Modeling language change:

An evaluation of Trudgill's theory of the emergence of New Zealand English

Gareth Baxter, *School of Mathematics, Statistics and Computer Science, Victoria University of Wellington, New Zealand*

Richard Blythe, *School of Physics, University of Edinburgh*

William Croft, *Department of Linguistics, University of New Mexico*

Alan McKane, *School of Physics and Astronomy, University of Manchester*

ABSTRACT

Trudgill (2004) proposes that the emergence of New Zealand English, and of isolated new dialects generally, is purely deterministic: it can be explained solely in terms of the frequency of occurrence of particular variants and the frequency of interactions between different speakers in the society. Trudgill's theory is closely related to usage-based models of language, in which frequency plays a role in the representation of linguistic knowledge and in language change. Trudgill's theory also corresponds to a neutral evolution model of language change. We use a mathematical model based on Croft's usage-based evolutionary framework for language change (Baxter et al. 2006), and investigate whether Trudgill's theory is a plausible model of the emergence of new dialects. The results of our modeling indicate that determinism cannot be a sufficient mechanism for the emergence of a new dialect. Our approach illustrates the utility of mathematical modeling of theories and of empirical data for the study of language change.

1. Introduction

In this paper, we use mathematical modeling to evaluate Trudgill's theory of new-dialect formation, as applied to the emergence of New Zealand English (Trudgill 2004; see also Gordon et al. 2004).* Our goals are twofold. The first is to investigate certain properties of the propagation of language change. Trudgill's theory of new-dialect formation has an interesting and theoretically controversial property, which Trudgill calls **determinism**: there is no role for social factors such as prestige or identity in the emergence of isolated new dialects such as New Zealand English. Instead, new-dialect formation is purely frequency-based, in terms of exposure to tokens of language use by the speakers with whom one interacts. Trudgill's theory is therefore related to usage-based theories of linguistic knowledge and language use (e.g., Bybee 2001; Pierrehumbert 2003). Trudgill's deterministic theory corresponds to a neutral evolution model in evolutionary theory (see §§5-6). Trudgill's theory contrasts with the classical sociohistorical linguistic theory, namely that propagation of linguistic variants is governed by associated social values of those variants in terms of factors like prestige or identity. The classical sociohistorical linguistic theory corresponds to a selection model of change by replication. Thus our evaluation of Trudgill's theory is a test of whether a neutral evolution model is a plausible model of the propagation of a language change in a speech community.

Our second goal is to bring mathematical models to bear on the explanation of empirical observations of language change, and thereby enrich sociohistorical linguistic theory. Implementation of theories, particularly quantitative theories such as those found in sociohistorical linguistics, in mathematical models is essential for evaluating their plausibility as well as their utility.

We begin by outlining the basic facts surrounding the emergence of New Zealand English (§2), and Trudgill's theory of new-dialect formation as he applies it to this situation (§3). We discuss some of the potential empirical difficulties in applying Trudgill's theory to New Zealand English (many of which are also raised in Kerswill 2007; see §4). However, the focus of our paper is on whether Trudgill's theory can account for what Trudgill predicts, even if there is doubt in some cases as to whether Trudgill's predictions actually hold for New Zealand English.

* Previous versions of this paper were presented at the First Scottish-Dutch Workshop on Language Evolution: Formal Modeling meets Empirical Data, University of Amsterdam, The Netherlands; the Departmental Colloquium Series, Department of Linguistics, University of New Mexico, USA; the 3rd Workshop on Approaches to Culture, Cognitive and Communication, Edinburgh, UK; the Workshop on Language Evolution: Computer Models for Empirical Data, Noordwijk, the Netherlands; the International Summer Atelier: Modeling Language Evolution with Computational Construction Grammar, Ettore Majorana Foundation and Center For Scientific Culture, Erice, Italy; and the 82nd Annual Meeting of the Linguistic Society of America, Chicago, USA. We thank members of the audiences for their comments; all responsibility for the final product remains with us. This research was partly supported by funding from the Royal Society of Edinburgh (Blythe) and the Engineering and Physical Sciences Research Council (UK) (grant GR/T11784; McKane); Blythe is an RCUK (Research Councils UK) Fellow.

We then turn to the modeling of Trudgill's theory. Our model is based on the usage-based theory of language change proposed by Croft (2000; see §§5-7), who is also cited by Trudgill in support of his deterministic theory (Trudgill 2004:150-51). This model is essentially an evolutionary one, which is natural for both a theory of innovation in language use and the sociohistorical linguistic theory of propagation of innovations in a speech community. We explore a variety of neutral evolution models which represent the essential characteristics of Trudgill's deterministic theory of new-dialect formation, and describe their properties with respect to theories of language change (§§8-9). We conclude, among other things, that none of the neutral evolution models we explore can account for the New Zealand English data under plausible assumptions, and that new-dialect formation, like other sociolinguistic situations of language change, likely requires a selection mechanism.

2. The creation of New Zealand English: demography and language

New Zealand English is the most recent variety of colonial English to emerge. Although the first European settlers arrived in the 1790s, it was not until the Treaty of Waitangi was signed in 1840 that substantial numbers of Europeans began to arrive in New Zealand (Gordon et al. 2004:38-39; Trudgill 2004:24). The European population increased dramatically from 1850 to around 1890, as indicated in the figures in Table 1:

Table 1. European population of New Zealand in the 19th century

Date	*Population*	*Source*
1831	300-330	Owens 1992:50
1839	2,000	Graham 1997:52
1841	5,000	Graham 1992:112
1851	27,000	Graham 1992:117
1858	59,000	Graham 1997:52
1861	99,000	Graham 1992:117
1864	172,000	Gordon et al. 2004:54
1871	256,000	Graham 1992:117
1881	490,000	Belich 1996:278
1886	579,000	Graham 1992:112
1891	627,000	Graham 1992:117
1896	750,000	Rice 1992:601, Graph 8
1901	850,000	Rice 1992:601, Graph 8
1906	1,000,000	Rice 1992:601, Graph 8

Between 1831 and 1881, 400,000 persons migrated to New Zealand; of these, 300,000 stayed in New Zealand. During the same period, 250,000 persons were born in New Zealand (Belich 1996:278). More precise figures for immigration between 1858 and 1890 are given by Graham (1992:125), and are compared in Table 2 with the population figures from Table 1 (the figures in italics at the right represent the approximate number of native-born New Zealanders from the preceding date; however, it understates the native-born because it does not include deaths since the preceding date):

Table 2. A comparison of immigration to total population in New Zealand, 1858-1890

	Net Immigration	Total Population		Total Population Increase Less Net Immigration
		59,000	1858	
1858-1860	21,000			
		99,000	1861	*19,000*
1861-1865	93,000			
		172,000	1864	*-20,000*
1866-1870	21,000			
		256,000	1871	*63,000*
1871-1875	82,000			
1876-1880	55,000			
		490,000	1881*	*97,000*
1881-1885	29,000			
		579,000	1886†	*60,000*
1886-1890	-9,000			
		627,000	1891	*350,000*

*250,000 native-born inhabitants in 1881 (Belich 1996:278)
†300,000 native-born inhabitants in 1886 (Graham 1992:112)

Table 2 indicates three phases of demographic change in 19th century New Zealand. Phase I, from the 1840s to around 1864, was a period when almost all the European population increase in New Zealand was by immigration. Phase II, from 1864 to around 1886, was the period when the first substantial New Zealand native generation was born; significant immigration continued through Phase II. Phase III, from 1886 until after 1900, was a period when immigration, or at least net immigration, largely ceased. Immigration did not pick up significantly again until after 1900 (http://www.nzhistory.net.nz/culture/immigration/home-away-from-home/summary); but by then, net new immigration was only a small percentage of the total population (McKinnon 1997:49). Native-born New Zealanders constituted around 48% of the total population in 1881, 52% in 1886, and 60% in 1901 (Hamer 1997:144).

The vast majority of immigrants at this time came from the British Isles (Gordon et al. 2004:44), although many of these came via Australia. The British immigrants largely came from southern England (Gordon et al. 2004:139; Trudgill 2004:16), Scotland and Ireland; the approximate proportion was 50% English, 27% Scottish and 23% Irish (Trudgill 2004:13, 16, citing McKinnon 1997). There were attempts to exclude or restrict Irish Catholic immigrants, but they were largely unsuccessful (Gordon et al. 2004:42, 43). Thus, a variety of British English dialects, though not northern English dialects, contributed to the formation of New Zealand English.

A distinctive, relatively stable New Zealand English dialect emerged with the generation of New Zealanders born around 1890 (Trudgill 2004:24-25, 113). This is approximately the second generation of native-born New Zealanders. Remarkably, there are recordings available of the first generation of native New Zealand speakers, that is, those born from the 1850s to the 1880s. In the late 1940s, the New Zealand National

Broadcasting Service sent a Mobile Disc Recording Unit to record music and oral histories from towns and rural settlements (Gordon et al. 2004:1-4). These speakers represent a generation before the emergence of the New Zealand English variety. The Origins of New Zealand English (ONZE) project used these recordings to analyze the origins of this colonial variety. The major report of this study is the monograph by Gordon et al. (2004); Trudgill was a part of the project, and published his theory in a separate monograph (2004). We now turn to Trudgill's theory.

3. Trudgill's deterministic theory

Trudgill's deterministic theory is intended to apply only to a very specific type of language change: new-dialect formation as a result of dialect mixture in a community isolated from other speakers of the same language (what Trudgill calls a 'tabula rasa' situation; Trudgill 2004:26). Thus, for example, it does not cover the new town koinés of towns like Milton Keynes in England (Kerswill 1996; Kerswill and Williams 2000, 2005), which is surrounded by older English-speaking communities. New Zealand English appears to satisfy these criteria. New Zealand was overwhelmingly populated by speakers of English varieties, and the influence of the indigenous language, Maori, was minimal (Gordon et al. 2004:69-70, 219; Trudgill 2004:4-5). However, New Zealand was populated by speakers of different English varieties, leading to dialect mixture. Finally, New Zealand is geographically isolated from other English-speaking communities, both the United Kingdom and Australia, and the creation of a new speech community eventually led to the creation of a new English dialect, New Zealand English. Trudgill argues that the process of new-dialect formation, at least in this type of situation, takes two generations, or approximately fifty years (Trudgill 2004:53).

Trudgill's theory involves several different mechanisms, but the central hypothesis is that, within certain bounds, the outcome of the new-dialect formation process can be predicted using a small number of principles, if one knows the linguistic variants of the input dialects and the proportions of immigrants speaking those dialects. The basic proportions of immigrants speaking the broad dialects of southern England, Scotland and Ireland were given in §2, and a large part of Trudgill's monograph is devoted to identifying the linguistic variants of the input dialects in mid to late nineteenth century Britain (Trudgill 2004:31-82). The most important principle of Trudgill's theory is that the most frequent variant, based on demography, survives to form part of the new dialect.

Trudgill divides the process of new-dialect formation into three stages, corresponding roughly to three discrete generations of speakers. Stage I represents the immigrant generation: native speakers of different British dialects who arrived in New Zealand after the middle of the nineteenth century. This generation basically speaks the dialect of their place of origin. However, in the dialect contact situation of early colonial New Zealand, Trudgill argues that variants that are in a very small minority or are geographically very restricted will be eliminated even at this early stage, giving plausible examples from a number of colonial dialects (his 'rudimentary leveling'; Trudgill 2004:89-93). Trudgill also suggests that interdialect forms may arise in Stage I (ibid., 94-99). Stage I corresponds basically to Phase I of the demographic history of New Zealand described in §2.

Stage II represents the first native-born generation; this is the generation recorded by the Mobile Disc Recording Unit. The availability of these recordings allow us to observe the features of a transitional generation that has been lost to linguistic analysis in other new-dialect formation situations (Trudgill 2004:100). Trudgill argues that Stage II is characterized in the ONZE project data by extreme variability, both in comparison to the source dialects in Britain and in comparison between speakers of the first native generation. Trudgill argues that children in this generation do not have a societal norm among their peers because of the absence of such norms in the new-dialect situation. Hence 'in diffuse dialect-contact situations the role of adults will be more significant than is usually the case' (ibid., 101). Since the adults came from a variety of dialect backgrounds, the children's language will contain a mixture of variants originally from different dialect areas in Britain. Trudgill gives the example of an ONZE speaker, Mrs. German, whose parents were from Suffolk, but lived in a community with a large number of Scottish immigrants: Mrs. German's speech included a mixture of East Anglian and Scottish features. The variation displayed by ONZE project speakers is of three types:

(i) the speakers 'select variants from different dialects at will' (Trudgill 2004:103), as in the case of Mrs. German, producing combinations of variants not found in Britain. This process does not seem to represent choice of a majority variant: for example, Mrs. German uses some East Anglian variants and some Scottish variants. Trudgill characterizes this phenomenon as highly individual, and adds that 'most of these combinations have had little permanent effect on the shape of modern New Zealand English' (ibid., 104).

(ii) the speakers use multiple variants more variably than in more stable speech communities (ibid., 105-6).

(iii) the speakers growing up in a single location 'may differ very markedly from one another', having selected different variants that are available in the same community (ibid., 106-8).

Thus there is little uniformity in the variants of the ONZE speakers taken as a whole, certainly in comparison to the next generation (see below). Nevertheless, Trudgill does argue that minority variants in the input from the Stage I speakers (the immigrants) will be eliminated at Stage II. He suggests a threshold of approximately 10% presence for a variant to survive in Stage II speech (Trudgill 2004:110-11). Stage II corresponds basically to Phase II in the demographic history of New Zealand in §2. However, Trudgill's Stage II does not include continued immigration, as actually took place in Phase II of the demographic history of New Zealand (compare Trudgill 2004:163, and see §4).

Stage III represents the second native-born generation; this is the first generation of speakers of an identifiable New Zealand English variety. Stage III corresponds basically to Phase III in the demographic history of New Zealand in §2, although in Phase III there remain a large proportion of immigrants (somewhat under half; see Table 2). This stage is characterized by focusing (LePage and Tabouret-Keller 1985). Trudgill describes focusing as 'the process by means of which the new variety acquires norms and stability'

(Trudgill 2004:88). The focusing process follows on a leveling of the variants from Stage II, such that the number of variant forms are reduced, usually to a single variant, 'as a result of group accommodation in face-to-face interaction' (ibid., 113-14).

Trudgill argues that the mechanism by which variants are chosen to be the norms of the focused variety is frequency-based: the majority form is the one that survives. Trudgill argues that again, it is the children—that is, the members of the second native-born generation—who do the choosing, from the input provided by speakers from the Stage II generation. In particular, the Stage III speakers do not choose the variant that is characterized by some social factor, such as an origin in the southeast of England, the prestige variant back in Britain. Trudgill argues that a number of features of Southern Hemisphere Englishes, including New Zealand English, are not of southeastern English origin, and they survived precisely because they outnumbered the southeastern variant in the demographic makeup of the immigrants. The southeastern variants that survived were those which were also present in other dialects of other immigrants. They survived because these variants were the majority variants, not because of any prestige attached to the southeastern English dialect.

Trudgill's model is summarized in Table 3:

Table 3. Stages in Trudgill's theory of new-dialect formation

Stage	*Generation*	*Variety*	*Process*
I	immigrants	essentially same as in region of origin	elimination of very low frequency and geographically restricted variants
II	first native-born generation	highly variable mixture of variants from different regions of origin; much intra-, inter-individual variation	elimination of low frequency variants (ca. 10% frequency?)
III	second native-born generation	focusing, leading to choice of usually one variant	propagation of majority variant

4. Some issues in applying Trudgill's theory to New Zealand English

Trudgill himself observes that two variants investigated in the ONZE project do not follow the deterministic principle that the majority variant in the demographic input becomes the New Zealand English variant. In New Zealand English, there was a merger of the weak vowels /ə/ and /ɪ/ to /ə/ (schwa), even though only 32% of the ONZE project speakers use schwa. Trudgill attributes this to the 'unmarked' status of schwa combined with the fact that it represents a fairly large minority (Trudgill 2004:119-20). However, 'markedness' is a questionable notion. It is not clear to us under what conditions the 'unmarked' variant wins out and under what conditions the majority variant wins out regardless of markedness.

The second problematic case is the fronting and lowering of the vowel /ʌ/ in the STRUT class of words (Trudgill 2004:136). Among the ONZE speakers born between 1850 and 1869, 34% have fronting and lowering of the STRUT vowel, while among the ONZE speakers born between 1870 and 1889, 40% have fronting and lowering of the STRUT vowel. Yet fronting and lowering of the STRUT vowel is characteristic of New Zealand English. Trudgill explains this in terms of the linguistic notion of drift, that is, 'inherited propensities for change' (Gordon et al. 2004:241). In particular, Trudgill points out that parallel changes took place across Southern Hemisphere Englishes that are all the result of dialect mixture from British English. The linguistic notion of drift, however, is very problematic. In the case of the fronting/lowering of the STRUT vowel, drift means that change is propagated in a parallel fashion in different speech communities descended from the same ancestral community. But what is the mechanism that drives the parallel propagation process in different speech communities? To us, drift is a non-explanation since there is no mechanism for propagation.

Nevertheless, against these two problematic cases, Trudgill provides seven positive cases where the deterministic theory makes the correct prediction. Gordon et al. further note that if the proportion of a variant is calculated in terms of tokens of the variants, rather than the proportion of speakers using the variant, then an eighth case, loss of rhoticity, also fits the deterministic theory: although most ONZE project speakers produced rhotic variants, only 9% of the tokens produced were rhotic (Gordon et al. 2004:240-41).

Another objection that has been raised is that there is evidence of social factors of the more traditional sociohistorical linguistic type in early colonial New Zealand. In particular, gender plays its usual sociolinguistic role, familiar from many other studies: women are more advanced than men in the shift towards the emerging New Zealand English conventions (Gordon et al. 2004:211-12, 276). Trudgill suggests that this could be due to female children interacting more frequently with older women than older men (Trudgill 2004:150). It could alternatively be due to the survival of traditional European gender roles in colonial New Zealand, and that those gender differences played the same sociolinguistic role in propagation of a change in colonial New Zealand as they have done in other Western societies.

Nevertheless, one could argue that this are not failings of Trudgill's theory but simply indicate that colonial New Zealand was not the perfect 'tabula rasa' situation that Trudgill's theory requires. Perhaps if New Zealand really were a tabula rasa, then it would fit Trudgill's theory better. This is the view taken by Kerswill (2007:661). This is the view that we will be examining in the remainder of this paper.

Finally, Trudgill's own theory contains a mysterious process: focusing. Focusing is essential in Trudgill's theory for the transition from the extreme variability of Stage II to the relative uniformity of Stage III. Exactly how focusing works for Trudgill is not entirely clear to us. The simplest interpretation of Trudgill's view of focusing is that it is a process of accommodation that requires some two generations to take place. Trudgill notes the high degree of mobility of early New Zealanders and argues that 'this was a society with relatively weak social network ties—precisely the sorts of ties that are the breeding ground for rapid supralocal linguistic change' (Trudgill 2004:161-62). Trudgill describes focusing as following a leveling process that reduced the Stage II variants to the one that became the New Zealand English variant (ibid., 113). This suggests that Trudgill

believes that the fixation of the New Zealand English variants was a consequence of the leveling process, and focusing merely solidified the outcome of leveling. If so, then the leveling process simply needs two generations of pliable language learners to lead to fixation of the new dialect variants.

What is clear, however, is that the process does not involve any social valuation of variants, even for a new identity as a member of the New Zealand English community. Trudgill argues against the Stage II generation (ONZE project speakers) signaling a mixed identity through their mixed language, and argues that Stage III did not focus on the emerging New Zealand English speakers via prestige factors based on England English (Trudgill 2004:157). Trudgill also writes that 'It would be ludicrous to suggest that New Zealand English speakers deliberately developed, say, closer front vowels in order to symbolize some kind of local or national New Zealand identity' (Trudgill 2004:157). Hence, in evaluating Trudgill's theory we investigate only a model of language change in which no social valuation of either speakers or variants exists.

5. A usage-based evolutionary framework for language change

Our aim in this paper is to attempt to model Trudgill's theory as summarized in Table 3, using only mechanisms corresponding to those proposed by Trudgill, and without appealing to any mechanism that would correspond to the introduction of social factors such as prestige, stigma or acts of identity. This goal is part of a general project of modeling language change following Croft's (2000, 2006, to appear) usage-based, evolutionary framework for language change. Croft's framework and its relationship to sociohistorical linguistic theory is described in this section; the following sections describe those aspects of the framework that we are modeling and their relevance to the emergence of New Zealand English.

Croft's framework treats language change as taking place through language use. Language change is a two-step process, the generation of variation (innovations), and the propagation of a variant through the speech community. Sociohistorical linguistics has focused on the latter step, while a variety of theories have been proposed for the former step. Croft's framework essentially incorporates the sociohistorical approach to the propagation of change, and develops a usage-based model of innovation.

The fundamental process of linguistic behavior is the **replication** of tokens of linguistic structure—sounds, words and constructions—in language use. Each time a speaker produces an utterance, she replicates linguistic structures that she has heard before, although the structures are often combined in novel ways. This replication process is of course mediated by the speaker and her knowledge about her language, which is based in turn on the language use she has been exposed to in face-to-face interaction. This knowledge takes the form of a usage-based mental representation: that is, the mental representation of linguistic structures includes a representation of their frequency in the input (see Bybee 2001; Pierrehumbert 2001, 2003). A usage-based model of linguistic knowledge can represent sociolinguistic variants and their frequencies as well.

When a speaker replicates a token of a linguistic structure, called a **lingueme** by Croft, the replication may be altered from prior replications—this would be an **innovation**. Thus, a lingueme may have two or more variants: in other words, a lingueme corresponds to a sociolinguistic variable. The rate of replication of one lingueme variant

over another may be influenced by a number of factors, including the rate of exposure to particular speakers using different lingueme variants (i.e. social network effects), or a social value attached to the lingueme variants (e.g.. classical social variables such as gender or socioeconomic class associated with the sociolinguistic variable).

Models of change by replication, as opposed to the inherent change of an object, are evolutionary models. Croft's framework is an instantiation of David Hull's General Analysis of Selection (GAS), which Hull applies to biological evolution and to conceptual change in science (Hull 1988, 2001). The central element in GAS is the **replicator**, a term which was coined by Dawkins (1976) in a similar approach. The replicator is the entity that is replicated in some process, preserving most of its structure (Hull 1988:408). In biological evolution, the canonical replicator is the gene, and the process is meiosis, which takes place in reproduction (sexual or asexual). In language change, the replicator is the lingueme, and the replication process is language use in face-to-face interaction. Replication must preserve much of the replicator's structure. For example, a speaker more or less accurately replicates the phonetic, grammatical and semantic structure of the sounds, words and constructions when she produces an utterance.

Hull argues that for selection to operate, a second entity (or set of entities) is required, the **interactor**. The interactor is an entity that interacts with its environment in such a way that it causes the differential replication of replicators (Hull 1988:408). In biological evolution, the canonical interactor is the organism. It interacts with its environment—the ecosystem and its fellow organisms—in such a way that it causes differential replication of its replicators. That is, the organism survives and reproduces, or doesn't survive or reproduce, and as a consequence its genes are differentially replicated—they are propagated through the population, or they ultimately go extinct. This process is selection. In language change, the speaker is the canonical interactor. The speaker interacts with her environment, in particular other speakers, and by virtue of that interaction causes differential replication of linguemes (variants). That is, in the context of social interaction the speaker will replicate some linguistic variants and not others, and thereby cause the differential replication of variants. The result is language change.

The selection process presented in the preceding paragraph is described in very abstract terms, because it is intended to apply to change by replication in any domain: biological evolution, conceptual change in science, language change, and so on. The General Analysis of Selection is inspired by research in biological evolution, but it abstracts away from those aspects of biological evolution that are irrelevant to change by replication, and hence do not carry over into other domains. In particular, the mechanism which causes variation to be generated, and the mechanism by which selection operates, are specific to each domain (see Croft 2006:96).

6. Replicator selection, neutral evolution and interactor selection

Since our concern is with propagation, we will not discuss mechanisms by which variation is generated (for mechanisms of innovation in sound change, see Ohala 1989; Bybee 2001; in grammatical change, see Croft 2000, to appear). As with most sociolinguistic research, we assume the existence of multiple variants, and we analyze how they are propagated (or not).

In the classical evolutionary selection model, variant replicators differ in fitness, a value associated with each replicator; and differences in fitness result in differential replication. We will call this classical selection model **replicator selection**: there is a value directly associated with the replicator that leads to its differential replication.

In language change, replicator selection models the classical sociohistorical linguistic model. Different variants have different social values associated with them (prestige, stigma, identity, etc.—not necessarily consciously, of course). By virtue of the social values associated with the variable, some variants are propagated at the expense of other variants. This is the theory that Trudgill rejects for new-dialect formation in an isolated speech community.

However, change by replication can happen without (replicator) selection operating. Since there is a degree of randomness in the replication process, there will be random fluctuations in replicator frequencies as reproduction takes place in a finite population. If the fluctuation happens to hit zero, then the replicator goes extinct. Thus, change takes place in the population simply by virtue of random processes; no selection has taken place. This process is called **genetic drift** (Crow and Kimura 1970); genetic drift is very different from linguistic drift (§4). The process is also called **neutral evolution**, and we will use this term in order to avoid confusion.

In language change, a neutral evolution model corresponds to Trudgill's deterministic model. A significant property of neutral evolution models is that the probability of fixation—fully successful propagation of a variant—is a function of the frequency of the variant. Neutral evolution is a probabilistic model, which is fundamentally different from a deterministic model that predicts that the majority variant will always win out. However, since Trudgill's theory is invoking the same usage-based processes as we are, namely that speakers alter their behavior in response to the language they hear around them, and those usage-based processes are probabilistic, then it is not implausible to consider Trudgill's theory in probabilistic terms, as we do below.

Neutral evolution models include processes that lead to differential replication. This can best be understood by considering the closest biological parallel, which is a group of islands which each contain a set of individuals of different species or alternatively individuals who have a particular gene which has different variants. Here speakers are equivalent to islands and utterances to migrations of individuals between islands (Blythe and McKane 2007). Even though no species or allele is favored over another, if one island is geographically closer to another, for instance, then the chances of interaction between these islands is greater. More generally, the nature of the network of islands and the strengths of the interactions between the islands would appear to have an influence on the evolutionary process. In the General Analysis of Selection, as defined in §5, this process is a type of selection: interaction of an interactor with its environment causes replication to be differential. What is unusual is that the environment in this case is another interactor. We will call this process **interactor selection**, to differentiate it from classical replicator selection.

In language change, structured social networks mean that a speaker is more likely to interact with certain speakers rather than others. In this sense, social distance replaces the geographical distance used in the biological example. Since the interaction results in the replication of some linguistic variants (replicators) over others in language use, it can bring about differential replication. We will call this type of interactor selection **neutral**

interactor selection, in that the only factor that influences replication is the frequency of interaction with the interactor.

Trudgill allows for neutral interactor selection in his deterministic theory. In §4, we noted that Trudgill proposes that women are more advanced in the use of New Zealand English variants because female children interact more with older women than older men. This is neutral interactor selection: women speak more to women than to men. Neutral interactor selection is a possible mechanism for another effect observed in the ONZE project speakers: the use of variants is partly correlated with the ethnicity of the speaker's parents (Gordon et al. 2004:263; see also Trudgill's example of Mrs. German in §3). Presumably this is because children speak more to their parents (at least at first) than to other members of the community.

There is another type of interactor selection that is possible in language change. In this type of interactor selection, interactors (interlocutors) are preferred or dispreferred by a speaker no matter how frequently or infrequently she interacts with them, and their linguistic replications (utterances) are weighted accordingly. Thus, variants of a speaker whose productions are weighted more heavily will be differentially replicated. This type of interactor selection is unlike network structure, where the weighting of variants is simply a consequence of frequency of interaction with different speakers producing different variants. We will call this **weighted interactor selection**, since interactors are weighting the productions (replicator replications) of particular interactors differently, no matter what the frequency of interaction with them is. In the biological parallel drawn above migration from one island to another might be favored over another because the soil conditions or the climate on that island were more conducive to the establishment or growth of a migrant.

Weighted interactor selection implies that a speaker's linguistic behavior is influenced not just by frequency of interaction but also a differential social valuation of particular speakers, possibly because of the social group to which they belong. It is not clear whether Trudgill would consider weighted interactor selection to conform to his model. However, Trudgill describes his theory as deterministic, and makes assertions such as 'the minority quite simply and mechanistically accommodated to the majority' (Trudgill 2004:148) and 'The new colonial dialects are a *statistical composite* of the dialect mixture' (ibid., 123, emphasis original). These statements imply that Trudgill would not allow weighted interactor selection in his deterministic theory.

Therefore, in order to test Trudgill's deterministic theory of new dialect formation, we will investigate models of neutral evolution and of neutral interactor selection only. Our aim is to push models of neutral evolution and neutral interactor selection as far as possible in order to see if they can model the emergence of a new dialect without having to resort to replicator selection, or even weighted interactor selection.

7. The description of the model

The model we have constructed was first described in Baxter et al (2006, section III). Here we summarize it in a slightly more qualitative way, and refer the reader to Baxter et al (2006) for a more mathematical treatment.

Trudgill's theory does not imply any interaction between sociolinguistic variables (linguemes), and we similarly assume that they are independent. This allows us to focus

on a single lingueme, which we assume can exist in several variants denoted by α, β, γ, … . The speech community consists of N speakers, labeled by an integer i which runs from 1 up to N. Each speaker will have a grammar – knowledge of their language – which includes her perceived frequency of the variants of the linguemes/sociolinguistic variables, in conformity with the principles of the usage-based model of linguistic representation (see §5). For example, the frequency of H-retention – the variant with the phoneme /h/ in particular contexts – in the English of early New Zealand was 75%, and the frequency of H-deletion is therefore 25%. If this fact represents what speakers are exposed to, then their knowledge of New Zealand English at the time includes this frequency information.

Since we are focusing on one lingueme, the grammar at a particular time, t, can be considered to be the set of perceived frequencies of the variants of that one lingueme at this time. To formulate this mathematically we assume, to make the description of the model easier, that there are only two variants, α and β. Then the grammar of speaker i at time t can be specified by $x_i(t)$, where x is the perceived fraction of time the α variant is used, and $(1 - x)$ the perceived fraction of time that the β variant is used. For example, the grammar of speaker 3 could at some time be such that $x_3 = 0.9$. This would mean that, since $1 - x_3 = 0.1$, she would perceive that the α variant is nine times more likely to be used than the β variant. The state of the language at time t is the set of the grammars of the individual speakers: $\{x_1(t), x_2(t)\ldots, x_N(t)\}$.

The aim of model is to see how the fractions x_i change with time due to the various interactions between the speakers. The first thing that has to be done is to set the initial values of these fractions. These could be taken to be random numbers between 0 and 1, or they could have a more specific form. For example, all speakers could have $x = 0.5$ initially, so that the α and β variants are perceived to be equally likely by all speakers. After choosing this initial condition, we allow the grammar of the population of speakers to evolve with time. We do this by repeating the following three steps in sequence:

1. Social interaction. A pair i,j of speakers is chosen to interact (hold a conversation). Some speakers could in principle be favored over others, simply because they are very well connected. Conversely, isolated speakers might have less chance of being picked. To model this we give a probability for each pair of speakers to be picked and make the choice on the basis of this. This probability is denoted by G_{ij}. G_{ij} represents simply the likelihood of occurrence of a linguistic interaction between two speakers; it is symmetric because if speaker i interacts with speaker j, then speaker j necessarily interacts with speaker i. G_{ij} reflects the geographical and social structure of the network of speakers; it models neutral interactor selection (§6). This is illustrated in Figure 1.

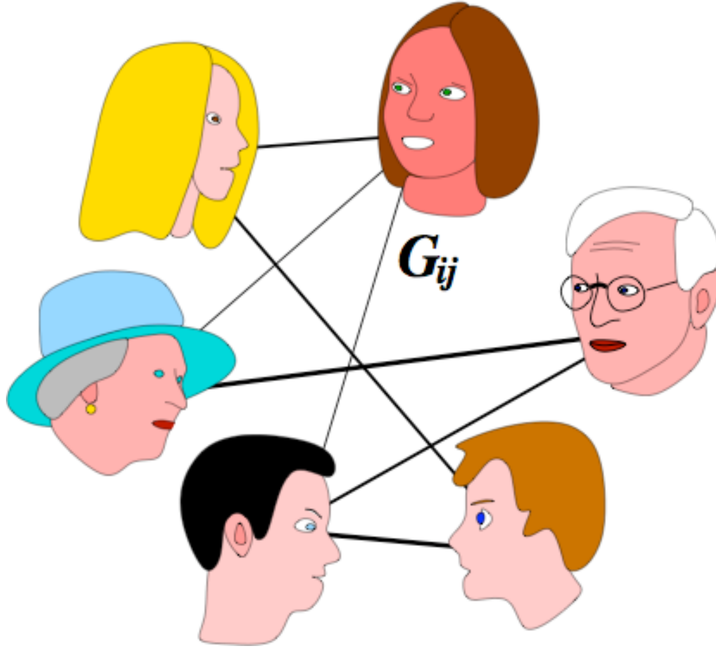

Figure 1: Speakers in the society interact with different frequencies (shown here schematically by different thicknesses of lines connecting them). The pair of speakers *i,j* is chosen to interact with probability G_{ij}.

So, in practice, a pair *i,j* of speakers are selected with a prescribed probability G_{ij}.

2. Reproduction. Each of the speakers selected in step 1 are now allowed to interact by first one of the individuals producing 10 or 20 (in general, *T*) *tokens* of the lingueme, and then the other doing the same. For example if *T* equals 10, 8 of these might be in the α form and 2 in the β form. To reflect the stochastic nature of the communicative process seen in real conversations (that is, that all language behavior is variable and probabilistic), the number of α or β variants will reflect the state of the grammar of the speaker, but will be random in the same way that a coin toss or the roll of a die is random. To take a simple example, suppose for speaker *i* that $x_i = 5/6$, then we can roll a die to see whether an α or β variant should be uttered; if a 1 was rolled the β variant would be uttered, on the other hand if any other number was rolled, the α variant would be uttered. If $T = 12$ for instance, then it would be most likely that 2 β variants would be uttered, and 10 α variants. But this would only be the most likely scenario. It could be that 9 or 11 αs were uttered, or even 12, with no βs being uttered at all. It is the stochastic nature of these events that ultimately leads to particular variants becoming extinct from the population. When only one variant is left we adopt the language of population genetics and say that it has become **fixed**. In general we can imagine tossing a (biased) coin – with a probability x_i of coming up heads – *T* times to see how many α variants are produced by speaker *i* in a particular utterance.

After speaker *i* has produced an utterance, speaker *j* produces a sequence of tokens according to the same prescription, but using their grammar. This is illustrated in Figure 2.

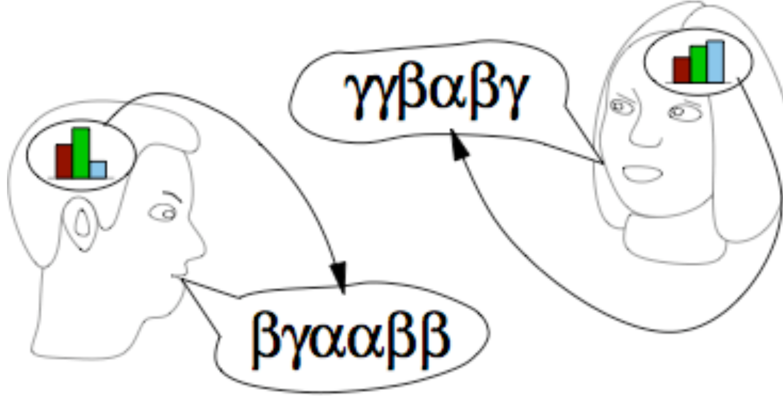

Figure 2: Both speakers *i* and *j* produce an utterance, with particular lingueme variants appearing with a frequency given by the value stored in the utterer's grammar. In this particular case three variants are shown (α, β and γ) and the number of tokens, *T*, is equal to 6.

We now need to describe the effect of these utterances on the listener – and also on the speaker who made the utterance.

3. Retention. The final step is to modify each speaker's grammar to reflect the actual language used in the course of the interaction. The simplest approach is to add to the existing speaker's grammar additional contributions which reflect both the tokens produced by herself and by her interlocutor. The weight given to these tokens, relative to the existing grammar, is given by a parameter λ. This parameter can be thought of as the receptiveness of the speaker to changing their grammar on the basis of the language they hear. Presumably, there will be a large amount of inertia in the retention of a grammar, and therefore λ will be small: the utterance that has just been produced will only have a small effect on the existing grammar. It is the cumulative effect of utterances over many interactions that will lead to a significant change in the grammars of the speakers.

The weighting of existing tokens relative to new tokens represented by λ brings about a reduction in the weight of existing tokens which continues every time new tokens are added to a speaker's memory. This reduction of weighting of older tokens is equivalent to their decay in memory. The rate of decay is controlled by λ, and if λ is small, as just described, the decay function is close to the exponential form widely used in psychological models of memory (see Anderson and Schooler 1991:396). After a token's memory has decayed to a value ε, the token is defined to be forgotten. The forgetting of tokens is what allows propagation/extinction to take place: after all tokens of one variant are forgotten, that variant is no longer produced and has gone extinct.

Conversely, retention in memory of a particular variant is dependent on continuously hearing new tokens of the variant. Our model of entrenchment in memory is based on three general results. The first is that exposure to tokens distributed over time rather than massed at once facilitates entrenchment or consolidation of the memory (see Rovee-Collier 1995:148 and references cited therein). We assume a uniform spacing of tokens over time, so that there is a time t_{mem} between exposures to tokens of the lingueme variants.

The second result is that there is a limit to the length of t_{mem} for consolidation of memory. That limit is called the **time window** for laying down a memory (Rovee-Collier 1995). If t_{mem} is longer than the time window, then the new tokens do not help to

consolidate the memory trace of the original tokens, and the memory is not entrenched (i.e. the older memory traces decay; Rovee-Collier 1995). The existence of the time window, combined with the memory decay function and the forgetting threshold, allow us to model the extinction of a lingueme variant and therefore the possibility of fixation of the variant that replaces it.

The third result is a linguistic generalization: linguistic forms with a higher token frequency are more likely to be retained while linguistic forms with a lower token frequency are more likely to be replaced with novel variants (see for example Bybee 1985 on the replacement of low token frequency forms in a morphological paradigm, Bybee and Slobin 1982 and Lieberman et al. 2007 on the replacement of irregular forms in low token frequency words, and Phillips 1984 on the replacement of the 'underlying forms' of low token frequency variants). This linguistic generalization follows from the first two results about memory: a lower frequency variant has a longer t_{mem}, and if t_{mem} is too long (exceeds the time window), new tokens will not contribute to the entrenchment of the variant and the memory of the variant based on prior tokens is therefore more likely to decay.

The weight, relative to her own utterances, that speaker i gives to speaker j's utterances is specified by a number H_{ij}. This is intended to reflect social status and other social effects, and models weighted interactor selection (§6). H_{ij} is asymmetric: speaker i may not give the same weight to speaker j's utterance that speaker j gives to speaker i's utterance. For the purpose of testing Trudgill's theory, we begin by taking H_{ij} to be the same for all pairs i and j are denoted by the constant H, to see if the data can be explained without having to resort to social effects of this kind.

To specify how exactly x_i changes during this time step, let us denote x_i at time t by $x_i(t)$, as before, and x_i at time $t + 1$ by $x_i(t + 1)$. Also suppose our random number generator (the die or coin described above), produces n α variants for speaker i and m α variants for speaker j. Then $x_i(t)$ will be modified by two factors: (i) we will have to add a factor of $\lambda n/T$ to $x_i(t)$, since n/T is the fraction of αs given in speaker i's utterance and λ is the weight we are giving to them, (ii) we will also have to add a factor of $\lambda H_{ij} m/T$ to $x_i(t)$, since m/T is the fraction of αs given in speaker j's utterance, λ is the weight we are giving to them and H_{ij} is the relative weight that speaker i gives to speaker j's utterances. In terms of equations

$$x_i(t+1) \propto \left(x_i(t) + \lambda \frac{n}{T} + \lambda H_{ij} \frac{m}{T} \right).$$

There will be a similar rule for speaker j, but with i and j interchanged, and the numbers n and m interchanged. Figure 3 illustrates this step.

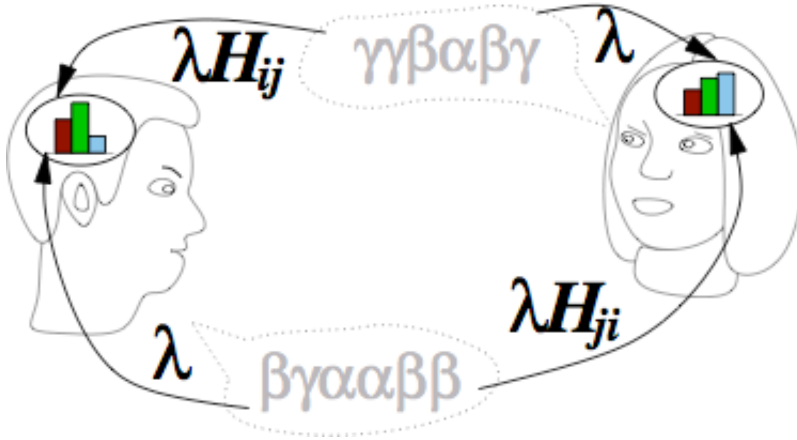

Figure 3: After the utterances have been produced, both speakers modify their grammars by adding to them the frequencies with which the variants were reproduced in the conversation. Note each speaker retains both her own utterances as well as those of her interlocutor, albeit with different weights.

We have used the symbol ∝ (proportional to) in this equation, rather than an equality. To see why this is, we imagine that we start from the fraction of the β variant in the grammar and carry out the same exercise. This would increase also, due to the addition of terms. So the fraction of αs and βs at time $t + 1$ when added together would be greater than 1. To remedy this, we have to divide the right-hand side of the equation by a factor of $1 + \lambda + \lambda H_{ij}$, which guarantees that the sum of the fraction of α and β variants add up to 1 at all times. Dividing by this factor means that we can replace the proportionality symbol by an equality.

Having carried out these three steps, we now pick another pair of speakers at random, and continue in this way for a large number of time steps – until the x values have been modified significantly. In the absence of other factors, the final state will be when either the α or the β variant dies out completely, that is, all the x_i are either 0 or 1. One of the questions of interest, for example, is: how long on average is it before this fixation happens? This may be investigated both mathematically and using computer simulations. To carry out the latter, a particular network structure has to be assumed. In the next section, we will see that it can be shown mathematically that in many cases the time to fixation is independent of the network structure. However, here we briefly consider two examples of speaker networks which we might expect to be applicable to the early colonial period in New Zealand.

The first is the "equal-groups" network, which might model several isolated homesteads in occasional contact with each other. The simplest case would consist of l groups of M speakers, so that $N=Ml$. Speakers within a particular group would all be interconnected and interact with the same strength, but there would only be weak interaction between the groups, perhaps mediated by a single speaker from each group. This is illustrated in Figure 4a. As townships formed, this network structure might be replaced by a "hub-and-spoke" network, which is as in the equal-group case, but with one group (the hub) being bigger than the others (not indicated in the figure below), and with the other $(l - 1)$ groups interacting with each other via the hub. This is shown in Figure 4b.

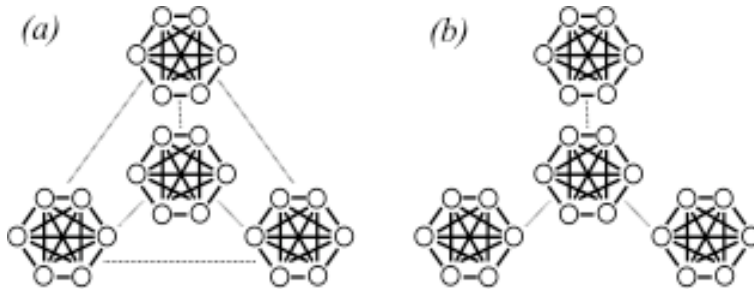

Figure 4. Illustration of two network structures. (a) A community divided into a number of equal groups within which interactions are frequent but between which are less so: the dotted lines are intended to indicate interactions between one or more pairs of speakers from different groups. (b) A 'hub-and-spoke' community in which there is one central (hub) group and several peripheral (spoke) groups. No interactions take place between different spoke groups, but all spokes interact with the hub.

Clearly many other network topologies may be postulated. Fortunately, as already mentioned, we will not need to specify the topology to find the mean time to fixation (see also the Mathematical Appendix for discussion of different network structures). Both this result and more details of our investigations of the model, are discussed in the next section.

8. Propagation by neutral interactor selection

In this section our aim is to explore whether the mathematical model described above reproduces behavior actually observed in the emergence of New Zealand English with empirically plausible parameter settings. Specifically, we wish to determine whether the main features of the empirical data – namely, that variants initially in the majority were in almost all cases adopted by all New Zealand English speakers in the space of about fifty years (Trudgill 2004:23) – can be explained by neutral interactor selection alone, as Trudgill has proposed.

We focus initially on an idealization in which the community of speakers has a fixed size over the relevant historical period. This is, of course, a simplification of the true situation, in which old speakers die and new speakers are born or continue to arrive through immigration. However, this simple case is still useful to us in evaluating Trudgill's deterministic theory because we can obtain quite general mathematical results and then see (e.g., by means of computer simulation) how these are affected as progressively more realistic features are introduced into the model. These extensions will be described in the next section.

As described in the previous section, the model is stochastic: every time it is run, a different outcome might be realized. In order to make comparison with the empirical data, we must therefore determine how likely a particular outcome is to be obtained within the model, or how such an outcome is typically reached.

In particular, we noted above that if initially a lingueme has multiple variants, eventually all but one of them will go extinct in the model. The probability that a specific variant, say α, fixes (i.e., is the only one remaining after a large number of interactions) can be calculated exactly. We focus specifically on the case of neutral interactor selection, in which all speakers ascribe the same weight to others' utterances: mathematically, this means setting all H_{ij} parameters to the same value. Then, the model

predicts that variant α fixes with a probability equal to its initial frequency across all speakers' grammars. For example, if some fraction x of speakers all initially use only variant α, and the remaining fraction initially do not use α at all, the probability that α fixes is x. The same fixation probability is also obtained if all speakers initially use α a fraction x of the time in their speech. This result (whose derivation is given in the Mathematical Appendix to this paper) holds no matter how large the community of speakers is, what its network structure is, how large the common value of H_{ij} is or how the variants are initially distributed among speakers *as long as* the only type of selection operating is neutral interactor selection. Therefore, it is not necessary for us to know any of these parameters in order to predict fixation probabilities from the model within the constraints imposed by neutral interactor selection.

The most likely outcome of the model is that all the variants that are initially in the majority go to fixation. If there are variants which are initially in a slight minority, their chances of fixation are almost as high as that of the majority variants. This provides some support for Trudgill's majority rule, and in particular the observation that the relevant initial frequency is not the number of speakers using a particular variant, but the token frequency (Gordon et al. 2004:239-241). In our model, the grammar frequency corresponds to the frequency with which a speaker uses a particular token at any given time. Furthermore, the stochastic nature of the interactions allows minority variants to fix without the need to invoke special mechanisms such as (linguistic) drift (Trudgill 2004, chapter 6). Two minority forms that did fix in the formation of the modern New Zealand English variety were the weak vowel schwa, initially used by 32% of the ONZE speakers (Trudgill, 2004:119) and fronted and lowered STRUT vowels, initially used by 34% of the ONZE speakers (Trudgill 2004:136; see discussion in §4 above). Unfortunately it is not clear what fraction of tokens these speaker percentages correspond to. Under the assumption that speakers were initially using these variants consistently (i.e., only ever produced tokens of a single variant), the token frequencies and speaker frequencies would coincide. Then, the model predicts that fixation of both minority forms through purely random effects would occur with probability $0.32 \times 0.34 = 0.1088$, i.e., about 11%. Therefore the hypothesis that propagation of these variants was due to some process other than neutral interactor selection alone cannot be accepted with a high degree of confidence.

We now turn to possibly the most striking aspect of the ONZE data, that a convention was established in a community of between 100,000 and 600,000 speakers in the course of fifty years (see Table 1, 1861-1890). Again, each run of the model yields a different time to fixation of a single variant. We therefore ask what the *mean* time to fixation is over many different runs, averaged over those runs in which the variant of interest actually became fixed. It is this value that we will compare with empirical data.

Analysis of the model, outlined in the Mathematical Appendix, leads to the result that for the vast majority of network structures, the number of *interactions* that need to take place on average until fixation occurs is independent of the network structure. Intuitively, one might expect a variant used by a speaker who interacts frequently to be more likely to fix, or fix more quickly, than one used by an infrequently-interacting speaker. However, fixation is a property of the whole system, not just a particular individual; so the infrequently-interacting speaker still has to have used or heard the variant enough times to reach the state of fixation. Changing the structure of the network does not affect this

time in those cases where the time it takes for the initial awareness of the variant to spread across the network is much less than the fixation time. These are in fact the vast majority of cases (see the Mathematical Appendix). Some extreme cases (such as chains of speakers, or very narrow bottlenecks between different groups) would in fact increase the fixation time (see the discussion of equation (2) below).

This number of interactions is given by the following combination of model parameters:

$$I_{fix} = \frac{N^2 T}{\lambda^2} f(hT)\omega(x) . \qquad (1)$$

In this equation N is the number of speakers, T the number of tokens exchanged per interaction, λ is the parameter controlling the rate at which speakers' grammars are allowed to change, x is the variant's initial frequency and all interaction parameters are set to a common value $H_{ij}=\lambda h$. (The appearance of λ here is due to technical reasons discussed in Baxter et al (2006)).

The notation $f(hT)$ indicates the presence of a factor contributing to I_{fix} through the product of the interaction parameter h and the number of tokens uttered per interaction T. This factor is plotted as a function of hT in Figure 5, and can be seen to be slowly varying and close to one when this product is large, but diverging rapidly if it is small.

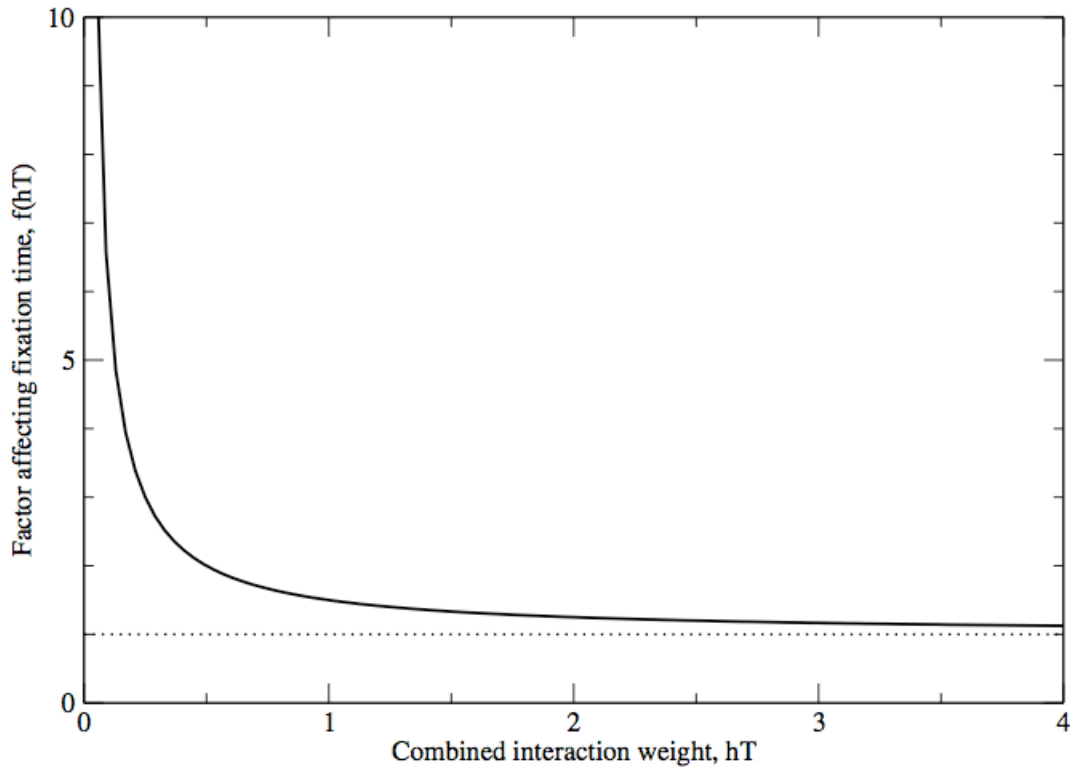

Figure 5: The factor $f(hT)$ appearing in Equation (1) for the number of interactions until fixation, plotted as a function of the combined interaction weight, hT, i.e., the product of the interaction weight h given to a single utterance from another speaker and the number of tokens per utterance T. As hT gets ever larger, the factor approaches the value one (the dotted line) from above.

In practice, the latter occurs only if h is very small, since T is by definition at least one.

This reflects the fact that the number of interactions needed until fixation occurs is very large if speakers give only a very small weight to others' utterances. Similarly, the initial variant frequency x contributes another factor $\omega(x)$ to I_{fix} and which we have plotted in Figure 6.

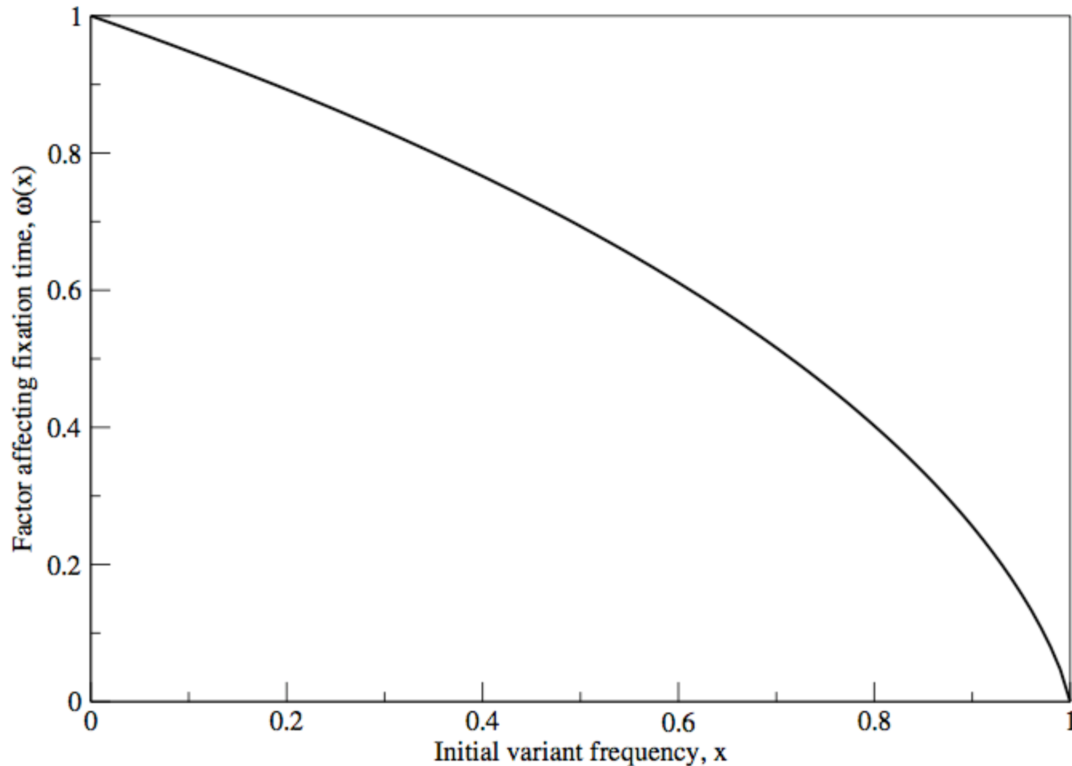

Figure 6: The factor $\omega(x)$ appearing in Equation (1) for the number of interactions until fixation, plotted as a function of the initial frequency x of the ultimately surviving variant.

The shape of the curve indicates that variants with an initially high frequency tend to fix more quickly in the model than those with a low initial frequency, as one might expect. The mathematical forms of both factors are presented in the Mathematical Appendix for the interested reader.

To make a comparison of Equation (1) with empirical data, there are two specific problems to overcome. First, we need to find a way to convert the number of interactions to a real time; and secondly we need to relate the abstract model parameters λ and h to real-world quantities. (We regard N, T and x as, at least in principle, measurable quantities).

The first task is achieved by estimating how many tokens of a linguistic variable a single speaker is likely to be exposed to in her lifetime. We call this number T^* (the asterisk distinguishing it from the number uttered per interaction). Then, since T tokens are uttered per interaction, the number of interactions per speaker lifetime is given by T^*/T.

Of course, one expects a huge variation in T^* over all possible variables. We focus therefore specifically on those analyzed in the ONZE project, which were mostly vowel

variables and a few other phonetic features. In Gordon et al (2004), token counts are presented for eight linguemes aggregated over interviews with 59 speakers of Early New Zealand English. These data are summarized in Table 4.

Table 4: Second column gives token counts for eight different linguemes analyzed quantitatively in the ONZE data aggregated over recordings of 59 different speakers (after Gordon et al, 2004). Third column converts this to an hourly rate per speaker, rounded to the nearest integer, based on a mean time per interview of 30'15" (see text).

Lingueme	***Aggregate token count***	***Hourly rate per speaker***
TRAP vowel	5706	192
DRESS lexical set	5709	192
KIT vowel	5990	201
START lexical set	2273	76
Unstressed vowels	2392	80
r variable	13700	460
h-dropping	3977	134
/hw/ ~ /h/ merger	2200	74

In order to estimate T^*, we need to convert these token counts into rates per speaker, which in turn demands knowledge of the length of the interviews conducted. This information is not presented in Gordon et al (2004), but can be estimated by identifying the relevant recordings in the catalog held by the New Zealand Radio Sound Archives (http://www.soundarchives.co.nz). We found the recordings of 55 of the 59 speakers listed in Gordon et al. (2004, Appendix 6). Taking into account that some recordings involved multiple speakers (by dividing the time of the recording equally between them), we found the mean recording length over these 55 speakers to be 30’15”. From this we can deduce the hourly per-speaker token rate also presented in Table 4. We have since been informed (Jennifer Hay, private communication) that, in fact, only a subset of the tokens uttered by each speaker were included in the token counts given. Therefore, the actual token rate would be higher than that given in Table 4, at least in the type of speech produced in an interview situation. As it happens, the mathematical analysis we are about to describe calls for us to use the smallest reasonable value for T^* so that neutral interactor selection has the greatest chance of explaining the empirical data. Thus we will take the smallest rate from Table 4, assume that on a typical day a speaker produces at least as much speech as they would in a one-hour interview, and further assume a fifty-year lifespan for each speaker (as data for life expectancy in the late 19th Century suggest, http://www.stats.govt.nz/analytical-reports/history-survival-new-zealand.htm).

This leads to a lifetime T^* estimate of approximately 1.3 million tokens (although one should bear in mind that the true value is likely to be much higher).

With the connection between the number of interactions and a real time fixed by T^*, our strategy now is to vary the unknown parameter λ so that the mean time to fixation given by Equation (1) corresponds with the observed fifty years. It then remains to decide if this value of λ is empirically plausible. Recall from §7 that λ controls the amount to which tokens uttered in the previous interaction affect the grammar. We can ask how many interactions are needed until the intensity of a stored token in the memory drops to some fraction ε of its initial value. For example, if ε =0.01, this would correspond to time taken for the initial intensity to decay to 1% of its original value. We call this time the memory time, and, as a fraction of a speaker's lifetime, will denote it t_{mem}. In the Mathematical Appendix, we show that for fixation to occur in fifty years, t_{mem} for at least one speaker in the community must satisfy the relationship

$$t_{mem} \leq \frac{-\ln \varepsilon}{\sqrt{2NT^* f(hT)\omega(x)}} . \qquad (2)$$

Again, this relationship holds for any social network structure comprising N speakers, as long as propagation is by neutral interactor selection alone. There are two reasons why this expression is written as an inequality rather than an equality. First, there are a small number of networks for which I_{fix} exceeds the value given by Equation (1). Secondly, the memory lifetime imposed by a particular value of λ varies from speaker to speaker in models in which different speakers interact with one another with different frequencies.

The question then arises: what is an acceptable range for t_{mem}? This is the time window within which exposures must re-occur for this consolidation to take place. For infants (aged 3-6 months), this time window is about 2-4 days for a range of learning tasks; for adults, these windows appear to be longer, in fact, much longer (Rovee-Collier 1995). Therefore, if a speaker's t_{mem} is required to be less than two days for community-wide fixation to occur in fifty years, we have a criterion for rejecting the model of propagation by neutral interactor selection.

So that this rejection criterion is invoked with the smallest likelihood, we choose parameter values that make the right-hand side of Equation (2) as large as is reasonably consistent with the empirical data. That is, we choose the smallest N and T^* suggested by the data. From Table 1, we see that the smallest N during the period of interest (1864-1890) is about 100,000. We argued above for an underestimate on T^* of 1.3×10^6. Since $\omega(x)$ is a decreasing function of x, we should also focus on the lingueme variants reported by Trudgill (2004) to have been in a healthy majority among the ONZE speakers. Two features, H retention (Trudgill, 2004:116) and diphthong shift (Trudgill, 2004:121), were reported as being initially used by 75% of speakers, so we shall take $x=0.75$ as a representative value.

In Figure 7 we plot the memory time appearing on the right-hand side of Equation (2) as a function of the product hT for various T^*, $N=100,000$, $x=0.75$ and ε =0.01. This means that every possible combination of the unknown parameters T and h is included somewhere on the graph. Note that the data have been plotted on logarithmic axes which means that even with the smallest T^* plotted – 10^5, i.e., an order of magnitude lower than the above underestimate – memory times imposed by the model are already an order of magnitude shorter than the allowable two-day time window, and even smaller for larger (and probably more realistic) values for T^*.

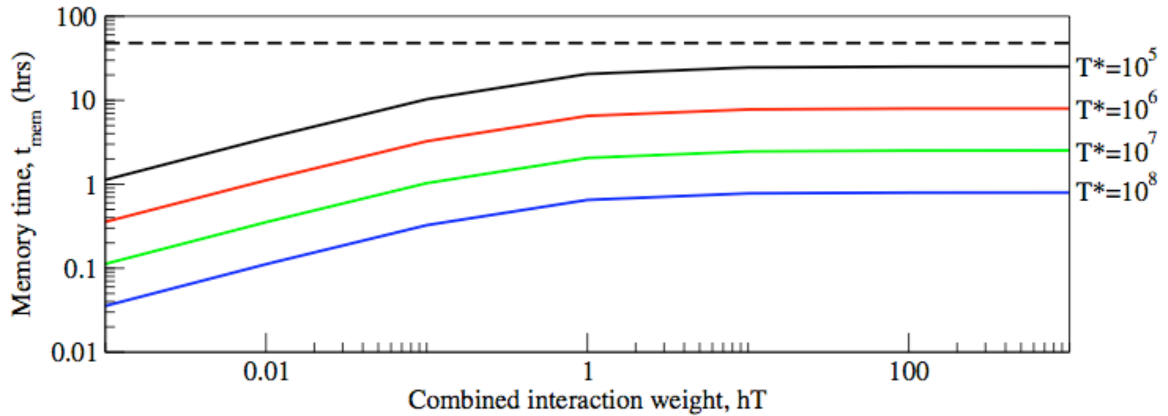

Figure 7: The memory time appearing on the right-hand side of Equation (2) that is imposed on the model by requiring fixation to occur within fifty years, as a function of the (unknown) combined interaction weight hT appearing in Figure 4. As described in the text, we have taken $N=100,000$, $x=0.75$ and $\varepsilon=0.01$ for values of T^* ranging from 10^5 (top curve) to 10^8 (bottom curve). For the linguemes relevant to the ONZE project, the data suggest that T^* of at least 10^6 is appropriate. The dashed line is the shortest memory time that we deem plausible (see text). Note that both axes are logarithmic: while the longest time window (largest h, smallest T^*) is only half that of a human infant, the shortest (smallest h, largest T^*) is 1000 times shorter.

Bearing in mind that the formula (2) is an overestimate of a speaker's memory time, and that we have deliberately erred on the size of caution with the values of N, T^* and x, we believe Figure 7 provides compelling evidence that neutral interactor selection is unlikely to be solely responsible for the fast convergence of the New Zealand dialect to a homogeneous, stable variety. Although this conclusion is based on our having set the forgetting parameter ε to the somewhat arbitrary value of 0.01 (1%), we note that for the case $T^*=10^6$ suggested by the data, the curve in Figure 7 can be made to approach the desired two-day memory window for large hT if ε is reduced to approximately 10^{-12}, an unfeasibly small number. So while we are confident that the above argument is correct, this aspect would benefit from further study. This is discussed once more in the final section, but for the moment we note that not only would larger, and more empirically reasonable, values of N and T^* demand an even smaller value for ε, but also that extensions to the model that include more realistic features also have the effect of extending fixation times, and hence further reduce the t_{mem} that is imposed. These effects are the subject of the next section.

9. Modeling generational replacement

In this section we add generational effects to our model of neutral interactor selection. The effect of the presence of speakers born at different times is an integral part of Trudgill's account. His three stages (see §3) correspond to three generations: immigrants, first-generation native New Zealanders, and second-generation New Zealanders. Trudgill uses the simplifying assumption that each generation is discrete and separate in time (Trudgill 2004:24). This is of course a stylized representation of demographic reality, but we will begin by modeling these three distinct generations. In this section we will show that this generational model – and even more realistic models of speakers of differing

ages – does not ameliorate the difficulty that fixation cannot occur within 50 years, and in fact makes the situation worse.

In §8 we used the most optimistic condition that all speakers had the same receptiveness to language use around them as children. Now we assume that children are more adaptable in their language use than adults – that is, adults are less adaptable than children. It seems clear that if some of the speakers in the population change more slowly than in the model of §8, the time taken for the population as a whole to reach fixation must be longer. This is borne out in simulations, which we now describe.

To represent the difference in susceptibilities between children and adults, in our first simulation speakers are assigned a constant value of the update parameter λ for the first half of their lifetime, which represents childhood. In the second half of their lives, the adult stage, speakers interact as normal, but their grammars are not updated at all, i.e. λ is set to 0 for these speakers. The generations overlap, so that each generation's child phase coincides with the previous generation's adult phase, as represented in Figure 8.

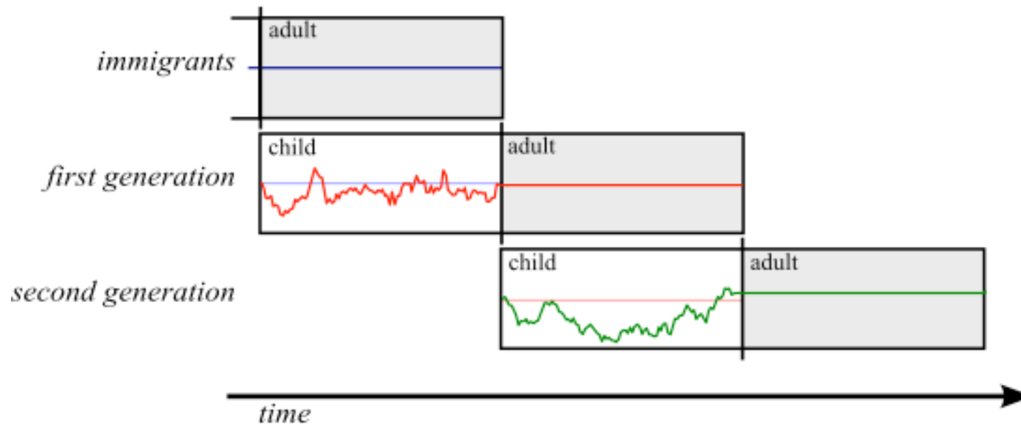

Figure 8. Schematic representation of the three discrete generations. The immigrant generation arrives as adult speakers. Each subsequent generation begins as children then changes into adults. At the same time the adults are removed. The colored traces represent the overall proportion with which each generation uses the first variant (α).

At any time, two generations coexist, one in an adult stage and one in a child stage. We make the further simplification that the original generation of immigrants is already in the adult phase at the beginning of the simulation, and they speak with a fixed grammar representing their original dialects. That is, the rudimentary leveling and other changes in Trudgill's first stage (Trudgill 2004:89-99) are considered to be sufficiently minor that we can neglect them for our current purposes. All other aspects of the model described in §7 and §8 remain unchanged.

In repeated simulations of this model, we found that the children as a group duplicated the overall proportion of variants used by the adults, but individually there was a large variation between speakers. This can be seen in Figure 9 in which the proportion of speakers who used variant α with a frequency x lying in a narrow range is represented by a vertical bar covering that range.

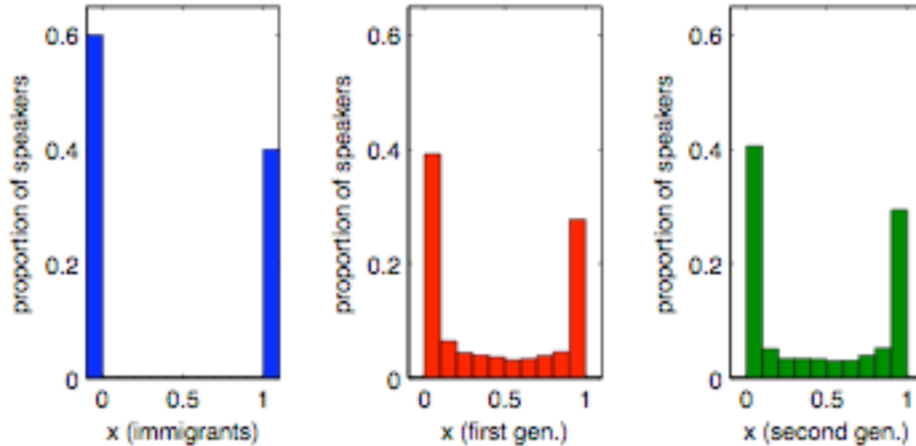

Figure 9. The distribution of speaker's grammar values x (the frequency with which variant α is used) in each of the three generations. The immigrant generation is fixed at the initial conditions (in this case 40% use exclusively variant α, 60% variant β) while subsequent generations conform to a distribution resulting from the interaction dynamics. Results shown are for a population of 60 speakers with interaction parameter $h=0.2$ and child susceptibility $\lambda=0.05$. Distributions have been averaged over 100 repeats of the experiment.

The bars just below zero and just above one represent the proportion of speakers that use exclusively variant β or exclusively variant α, respectively, at the end of each generation. Forty percent of the immigrant generation used exclusively variant α, while the remainder used exclusively variant β. The results shown in the figure are for a relatively small population, but the shape of the distribution shown depends only on the interaction parameter h and not on the population size (it is very similar to the quasi-stationary distribution described in Baxter et al. 2006). Several different values of T^*, the number of tokens each speaker hears in a lifetime, were tried but it was found that an equilibrium distribution was reached relatively quickly in each generation, and did not change when generations ran for a larger number of interactions.

In the first generation all speakers used both variants (the proportion of speakers who use exclusively one variant is zero), but most speakers used one variant a large majority of the time: some speakers used mostly variant β, while others used mostly variant α. This distribution arises because the children's grammars drift due to the stochastic nature of the model and their interactions with each other, with the feedback from her own utterances tending to drive an individual speaker's grammar towards a majority of one variant. On the other hand, because the children continue to hear examples of both variants from the immigrant generation adults (and in the original proportions), the overall proportion of variants used by the first generation tends to stay fairly close to that used by the immigrant generation. This situation is qualitatively similar to the intra- and inter-individual variability of Trudgill's Stage II (Trudgill 2004:101-108): individual speakers differ from one another, their usage can be variable, and the overall mixture reflects the range and frequency of variants available. It should also be noted that this variation appears in the early stages of any model in which speakers can drift, even if there is no inclusion of different generations of speakers.

More problematic for Trudgill's thesis is that the second generation repeats this process in relation to the first generation, learning a similar overall frequency of variant use and settling on a very similar distribution of frequencies of use among different

speakers, as can be seen in the rightmost panel of Figure 9. In Figure 8 the overall proportion of variant α used by each generation is represented by a colored trace. We see that each generation shows some variation during their childhood phase, but the frequency does not move very far from that used by the previous generation in their adult phase. This pattern continues over subsequent generations, with a small random variation between generations until one variant finally becomes fixed. The amount of change from one generation to the next is independent of the number of tokens each speaker hears in her lifetime (T^*) and is instead governed by the size of the population and the strength of the interaction between speakers (the parameter h of §8). The effect of the presence of a 'frozen' adult generation is to inhibit fixation. The time taken to reach fixation for a population of 100,000 speakers and small values of h is typically hundreds of generations. The number of generations required to reach fixation actually increases with larger values of h.

We also investigated allowing adult speakers to modify their grammars a little. The value of λ used for the adults should be much smaller than that for children to represent the relative inflexibility of adult speakers. (For example, assigning the adults a value of t_{mem} of 1 year and children a value of 2 days would mean the λ values for adults would be 1/200 of that used for children.) This allows the population as a whole to drift a little in each generation, making fixation times shorter than if the adults are 'frozen', but still significantly slower than in the undifferentiated population of §8 (where we used the optimal learning rates of children for the entire population). Similarly we considered variations of the model in which the time delay between generations and the relative proportions of the lifespan spent as children and adults were varied. In any version in which the child generation is always coexisting with an adult generation, the outcomes are qualitatively the same. In versions in which the children spent some time in isolation – the adults were 'killed off' before the younger generation switched to being adults – each generation could drift somewhat away from the previous generations mean value in the gap between generations. Even in the extreme case in which children spend almost their entire lives without contact with adults, fixation is still slower than in the generation-less model we described in the previous section. Furthermore, this scenario is very unrealistic, as it requires all the adults to vanish while their offspring are still children, creating a society of orphans that grows up independently from their parents' influence.

Finally we explored models in which the idealization of three distinct generations of speakers was relaxed. In particular, we investigated continuous generations in which speakers are removed and new speakers introduced individually at different times, rather than an entire generation being replaced at once. We also allowed for a gradual decrease in receptiveness instead of the all-or-none model we began with. Of course such a model is much more realistic than the previous one, though now it is much more difficult to identify speakers as belonging to a particular generation. In these models, speakers begin with a high receptiveness (high λ), which declines in a smooth way as they get older (see Tomasello 2003:286-87). The rate at which λ declines was varied, from a situation in which speakers remained plastic throughout their lifespan, to one in which they very quickly became 'set in their ways' and did not change at all after a few interactions. If the speakers' receptiveness decays quickly, the population contains a group of 'adults' whose grammars are essentially fixed (they cannot change enough to drop either variant within

their lifetime). These speakers still have an effect on the other members of the population, and so we have a situation similar to the original model with discrete generations: variants take a very long time to be eliminated because examples of it are always produced by adults with unchangeable grammars. The slower the decay in receptiveness, the shorter the mean time to fixation becomes. Nevertheless, even with very little decay in λ the average time that it takes for fixation to occur is still longer than we would find if all speakers used the same (children's) value of λ. The same is true of all the variations investigated: as adult learning rates are necessarily slower than those of children, so the overall rate at which change can happen *must* be lower (i.e. fixation times longer) than those estimated in §8.

Trudgill makes the simplifying assumption that immigration does not continue after Stage I, or at least that this does not have a significant effect on the process (Trudgill 2004: 163). In fact, because time to fixation increases with population size the natural growth of the population and continuing immigration will further slow the fixation process. New immigrants in Stage II and III would also slow the approach to consensus through their continued use of minority variants that may have already become less frequent or lost completely amongst the native speakers. Under neutral interactor selection, the usage patterns of all speakers are equally influential, even if the speakers in question arrive later – the 'founder effect' (Mufwene 2001) does not occur because we have explicitly excluded any social valuation effects. Finally in these generational variations to the model the result still holds that the structure of the network of interactions does not affect the outcome at all.

This exhausts the possibilities for modeling the NZE situation with neutral interactor selection only, as suggested by Trudgill's theory. Although the introduction of generations appears to model some of the basic features of Stage II of Trudgill's theory – the extreme variability of the linguistic behavior of the first native-born generation – it does not by itself allow for the transition to a leveled, uniform new dialect. In making the model more realistic by introducing adults resistant to influence, a decrease in receptiveness as speakers age, the outcome is always that fixation is slower than in the non-aging model of section 8. It is clear that such a situation is insufficient to achieve a common dialect in the time frame that is observed in the ONZE data.

10. Conclusion

Trudgill's deterministic theory of new-dialect formation is appealing. In the absence of social structures from the original society, the children of immigrants to colonies are free, so to speak, to select any linguistic variants around them, without any social value attached to any particular variant. A simple probabilistic effect leads to the majority variant being chosen. Variation in the uptake of variants are attributed to variation in network density.

Trudgill's theory corresponds to a neutral evolution model of propagation, with allowance for what we have called neutral interactor selection (i.e. effects of social network structure). In this paper, we have described the implementation of a mathematical model of Trudgill's theory, and applying generous empirical values from New Zealand English for crucial variables, including memory retention of tokens. We conclude that simple factors of frequency of exposure to language use and social network

structure are insufficient in themselves to account for the emergence of a new dialect in an isolated society, as proposed by Trudgill and suggested by the evidence from the history of New Zealand English. We have also demonstrated the necessity as well as utility of employing mathematical modeling in language change: one cannot be certain that an intuitively plausible model actually works without a precise quantitative model.

Our results do not imply that social network structure plays no role in language change. Croft's framework, and our mathematical model based on it, take as a starting point Trudgill's principle that our language behavior is influenced by the language behavior of our interlocutors. Demographic and probabilistic factors clearly play a role, in that most of the time, the majority variant is indeed selected, and also in regional variation in New Zealand, for example the persistence of Scottish English variants in the Southland of New Zealand, which was settled by a disproportionately large number of Scottish English speakers. It is just that this cannot be the entire story; in fact, it is mathematically impossible, given realistic assumptions about linguistic interactions, receptiveness to language use over age, generational replacement, social network structure and population size.

We conclude with some brief speculations about social and demographic factors that may have contributed to the emergence of a new dialect in New Zealand around 1890. There are three factors that might be relevant: the dramatic decline of immigration, the loss of isolation of settlements, and the establishment of indigenous social differences within New Zealand society. The first two factors contribute to the creation of an independent yet integrated speech community in New Zealand, and the third provides a mechanism to drive social selection processes.

Immigration had declined significantly by the 1890s (Hamer 1997:143; see also §2). In §2, we noted that during Phase II of the demographic history of New Zealand (1864-1886), corresponding to the ONZE project speaker generation, substantial immigration continued, with net immigration of approximately 200,000 as against an approximately 300,000 natural increase. Of course, the latter group is entirely made up of children and adolescents, while the immigrants are mostly adults (though many immigrants brought families with them). But in Phase III, corresponding to Stage III in Trudgill's theory, immigration dramatically declined. Thus during the period when the first New Zealand English speaking generation was growing up, there was very little immigration to New Zealand, and thus no continual major disruption of the local linguistic situation.

In the earliest days of New Zealand, the 1840s, the six major settlement communities were very isolated from each other (Graham 1992:119). However, the isolation of communities and of families decreased significantly between the first and second native-born New Zealand generations. From 1870 to 1879, the public railways increased from 46 miles to around 1150 miles, and from 1871 to 1881, the government built 2000 miles of roads and tracks in the country (Belich 1996:352-53). The quickest time to travel from Dunedin, near the south end of the South Island, to Auckland, near the north end of the North Island, decreased from 15 days in 1859 to 5.5 days in 1879 (and to 3 days in 1898; McKinnon 1997:52). Investment in public works declined in the 1880s due to economic depression (Dalziel 1997:111), but the previous improvements in infrastructure allowed social networks to become wider, especially given the aforementioned high mobility of the population. Finally, between 1871 and 1891, the proportion of children aged 5-14 in public schools rose from 27% to 80% (Gordon et al. 2004:56; Graham 1992:132); this

does not include children in private schools, which were 20% of total schools in 1890 (Dalziel 1997:122). Thus, within communities, children were brought into greater contact with each other through schools. These processes imply a broadening of social networks at both a local and a national level, particularly among children.

In fact, these two factors – absence of immigration and a fully connected social network – are already incorporated into the neutral interactor selection model we tested in this paper. Our results reinforce the likelihood that even under these circumstances, a more powerful selection mechanism such as replicator selection is required. How would such a mechanism arise? As we noted in §4, Trudgill writes that 'It would be ludicrous to suggest that New Zealand English speakers deliberately developed, say, closer front vowels in order to symbolize some kind of local or national New Zealand identity' (Trudgill 2004:157). However, it could be that social differences *within* New Zealand, combined with replicator selection, led to the propagation of linguistic variants that became the New Zealand English dialect. Olssen writes, 'By the 1890s [that is, by the time of the birth of the first generation of New Zealand English speakers] stratification within New Zealand society was already well-established' (Olssen 1992:272). For example, by 1882 1% of landowners possessed 32% of the assessed value of landholdings; 250 people or companies owned 7.5 million acres of land (Simpson 1997:185). We also noted the gender differences that pattern in the traditional way (§4); it is unlikely that New Zealanders abandoned the attitudes towards gender that are characteristic of European societies.

The prevalence of the majority variants means that they were more likely to be propagated; but the mechanisms by which they were propagated were probably social. Nevertheless, only mathematical modeling of replicator selection in the context of the social and demographic changes described in the preceding paragraphs can tell us whether this is indeed the case.

Mathematical Appendix

In this appendix we give some additional details of how the major mathematical results presented in the main text are obtained. To prevent over-burdening this section with equations, we shall make frequent reference to our earlier work (Baxter et al. 2006) that contains further mathematical details. In footnotes we place comments of a technical nature that can be skipped by less mathematically inclined readers.

Probability of fixation – As discussed in the main text, the model defined in Section 7 is stochastic. That is, each realization of the sequence of steps described results in a different outcome because the choice of speakers that interact and tokens they produce is random. However, one can ask how *average* quantities change with time, and these vary in a predictable way. In the present context, we mean an average taken over the distribution generated by many independent realizations of the same process.

Using the results of Baxter et al. (2006), one can show[1] that the average of the lingueme variant frequency x_i stored in speaker *i*'s grammar, a quantity denoted $\bar{x}_i$, changes with time according to the equation

$$\frac{d}{dt}\bar{x}_i = h\sum_{j=1}^{N} G_{ij}\left(\bar{x}_j - \bar{x}_i\right) \qquad \text{(A1)}$$

within a model of neutral interactor selection, where the speaker interaction weights H_{ij} are all equal to the common value λh. In words, this equation says that, at any given instant, the rate at which the average quantity $\bar{x}_i$ changes with time depends on the values of other averages $\bar{x}_j$ through the sum on the right-hand side. In this formulation, one unit of time corresponds to $1/\lambda^2$ interactions between pairs of speakers.

To determine the probability that the variant whose frequency is given by x_i eventually fixes, i.e., becomes the sole remaining variant used in the entire community, it is useful to focus on its total usage frequency across the entire community. This quantity we simply call x (with no subscript) and define as follows:

$$x = \frac{1}{N}\sum_{i=1}^{N} x_i \ .$$

This frequency can equal zero only if no speakers are using this variant (i.e., it has gone extinct), and similarly can equal one only if all speakers use only this variant (i.e., it has fixed). The key point here is that, using Equation (A1), one can show that the rate of change of the frequency $\bar{x}$, i.e., the average value of the community frequency x, is zero[2]. That is, the value $\bar{x}$ does not change with time.

The significance of this result is as follows. Consider a very large number of independent realizations of the process, all starting from the same initial condition, i.e., the same set of grammar frequencies x_i. The initial value of the quantity x is then the combined usage frequency across the entire community, and has the same (prescribed) value in all realizations. Therefore its average $\bar{x}$ across all realizations is equal to this prescribed value. Now look sufficiently far ahead that fixation or extinction is

---

[1] Specifically, the Fokker-Planck equation is multiplied by x_i and all x coordinates are integrated over to obtain the average.

[2] This is achieved by summing Equation (A1) over all speakers *i*.

guaranteed to have occurred. If in each independent realization fixation occurs with probability p, then in a fraction p of realizations the value of x=1, and in the remaining fraction x=0. Thus, at very late times, $\bar{x} = p \times 1 + (1 - p) \times 0 = p$. Therefore, the fixation probability p is equal to $\bar{x}$, which in turn is equal to the variant's initial usage frequency.

Mean time to fixation – In Sections 8 and 9 we were mostly concerned with the number of interactions between speakers[3] that are needed on average until a state of fixation is reached. With only a single generation of speakers, as described in Section 8, computer simulations suggest (see also Baxter et al, 2006) that fixation is a two-stage process:

1. The average quantity $\bar{x}_i$, which may initially be different for each speaker, approaches the community average $\bar{x}$.
2. The x_i values fluctuate until such a time that all speakers all spontaneously fix on– or stop using–the variant.

This behavior can be seen from Figure A1. The average $\bar{x}_i$ for two speakers, one initially always using a variant and one initially never using it, is plotted, along with the variance in a speaker average and the covariance between two of them. In both plots, all speakers interact with each other equally often, and in the upper plot *N=20* whilst in the lower plot *N=40*. In both cases half the speakers initially used one variant, and the other half the other. Superposed on these plots is the probability distribution for the fixation time, and the solid vertical line is the mean fixation time.

[3] Here, and in Section 8, we mean by interactions the *total* number of interactions that take place in the whole community; to convert this to the number of interactions experienced on average by a single speaker, one should divide by *N*.

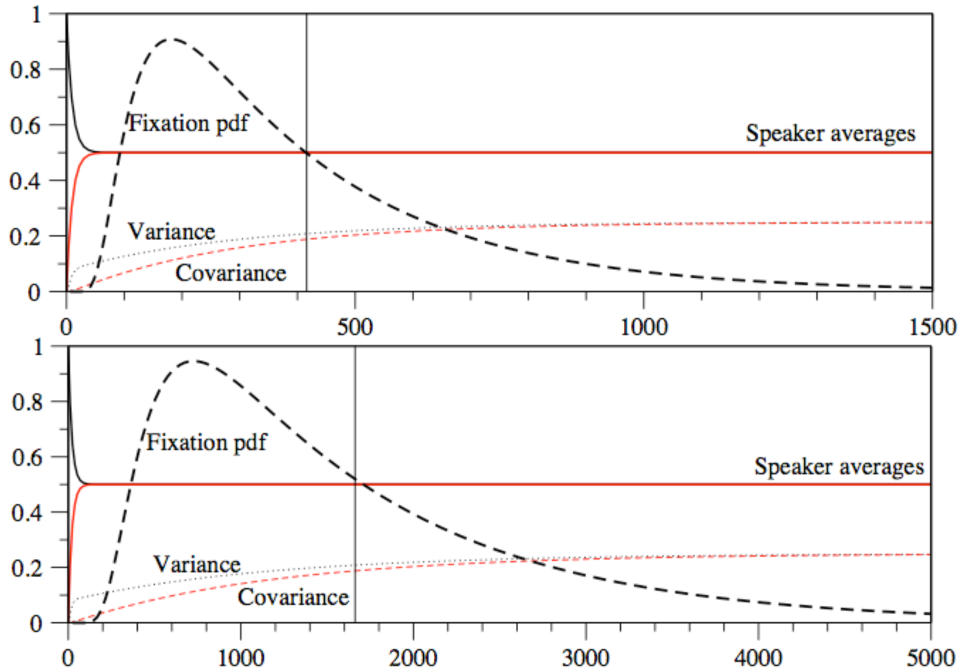

Figure A1: Variation in averages, variance and covariance of x_i for systems of $N=20$ (upper figure) and $N=40$ (lower figure) speakers, all of whom interact equally often with each other. These data were obtained by direct numerical integration of exact differential equations governing them. The heavy dashed curves superimposed on these plots (vertical scale not shown) are the probability distribution functions (pdfs) of fixation time, obtained using the assumptions given in this Appendix. The solid vertical lines indicate the mean of these distributions which have been verified by computer simulation (Baxter et al, 2006).

The main point to notice is that the initial relaxation of $\bar{x}_i$ to the community mean of 0.5 happens very quickly compared to the mean fixation time (or the peak in the fixation time distribution). Furthermore, the duration of the initial relaxation relative to the mean fixation time (or the peak of the fixation time distribution) is shorter in the system with the larger number of speakers. In fact, it turns out that, under neutral interactor selection, the first stage (relaxation) lasts for a time which–for the vast majority of networks–is proportional to the number of speakers N. Meanwhile, on all networks, the second stage lasts for a time proportional to N^2. Therefore, if the number of speakers is large, the second stage typically lasts very much longer than the first, and so one can–to a good approximation–ignore the expected length of the first stage, and focus on that of the second. There are some networks where the first stage lasts much longer then the second–for example, if the community comprises a very long "chain" of speakers, each talking only to its neighbors. Such networks lack the "small-world" property, whereby the shortest path between any pair of speakers is much less than the total number N, which is believed to be present in real social networks (Watts 1999). We will not discuss these unusual cases further here.

The lifetime of the second stage is estimated by making two assumptions on the variation in grammar frequencies x_i across the community at the start of the second stage. First it is assumed that the difference between x_i and the community average x is uncorrelated with the difference for some other speaker j. In other words, the covariance of x_i and x_j is assumed to vanish. Support for this assumption is given in Figure A1,

which shows this covariance to be small at the onset of the second stage. Secondly, it is assumed that at this time, the rate of change of the variance is so slow that it can be neglected–this can also be seen from Figure A1, where the rate of growth of the variance slows dramatically at the start of the second stage.

Following a procedure similar to that used to obtain (A1) from the results of Baxter et al (2006), one can show that under neutral interactor selection, a time-independent variance in x_i is obtained if the condition

$$G_i\left[\bar{x}_i(1-\bar{x}_i)-(2h+\tfrac{1}{T})\,\mathrm{Var}(x_i)\right]=0$$

is satisfied. In this equation, $\mathrm{Var}(x_i)$ denotes the variance of x_i and G_i is the total fraction of all interactions in which speaker i participates. The main point to notice about this equation is that it can be satisfied only if $G_i = 0$ (which is never true as all speakers are involved in some fraction of interactions) or the quantity in square brackets is zero. Since this latter quantity does not involve the interaction frequencies G_{ij}, it follows that the variance of x_i is independent of the network structure. If we further assume that the lifetime of the second stage depends only on the mean and variance of the grammar frequencies at the start of the second stage, it follows that this lifetime is also independent of the network structure under neutral interactor selection.

Specifically, this means that the fixation time given for the case of a network in which all speakers interact equally often with each other (Baxter et al, 2006) applies for any network structure where the duration of stage one (described above) is much shorter than that of stage two. Expressed in terms of the total number of interactions that need to have taken place in the whole community, this fixation time is[4]

$$I_{fix}=\frac{N^2T}{\lambda^2}\left(1+\frac{1}{2Th}\right)\frac{(x-1)\ln(1-x)}{x}\ . \qquad \text{(A2)}$$

In Figure A2 we demonstrate the fact that the fixation time is independent of the network structure by plotting its value (divided by the number of speakers) obtained by computer simulation for a range of structures (see §7). These are: a flat society, where all speakers interact equally with all other speakers; a hub-and-spoke society divided into groups of 10 speakers, one of which is a central hub whose speakers interact with those in the remaining spoke groups, between which there are no interactions; an equal group network, where pairs of speakers within the groups of ten interact 10 times more frequently than pairs drawn from two different groups; a ring of speakers with 10% of the the possible remaining long-range connections between them established; and a network in which only 20% of all possible interactions occur. We see that for any given system size, the fixation times across all structures are in agreement with each other and the theoretical prediction given above (at least within the errors that arise from statistical fluctuations inherent in a finite sample of simulated realizations of the process).

[4] This differs from the expression in Baxter et al (2006) by a factor $1/\lambda^2$ which is due to a conversion between continuous time units and interactions; furthermore, we have only kept terms proportional to N^2 here, consistent with our having neglected the lifetime of stage one.

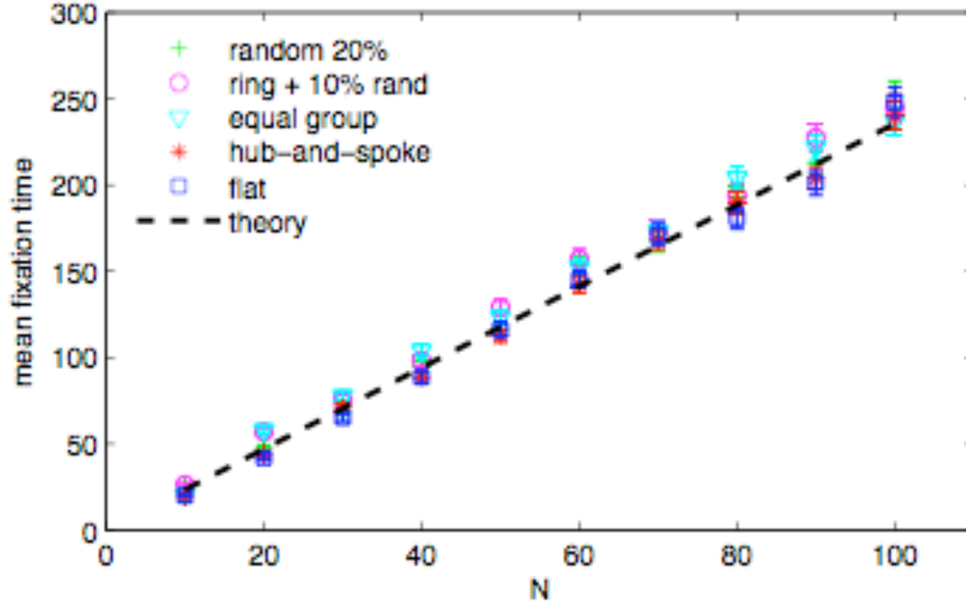

Figure A2: Mean fixation per speaker obtained by computer simulation for a range of different networks and sizes under neutral interactor selection. The initial lingueme frequency $x=0.4$; the interaction weight h=0.2; and the weight given to the most recently-used utterances $\lambda=0.03$. The different symbols correspond to different network structures (see text) and the dashed line the theoretical prediction, Eq (A2).

We finally remark in this section that the fixation time predicted by Equation (A2) can actually be shown by a more rigorous argument to underestimate its true value under neutral interactor selection. The details of this calculation will be presented elsewhere (Baxter et al. 2008).

Memory lifetimes implied by the model – For a given set of parameters, the model predicts (under neutral interactor selection) that the total number of interactions between all speakers that take place before fixation occurs is

$$I_{fix} = \frac{N^2 T}{\lambda^2} f(hT)\omega(x)$$

where $f(hT) = 1 + \frac{1}{2hT}$ and $\omega(x) = (x-1)\ln(1-x)/x$ (see equation A2). We will assume that this number of interactions takes place over a single human's lifetime, and ask what value the parameter λ would need to be to achieve this. In turn, this needs to be converted to a real time so that it may be compared against empirically-determined memory time windows.

The strategy is the following. We focus on speaker *i* who participates in a fraction G_i of all interactions. Therefore, by the time fixation has occured (on average) she has uttered $G_i I_{fix} T$ tokens. If fixation occurs within a single speaker's lifetime, this number must be equal to T^*. Substituting in the full expression for I_{fix} given above, we find that

$$\frac{G_i N^2 T^2}{\lambda^2} f(hT)\omega(x) = T^* .$$

Rearrangement of this expression tells us the value of λ that is needed to achieve fixation in speaker *i*'s lifetime:

$$\lambda = \sqrt{\frac{G_i f(hT)\omega(x)}{T^*}} NT \ . \qquad \text{(A3)}$$

The parameter λ controls what fraction of a speaker's grammar contains information about the most recent interaction. The intensity of this information thus decays by a factor of approximately 1- λ in each interaction. Therefore, the larger the value of λ that is imposed, the more quickly a speaker forgets information about earlier utterances. We see that λ is large when N is large and/or T^* is small. As fixation times grow with N, it follows that to allow fixation to occur in a (fixed) human lifetime as N is increased, correspondingly faster turnover of the grammar is required to achieve this. On the other hand, if T^* is small, the time between successive tokens heard by a speaker is long, and λ has to be increased in order that the tokens heard near the start of a speaker's life are forgotten before the end.

To convert λ to a fraction of a human lifetime, we need to introduce a threshold intensity of the memory of a token at which it is considered forgotten. This we set at a fraction ε of its initial intensity. Since this intensity decreases by a factor 1-λ at each interaction, the number of interactions n until the threshold ε is reached is found from:

$$(1-\lambda)^n = \varepsilon \ .$$

To rearrange this expression to give n as a function of λ, we take the logarithm of both sides

$$n \ln(1-\lambda) = \ln \varepsilon \ .$$

We use the fact that, for small λ, $\ln(1-\lambda) \approx -\lambda$ to obtain the simpler, approximate expression

$$n = -\frac{\ln \varepsilon}{\lambda} .$$

Note that since $\varepsilon < 1$, its logarithm is negative and hence n is a positive number as required.

It now remains to write this number of interactions as a fraction of speaker i's lifetime. This is achieved by dividing it by the number of interactions in which a speaker participates in her lifetime, which is simply T^*/T. Thus the memory time, as a fraction of a speaker's lifetime, is

$$t_{mem} = -\frac{\ln \varepsilon}{\lambda} \frac{T}{T^*} .$$

Substituting in the value of λ from (A3) we obtain

$$t_{mem} = -\frac{\ln \varepsilon}{\sqrt{G_i N^2 T^* f(hT)\omega(x)}} .$$

An inequality that holds for any network is obtained by noticing that the smallest t_{mem} applies to the most frequently interacting speaker, i.e., that with the largest G_i. In turn, the largest G_i must be at least $2/N$, since all the G_i are required to sum to 2. Hence, by putting $G_i=2/N$ in the previous expression, we find an upper bound on the shortest memory time imposed on any speaker in the model. This bound is presented as Equation (2) in Section 8.

References

Anderson, John R. and Lael J. Schooler. 1991. Reflections of the environment in memory. *Psychological Science* 2.396-408.

Baxter, Gareth J., Richard A. Blythe, William Croft and Alan J. McKane. 2006. Utterance selection model of language change. *Physical Review E* 73.046118.

Baxter, Gareth J., Richard A. Blythe, and Alan J. McKane. 2008. Fixation and consensus times in a network: a unified approach. Available on the physics preprint archive at http://arxiv.org/abs/0801.3083.

Belich, James. 1996. *Making peoples: a history of New Zealanders*. Honolulu: University of Hawai'i Press.

Blythe, Richard A. and Alan J. McKane. 2007. Stochastic models of evolution in genetics, ecology and linguistics. *Journal of Statistical Mechanics* 07-P07018.

Bybee, Joan L. 1985. *Morphology: a study into the relation between meaning and form*. Amsterdam: John Benjamins.

Bybee, Joan L. 2001. *Phonology and language use*. Cambridge: Cambridge University Press.

Bybee, Joan L. and Dan I. Slobin. 1982. Rules and schemas in the development and use of the English past tense. *Language* 58.265-89.

Croft, William. 2000. *Explaining language change: an evolutionary approach*. Harlow, Essex: Longman.

Croft, William. 2006. The relevance of an evolutionary model to historical linguistics. *Competing models of linguistic change: evolution and beyond*, ed. Ole Nedergård Thomsen, 91-132. Amsterdam: John Benjamins.

Croft, William. To appear. The origins of grammaticalization in the verbalization of experience. *Linguistics*.

Crow, James F. and Motoo Kimura. 1970. *An introduction to population genetics theory*. New York: Harper and Row.

Dalziel, Raewyn. 1997. Railways and relief centres (1870-1890). Sinclair 1997, 99-124.

Dawkins, Richard. 1976. *The selfish gene*. New York: Oxford University Press.

Gordon, Elizabeth, Lyle Campbell, Jennifer Hay, Margaret MacLagan, Andrea Sudbury and Peter Trudgill. 2004. *New Zealand English: its origins and evolution*. Cambridge: Cambridge University Press.

Graham, Jennie. 1992. Settler society. Rice 1992, 112-40.

Graham, Jennie. 1997. The pioneers (1840-1870). Sinclair 1997, 49-74.

Hamer, David. 1997. Centralization and nationalism (1891-1912). Sinclair 1997, 125-52.

Huggins, A. W. F. (1964). Distortion of the temporal pattern of speech: Interruption and alternation. *Journal of the Acoustical Society of America* 36.1055-1064.

Hull, David L. 1988. *Science as a process: an evolutionary account of the social and conceptual development of science*. Chicago: University of Chicago Press.

Hull, David L. 2001. *Science and selection: essays on biological evolution and the philosophy of science*. Cambridge: Cambridge University Press.

Kerswill, Paul. 1996. Children, adolescents, and language change. *Language Variation and Change* 8.177-202.

Kerswill, Paul. 2007. Review of Trudgill 2004. *Language* 83.657-61.

Kerswill, Paul and Ann Williams. 2000. Creating a New Town koine: children and language change in Milton Keynes. *Language in Society* 29.65-115.

Kerswill, Paul and Ann Williams. 2005. New towns and koineization: linguistic and social correlates. *Linguistics* 43.1023-48.

Lieberman, Erez, Jean-Baptiste Michel, Joe Jackson, Tina Tang and Martin A. Nowak. 2007. Quantifying the evolutionary dynamics of language. *Nature* 449.713-16.

McKinnon, Malcolm (ed.). 1997. *New Zealand historical atlas*. Auckland: David Bateman.

Mufwene, Salikoko. 2001. *The ecology of language evolution*. Cambridge: Cambridge University Press.

Ohala, John. 1989. Sound change is drawn from a pool of synchronic variation. *Language change: Contributions to the study of its causes*, ed. Leiv Egil Breivik & Ernst Håkon Jahr, 173-98. Berlin: Mouton de Gruyter.

Olssen, Erik. 1992. Towards a new society. Rice 1992, 254-84.

Owens, J. M. R. 1992. New Zealand before annexation. Rice 1992, 28-53.

Phillips, Betty S. 1984. Word frequency and the actuation of sound change. *Language* 60.320-342.

Pierrehumbert, Janet B. 2003. Probabilistic phonology: discrimination and robustness. *Probabilistic linguistics*, ed. Rens Bod, Jennifer Hay and Stefanie Jannedy, 177-228. Cambridge, Mass.: MIT Press.

Rice, Geoffrey W. (ed.). 1992. *The Oxford history of New Zealand,* 2nd ed. Oxford: Oxford University Press.

Simpson, Tony. 1997. *The immigrants: the great migration from Britain to New Zealand, 1830-1890*. Auckland: Godwit Publishing.

Sinclair, Keith (ed.). 1997. *The Oxford illustrated history of New Zealand*. Oxford: Oxford University Press.

Tomasello, Michael. 2003. *Constructing a language: a usage-based theory of language acquisition*. Cambridge, Mass.: Harvard University Press.

Trudgill, Peter. 1986. *Dialects in contact*. Oxford: Basil Blackwell.

Trudgill, Peter. 2004. *New-dialect formation: the inevitability of colonial Englishes*. Edinburgh: Edinburgh University Press.

Watts, Duncan J. 1999. *Small worlds: the dynamics of networks between order and randomness*. Princeton: Princeton University Press.