." / used "..." more often / which one was
       used before
62
```

```
63    # people use more A these days <> the mix of the two variants isn't
        changing <> people use more B these days
64
65    # apparent age (Do you think the first/second variant is used more often
        by younger/older speakers? (i.e. explicitly asking for their knowledge
        about apparent age))
66    likert(usage(paste(left$age, "use"), left$use, right$use), paste("How much
         do you think\\emph{", left$age, "}use either of the variants?"))
67
68    likert(usage(paste(right$age, "use"), left$use, right$use), paste("How
        much do you think\\emph{", right$age, "}use either of the variants?"))
69  }
70
71  variable <- function(long, short) {
72    long <- paste("``\\emph{", long, "}''", sep="")
73    short <- paste("`", short, "'", sep="")
74    r <- expand.grid(use=short, realtime=short, age=c("younger speakers", "
        older speakers"), stringsAsFactors=FALSE)
75    r$presentation <- rep_len(long, nrow(r))
76    return(r)
77  }
78
79  randomisation <- function(v, i, description="saying the same thing") {
80    questionnaire(v[i,], v[1+nrow(v)-i,], paste(description, ":", sep=""))
81  }
82
83  negation <- variable(c("He didna go", "He didnoo go"), c("didna", "didnoo"))
84
85  # Tak du a peerie sweetie!
86  imperative <- variable(c("Mak du dy ain denner!", "Du mak dy ain denner!"),
        c("Mak~du..", "Du~mak.."))
87
88  # Does du hear yun?
89  yesnoquestion <- variable(c("Kens du Sarah?", "Does du ken Sarah?"), c("Kens
        ~du..?", "Does~du~ken..?"))
90
91  # Whaar cam he fae?
92  whquestion <- variable(c("Whit gae du him?", "Whit did du gie him?"), c("
        Whit~gae~du..?", "Whit~did~du~gie..?"))
93
94  for (i in 1:nrow(imperative)) {
95    randomisation(imperative, i, "asking somebody to do something")
96    cat("\\setcounter{page}{1}")
97    randomisation(negation, i, "negating a sentence") # saying something didn'
        t happen
98    # extended questionnaire
```

```
99    randomisation(yesnoquestion, i, "asking somebody a question")
100   randomisation(whquestion, i, "asking somebody a question")
101 }
102 ```
```

C.3   Questionnaire data processing code

<div align="center">questionnaire/data/shetland–data.R</div>

```
1  vars <- c("imp", "ynq", "whq", "neg")
2  varcolors <- c("orange", "blue", "green", "darkgrey")
3  names(varcolors) <- vars
4  changingcolors <- varcolors[1:3]
5  longvars <- c(imp="imperatives", ynq="yes/no questions", whq="wh questions",
        neg="negation")
6  vardesc <- c(imp="imperatives", ynq="yes/no questions", whq="wh questions",
      neg="negation (stable control)")
7
8  lvls <- c("onlyout", "moreout", "both", "morein", "onlyin")
9  collapsedlvls <- c("fewin", "morein", "onlyin")
10 collapsedmapping <- c("fewin", "fewin", "fewin", "morein", "onlyin")
11 rellvls <- c("behind", "level", "ahead")
12 adddiff <- function(d, v1, v2="other", name=v1) {
13   diffname <- paste(name, "diff", sep="")
14   diffs <- as.numeric(d[[v1]]) - as.numeric(d[[v2]])
15   d[[paste(name, "relative", sep="")]] <- factor(rellvls[2+sign(diffs)],
      levels=rellvls, ordered=TRUE)
16 #  d[[diffname]] <- factor(diffs, levels=do.call(seq, as.list(range(diffs)))
      , ordered=TRUE)
17   d[[diffname]] <- factor(diffs, levels=-4:4, ordered=TRUE)
18   levels(d[[diffname]])[6:9] <- sprintf("%+i", 1:4)
19   return(d)
20 }
21
22 nicelvls <- c("only out", "more out", "both", "more in", "only in")
23 helmertvars <- c("whq", "question", "notchanging")
24 helmertlevels <- function(var) contr.helmert(4)[5 - as.numeric(var),]
25 arrangedata <- function(d) {
26   d$gender <- factor(c(F="female", M="male")[d$gender], levels=c("female", "
      male"))
27
28   d$self <- factor(d$self, levels=lvls, ordered=TRUE)
29   d$jitteredself <- jitter(as.numeric(d$self))
30   d$other <- factor(d$other, levels=lvls, ordered=TRUE)
31   d$var <- factor(d$var, levels=vars)
32   d$longvar <- factor(longvars[d$var], levels=longvars)
```

```
33    # helmert coding (not reverse, as returned by contr.helmert())
34    d[,helmertvars] <- helmertlevels(d$var)
35    d$oldervar <- factor(d$oldervar, levels=c("out", "same", "in")) # TODO
        consider this ordered or not?
36    d$young <- factor(d$young, levels=lvls, ordered=TRUE)
37    d$old <- factor(d$old, levels=lvls, ordered=TRUE)
38    levels(d$self) <- nicelvls
39    levels(d$other) <- nicelvls
40    levels(d$young) <- nicelvls
41    levels(d$old) <- nicelvls
42
43    # collapse first three levels to avoid empty cells in ordered logit model
44  # flattenedlevels <- c("lessin", "lessin", "lessin", "morein", "onlyin")
45  # d$selftrunc <- factor(flattenedlevels[d$self], levels=unique(
        flattenedlevels), ordered=TRUE)
46  # d$othertrunc <- factor(flattenedlevels[d$other], levels=unique(
        flattenedlevels), ordered=TRUE)
47
48    # dummy variable for between-changing/stable tests
49    d$stable <- d$var == "neg"
50
51    # swap "in" and "out" for the stable "neg" variable so that "in"/"out"
52    # responses can be meaningfully interpreted as 'majority'/'minority'
        variant
53    d[d$var=="neg","oldervar"] <- levels(d$oldervar)[4-as.numeric(d[d$var=="
        neg","oldervar"])]
54    for (col in c("firstvar", "self", "other", "firstolder", "young", "old"))
        {
55      lvls <- levels(d[[col]])
56      d[d$var=="neg",col] <- lvls[1+length(lvls)-as.numeric(d[d$var=="neg",col
        ])]
57    }
58
59    # identical measures (under different labels)
60    #d$selffreq <- factor(d$self, levels=freqlvls)
61    d$oldervar <- factor(c("outgoing older", "always both", "incoming older")[
        as.numeric(d$oldervar)], levels=c("outgoing older", "always both", "
        incoming older"))
62    d$oldervarfreq <- factor(c("minority older", "always both", "majority
        older")[as.numeric(d$oldervar)], levels=c("minority older", "always both
        ", "majority older"))
63
64    # derivative measures
65    d <- adddiff(d, "self")
66    d <- adddiff(d, "old")
67    d <- adddiff(d, "young")
```

```r
68    d <- adddiff(d, "young", "old", "apparent")
69    d$apparentdiffN <- as.numeric(d$apparentdiff)-5
70  #   d$apparentdiff <- as.numeric(d$young)-as.numeric(d$old)
71  #   d$apparentrelative <- factor(rellvls[2+sign(d$apparentdiff)], levels=
        rellvls, ordered=TRUE)
72  #   d$apparentdiff <- factor(d$apparentdiff, levels=difflvls(d$apparentdiff),
         ordered=TRUE)
73    return(d)
74  }
75
76  da <- arrangedata(read.table("shetland.csv", header=TRUE, sep=" "))
77  da$condition <- as.factor("paper")
78  do <- arrangedata(read.csv("shetland-online.csv"))
79  do$id <- factor(do$id)
80  do$condition <- as.factor("online")
81  d <- rbind(do,da)
82
83  #d$loc <- factor(d$loc, levels=c("Central", "Lerwick", "South", "West", "
       Bressay", "North", "Whalsay"))
84  d$loc <- relevel(relevel(d$loc, "Bressay"), "Lerwick")
85
86  #agecats <- c("young", "middle", "old")
87  agecats <- c("age <= 32", "age > 32")
88  # split participants in 3 evenly sized categories: <=27 (N=26), 28-49 (N=26)
       , >=50 (N=25)
89  #d$agecat <- factor(agecats[2 - (d$age <= 27) + (d$age>=50)], levels=agecats
       , ordered=TRUE)
90  d$agecat <- factor(agecats[1+(d$age>32)], levels=agecats)
91  participants <- unique(d[c("id", "gender", "age", "agecat", "loc", "
       condition")])
92  #agecatmeans <- sapply(agecats, function(a) mean(subset(participants, agecat
        == a)$age))
93  #agecatmeans <- aggregate(participants$age, by=list(participants$agecat),
       FUN=mean)$x
94
95  variable <- function(v, data=d) # "imp", "neg", "whq", "ynq"
96    subset(data, var == v)
97
98  changing <- subset(d, var != "neg")
99  changing$var <- factor(changing$var)
100 neg <- subset(d, var == "neg")
101 imp <- subset(d, var == "imp")
102 whq <- subset(d, var == "whq")
103 ynq <- subset(d, var == "ynq")
104
105 # given the two distribution of responses to younger/older, what is the
```

```r
106  # baseline distribution of differences between them that could have occurred
          by
107  # chance?
108
109  diffchancelevel <- function(p1, p2, diff) {
110    nlevels <- length(p1)
111    sum(sapply((1 + max(0, -diff)) : (nlevels - max(0, diff)), function(first)
        p1[[first]] * p2[[first+diff]]))
112  }
113
114  diffchancelevels <- function(p1, p2=p1) {
115    nlevels <- length(p1)
116    sapply((-nlevels+1):(nlevels-1), function(diff) diffchancelevel(p1, p2,
        diff))
117  }
118  # plot baseline distribution of derived responses assuming uniform responses
119  #barplot(diffchancelevels(rep(0.2, 5)), col=temp.colors(9))
120
121  #diffchancelevels(rep(0.2, 5), rep(0.2, 5))
122  #diffchancelevels(c(0,0,1,0,0), rep(0.2, 5))
123
124  # grammaticality judgments
125
126  ids <- as.character(unique(da$id))
127  suppressMessages(library(foreach))
128  jdg <- read.csv("judgments.csv", sep="\t")
129  jdg <- foreach (i = 1:nrow(jdg), .combine=rbind) %do%
130    data.frame(id=ids, var=as.character(jdg$var[i]), verb=as.character(jdg$
        verb[i]), incoming=jdg$variant[i]=="incoming", judgment=as.numeric(jdg[i
        ,paste("X", ids, sep="")]), stringsAsFactors=FALSE)
131  # make var a factor (with identical ordering as the other dataset)
132  jdg$var <- factor(jdg$var, levels(d$var))
133
134  # filter single NA
135  jdg <- jdg[complete.cases(jdg),]
136
137  # we have to do some sort of aggregation first because not every verb was
138  # presented in both the new and old variant context. here we just take the
        mean
139  # across all verbs contexts, but could filter down to only those presented
        with
140  # both variants etc.
141  meanjudgments <- aggregate(jdg$judgment, by=list(id=jdg$id, var=jdg$var,
        incoming=jdg$incoming), FUN=mean)
142  # now put incoming+outgoing judgments on same row
143  meanjudgments <- aggregate(meanjudgments$x, by=list(id=meanjudgments$id, var
```

```
        =meanjudgments$var), FUN=c)
144 # positive values -> incoming variant rated more highly
145 meanjudgments$absdiff <- meanjudgments$x[,2] - meanjudgments$x[,1]
146 meanjudgments$reldiff <- meanjudgments$x[,2] / meanjudgments$x[,1]
147 # merge grammaticality judgments and usage level estimates by speaker "id"
        and "var": intersect(names(meanjudgments), names(da))
148
149 # approach #2: only average ratings given for verbs rated in both
150 # incoming+outgoing context
151 matchedjudgments <- merge(subset(jdg, incoming)[-4], subset(jdg, !incoming)
        [-4], by=c("id", "var", "verb"))
152 colnames(matchedjudgments)[4:5] <- c("incoming", "outgoing")
153
154 meanmatchedjudgments <- aggregate(matchedjudgments[c("incoming", "outgoing")
        ], by=list(id=matchedjudgments$id, var=matchedjudgments$var), FUN=mean)
155 meanmatchedjudgments$absdiff <- meanmatchedjudgments$incoming -
        meanmatchedjudgments$outgoing
156 meanmatchedjudgments$reldiff <- meanmatchedjudgments$incoming /
        meanmatchedjudgments$outgoing
157
158 matchedjudgments$absdiff <- matchedjudgments$incoming - matchedjudgments$
        outgoing
159 matchedjudgments$reldiff <- matchedjudgments$incoming / matchedjudgments$
        outgoing
160 #aggregate(jdg$judgment, by=list(id=jdg$id, var=jdg$var, verb=jdg$verb), FUN
        =c)$x
161
162 plotjudgmentcor <- function(judgments, c1, c2, render=c1, xlab=paste("
        estimated usage level (", c1, ")", sep=""), ylab=paste("relative
        acceptability (incoming", if (c2=="absdiff") "-" else "/", "outgoing)"),
         main=NULL, ...) {
163   d <- merge(judgments, da)
164   d$var <- factor(d$var)
165   # R's core cor.test can't do p values for ties
166 #  print(cor(as.numeric(d[[c1]]), d[[c2]], method="kendall"))
167 #  kendall <- cor.test(as.numeric(d[[c1]]), d[[c2]], method="kendall")
168   # rpudplus can (rpucor.test()), but requires a license for the test
169 #  print(rpud::rpucor(cbind(as.numeric(d[[c1]]), d[[c2]]), method="kendall",
         use="pairwise"))
170   # calculate 95% confidence interval of tau (tau slightly different here)
171 #  print(DescTools::KendallTauB(as.numeric(d[[c1]]), d[[c2]], 0.95))
172   # apparently the best package (with p values accounting for ties) is
        pvrank:
173   kendall <- pvrank::rankor(as.numeric(d[[c1]]), d[[c2]], "kendall", print=
        FALSE, type="greater")
174   plot(as.numeric(d[[render]]), d[[c2]], xlim=c(0.8, 5.2), xlab=xlab, ylab=
```

```
       ylab, main=if (is.null(main)) bquote(tau[B] == .(round(kendall$Value, 3)
         ) ~ "(p" == .(paste(round(kendall$Cpv, 3), ")", sep=""))) else main, col
         =changingcolors[d$var], xaxt="n", pch=4, pty="s", ...)
175    axis(1, 1:5, levels(d[[c1]]), las=3)
176    legend("topleft", levels(d$var), fill=changingcolors)
177    # plot relative judgment ('equal acceptability') baseline
178    abline(h=if(c2=="absdiff") 0 else 1, lty=2)
179  }
180
181  # usageestimate could be "self", "other" or "selfdiff"
182  # for the 'judgments' argument, see below
183  plotjudgments <- function(judgments=meanmatchedjudgments, usageestimate="
         jitteredself", ...) {
184    par(mfrow=c(1,2))
185    plotjudgmentcor(judgments, usageestimate, "absdiff", ...)
186    plotjudgmentcor(judgments, usageestimate, "reldiff", ...)
187  }
188  # this is the plot averaged over all lexical items
189  #plotjudgments(meanjudgments)
190  # averaged only over matched lexical items (slightly higher)
191  #plotjudgments(meanmatchedjudgments)
192  #plotjudgments(meanmatchedjudgments, "other")
193  # this plot over all individual items (strong lexical effects)
194  #plotjudgments(matchedjudgments)
195
196  # intensity in (0,1]
197  temp.colors <- function(mn, mx=NULL, intensity=1) {
198    if (is.null(mx)) {
199      mx <- floor(mn/2)
200      mn <- ceiling(-mn/2)
201    }
202    hsv(c(rep(0.65, abs(mn)), FALSE, rep(0, abs(mx))), intensity*abs(mn:mx)/
         max(abs(c(mn,mx))))
203  }
```