

anticipatory/regressive (with the affected sound coming first and the conditioning sound later), or progressive/perseverative, where the conditioning sound precedes the affected one (1b). Finally, the affected and conditioning sounds in distant assimilations are separated by other segments (see (1c)). Note that the arrowhead > indicates that one form becomes another, while the asterisk marks a reconstructed form.

- (1) Assimilation
- partial: OE *efn* 'even' > West-Saxon *em*
complete: Latin *septem* 'seven' > Italian *sette*
 - anticipatory/regressive: *septem* > *sette*
perseverative/progressive: Proto-Germanic **wulno* > OE *wull* 'wool'
 - distant: **penk^we* > **k^wenk^we* > Latin *quinque* 'five'

Unlike assimilation, dissimilation tends to be sporadic and occurs only in isolated words, although it seems to happen more frequently with liquids, the cover-term for /r/ and /l/ sounds (2).

- (2) Dissimilation
- | | | |
|------------------------------------|------------|---|
| Latin | Old French | English loan |
| <i>peregrinus</i> > <i>pelerin</i> | | <i>pilgrim</i> |
| <i>purpura</i> > <i>purpre</i> | | <i>purple</i> > <i>purpel</i> > <i>purple</i> |

Segments may also be inserted, a process known as epenthesis, or lost. Epenthetic vowels often break up 'difficult' consonant clusters, while epenthetic consonants frequently arise due to mistiming (3). For instance, in the words *hamster* and *prince*, a nasal is followed immediately by a voiceless fricative, meaning that the vocal folds must be moved apart, the soft palate raised, and the complete oral closure relaxed simultaneously. However, the first two processes may precede the last, leading to an epenthetic voiceless stop.

- (3) Epenthesis
- Latin *schola* 'school' > Old French *escole* (> French *école*)
 - hamster* [hæmpstə(ɪ)], *prince* [prints]

The terms for loss of segments depend on the unit being lost and the position it vanishes from: apocope is loss of a final vowel (4a), syncope affects medial vowels (4b), and haplogy is the loss of a whole syllable from a sequence of similar syllables (4c).

- (4) Loss
- apocope: Middle English [nɑ:mə] > Modern English *name*
 - syncope: OE *munecas* > ModE *monks*
 - haplogy: Old Latin **stipi-pendium* > Latin *stipendium* 'tribute, soldiers' pay'
OE *Engla-lond* > ModE *England*

The types of loss shown in (4) tend to be sporadic, but loss may also be more regular. Regular cases involve the progress of a segment through a

2 Three views of sound change

2.1 Introduction: types of sound change

In this chapter we shall investigate change in phonetics and phonology, primarily by outlining the approaches of three schools of linguists, in chronological order the Neogrammarians, Structuralists and early Generativists. This provides an opportunity to explore the development of sound change theory and introduce schools of linguistic thought which will be mentioned intermittently throughout the book. First, however, it may be useful to outline some of the types of sound change one might expect to encounter in any language.

Many sound changes affect single sounds, or vowel or consonant segments, and we shall concentrate on these below. However, some sound changes involve larger units, such as clusters of consonants, or diphthongs; and supra-segmentals such as rhythm, stress and intonation may also change. For instance, students of Germanic languages propose an early Germanic Accent Shift, which in general placed the main word stress on the first syllable of the stem; and English and German intonation patterns differ, although these languages are descended from a common source, so that we must assume changes in the intonational system of one or both daughters. It is certainly harder to study changes in suprasegmental features, since stress and intonation are rarely recorded in writing; but they do nonetheless occur.

Segmental sound changes can be conditioned, occurring in particular, specifiable environments, or unconditioned, applying to all occurrences of a particular sound. They may also be regular, or irregular and sporadic. One typically regular, conditioned type of sound change is assimilation, which involves one sound becoming more like another in its environment. Assimilation may be partial, in which case two sounds come to share certain features, or complete, producing two identical segments. Both possibilities are illustrated in (1a); in the West-Saxon dialect of Old English, the [v] of *efn* (or *efen*) [ev(ə)n] becomes nasal [m] before nasal [n], while the sequence [pt] becomes [t] in Italian *sette*. Assimilations may be

sequence of weakening changes, as shown in (5). The so-called sonority hierarchy, which runs from the (strongest) optimal consonants to the (weakest) optimal vowels, appears in (5a); weakening can be analysed as movement along this hierarchy in the direction of the arrow, generally through the stages in (5b).



b. weakening: voicing > fricative / glide > vowel > zero

Finally, segments may be reordered, often by metathesis of adjacent segments, which is again sporadic and again tends to involve liquids (6). It is also possible to reorder segments in different syllables or words, producing Spoonerisms like *You have tasted the whole worm* (for *wasted the whole term*), but these are highly irregular.

- (6) Metathesis
 OE *ācsian* > *ask*
 OE *brid* > *bird*

The majority of the changes we have considered so far are conditioned, taking place in a restricted set of environments, and some seem to result from the actions of speakers, while others are arguably best explained with reference to hearers. For instance, epenthetic vowels may break up an articulatorily taxing sequence of consonants, while assimilation (see 2.2 below) is often explained in terms of ease of articulation, the theory that speakers prefer sequences of similar sounds, which are thought to be easier to pronounce than sequences of very different ones. Conversely, Ohala (1981, 1987) argues that listeners are responsible for dissimilation. Listeners learn by experience that speakers may distort certain clusters of sounds by assimilation, and work out corrective rules to help disentangle the intended pronunciation. However, invoking these corrective rules inappropriately leads to dissimilation.

For example (Ohala 1981: 188), Proto-Bantu *-bua 'dog' became Pre-Shona *bwa and Shona [bʷa], where the Pre-Shona labio-velar *w has become velar [v] after a labial. By this dissimilatory change, a sequence of labial plus labial is replaced by labial plus velar. Ohala argues that listeners must have assumed that the preceding labial had distorted the next sound, and compensated by removing the labial component, leaving a plain velar, which they would then produce when acting as speakers. In other words, listeners are suspicious of sequences of similar sounds, and tend to alter one to undo the supposed assimilation; but sometimes their suspicion is

misplaced and by trying to undo a non-existent assimilation they cause dissimilation instead.

Ohala's claims extend to distant dissimilation. For example, Grassmann's Law in Sanskrit de-aspirates the first of two aspirated sounds in a word. Ohala argues that aspiration spread from the two aspirated sounds to intermediate segments, but listeners reinterpreted this spread as emanating from a single, final aspirate, de-aspirating the rest of the word.

Ohala's work predicts that, in dissimilation, the first of two similar sounds will generally be the one affected, because assimilation is predominantly anticipatory; and that in cases of distant assimilation, the features affected will be those (like retroflexion, aspiration, and place of articulation features) which tend to spread over adjacent segments by assimilation. Both predictions seem broadly to be borne out.

I have emphasised this account of dissimilation to provide a counter-balance to the fact that 'in theories on the origin of sound change the speaker has usually been assigned the leading role: the speaker is claimed to have modified his pronunciation in order to reduce the energy expended in speaking, to have made his speech more distinct in order to make it more intelligible, etc.' (Ohala 1981: 178). As we have seen, certain types of sound change may prove more amenable to explanation from the speaker's point of view than the hearer's, while the opposite is true of others. However, since we all act as both speakers and hearers in normal conversation, the interplay of production and perception should not be neglected in our attempts to explain language change. We shall return to this periodically below, notably in the account of abduction in Chapter 4.

As well as the conditioned changes discussed so far, there are also unconditioned changes, the best-known being sound shifts, large-scale changes sometimes involving large numbers of sounds. In the next two sections, we shall encounter two famous sound shifts: Grimm's Law, which provides us with a good example of Neogrammarian theory; and the English Great Vowel Shift, which usefully illustrates Structuralist ideas.

2.2 The Neogrammarians

There must ... exist a rule for the irregularities; the task is to find this rule
 (Verner 1978: 36)

2.2.1 Introduction

The Neogrammarians (*Junggrammatiker* or 'young grammarians' in German) were a group of scholars including Paul, Brugmann and Osthoff working mostly on Indo-European languages in and around Leipzig in the last quarter of the nineteenth century. By this time, attempts had been

made at reconstructing Proto-Indo-European (PIE), and some changes affecting IE languages or subfamilies had been described. A beginning had also been made in the recognition of regularities in change. For example, the First Germanic Consonant Shift, which changed reconstructed Proto-Indo-European voiceless stops into Germanic voiceless fricatives, PIE voiced stops into voiceless stops, and PIE voiced aspirated stops into plain voiced ones, was first described by Rasmus Rask in 1814 (although we now know this sound shift as Grimm's Law, after Jacob Grimm who effectively publicised it). Similarly, Bopp seems to have been the first to use the term *Lautgesetz* 'sound law', which later became a mainstay of the Neogrammarian approach, in the 1820s. In work of this kind, we see clear forerunners of Neogrammarian ideas.

However, most discussion of sound change to date had been unsystematic and atomistic, consisting of a mere cataloguing of changes, or the rather directionless pursuit of individual forms down the branches of the family tree. Only rudimentary explanations, involving the Biblical Tower of Babel story, or the alleged effect of climate, diet or race on language, had been offered – for instance, frication of stops might result from speakers moving into mountainous regions, where the thin air made it harder to catch one's breath and the exertion of running up and down mountains promoted heavy breathing (Meyer 1901). Similarly, Grimm connects Grimm's Law with 'the Germans' mighty progress and struggle for freedom which inaugurated the Middle Ages and was to lead to the transformation of Europe' (1848: 417). 'Does there not', he asks us, 'lie a certain courage and pride in the strengthening of voiced stop into voiceless stop and voiceless stop into fricative?' (1848: 437; translation from Sampson 1980: 30).

Other pre-Neogrammarian historical linguists equated linguistic change with decay: language was seen as an organism, which is born, matures, then grows old and dies; and heavily inflected languages like Latin were considered to be highly developed, whereas later stages of Romance, like French, with its reduced inflectional system, were regarded as degraded and debased (see Chapter 12). The Neogrammarians, however, began to see that languages were not decaying at all. It is true that some languages sometimes lose elements of their morphological systems, as many of the inflectional endings have been lost from English and French. But this does not mean that they lose the capacity to express categories like tense; they simply evolve alternative strategies, like the use of auxiliary verbs. Furthermore, morphology may be gained as well as lost; so for instance in French we now have the productive adverbial ending *-ment*, from Latin *mens*, *mentis* 'mind' – forms like Old French *devotement* 'devotedly' therefore started out as Latin *devota mente* 'with a devoted mind'. Modern

stages of languages are in fact just as useful as tools of communication as earlier ones, and no more debased than their ancestors. Instead of seeing changes as slovenly habits destroying the rules set down in some apocryphal Golden Age, the Neogrammarians attempted to rigorously define, describe and explain them.

2.2.2 The regularity hypothesis

In the late nineteenth century, science was becoming a legitimate pursuit and rather an exciting one, with the emergence, for instance, of Darwin's theory of evolution. The world was increasingly regarded as an orderly and law-governed place, and the Neogrammarians tried to show that this regularity extended to language by formulating principles and methodological assumptions about language change, and by studying sound change scientifically.

The Neogrammarians believed that only historical linguistics could be truly scientific, an attitude unfamiliar to late twentieth century linguists brought up to regard synchronic linguistics as primary. The Neogrammarians' view of the primacy of diachrony arose from their two main concerns. First, they were interested in explaining the regular, common similarities that exist among related languages. This concern encouraged the development of reconstruction, as divergent modern forms were traced back to a single form in the proto-language. Neogrammarian methodology involved first finding cognate words sharing similar sounds and meanings, like German *zu*, *zehn* and English *to*, *ten*, or Modern English *house*, *mouse* and Old English *hus*, *mus*. Next, correspondences of sounds such as those between English initial [t] and German [ts], or Modern English [aʊ] and Old English [u:], were analysed. The differences between these sounds were then ascribed to the operation of sound change.

Secondly, the Neogrammarians were worried about synchronic irregularities in language. For instance, the vast majority of Modern English nouns form their plural by affixing *-s*, but *foot* is irregular in having the plural *feet* rather than ***foots* (note that * marks reconstructions, while ** indicates ill-formed, non-occurring forms). Labelling *feet* as a synchronic irregularity indicates that it is inexplicable given our knowledge about the present-day language. However, the Neogrammarians argued that such forms can be accounted for neatly if we assume that they were once quite regular, but that sound change(s) then operated and caused the apparent irregularity.

(7)	pre-Old English i-Mutation	*fōt	*fōti
	Old English	*fōt	*fōti
		fōt	fēt

- (8) a. OE *fōt* – *fēt* *bōc* – *bēc*
 b. *stone* – *stones* = *book* – *X*
 X = *books*

Analogy makes some irregular form conform to a regular pattern, here illustrated by *stones*, which takes the weak *-s* suffix. This interaction between sound change and analogy is encapsulated in Sturtevant's Paradox: sound change is regular but creates irregularity, whereas analogy is irregular but creates regularity.

Finally, the Neogrammarians used the term sound change only for regular changes like assimilation and weakening. Changes like metathesis, haplogy and dissimilation tend to violate the regularity hypothesis, since they are sporadic, applying in some words but not others with the same context, and are also observable: an interchange of two sounds can hardly happen gradually and imperceptibly.

2.2.3 The motivation for sound change

It is often said that the Neogrammarians saw all regular sound change as mechanical and physiologically motivated: for instance, they frequently invoked explanations like ease of articulation, explaining the assimilation from Latin *septem* to Italian *sette* by the assertion that a sequence of [t] plus [t] is easier to pronounce than a sequence [p] plus [t], with articulatory movements between the segments. However, the Neogrammarians also made use of some rudimentary psychology.

The Neogrammarian who speculates most on psychological causes of change is Hermann Paul (1978). His ideas make crucial reference to variation in a speech community, a notion clearly ahead of his time (see Chapter 9). Paul assumes that we produce sounds using muscles and motor nerves, and that this produces a motory sensation. The sensation is physical, but after the muscle activity dies away, a residual mental sound picture allows us to repeat the same sound again by matching the motory sensation with the sound picture. Normal speech involves a fair amount of deviation from the norm – as Paul says, 'even the most practised marksman misses his mark sometimes' (1978: 8) – but these deviations are generally too minute to hear, and will in any case proceed in both directions from the norm, ultimately cancelling one another out.

Occasionally, perhaps for reasons of convenience or ease of articulation, an imbalance will develop. If this happens for one speaker, it will be noticed and stopped, since Paul assumes communication will be impaired if some speakers are out of step with the rest of the speech community. But sometimes there will be a consensus, with all speakers producing a minute shift away from the target. If this imbalance outweighs deviations to the

20 Three views of sound change

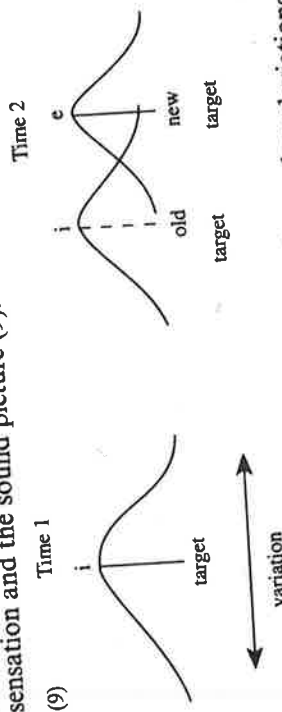
Great Vowel Shift	[fɪ:t]	[fɪ:t]
Modern English	[fot]	[fɪ:t]

The development of *foot* – *feet* from pre-Old English is shown in (7). At this very early stage, we postulate a suffix on the plural form; the suffix vowel, high front [i], triggered a sound change called i-Mutation, a distant assimilation fronting the preceding stem vowel. The suffix was lost by Old English; the stem vowel remained front but had also unrounded, giving us English; the Modern English forms, with a back vowel in the present tense form, a front vowel in the past, and no suffix on either. The Great Vowel Shift, which we shall meet again in the next section, raised both stem vowels, and [u:] later shortened to [o] in *foot*. Regular Modern English plurals, which lacked the high front vowel suffix, have a different history.

The main tool the Neogrammarians used in their account of cross-linguistic similarities and synchronic irregularities is the regularity hypothesis, a strong claim which in its most extreme form asserts that sound change is regular and exceptionless. Sound change was also thought to simultaneously affect all words with the same context in the language in question, and all speakers in the same speech community. The Neogrammarians also claimed that sound changes were purely phonetically conditioned, and could not refer to non-phonetic factors, such as the morphology, syntax and semantics. Sound change was therefore seen as operating with 'blind necessity', without concern for the grammatical consequences of its actions: if a sound change deleted a suffix distinguishing past from present tense forms, this was unavoidable and at best could be resolved later by alterations elsewhere in the language. The further claim that sound change proceeds by tiny, unobservable increments is probably related to this, since it was thought that if speakers knew a sound change was in operation and might have undesirable consequences for the grammar, they would try to stop it.

Although speakers do not generally seem to stop sound changes in progress, there is a 'housekeeping device' in the grammar which sometimes steps in after a sound change has occurred: this is analogy (see Chapter 4). The Neogrammarians saw sound change and analogy as opposing but interdependent processes. Since sound change operates regardless of the consequences for grammar or meaning, phonological and grammatical structure can get out of line; analogy is the process which effects realignment. However, analogy is sporadic; it clears up after sound change, rectifying mismatches of sound and grammar, but it does not apply in all such cases. Thus, only one of the irregular Old English nouns shown in (8a), *foot* – *feet*, is still irregular in Modern English. The other, if it had followed the same course, would now be *book* – *beek* or *book* – *beech*, but is in fact *book* – *books* due to the operation of analogy (8b).

other side, the target itself will gradually shift, altering both the motory sensation and the sound picture (9).



Unfortunately, Paul fails to explain why random deviations should ever become cumulative and directional. In addition, this mechanism is not intended to cover sporadic changes like metathesis and dissimilation; instead, Paul suggests that these are akin to speech errors and arise due to repeated mispronunciations which proliferate and ultimately become the norm. Thus, OE *ācsian* was presumably mispronounced as *ācsian* so frequently between OE and ME that children learning the language heard the 'error' more frequently and reinterpreted it as the target – hence the modern form *ask*.

2.2.4 Grimm's Law and the regularity hypothesis

The most influential contribution of the Neogrammarians is arguably not any single explanation of sound change, but rather the regularity hypothesis itself. This has frequently been criticised on the grounds that absolute regularity is a myth, but the critics betray a failure to read the small print, since the Neogrammarians did not claim any such thing. In fact, the Neogrammarian notion of regularity is quite restricted. First, it excludes sporadic changes. Secondly, it is restricted to a particular speech community at a particular time; the Neogrammarians did not claim that the same sound in the same context in different languages or communities would always undergo the same change. Even with these exclusions, however, problems did arise.

The Neogrammarians seem to have seen their regularity hypothesis as an equivalent of one of the physical laws like the law of gravity, and therefore as an immutable consequence of the way the world is. Pursuing the same analogy briefly, imagine Sir Isaac Newton throwing an apple into the air and catching it, in the interests of testing his theory, until on one occasion the apple fails to come down. Sir Isaac looks around for a reason, sees that the apple is stuck in a tree, or has been carried off by a passing seagull, and the exception is explained. It is this kind of theorising which marks the greatest achievement of the Neogrammarians. Not every sound

change they formulated turned out to be exceptionless. Instead, they tackled apparent exceptions in two ways: they assumed that the change had been inadequately formulated, and altered it; or they proposed an additional change. Thus, by minutely examining the data, they accounted for the counter-examples.

Let us now look at an extended example of this reasoning involving the First Germanic Consonant Shift, which appears in an informal and slightly simplified version in (10). The change is an unconditioned sound shift, affecting *all* occurrences of these sounds, not only those in specific contexts.

- (10) Grimm's Law
- a. voiceless stop > voiceless fricative
PIE *p t k > Gmc. /f θ x/
 - b. voiced stop > voiceless stop
PIE *b d g > Gmc. /p t k/
 - c. voiced aspirated stop > voiced stop
PIE *bh, dh, gh > Gmc. /b d g/

There are two main sets of exceptions to Grimm's Law as it stands in (10). First, as shown in (11), there is a set of items in which voiceless stops do not shift to fricatives.

(11)	Latin	Gothic	Old English
	<i>captus</i>	<i>hafis</i>	<i>hæft</i>
	<i>piscis</i>	<i>fisks</i>	<i>fisc</i>
	<i>squā</i>	<i>spreiwan</i>	<i>spreiwan</i>
			'prisoner'
			'fish'
			'spit'

The words themselves are not exceptional, since the initial /p/ of *piscis*, for instance, has quite regularly become a fricative. On careful examination, however, it becomes clear that Grimm's Law simply has not been formulated carefully enough, and is missing the generalisation that voiceless stops *never* become fricatives when they follow another stop or a fricative. We do not, in fact, have exceptions, but rather a subregularity; (10a) can therefore be reformulated as in (12), reworking a basically sound idea to iron out the residual problems.

- (12) voiceless stop > voiceless fricative

UNLESS directly following a stop or fricative

The second problem is rather more complex: although Grimm's Law predicts that PIE voiceless stops should become Germanic voiceless fricatives, they sometimes appear as Germanic voiced stops or voiced fricatives. PIE *bhrātēr- 'brother' becomes Gothic *brōþar*, with medial *t > /θ/, but PIE *pāter- 'father' becomes Gothic *fadar* with a medial voiced /d/. The context causing this aberrant change is extremely hard to find, making the voicing development look truly irregular. However, Karl

24 Three views of sound change

Verner noted a connection with the position of the accent in PIE: voicing only occurred if the PIE voiceless stop occurred between voiced segments, and the previous syllable was unaccented; otherwise, the regular Grimm's Law change to voiceless fricatives took place.

One reaction to these discoveries might be to alter Grimm's Law further. Verner, however, considered this the wrong course to take, since the voicing he noted affected also the single fricative, *s, which existed in PIE before the operation of Grimm's Law. He therefore formulated an additional change, Verner's Law (13).

- (13) Verner's Law
voiceless stops between vowels, when the preceding vowel is unaccented > voiced stops or fricatives

We can also work out the relative chronology of Grimm's and Verner's Law: Verner's Law must have operated after Grimm's Law historically, since it voiced the new fricatives /f θ x/ created by Grimm's Law as well as the pre-existing *s.

The Neogrammarian regularity hypothesis may therefore be generally upheld, although the sound changes first suggested might have to be supplemented or reformulated. This notion of regularity allowed the Neogrammarians to pioneer the explanation of apparent exceptions by rigorous examination of the data. The intention was not to deny the existence of exceptions, but to account for them: 'There must... exist a rule for the irregularities; the task is to find this rule' (Verner 1978: 36).

2.3 The Structuralists

Every modification must be treated as a function of the system of which it is a part
(Jakobson 1978: 103)

2.3.1 Saussure and the basis of Structuralism

The tenets of Structuralism were first laid down by the Swiss linguist Ferdinand de Saussure in the early twentieth century, and subsequently borrowed into other fields, including literary criticism. Saussure was trained in Neogrammarian circles in Leipzig, but he taught mainly in Paris and later in Geneva where, between 1907 and 1913, he gave three courses of lectures on general linguistics. This was a new departure, given the Neogrammarian, historicist domination of contemporary linguistics, and given that Saussure's own published work was philological. After Saussure's death, two of his students, working from the notes of those attending the lectures, constructed the book which we know as the *Course in General Linguistics*, and which outlines Structuralist linguistic thought.

A number of different schools of Structuralist linguists developed during the first half of the twentieth century (Lepschy 1970), notably the Prague School, including Trubetzkoy and Jakobson, and the American Descriptivists, like Bloomfield and Hockett; but these share an adherence to Saussure's basic ideas. An account of Structuralist views of language change is therefore best prefaced by an introduction to Saussure's general theory of language.

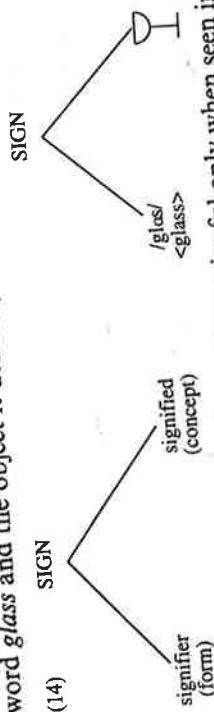
Saussure's thought is often presented in the form of a series of dichotomies. First, he distinguished *langue* from *parole*. *Parole* is the individual, executive side of language, and the actual language, spoken or written, which results; while *langue* is the abstract system of units and rules underlying the surface variation. Saussure assumes that the linguist should be concerned primarily with *langue*. Secondly, Saussure distinguished synchronic from diachronic linguistics; but whereas the Neogrammarians had held that only historical linguistics was truly scientific, Saussure believed synchronic study to be primary, on the grounds that we can only understand language by considering how a particular linguistic system functions at a particular time; diachronic work involves comparing pre-established synchronic systems for two or more stages of a language. Saussure's third dichotomy involves syntagmatic as opposed to associative relations between units. Syntagmatic relations hold between sequentially ordered units, like the three segments in *cat* [kæt], or the words in the sentence *Anna is a teacher*. Associative relationships involve any relationships which speakers may perceive among words which are *not* sequentially arranged. For instance, speakers of French may associate all words ending in *-ment*, while an English speaker might associate the word *wine* with *glass* and *bottle*. In modern linguistics, one class of associative relations is of particular importance: these are now known as paradigmatic relations, and involve a choice between alternative elements at a single structural point. So, [bæt], [pæt], [mæt], [sæt] as opposed to [kæt] involve a substitution of the initial consonant; all these initial consonants are therefore paradigmatically related, in that they can all optionally fill the same structural slot. The same holds of the alternative agent nouns if *teacher* in *Anna is a teacher* is replaced by *carpenter* / *lawyer* / *doctor* / *taxi driver*.

The theory of syntagmatic and paradigmatic relations presupposes that, on each linguistic level, units will contrast or be opposed to one another, but will also combine in linear order to make up higher-order units. We might then expect that language could be analysed on the syntagmatic and paradigmatic dimensions both synchronically and diachronically. The Neogrammarians were particularly concerned with syntagmatic relations; for instance, they interpreted cases of assimilation as one sound influencing

26 Three views of sound change

another, adjacent sound. However, as we shall see, the Structuralists were more interested in paradigmatic relationships.

Perhaps the central element of Saussure's theory is the idea of the linguistic sign. The sign has two parts, as shown in (14) – the signifier, or form, and the signified, or concept. The sign itself, the union of these two parts, is essentially arbitrary; there is no necessary connection between the word *glass* and the object it denotes, and either may change over time.



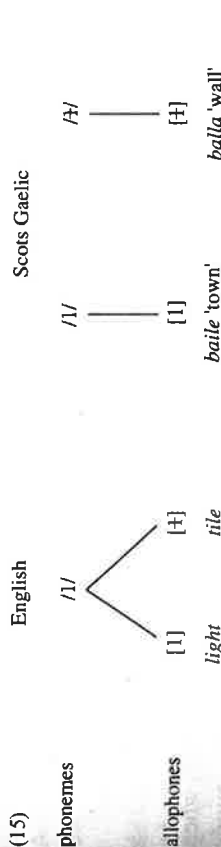
Saussure claims that signs are meaningful only when seen in their place in the system of a certain language at a certain time. Only the system is important, as we can see using the example of colours (Culler 1976). The spectrum is divided differently in different languages; non-native speakers of English, for instance, must therefore be taught the English colour terms. However, they will not learn what *blue* means by being put in a room with one hundred blue objects and told that this is the meaning of *blue*, since they will then be unable to generalise and to classify novel objects as blue or not-blue. This is because *blue* has no meaning in isolation; it is what is not-green, not-purple, not-black, not-white and so on, and such distinctions can only be grasped by considering all the colours together. The crucial thing is not the individual terms, but the way they contrast and interact within the system.

The units of language are likewise contrastive and relational, and can only be understood by considering their place within the language system. This assumption leads to a viewpoint rather different from that of the Neogrammarians, who were still essentially atomistic in their approach, considering individual units and changes. For the Structuralists, a language is crucially, in Meillet's terms, 'un système où tout se tient' – a system with a place for everything, and everything in its place. This must be borne in mind now, as we turn to the Structuralist view of sound change.

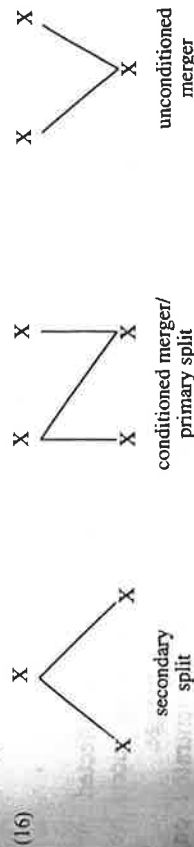
2.3.2 Phonemic change: split and merger

For the Structuralists, the central units of the sound system were phonemes. Like most other Structuralist terms, the phoneme is based on contrast. Sounds are phonemic if they make a meaning difference between words. In English, /p/ and /b/ do contrast, as in *pat* versus *bat*, while the two /l/ sounds, clear [l] in *light* and dark [ɫ] in *hall*, do not. Native speakers are

generally more aware of the first difference than the second, and the sounds which contrast, like /p/ and /b/, are phonemes, while the others, like [l] and [ɫ], which appear in predictable contexts and are variants of a single phoneme (in this case /l/) are called allophones. Each language has its own phonemic system; as (15) shows, Scots Gaelic does have a contrast between clear /l/ and dark /ɫ/.



The Structuralist typology of phonemic change (Hoenigswald 1960) crucially involves split and merger: split involves one phoneme becoming more than one, while in merger, two or more phonemes collapse into one (see (16)).

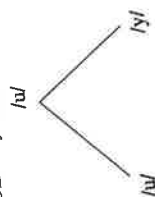


Unconditioned merger happens when the realisations of two phonemes fall together in all contexts, and is irreversible; a phoneme resulting from merger behaves exactly like a phoneme with a single source. In Sanskrit, for instance, PIE *a and *e merge to give /a/ in all environments; [e] becomes [a], and /e/ is lost. There is also conditioned merger, which is associated with primary split. Here, one phoneme develops a new allophone in a certain environment; this matches the realisation of some other phoneme, with which it merges, splitting from the phoneme to which it originally belonged. For instance, in Latin /s/ > [ʃ] between vowels; this [ʃ] sounded, we can assume, like [ʃ] from the phoneme /r/, and was reinterpreted as a realisation of /r/ rather than of /s/ (some occurrences of [ʃ] between vowels in later Latin, as in *ambrosia*, are the result of a subsequent change of /t/ > [ʃ]; other cases appear in loanwords, for instance). Conditioned merger with primary split, unlike unconditioned merger, does not involve the loss of a phoneme; the phoneme system remains unchanged, but the distribution of its members alters. Secondary split, on the other hand, introduces a new phoneme into the system: here, a phoneme again develops a new allophone, but this does not correspond

Three views of sound change

28 The sound will remain allophonic to an allophone of any other phoneme. The conditioning context which caused the original change is lost, in which case it becomes phonemic by primary split. An example is given in (17).

- (17) Pre-OE: i-Mutation /u/ > [y] / -i j/
 *trumian > *trymian 'strengthen'
 *tryman > OE *trymanian*
 /u/



The pre-Old English sound change of i-Mutation fronted vowels when followed by high front /i/ or /j/. Later, however, the high front sounds conditioning the change were lost, making it impossible to predict where [y] might occur rather than [u]; earlier /u/ then split to /u/ versus /y/ in Old English.

2.3.3 Structure and function

The Structuralists, then, were interested in change as it affected the phoneme system, and studied such change by comparing successive systems. However, they also attempted to explain sound change.

As we saw in 2.2 above, the Neogrammarians concentrated on conditioned sound changes like assimilation, where one segment affects another to which it is related syntagmatically. Their explanations of ease of articulation or convenience might hold for these changes, but were unsuitable for unconditioned sound changes, or sound shifts, like Grimm's Law: the Neogrammarians formulated laws to say *what* happened, but were unable to say *why* the sounds should have followed their rather circular course. Why, for instance, would it be 'easier' to substitute voiceless fricatives for voiceless stops, and voiceless stops for voiced ones, than to maintain the status quo?

The Structuralists attempted to explain sound shifts by invoking the structure of systems and the function of language; and we shall look first at structure.

One important aspect of structure is symmetry; gaps in sound systems seem to be disfavoured. For instance, the pre-eighteenth century English fricative system appears in (18); there are three voiced - voiceless pairs, but two voiceless fricatives without voiced partners.

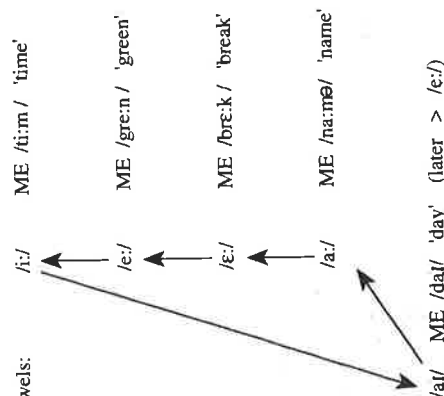
- (18) f θ s f h
 v ð z

A new voiced partner, /ʒ/, has now evolved for /ʃ/. /ʒ/ arose from the combination of [z] + [j] in words like *treasure*, *pleasure*, and was also borrowed from French in loans like *beige*, *azure*, *rouge*. /h/, however, is not gaining a voiced equivalent, but may be disappearing from the language. Many English accents now have /h/-dropping, making *ham* and *am*, or *hit* and *it* homophonous; a corroborative story tells of an address dictated by a Cockney in which Harwich Harbour is heard and written as Arijaba. The social pressure maintaining /h/ may not hold out forever, so that the English fricative system may be gap-free relatively soon.

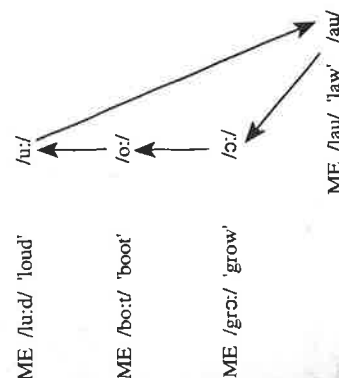
The idea of symmetry may also explain some unconditioned sound shifts, like Grimm's Law or the Middle English Great Vowel Shift, in which sounds (this time long vowels) again seem to move round in a circle. A rather simplified version of the Great Vowel Shift appears in (19).

(19) The Great Vowel Shift

a. Front vowels:



b. Back vowels:



30 Three views of sound change

It is possible to see this shift as a drag chain, which involves one phoneme shifting, and leaving a space into which others will be pulled. In the case of the Great Vowel Shift, this analysis assumes that the high vowels diphthongised for some reason, and lower vowels were progressively raised to fill the abhorrent gaps: the diphthongising high vowels drag the rest of the long vowel system along behind them.

However, an alternative view might invoke function rather than structure. Martinet (1952, 1955), for instance, notes that the primary function of language is communication, and that change must not be allowed to jeopardise communication needs.

One change which might seriously disrupt communication is phoneme merger. If two phonemes fall together, an inevitable consequence is homophony; words with different meanings will begin to sound the same, and this might cause communicative difficulties if widespread. Martinet here refers to the notion of functional load, or the number of minimal pairs which exists for two phonemes. The opposition between /t/ and /d/, for instance, distinguishes a large number of pairs of words in English (including *tip* – *dip*, *at* – *add*, *hat* – *had*, *tear* – *deer*, *try* – *dry*, *writer* – *rider*), while the contrast between /n/ and /ŋ/ does relatively little work. The first opposition therefore has a high, and the second a low functional load, and pairs of phonemes with low functional load are more likely to merge, since this would cause less communicative damage. This conclusion, however, depends crucially on the structure of the language concerned; so, in English, although the contrast of /θ/ and /ð/ has a very low functional load, /θ/ and /ð/ are relatively unlikely to merge since they are distinguished by only one feature, voicing, which does a good deal of work elsewhere in the consonant system; nothing would therefore be gained in terms of economy if the /θ/ – /ð/ contrast were lost.

This notion of economy is also functionally relevant; sound systems tend toward economy as well as symmetry. Systems are most economical when they derive the maximal number of contrasts from the minimal number of features. Furthermore, systems with a wide margin of safety between sounds seem to be preferred, since well-defined distinctions can be more readily perceived. This might explain the frequent occurrence of triangular vowel systems like those in (20), where vowels are maximally distinct acoustically and articulatorily.

- (20) $\begin{matrix} & i & u & & i & u \\ & & & & e & o \\ & & & & & a \end{matrix}$

Martinet's attempts to explain sound change in terms of maximum differentiation of phonemes, margin of safety, the best use of phonetic space and the avoidance of wholesale mergers, might provide a second

possible explanation for the Great Vowel Shift. Instead of a drag chain, we might propose a push chain, where one phoneme shifts and encroaches on the territory of some adjacent phoneme; this moves in its turn, again setting off a chain reaction. In the present case, we assume that one of the lower vowels was gradually raised; to avoid a merger, the next highest vowel shifted in the same direction, and the process was repeated until the original system of oppositions was restored, albeit with different vowels in different sets of lexical items.

It is certainly unwise to see the notions of structure and function as mutually exclusive, and indeed the preferable account of the Great Vowel Shift may involve both drag and push mechanisms, as data from Scottish English and other northern varieties suggest. The Middle English ancestors of these northern dialects did not have quite the same system as the southern dialects where the Great Vowel Shift began. Instead, they had the system in (21).

- (21) $\begin{matrix} & i: & & u: & & \\ & e: & \emptyset: & & & \\ & \varepsilon: & & \text{ɔ:} & & \\ & a: & & & & \text{au} \end{matrix}$

The main difference is that /o:/ had fronted to /ø:/ in the north. In these northern areas, /u:/ also failed to participate in the Great Vowel Shift, so that modern Scots dialects still have /hus/, /mus/ rather than diphthongised /haus/, /maus/. /u:/ may have failed to diphthongise in these varieties because there was no /o:/ to push it, suggesting that this 'top' part of the Great Vowel Shift was a push chain. However, the other vowels did shift, indicating that the vowel to start the whole process may have been /ɔ:/ (and similarly front /ε:/), which moved up towards /o:/ where such a vowel existed, and dragged the other vowels along behind.

Structuralist notions of structure and function with respect to particular linguistic systems can therefore help us understand sound shifts as well as the smaller-scale conditioned changes discussed by the Neogrammarians. In addition, while the Neogrammarians considered each sound change as an independent process, operating blindly and without concern for the consequences, the Structuralists believed that one change might disturb the equilibrium of a system, provoking further changes until equilibrium was restored. Structuralist theory also permits a partial integration of synchrony and diachrony, since changes are seen as a function of the synchronic system in which they occur, and any laws governing the structure of synchronic systems will also determine and constrain change. These notions of the interaction and interdependence of synchronic principles and diachronic developments are crucial to much modern historical linguistics.

32 Three views of sound change

Although we have concentrated in this section on language-internal motivations for change, the Structuralists were very much aware that external considerations could also influence the onset and direction of linguistic changes. For example, Martinet notes that gaps in a system may remain empty for lengthy periods, and may be filled at a specific time due to external factors such as language contact. Similarly, Meillet accepts that some similarities between related languages appear only after divergence from the common ancestor, so that resemblances among languages may reflect parallel development rather than genetic relationship alone. Finally, Jakobson (in Keiler 1972) is at pains to point out that borrowing and linguistic convergence may have profound effects on language structure. We shall return to matters of language contact in Chapters 7-11, but should be aware that the Structuralists invoked external as well as internal causation in their search for explanations of change.

There are, however, two problems for Structuralist accounts of sound change. First, it is unclear how change is ever to happen, if every element in the system is dependent on every other element; and second, if units have no meaning in isolation, but only gain their significance from their place in the system of a certain language at a certain time, how can we compare different languages or different stages of the same language? Presumably comparison presupposes complete structural equivalence, which is extremely unlikely to exist; to pursue our analogy of colours, *blue* in English cannot, in strict Structuralist terms, be compared with the word for 'blue' in Russian or Welsh, since these cover slightly different areas of the spectrum, and are opposed to different sets of terms.

2.4 The Generativists

(Postal 1968: 270)

What really changes is not sounds, but grammars

2.4.1 *Generative Grammar*

In the last two sections, we have focussed on schools of linguistic thought which essentially constitute closed chapters in the history of the discipline. Diachronic (and synchronic) linguists have learned, and still can learn much from the approaches of the Neogrammarians and Structuralists; and indeed, the Generativists themselves might be described as Structuralist in their concentration on linguistic systems. However, there are now very few linguists working entirely within the Neogrammarian or Structuralist paradigms.

The Generative school, however, has not so far been superseded by a new, dominant paradigm, so that numerous linguists would still describe themselves as Generativists. In the last quarter-century, Generative syntax

and phonology have both fragmented into a number of related but distinct synchronic and diachronic models. Since not all of these are yet fully developed, an up-to-date survey would necessarily be incomplete and also very complex theoretically. Consequently, this section will focus on early Generative work in the so-called Standard Model, as practised primarily in the 1960s. This has the advantage of allowing a relatively finished picture to be drawn, since Generativists today only very rarely adhere to the Standard Model. The picture might be seen as a rather old-fashioned one, but only if it is taken as an attempt to portray the current state of Generative linguistics, rather than a study of the beginnings of Generative theory included to maintain the historical perspective of this chapter. More recent aspects of Generative theory will be explored briefly in Chapter 3 (for phonological change) and Chapter 5 (for historical syntax).

Generative Grammar, with Noam Chomsky as its leading figure, came to prominence from the late 1950s with works on synchronic syntax (Chomsky 1957, 1965); the Generative approach spread to phonology (Halle 1959, Chomsky and Halle 1968), and later to sound change (King 1969a). One of its primary innovations was the realisation that the class of well-formed or grammatical sentences in any language is potentially infinite; consider the English sentence *I know Sarah and Clive and Aidan and Kathy...*, which obviously could be extended *ad infinitum*. Native speakers are also constantly producing novel sentences which they have not previously uttered. If I say *Alexandra fed forty-two lettuces to the llama*, I can feel reasonably confident that I have not uttered this sentence before; indeed, this may be the first time anyone has produced it. Even so, other native speakers of the same language can readily understand such newly-minted sentences, provided that they are grammatical.

Sentences in a human language will be grammatical if they follow particular patterns; they are not simply random collections of words thrown together in random order. Furthermore, the repeated patterns found in a given language will tend to be rather few, and this insight lies behind the Generative analysis of language and language acquisition. Many earlier studies had assumed that children acquire their native language by listening, memorising and repeating; however, this would mean that children could not move beyond the repertoire of sentences they had already heard, whereas in fact they often produce novel utterances. The Generative approach to this problem crucially involved regarding language as a rule-governed phenomenon: in acquiring a language, children would have to learn a small set of rules, and a larger but still finite set of words, and could then from these finite resources produce or generate a potentially infinitely large set of grammatical sentences.

Acquisition of even these rules or patterns, however, is not strictly a

learning task, given Chomsky's further assumption that all human languages are analysable using the same sorts of rules and grammars. If the same types of structures are common to all languages, Chomsky reasons, then why should they not be already present in the mind at birth? A genetically inherited outline of language universals is consequently assumed to be innate. Children learn their own language by listening to linguistic data and filling language-specific details into this largely pre-set plan. This interaction between the innate Universal Grammar and heard, language-specific data allows the child to build a personal, internalised mental grammar. This internal grammar, which Chomsky calls competence, allows native speakers to understand and produce sentences, and represents their tacit or subconscious knowledge of their native language. The Generative linguist's task is then to build an explicit rule system which does the same job as this internal grammar. The Generativists were consequently concerned with the notion of a linguistic system, but while the Structuralists considered actual linguistic data, the Generativists concentrated on the underlying system of rules; whereas the Structuralists said 'phonemes change', the Generativists said 'rules change'. Indeed, actual language data, which Chomsky calls performance, was of strictly limited interest to the early Generativists, precisely because slips of the tongue, tiredness, lack of concentration and numerous other such factors ensure that many actual utterances do not entirely follow the rules which generate the idealised forms of competence. Chomsky hypothesises that children can abstract away from the limited and imperfect data they receive, in constructing their grammar. However, the primacy of competence is clear from the simplifying assumption that 'Linguistic theory is concerned primarily with an ideal speaker-listener, in a completely homogeneous speech-community, who knows its language perfectly and is unaffected by such grammatically irrelevant conditions as memory limitations, distractions, shifts of attention and interest, and errors... in applying his knowledge of the language in actual performance' (Chomsky 1965: 3). It is easy to see that issues of variation were not addressed in early Generative theory.

Two short illustrative examples might be useful at this point. A Generative syntactician would assume that a native English speaker would regard the sentences in (22) as related.

- (22) *Alexandra fed the llama.*
Did Alexandra feed the llama?
Alexandra did not feed the llama.
The llama was fed by Alexandra.
Was the llama fed by Alexandra?
The llama was not fed by Alexandra.

Early Generative syntax assumed that these sentences share essentially the same, rather abstract deep structure form, which might look something like ALEXANDRA FEED LLAMA, although each would also have markers to specify whether it would end up as a statement, negative or question on the surface. Transformational rules then turned the deep structures into the different surface structures; for instance, the rule for making questions said roughly 'add an appropriate form of the verb *do* at the beginning of the sentence'.

Similarly, in Generative phonology the goal was to assign to each morpheme, each meaningful part of a word, a single unique shape called the underlying form. The surface forms which actually appear in different contexts were then derived from this by rule. For instance, we might assume that native speakers of English know a unit meaning 'negative', with the underlying form /n/. When this prefix is added to different adjectives, the nasal assimilates partly or completely to the first consonant of the adjective stem, as shown in (23). This variation is secondary and rather superficial; what really matters is that all these ostensibly different prefixes can be related, and the variations derived by rule.

- (23) /n/ = negative prefix
 i[n] temperate
 i[m] plausible
 i[ɔ] coherent
 i[r] relevant
 i[l] literate

Such alternations of sounds are very common; another example from English is *electr[i]k* - *electr[i]sity* - *electr[i]cian*, where the stem morpheme would again have a single underlying form, say /elɛktrik/. Generative phonological reference to the morpheme, a non-phonological unit, also illustrates the Generative idea that phonology interacts with other components of the grammar, such as the morphology or syntax, so that sound change may be non-phonetically conditioned.

The Generative linguist, then, aims to write a grammar mirroring the native speaker's competence. However, the number of grammars which could generate the same set of sentences is potentially infinite, and it is unclear which we should prefer. The primary early Generative criterion was simplicity: we should assume that native speakers are creatures whose minds work on a principle of least effort. Their internalised grammars must therefore operate with minimal apparatus and fuss, generating all the necessary data with the minimum complexity; that is, they must be maximally simple. This requirement is a version of Occam's Razor, which states that *entia non sunt multiplicanda praeter necessitatem*, 'entities are not to be multiplied beyond necessity' - or, less formally, don't ask for

Three views of sound change

36

more units than you need. If the linguist's formal grammar is to match the internal one, it must also be maximally simple, and in early Generative terms, simplicity was measured by counting the number of rules and units in the grammar, with lower numbers preferred (this, of course, is itself a slightly simplified outline).

2.4.2 The Generative theory of sound change

2.4.2.1 Introduction

In Generative theory, then, all sound change was seen as change in the grammar. Change between two related languages or varieties was established by constructing and comparing systems of rules and underlying forms for each stage or language. Logically, phonological change could only occur in the form, order or inventory of rules, or in the underlying representations.

The early Generative belief in evaluation procedures based on measurements of simplicity also led to the belief that all change must be simplifactory, translating more complex to simpler and more economical grammars. The Generativists therefore denied the functional motivation advocated by the Structuralists. For instance, King (1967) measured the functional load of various oppositions of sounds (that is, roughly, the number of pairs of words distinguished by the difference between these sounds). He used data from a number of Germanic languages, concentrating on sounds which are known to have subsequently undergone sound change, and his results contradict in part the predictions of the functional load hypothesis. For instance, /y/ has been lost by merger in Icelandic. There are two other Icelandic vowel phonemes close to /y/ in terms of their component features, with which /y/-/i/ had a functional load more than four times higher than that of /y/-/u/, but /y/ nevertheless merged with /i/.

In the next two subsections, we shall consider the mechanisms of phonological change proposed by the Generativists: first, the different types of rule change; then, restructuring of the underlying forms.

2.4.2.2 Rule change

2.4.2.2.1 Rule addition

Rule addition, which is also simply called innovation, was the most basic type of change recognised by the Generativists, and also the only sort of change which could affect the grammar of adult speakers; all the other changes to be discussed here take place between generations, as children acquiring language internalise a slightly different grammar from their

parents'. In rule addition, the speaker adds a new rule to the end of his phonological rule system; this constitutes a response to the introduction of a sound change, and the change and the resulting phonological rule will generally be identical. Rule addition only changes the synchronic rule system, not the underlying forms of morphemes; any change at the underlying level will only take place later, if at all.

An example of an added rule is given in (24).

$$(24) \quad \left[\begin{array}{l} + \text{ obstruent} \\ - \text{ continuant} \\ + \text{ voice} \end{array} \right] \longrightarrow [- \text{ voice}]$$

This rule, although formulated using Generative phonological notation, has identical results to the second part of Grimm's Law, in (10b) above, and states that voiced stops (or voiced, obstruent non-continuants) become voiceless; their value for voicing changes. A Generative statement of the diachronic sound change would differ from the synchronic phonological rule only in the conventional use of an arrow ($\longrightarrow$) in the latter, and a shaftless arrowhead ($>$) in the former.

The consensus view in later Generative work was that such rules would only be added at the end of the phonological rule system. Earlier, a process called rule insertion (King 1969a: 43) had allowed new rules to be added anywhere in the grammar. This was rejected because a rule introduced early in the inventory might operate cumulatively with later phonological rules to produce an output rather remote from the previous pronunciation, potentially impairing communication between generations. However, a rule added at the end of the rule component would alter pronunciations only according to its own effect, allowing communication with speakers as yet lacking the change.

2.4.2.2.2 Rule loss

As well as appearing in a grammar, rules were said to disappear. King's (1969a: 46ff) example of rule loss involves final devoicing in Yiddish.

Yiddish ultimately derives from Middle High German, which underwent a sound change (and therefore added a rule) devoicing final stops and fricatives. This produced the alternations shown in (25b), which were not present in Old High German (25a) when the change had not yet operated. Most Modern German dialects retain this final devoicing rule, but Yiddish does not; as (25c) shows, voiced obstruents may appear finally in Yiddish.

- (25) a. Old High German
gab 'he gave' - gabum 'we gave'
tag 'day' - tage 'days'

- b. Middle High German
gap - *gaben*
tac - *tage*
- c. Standard Yiddish
log 'day' - *leg* 'days'
noz 'nose' - *nezer* 'noses'

We can hypothesise that final devoicing *did* apply in Yiddish, since isolated words like *avek* 'away' and *gelt* 'money' have voiceless final stops in Modern Yiddish, but had voiced ones in Old High German. Some Yiddish dialects also maintain final devoicing. In Generative terms, then, Yiddish *did* have a devoicing rule, but has lost it. Such loss probably should not be interpreted as the instantaneous disappearance of a rule, but rather as a gradual reduction in its scope; the process applies to fewer and fewer forms and ultimately fades out of the system.

2.4.2.2.3 Rule reordering

Two successive stages of a language, or two different dialects, sometimes seem to have the same two phonological rules applying in different orders. For instance, after German final devoicing was introduced, around AD 1000, another change lengthened vowels before voiced obstruents; so, in the fifteenth and sixteenth centuries, words like *lob* 'praise' and *veg* 'path' surfaced with short vowels in the nominative forms and long ones in the genitive (26a). This distribution is best accounted for if we assume that final devoicing preceded vowel lengthening, and that the synchronic order of rule application matched the chronological order of introduction of the sound changes. In Modern German, however, both nominatives and genitives have long vowels, suggesting that the rules have come to apply in the opposite order.

- (26) a. underlier: lob lobəs veg vegə
 final devoicing: lop — vek —
 V lengthening: — lo:bəs — ve:gə
- b. underlier: lob lobəs veg vegə
 V lengthening: lo:b lo:bəs ve:g ve:gə
 final devoicing: lo:p — ve:k —

One possible reason for such a rule reordering is suggested by Kiparsky (1978), who invokes the notion of rule simplification. Generative rules are regarded as simpler if they apply maximally and can be stated economically: an obvious case of simplification is shown in (27), where the

simplified version has fewer features and applies potentially to more segments. The symbol # indicates a word boundary.

- (27) a. $\left[\begin{array}{l} + \text{obstruent} \\ + \text{continuant} \end{array} \right] \longrightarrow [- \text{voice}] / \text{ — } \#$
 b. $[- \text{obstruent}] \longrightarrow [- \text{voice}] / \text{ — } \#$

Kiparsky argues that rule reordering also depends on the notion of maximal application, the requirement that each rule should do as much work as possible. In Kiparsky's terms, two rules may be ordered in feeding order, or in bleeding order. In feeding order, the first rule provides contexts for the next, as shown in (28a). But in bleeding order, the first rule robs the second of segments to which it could otherwise have applied, as in (28b).

- (28) a. Feeding order:

1. $v \longrightarrow f$

2. $f \longrightarrow \phi$

- b. Bleeding order:

1. $v \longrightarrow f$

2. $v \longrightarrow w$

In (28a), rule 1 provides cases of /f/ for rule 2 to delete, but in (28b), rule 1 shifts all instances of /v/ to /f/, leaving no /v/s for rule 2 to apply to. Kiparsky argues that, to allow maximal application and thus increased simplicity, rules will tend to be reordered either into feeding order, or out of bleeding order. In the German final devoicing case, we see reordering out of bleeding order.

2.4.2.2.4 Rule inversion

The final type of rule change is rule inversion (Vennemann 1972), which involves the reinterpretation of original surface forms as underlying forms. One example involves the treatment of /r/ in certain accents of English, including Southern British English Received Pronunciation (RP). The operation of a number of eighteenth century sound changes in the ancestor of this variety meant that /r/ appeared after a restricted set of vowels, namely /ɔ:/, /ɑ:/, /ə/ and /ɜ:/, which we can regard as a long schwa (/ə/). Some examples are shown in (29).

- (29) /ɜ:/ oar, floor, for, lore, shore ...
 /ɑ:/ star, bazaar, far ...
 /ə/ beer, fear, near, here ...
 /ɜ:/ care, there, air, square ...
 /ɑ:/ - /ɜ:/ assure, pure, lure, poor ...
 /ɜ:/ stir, fir, fur, word, err, heard ...

fixed. Children were also said to be responsible for the second major type of phonological change, restructuring of the underlying representations. King (1969a) provides an extended example of restructuring involving English /hw/ versus /w/. At an earlier stage of English (36a), adult speakers had contrasting /hw/ in *whales*, *which* and /w/ in *Wales*, *witch*, and this situation still persists in certain Modern English dialects, such as Scottish English. However, /hw/ gradually merged with /w/ in varieties like Southern British English, leading to the addition of a rule as in (36b). Adults, who can only add rules, would retain /hw/ in the underlying forms of *whales*, *which*, but cease to produce it.

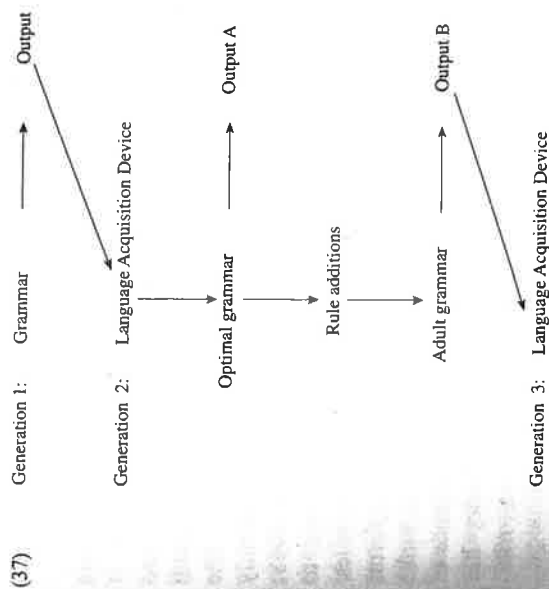
The next generation of children have the target of constructing a maximally simple grammar of their variety of English. Since they will never hear [hw] from their parents, there is no motivation for them to postulate contrastive /hw/ and /w/, plus a merger rule; instead, they will hypothesise underlying /w/ in all previous /w/ and /hw/ words, as in (36c). This restructuring produces a revised, simpler grammar, without the rule merging /hw/ with /w/, but with the effects of that rule transferred into the underlying representations.

- (36) a. Adults: /hw/ [hw] /w/ [w]
 whales *which* *Wales* *witch*
- b. New rule: /hw/ → [w]
 Underlier: /hwɪf/ → [wɪf] *which*
 Surface: [wɪf]
- c. Children: /w/ [w]
 which *witch* *Wales* *whales*

2.4.2.4 Summary

A schematic representation of the Generative theory of sound change is given in (37).

The Generativists



This Generativist model of change raises a number of difficulties. First, early Generative theorists tended not to consider the effects of change on the system of sounds in a language (the phoneme system in Structuralist terms). They consequently omitted from consideration Structuralist explanations of change, which centred on the structure and function of the system. Instead, the Generativists wrote formal phonological rules which reflect completed changes; but these are only restatements of the effects of the change, and are essentially non-explanatory.

Secondly, the notion of simplification, although it might be a candidate for explanation in the case of some changes, cannot be the cure-all the Generativists claimed it to be. Not all changes can be construed as simplificatory; it is hard, for instance, to see what a sound shift like Grimm's Law or the Great Vowel Shift might simplify, and Neogrammarian changes of the irregularity-creating variety certainly introduce complexity into the grammar. Even more basically, how can the addition of a new rule ever be considered a simplification? Furthermore, we have no entirely clear idea of what constitutes simplification, which is a relative rather than an absolute term. In other words, a particular change might be simplificatory under certain circumstances, but induce complexity in others. Finally, the characterisation of all change as simplificatory presumes a view of change as constantly creating ever simpler grammars. Languages, on this view, must have been getting gradually simpler ever since language began, a claim for which there is no evidence at all. Models promoting a single explanation in this way can all too easily be hijacked to promote a view of some languages as simpler or better than others, and of

44 Three views of sound change

change as directed, cumulative and purposive, a view which we shall consider and reject in Chapter 12.

Even more serious problems are raised by Generative assumptions on the relationship of sound changes and synchronic phonological rules. If each change corresponds to one added rule, and instances of rule loss and reordering are infrequent, then the historical phonology of a language will be almost directly mirrored in the order of its phonological rules. The mechanism intended to deal with this was restructuring, but this was only sporadically invoked in the literature, allowing Chomsky and Halle (1968: 49) to claim that '... underlying representations are fairly resistant to historical change, which tends, by and large, to involve late phonetic rules. If this is true, then the same system of representation for underlying forms will be found over long stretches of space and time.' Such attitudes make for an extremely static model, and leave little scope for the divergence of dialects and languages to occur. Nonetheless, Generative theory has inspired a considerable amount of more recent work in historical linguistics, some of which we shall pursue in Chapters 3, 4 and 5.

2.5 The question of explanation

The three schools surveyed above have one common characteristic: they are all more successful at describing *what* happened than *why* it happened, although they all claim, with varying degrees of conviction, that they are explaining language change. Explanation can focus on one of two areas: the actuation problem, or the issue of why a change might start; or the transmission problem, the question of how a change, once initiated, spreads. The Neogrammarians attempted to tackle the actuation problem, suggesting physical causes such as ease of articulation, and psychological analogues, notably the notion of progressive directional displacement of pronunciation from an internalised target or sound picture. However, there is no explanation of why initially random displacements should become directional, nor any solution to the problem that assimilation, for instance, may operate in a given context in one language but fail in an identical environment in another. As for transmission, the Neogrammarians assumed only that change is lexically abrupt but phonetically gradual, and operates simultaneously for all members of a speech community; some problems for this hypothesis have already been raised in 2.2, and will be pursued in Chapter 3 below. The Structuralists did not really take issue with transmission, and failed to account adequately for actuation; although their notions of structure and function may explain the shape of a sound shift like the Great Vowel Shift once it has begun, there is no explanation for the movement of the first vowel. Finally, the

Generativists formulated changes, but this is descriptive rather than explanatory. Early Generative theory, with its emphasis on simplification, provided no adequate general account of actuation or transmission.

These inadequacies may be characteristic of the theories of change we have examined so far, or may be symptomatic of the inexplicable nature of linguistic change. Lass (1980) contends that language change simply is not amenable to explanation, if explanations are defined in deductive – nomological terms as strictly causal, universally valid covering laws which predict both that something will happen and how. Such an explanation has the structure '“X because it couldn't have been otherwise (because Y)”' (Aitchison 1981: 172); for instance, if someone's head is cut off, we can predict absolutely that he will die, since the heart cannot beat if it is disconnected from the brain, and a functioning heart is necessary for life. Such explanations, according to Lass, are unattainable in linguistics. For instance, the ease of articulation 'explanation' does not count since the same context may not always trigger the same change; similarly, functional 'explanations' are faulty because they are not universally valid.

It is certainly true that we cannot hope to explain linguistic change if we define explanation as rigidly as Lass does. However, seeking explanations of this type may be inappropriate in linguistics, which is not strictly comparable to the 'hard' physical sciences, but has a good deal more in common with biology. In the biological sciences, single absolute causes are rarely if ever to be found; sometimes, however, a cluster of interacting factors causing some phenomenon may be identified, although these are typically non-predictive and particularistic. Even in physics or chemistry, strongly predictive explanations may be a target rather than an achieved reality. Thus, Ohala (1987) argues that a physicist asked to predict the trajectory of a ball hit with a bat can only do so if she knows the angle at which the bat strikes the ball, the force of the blow, the speed and direction of the wind (and perhaps, if we now introduce chaos theory, even the whereabouts of all the butterflies in the world). In linguistics, we also have to specify contextual factors in this way, and consequently explanations in both fields can be probabilistic or statistical at best.

Arguably then, while seeking grand, unified explanations for language change, we should not reject intermediate, partial explanations specific to particular phenomena. For the moment, we may have to accept a lower-key definition of explanation at a less elevated but more commonsense level: explanation might then constitute 'relief from puzzlement about some phenomenon' (Bach 1974, quoted Greenberg 1979: 279). Our grounds for assessing explanations are not currently so well-developed as we might ultimately wish them to be, and we may sometimes fall back on insight and intuition; and the definition of explanation as providing relief

from puzzlement with respect to a particular phenomenon does not reach the epistemological high ground occupied by Lass's deductive-nomological model. But in a discipline where the three theories of change we have explored so far frequently cannot even offer relief from puzzlement, we might profitably regard this less elevated type of explanation as a useful intermediate goal, and perhaps even as an ultimate one. As Lass says (1980: 146-7), '... even second-best is not the same as universal darkness, and there may well be areas in which second-best is best, because first-best is simply not possible in principle'.

In the chapters to follow, we shall consider further, generally particularistic and non-predictive, explanations of changes in all components of the grammar, while still striving to find general causes and motivations for change. In the next chapter, we shall look at two more recent bodies of work on sound change. The first, lexical diffusion theory, attacks the Neogrammarian solution to the transition problem. The second section focusses on the extension of Lexical Phonology, originally a synchronic phonological model, to the diachronic domain, and attempts to show that, despite the shortcomings of Standard Generative Phonology, its successors may cast light on sound change and on the relationship of synchrony and diachrony.

3 Sound change 2: the implementation problem

3.1 Introduction

In the last chapter, we considered the attempts of three successive schools of linguists, the Neogrammarians, Structuralists and early Generativists, to describe and explain sound change. We saw that all these theories are inadequate for much the same reason: although they may produce very elegant and ingenious descriptions, and although they all propose explanations for change, none comes very close to being truly explanatory. This is because none successfully tackles either the problem of actuation (why sound change begins) or the issue of implementation or transmission (how the change spreads): of course, these issues do overlap, since transmission can be interpreted as actuation for previously unaffected speakers. We shall return to the actuation issue in Chapter 9, when we consider the relationship between change and variation, and the transmission of a change through a speech community.

In this chapter, our main concern in Section 2 is the implementation of change in the grammar of the individual native speaker; this leads naturally to a discussion of the relationship between synchrony and diachrony, and specifically between synchronic phonological rules and diachronic sound changes, in Section 3.

3.2 Lexical diffusion

3.2.1 *The issue of transmission*

The issue of transmission involves the spread of a change, through the speech output and internalised grammar of the native speaker, and through the speech community as a whole. As we have seen, the Neogrammarians, with their methodological assumption of the imperceptibility of sound change, believed implementation to be beyond the scope of investigation. In the Neogrammarian view, sound change was lexically abrupt but phonetically gradual, proceeding by minute and

A correlation of two properties may indicate causality; on the other hand, it may indicate that some third property exists which explains both, or that there is no relationship beyond chance. For instance, the crime rate in America seems to correlate with the number of church-goers (Moroney 1951): that is, the more people attend church, the higher the incidence of crime. It seems unlikely that this correlation reflects direct causation; however, there may be a third factor underlying both. In this case, this seems likely to be the incidence of large families from deprived backgrounds, which has been shown to produce both a large proportion of churchgoers and an elevated crime-rate, both reactions, albeit of different kinds. Similarly, a story which may be apocryphal but deserves to be true involves research in an Austrian village in the 1960s, which indicated a statistical correlation between the number of storks nesting and the birth-rate: during the relevant period the town was expanding, producing a contemporaneous increase in families and in rooftops congenial to storks.

In Brosnahan's case, the prevalence of blood group O is unlikely to cause a change in the incidence of dental fricatives for the reasons given above. There may be an underlying common cause, unique to the European speech community, but this is unknown and perhaps, given the time depth involved, unknowable. Under these circumstances, we should reject such hypotheses in favour of explanations invoking linguistic contact, at least until they can be better tested and evaluated: at present we can only conclude that Brosnahan is violating Occam's Razor by introducing an extraneous additional factor.

8.5 Conclusion

In this chapter we have seen that language contact and bilingualism may precipitate the transfer of linguistic units and patterns from one system to another. The more stable and prolonged such contact is, the more likely the resulting influence is to be grammatical as well as lexical, and mutual rather than unidirectional, distinguishing convergence from borrowing. However, all these cases of linguistic contact may be modified by social factors; thus, languages roughly equal in prestige are likely to show mutual influence, while a less prestigious language is more likely to borrow from a more prestigious one than vice versa. In the next chapter, we shall explore the social context of language change, and the connections of change with variation.

9 Linguistic variation

9.1 Introduction

In this chapter, we return to the problems of the actuation (or initiation) and implementation (or transmission) of change which have been at least a background concern throughout this book. In Chapter 3, we considered a partial solution to the implementation problem, namely the model of lexical diffusion which assumes that at least some changes spread gradually across the set of eligible lexical items. However, this still does not resolve the issue of how or why changes move from speaker to speaker. We have also failed so far to find any satisfactory account of actuation; the schools of theoretical linguistics we have studied have all attempted, more or less successfully, to explain why changes take the particular courses they do once they have begun, but none is able to explain that beginning. For instance, the Structuralist notions of function and structure may tell us why the Great Vowel Shift proceeded through the vowel system in the way it did, but not why the first vowel shifted in the first place. In this chapter, we shall attempt to solve these problems by invoking linguistic variation, which many historical linguists now see as inextricably linked with language change. Our conclusion will be that studies of variation do enlighten us as far as transmission is concerned: the actuation problem, sadly, will remain as mysterious as ever, and we shall consider arguments that this issue may even be outside the domain of historical linguistics.

The issue of variation has until relatively recently been rather neglected within linguistic theory, which has been concerned predominantly with idealised synchrony; variability is generally consigned to the tray marked 'future research'. This kind of temporary evasion may be unobjectionable enough in itself – we cannot, after all, expect to produce a Theory of Everything instantaneously, and must therefore operate with a certain list of priorities – but it becomes dangerous when it is enshrined as an unstated principle: variation studies have become a neglected and perhaps unfashionable area of linguistics, and this has had profound consequences for historical linguistics, as we shall see.

Different schools of linguistics have factored out variation in different ways. Paul, for instance, assumes that change can only be considered with respect to the individual idiolect, while Saussure idealises, seeing languages as essentially homogeneous entities. Bloomfield and the Descriptivists showed more interest in variation, partly because one of their central tenets was the assumption that languages could differ in unpredictable and unconstrained ways; but their interest was primarily cross-linguistic rather than intra-linguistic, and in any case they lacked the theoretical apparatus to tackle problems of variation. Finally, Chomsky made assumed homogeneity a central and explicit part of Generative theory. Since the goal of linguistics in his view is to describe competence and not performance, 'Linguistic theory is concerned with an ideal speaker-listener, in a completely homogeneous speech-community, who knows its language perfectly and is unaffected by such grammatically irrelevant conditions as memory limitations, distractions, shifts of attention and interest, and errors (random or characteristic) in applying his knowledge of the language in actual performance' (Chomsky 1965: 3-4). Variation, in the speaker or in the speech community, is seen as a peripheral performance factor.

We shall see below how the re-centralising of linguistic variation may affect the study of language change; and rather than considering variation as random, inexplicable and dangerous, we shall adopt Weinreich, Labov and Herzog's picture of language as a system containing 'orderly heterogeneity' (1968: 100). That is, variation is not random but strictly controlled, often by extra-linguistic factors, and the specification of these factors may help us account for change. Dialectological and sociolinguistic studies, which we shall consider in 9.2 and 9.3 respectively, reveal correlations of language variation with geographical region, sex, age, social class and ethnic group; and shifts in these correlations may tell us why linguistic features spread through the speech community, and conceivably how this spread begins. We may thus escape from the pessimism regarding the discovery of causes of change which prevailed before the advent of sociolinguistics, neatly summed up by Paul Postal (1968: 283) in his assertion that 'there is no more reason for language to change than ... there is for jackets to have three buttons one year and two the next'.

9.2 Dialectology

The observation of dialect differences within a language is by no means new, and by the mid nineteenth century there were already numerous studies, for instance in Romance linguistics, for which the Neogram-

marians attempted to provide a theoretical framework. The Neogrammarian regularity hypothesis, claiming that sound change is regular and exceptionless, was first promoted in 1876; but it clearly had to remain a hypothesis until some way of testing it could be found. A possible experiment was attempted by Georg Wenker, a young German scholar working on West Germanic consonants, also starting in 1876. Wenker's work marks the beginning of the systematic study of the regional distribution of language variation, using the methods of dialect geography.

Grimm's and Verner's Laws (see Chapter 2) together form the First Germanic Consonant Shift. A second, and chronologically later Second Germanic Consonant Shift also operated in German around AD 500. It affected only Proto-Germanic voiceless stops, as shown in (1), and split German into two sets of dialects, Low German in the north, where the change did not take place, and High German further south, where it did operate.

- (1) Proto-Germanic *p *t *k in Low German remain /p t k/

Proto-Germanic *p *t *k in High German:

- remain /p t k/ after a fricative
- shift to affricates ([pf ts kx]) initially, after consonants other than fricatives, and when geminate ([kx]) later becomes [kx], [k] or [x] dialectally)
- shift to fricatives /f s x/ after vowels and finally.

Some examples of the results of this change, specifically as it affected medial and final reflexes, are given in (2).

- (2)
- | Low German | High German |
|------------|-------------|
| Dorf[p] | Dorf[f] |
| da[t] | da[s] |
| ma[k]len | ma[x]len |
- 'village'
'the'
'make'

Wenker's approach to the regularity hypothesis relies on the assumption that sound change is instantaneous as well as exceptionless, operating for all speakers in a speech community, and in all words with the appropriate phonetic environment, at the same time. Consequently, if the regularity hypothesis is to be upheld, a sound change which has operated in one dialect but not another should have left a clear and definite boundary between the affected and unaffected varieties. Wenker therefore set out to find the boundary created by the Second Germanic Consonant Shift between High and Low German.

Wenker developed a questionnaire consisting of approximately forty sentences, each containing large numbers of Proto-Germanic voiceless stops – the diagnostic words are easily recognisable in an example like *Im Winter fliegen die trocknen Blätter durch die Luft herum* ('in winter the dry

leaves fly around through the air'). He sent these, written in conventional Standard German orthography, to schoolteachers in every village along the supposed High/Low German boundary; an initial pilot study was conducted around Düsseldorf, but between 1877 and 1887 around 50,000 localities were covered. The schoolteachers were asked to transcribe the sentences into the local dialect; and equipped with these results, Wenker sat down to trace the boundary and to prove the regularity hypothesis.

Unfortunately, 'from the beginning, the findings of linguistic geography have been used by historical linguists to bolster their theoretical viewpoints, but seldom has the evidence provided the proof that was desired' (Weinreich, Labov and Herzog 1968: 151). Wenker's goal was to find a single isogloss, or bundle of coincident isoglosses, dividing High from Low German, where an isogloss is an imaginary line on a map marking the boundary between two linguistic variants. This is precisely what he found for most of the way across Germany; however, in the area of the Rhine, all the isoglosses for individual eligible lexical items seemed to go their separate ways, so that the *dat/das* isogloss crosses the Rhine south of Koblenz, *Dorp/Dorf* south of Bonn, *maken/machen* between Düsseldorf and Köln, and *ik/ich* a good deal further north at Urdingen.

The discovery of this so-called Rhenish Fan, and its theoretical implication that isoglosses reflecting the same sound change could part company, seemed to indicate that the regularity hypothesis was entirely misconceived, and for a while it was rivalled by the opposing maxim that 'every word has its own history'. However, this return to atomism was too extreme a reaction, and regularity can be discerned in the pattern of the Rhenish Fan, which in fact tells us two rather important things.

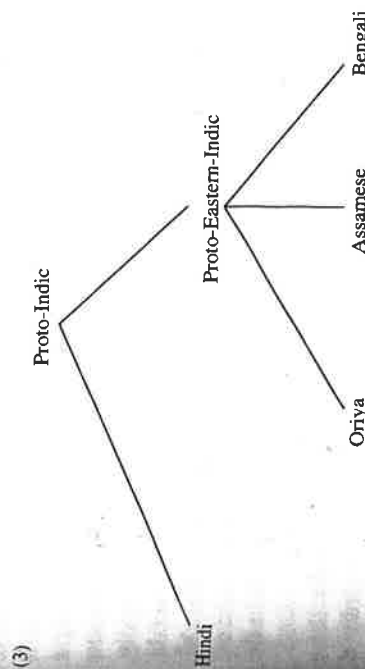
First, 'the network of isoglosses which proceeds from a study of dialect geography often represents the synchronic equivalent of the *transition* problem – that is, the route by which a linguistic change is proceeding to completion' (Weinreich, Labov and Herzog 1968: 153). In other words, the synchronic distribution of the isoglosses, revealed by dialect geography, presents a picture of the stages of the change, showing its spread in space and time. It seems that the Second Germanic Consonant Shift moved from south to north, affecting certain categories of sounds and certain words earlier than others. This means that the regularity hypothesis need not be discarded, but only modified; we can no longer make the claim that all sound changes are instantaneous, phonetically gradual and lexically abrupt – instead, we must allow some changes to operate by gradual diffusion.

Diffusion theory (see Chapter 3) then gives some theoretical status to the widely separated isoglosses found in cases like the Rhenish Fan.

Secondly, this kind of distribution of isoglosses begins to indicate that linguistic change is not purely linguistic, but instead may depend on social,

political and environmental factors. For instance, the individual isoglosses in the Rhenish Fan reflect to some extent the political boundaries of the city-states which existed at the time of the shift. Sound changes seem to originate in cultural centres or areas of political dominance, which we might call focal areas. Changes diffuse outwards from these focal areas, but may not affect relic areas, which are isolated for social or geographical reasons. Finally, between focal areas there are transition areas, which may be subject to pressure from more than one direction; and it is in areas like these, which have also had a long history of settlement, that complex patterns of isoglosses like the Rhenish Fan are characteristically found. Areas which have been settled more recently will tend to have fewer dialect differences; it follows that dialectal variation in British English is far more pronounced than in American English.

The recognition that sound changes take time to operate and can diffuse from one area, speaker or word to others also required a fundamental revision in historical linguistic theory. The Neogrammarians had traditionally represented linguistic relationships and language history by means of family trees (3), introduced into linguistics by Schleicher in 1871.



Family trees, like the one in (3), represent languages as uniform entities, and show splits of languages as clear-cut and immediate. If we are to accept Wenker's results, and incorporate notions like diffusion, gradual sound change, and influence of one variety on another into our theory, then these simplistic assumptions are untenable, certainly if we take trees literally and perhaps even if we are careful to see them as no more than shorthand.

The alternative model introduced to supplement or supplant family trees is the wave model, suggested by Schmidt in 1872. Wave theory recognises that innovations may arise in one variety and spread to others to which it may or may not be related, accounting for the influence of one dialect on another within a language, as well as borrowing and convergence (the topic of the last chapter). Innovations can spread over dialect and language

boundaries, but may be constrained by socio-political factors, allowing both for the development of new isoglosses which may follow the creation of a new political or cultural centre, and the levelling of dialects which may result from the removal of a political boundary.

Isogloss maps are used in wave theory to indicate the features shared by particular varieties, but may become extremely clumsy and hard to read if large numbers of features are included, and are essentially synchronic rather than diachronic – a map reflects the situation at a particular point in time, and further description will be required to ascertain how that situation arose. Although the family tree incorporates problematic assumptions, it is accessible, convenient and clearly shows the historical dimension. One solution might be to find a way of combining the two (Hock 1986, Southworth 1964), or alternatively one might supplement a tree diagram with an isogloss map in cases where the spread of a particular feature rather than the broad historical perspective is at issue.

Wenker's techniques of data-gathering were also problematic in several respects. First, he used postal questionnaires, so that answers were not collected in person. Secondly, the sentences were written in standard German orthography; Wenker's informants obviously had no knowledge of any form of phonetic transcription, so this is inevitable, but it is notoriously difficult to represent dialect and accent differences using a standardised spelling system, and the methods of transcription used must have varied from person to person, leading to inconsistency. Finally, Wenker sent his questionnaires to schoolteachers, who are often seen as the arbiters of 'good' language use in a community. It is certainly conceivable that such people might be unwilling to accept the existence of non-standard linguistic behaviour in their villages, or at least inclined to modify it in the direction of the standard in their written responses.

Dialect surveys following Wenker have attempted to address these difficulties, generally by introducing personal interviews by a trained fieldworker using the International Phonetic Alphabet or some other form of phonetic transcription, to replace or supplement written questionnaires. The first work of this kind was the Linguistic Atlas of France, co-ordinated by Jules Gilliéron from 1896 onwards. Gilliéron employed as his principal fieldworker one Edmond Edmont, a grocer by trade, who was trained in phonetic transcription and spent some years cycling around France, interviewing villagers, and sending the results back to base. From these results Gilliéron discovered a bundle of coincident isoglosses separating French and Provençal, of the sort Wenker had looked for in vain with respect to High and Low German. This bundle runs from the west, north of Bordeaux, around Clermont-Ferrand and off to the east, and again correlates with non-linguistic features. For instance, houses to the north of

this line generally have pitched roofs, while those to the south have the flat roofs characteristic of Mediterranean areas; and agricultural practices differ markedly between the two sectors.

Giliéron's main interest was in what he called phonetic etymologies. He applied to Latin forms the changes he knew should have operated between Latin and modern French, then checked various modern dialects to see whether they exhibited the expected outcome. If a form in a particular dialect was different from that predicted from Giliéron's knowledge of historical developments, then this discrepancy had to be explained. For instance, the word for 'cockrel' in modern French should be a descendant of Latin *gallus* 'cockrel'. In Gascony, however, we find instead the words *faisan* (usually 'pheasant') or *vicaire* (usually 'curate'). Giliéron pointed out that *gallus* should have undergone certain changes specific to Gascon, becoming ***gat*; however, this is also the expected reflex of Latin *cattus* 'cat', and in fact Gascon *gat* 'cat' does occur. Giliéron's suggested explanation involves homonymic clash; that is, homophony involving items from the same semantic field is not permitted, since unacceptable confusion might result. Since 'cat' and 'cockrel' are both farmyard animals, and since the expected developments in Gascon would leave *gat* for both, this item has been restricted to mean 'cat', and an alternative has been substituted for 'cockrel' – although it is not clear why the particular substitutes we find have been selected, or why 'cockrel' should have been replaced rather than 'cat'.

Since these initial dialect surveys in Germany and France, many others have been carried out, in England, Scotland, Italy, the USA and elsewhere. They all involve essentially the same methods of dialect geography, using direct interviewing, questionnaires or a combination of the two, and the results are usually published in atlas form, with each map showing the particular variants of a word used in different areas of the country. These are frequently of limited interest, notably because data are often presented without analysis or attempts at explanation, and especially because of their emphasis on 'purity' – the recording of a dialect before it becomes corrupted by the young or by incomers from other areas. To obtain information on this unsullied state, early dialect surveys concentrated on only one type of informant, the NORMS (Chambers and Trudgill 1980) or Non-mobile, Older, Rural Males. Speakers falling into this category were generally thought to speak a 'purer' version of the dialect than mobile individuals, who might have had their speech contaminated by other varieties; young people, who more easily incorporate innovations into their language; urban informants, who are surrounded by speakers of different varieties; and women, who, as we shall see in 9.3, tend to produce more standard speech than men. We shall see in the next section what

attempts have been made to redress the balance and include a wider range of informants in sociolinguistic studies.

There are also bodies of dialectological work within the Structuralist and Generativist paradigms, which we can safely bypass here since they are of no particular consequence for historical linguistics. The Structuralist approach (Weinreich 1954; Moulton 1960) relies on the independent establishment of a phoneme system for each of the varieties being examined; from these individual systems, a single, higher order comparative system is constructed. Such so-called diastystems allow dialects to be compared, but can really only indicate differences of phoneme inventory. On the other hand, the Standard Generative approach to dialectology rests on an assumption of identity: dialects of one language are said to share the same underlying representations, and differ only in the form, ordering and/or inventory of rules. The underlying representations may be supplied by a single dialect (Brown 1972), or be neutral between various dialects (Thomas 1967). However, this model is again inadequate, since the status of these basal forms for speakers is unclear. Furthermore, the view that dialects of a single language share common underlying forms whereas different languages differ at this level prevents us from seeing dialect and language variation as the continuum which geographical and social investigation have shown it to be. The change from shared underliers to discrepant ones would have to be sudden and catastrophic, conflicting with the growing body of evidence that suggests language change is predominantly gradual. In view of these difficulties, we shall not pursue Structuralist and Generative dialectology further here, turning instead to the impact of sociolinguistics on the study of language change.

9.3 Sociolinguistics

9.3.1 *The methodology of sociolinguistic investigation*

The study of language change in its social context has been described by some as a virgin field; by others, as a barren territory. A brief examination of what has been written in the past on this subject shows that it is more like an abandoned back yard, overgrown with various kinds of tangled, secondary scholarship' (Labov 1972: 260).

Labov is quite right in his observation that some weird and wonderful work has been done in sociolinguistics. In this section, we shall concentrate on the relevance of some better-substantiated sociolinguistic studies for historical linguistics, and will refer in particular to the work of Labov in America, and Trudgill and the Milroys in Britain, describing their techniques and findings and their contribution to the actuation and transmission problems. The historical consequence of this work arises

from the assumption that variation in the speech community is not random, but structured, and that it may in some cases represent change in progress. Since the social context of ongoing change is clearly observable, its mechanisms and causes may be easier to perceive than those of completed changes, for which the context is often not recoverable. Findings may then be generalised from changes in progress to completed changes, provided that we accept the Uniformitarian Principle, '... the claim that the same mechanisms which operated to produce the large-scale changes of the past may be observed operating in the current changes taking place around us' (Labov 1972: 161).

Of course, sociolinguistics was not always such a lively field: as we saw in the last section, studies of variation before the 1960s, when they were done at all, focussed on regional, and specifically rural dialectology. By the middle of this century, 'some dialectologists began to recognise that the spatial dimension of linguistic variation had been concentrated on to the exclusion of the social dimension' (Chambers and Trudgill 1980: 54). The emphasis on rural investigations followed from the early dialectological preoccupation with 'pure' varieties, since rural areas tend to be more linguistically uniform than urban ones, and older rural speakers in particular are likely to have been in contact with fewer other varieties than their town-dwelling counterparts. However, with the advent of descriptive rather than prescriptive linguistics, and the recognition that the majority of potential informants in America and Europe now live in the cities, it was realised that a great deal of potentially interesting data was simply being ignored. In the 1960s, urban dialectology, or sociolinguistics, was born.

The earliest work on urban variation simply represented a transfer of dialectological methodology to the towns – for instance, Eva Sivertsen's *Cockney Phonology* (1960), although ostensibly dealing with a variety spoken by thousands of speakers, drew all its data from four old ladies living in Bethnal Green (Chambers and Trudgill 1980). However, various strategies for achieving a wider distribution of data, from a good range of speakers, have been developed since, and are now generally used in sociolinguistic investigations. Notably, it is essential not to select informants in a subjective way, since the experimenter's own prejudices could have an unwelcome influence on the results. Instead, random, representative samples are required. Of course, it is impossible to interview every member of a speech community, especially one as large as the Lower East Side of New York City (Labov 1972), Norwich (Trudgill 1974) or Belfast (Milroy and Milroy 1985), but every speaker must have an equal chance of being chosen. Frequently, random sampling is achieved by selecting every hundredth, or thousandth name from the electoral register or some similar list.

Once the informants have been selected, they are interviewed. At this point, experimenters encounter the problem of eliciting natural, informal speech (which may well be where changes begin) from their informants, who may be made nervous by the presence of an unfamiliar interviewer and a taperecorder; this may cause them to concentrate on speaking 'correctly', and therefore formally. The difficulty this presents is summed up in the Observer's Paradox: sociolinguists would like to know how speakers speak when they are not being observed, but the only way to find out is by observing them. To use Labov's analogy, everybody knows that the light inside the fridge is on when the door is open, but how do we find out whether it is on or off when the door is closed, when checking might itself interfere?

Various techniques have been developed to deal with the Observer's Paradox, including recording informants in a group and observing speech which is strictly outside the interview situation. The most common technique in earlier sociolinguistic studies (see Trudgill 1974) is the structured interview, which involves asking the informant to perform various tasks at decreasing levels of formality. First, the informant reads a list of minimal pairs, a word list and a short passage. Next comes a question and answer session, and finally the interviewer encourages the informant, who by now has probably relaxed a little, to produce informal conversational speech by asking about childhood rhymes and sayings, encouraging diversions to make the informant talk for longer continuous periods, and perhaps asking questions which are likely to elicit an emotional response. This structured interview technique therefore enables a comparison of formal and casual styles within the same speaker.

Later experimenters have often followed Milroy (1980), who argues that it is easier to gain access to the vernacular if the informant feels at ease, and that this can be achieved by using an interviewer who is integrated into the speech community. This technique is known as participant observation, or more colloquially, as the 'friend-of-a-friend technique'. Milroy, investigating the speech of three housing-estates in Belfast, integrated herself into the various communities by making friends with one person, and having that person introduce her to other contacts. After visiting the communities over a certain period of time, making herself known and doing various favours for her potential informants (such as giving them lifts in her car), she succeeded in recording a good deal of high-quality informal speech.

The recorded data next have to be analysed. In general, the experimenter will have had a fairly good idea to start with of the features which are in variation, and therefore perhaps changing, and the interview is designed to elicit plentiful examples of these features. Sociolinguistic research is centred on the notion of the linguistic variable, an element which has a number of

realisations, or variants, in speech, but a constant meaning. For instance, some English speakers pronounce the word *hat* as [hæt], and others as [æʔ], but the meaning is the same in both cases. The variants of each variable are typically not randomly distributed, but instead correlate with extralinguistic factors like the age, sex and social class of the speaker, and the level of formality. To pursue the example of [h]-dropping, a sociolinguistic investigation would recognise a variable, written (h), with two variants, (h)·[h] and (h)·Ø. When analysing the data, the experimenter would assign a particular score to each occurrence of each variant, perhaps giving 1 for (h)·[h] and 0 for (h)·Ø. The resulting figures are then plotted against age, sex and so on, to see whether any significant pattern emerges. This kind of quantitative, statistical approach is necessary because most speakers will use both variants; however, their frequency of use will differ from speaker to speaker on the basis of non-linguistic factors. Most variables investigated have been phonological, and the examples discussed below will follow this trend, although it is certainly possible to study syntactic variables. For instance, the variation between single and multiple negation in English might be one such example; Labov (1972) also includes studies of the presence or absence of *be* in Black Vernacular English in sentences like *She's real nice* versus *She real nice*, and of negatives with *no* and *any* in the same dialect, as in *I ain't got no money* versus *I don't have any money*.

One rather common type of variation can be illustrated with Trudgill's (1974) results on the (ng) variable in Norwich. (ng) is the final consonant in words ending *-ing*, such as *seeing, helping, sitting*, and has two variants, (ng)·[ŋ] with the velar nasal, and (ng)·[n] with the alveolar. In (4), the score 100 indicates consistent use of [ŋ], while 0 shows consistent use of [n].

(4) (ng) plotted against class and style
(after Trudgill 1974: 92)

Class	Word list	Style		
		Reading passage	Formal speech	Casual speech
Middle Middle Class	0	0	3	28
Lower Middle Class	0	10	15	42
Upper Working Class	5	15	74	87
Middle Working Class	23	44	88	95
Lower Working Class	29	66	98	100

The occurrence of the two variants clearly correlates with two non-linguistic factors: one is style, with all speakers producing progressively more instances of [ŋ] as formality decreases, and the other is social class; the higher the class to which an informant belongs, the more instances of [ŋ] he or she will produce in all styles. Of course, the concept of social class

Levels of centralisation for both diphthongs clearly increase as age decreases, reaching an apparent peak in the 31- to 45-year-olds, with less centralisation in the youngest age group. We shall return to this example, and to the link between age variation and linguistic change, in the next section.

Finally, it should be stressed that linguistic features may also correlate with ethnic or religious group, and with other, less easily definable characteristics like the speaker's level of ambition, for instance. Furthermore, the recognition that some linguistic variation may be conditioned by social factors does not mean that linguistic factors cease to be important. For example, the two diphthongs in (6) pattern in slightly different ways, and this is likely to have a linguistic rather than a social explanation. Similarly, the Milroys' work in Belfast revealed that the variable (u) has a variant [ʌ], characteristic of lower-class speech. However, the percentage of [ʌ] found varied quite noticeably from word to word, as shown in (7).

(7) Percentage of [ʌ] in Belfast
(after Hudson 1980: 169)

pull	74 %
put	39 %
look	27 %
should	8 %

This variation may be a function of linguistic context, or may show that [ʌ] is spreading by lexical diffusion. In other words, linguistic changes in progress do not only spread from speaker to speaker or from group to group; they may also spread from word to word or from context to context.

Figures like those given above seem to indicate that speakers are, consciously or subconsciously, obeying certain fixed norms of behaviour; indeed, Labov (1972) defines a speech community as a group of speakers sharing a set of norms. Certain follow-up tests can be carried out after the primary data from a sociolinguistic experiment have been analysed, to check for the existence of, and informants' adherence to, these norms. First, speakers can be asked to participate in self-evaluation tests, during which they are played recordings of sentences containing the variants under investigation, and asked to assess which they would tend to produce themselves. In general, 'their answers reflect the form which they believe has prestige or is 'correct' rather than the form they actually use' (Labov 1972: 213). Only working-class men tend to under-report their level of use of prestige variants; men wishing to stress their membership of lower-class, urban male groups are unlikely to publicly assert their adherence to values,

is a notoriously difficult one to define, and correlations of linguistic variation with class must therefore be treated with some care; but even so, the pattern revealed by Trudgill's data is striking.

It should also be borne in mind that tables like (4), where speakers are treated as representatives of a particular group, do not reveal what individuals are doing; they simply indicate trends in the speech community. Furthermore, a given individual may belong to more than one group. For instance, if we further split Trudgill's informants according to gender as well as social class, we find (see (5)) that quite consistently, across classes and styles, women use more of the higher-status variant [ɪ] than men. The anomalous LMC figures for casual speech arise from the small number of cases of this variable Trudgill succeeded in recording for this group and context; small amounts of data can therefore skew experimental results. We shall return in 9.3.2 to the general claim that women typically produce more high-prestige variants than men of the same class.

(5) (ng) plotted against class, sex and style
(after Trudgill 1974: 94)

Class	Word list	Style			Casual speech
		Reading passage	Formal speech		
MMC	M	0	4		31
	F	0	0		0
LMC	M	0	27		17
	F	0	3		67
UWC	M	0	18		95
	F	11	13		77
MWC	M	24	91		97
	F	20	81		88
LWC	M	66	100		100
	F	17	97		100

Linguistic variation may also correlate with the speaker's age. Labov (1972: Chapter 1) reports a study of vowel centralisation of the diphthongs (ai) and (au) on the island of Martha's Vineyard, off Massachusetts. Some of his results are shown in (6).

(6) Centralisation plotted against age
(after Labov 1972: 22)

Age	(ai)	(au)
75-	25	22
61-75	35	37
46-60	62	44
31-45	81	88
14-30	37	46

in language or other aspects of behaviour, largely determined by the middle class.

A second option is a matched guise test (Lambert *et al.* 1967), in which one speaker who is able to use two (or more) accents is recorded twice, once using each variety. These recordings are played to the informants, who are asked to evaluate 'Speaker A' and 'Speaker B' according to a number of criteria, including the speaker's occupation, likely level of education, salary, family background, leadership qualities, degree of honesty, and so on. It seems that most informants judge the voice with the higher percentage of prestige variants as belonging to a more educated speaker from a higher social group, who is likely to have a higher-status job – although speakers with regional accents, especially rural ones, are sometimes judged to be more honest and friendly! Such tests again evidence the existence of shared norms of linguistic behaviour, which govern informants' judgments of other speakers as well as their own actions.

Finally, the results obtained from sociolinguistic experiments may be skewed by the Observer's Paradox. It can be useful, therefore, to conduct a follow-up experiment using the technique of rapid anonymous observation: here, very short interviews are conducted with a number of informants who did not form part of the original survey, who will be encountered by chance (maintaining the random nature of the experiment) and who really will not know that their speech is being observed, so that the Observer's Paradox should not come into play. If the results are comparable with those from the main experiment, it can be concluded that the Observer's Paradox did not have too detrimental an effect on the original interviews. The best-known rapid anonymous survey is Labov's (1972) New York department store experiment. As part of his sociolinguistic survey of the Lower East Side, Labov examined the (r) variable, which has variants (r):[r] and (r):Ø before consonants and pauses; his results for (r) were generally comparable to Trudgill's figures for (ng) in Norwich (see (4) above). To verify these results, Labov selected three New York department stores, one high-class, one middle-class and one lower-class. Labov assumed that sales assistants in these stores were likely to imitate or accommodate linguistically towards their customers, in order to 'fit in': staff at the higher-class store should therefore be predicted to produce most occurrences of the prestige variant (r):[r], and those in the lower-class store should produce fewest. Labov approached shop assistants in all three stores and asked the location of particular goods, selected to elicit the phrase *fourth floor*, which contains two instances of (r). The results showed exactly the predicted variation, if the staff at the three stores are compared with the classes of informants from Labov's main experiment. A further example of rapid anonymous observation, again designed

by Labov and again rather ingenious, involved the variation between [s] and [ʃ] in *str*-clusters in Philadelphia. Labov and his co-experimenters stopped people at random around Philadelphia and asked 'Can you tell me how to get to X Avenue?' The answer was generally 'X Street?', allowing observation of the required variable (L. Milroy 1987: 74).

9.3.2 Sociolinguistics and language change

9.3.2.1 Introduction

We now return to the question of how the results of sociolinguistic surveys can enlighten us on the subject of linguistic change in progress. Just as separated isoglosses in dialectology could be taken as showing the spread of a change in space and time, so the correlations of linguistic features and social factors, if they are seen not as fixed but as shifting, can indicate the spread of an innovation through the speech community. It may be that our quantitative data can even tell us which group has started the spread of the change.

9.3.2.2 Real and apparent time

Let us begin our investigation by returning to the variable centralisation of the diphthongs /ai/ in words like *fire*, *high*, *sight* and /au/ in *loud*, *mouth*, *how* among inhabitants of Martha's Vineyard, a small island with a population of around 6,000, off Massachusetts. Some data from Labov's (1972, Chapter 1) investigation were given above as (6), and are repeated for convenience as (8).

(8) Centralisation plotted against age
(after Labov 1972: 22)

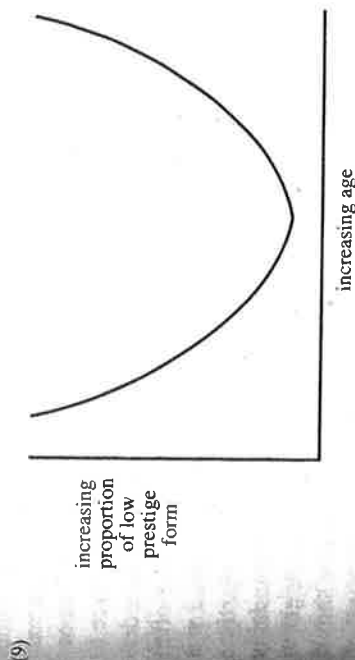
Age	(ai)	(au)
75–	25	22
61–75	35	37
46–60	62	44
31–45	81	88
14–30	37	46

It is clear from (8) that younger people, with the exception of the very youngest group, show a higher degree of centralisation of both diphthongs than older ones. Labov hypothesises that this distribution of variation indicates a change in progress, and moreover a change which is being introduced by middle-aged speakers and being adopted strongly by 31- to 45-year-olds, and variably by younger speakers. However, Labov's proposal relies on the validity of studies in apparent time.

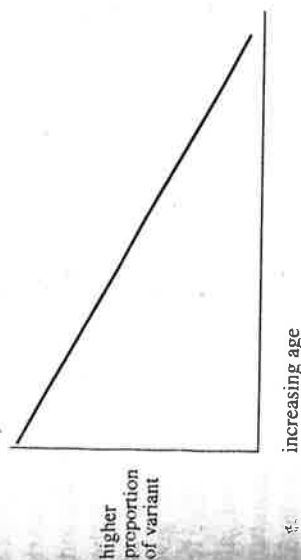
The ideal historical linguistic survey would take place in real time; a representative sample of informants would be interviewed, then re-

Linguistic Atlas of New England (LANE). These revealed that, at that time, inhabitants of Martha's Vineyard did centralise (ai) to a moderate extent, but scarcely if ever centralised (au). Comparing these LANE records with Labov's findings therefore indicates clearly that a change is under way.

Such cases, where apparent-time experiments can be checked against earlier dialectological records, may predispose us more favourably towards accepting the results of surveys conducted in apparent time, even when no real-time back-up is available. We can then identify the distribution of variants we would expect to find in a speech community when some linguistic change is in progress. The 'normal' distribution, expected when change is not affecting the variable under consideration, would be roughly that shown in (9), where young and old speakers characteristically use higher proportions of lower-status variants. Chambers and Trudgill (1980) suggest that this may be because 'middle-aged' speakers are generally trying to succeed in their careers, and may be in contact with a wide range of contacts who they wish to impress: speakers in this central age band are therefore most likely to be responsive to the norms of linguistic behaviour in their speech community.



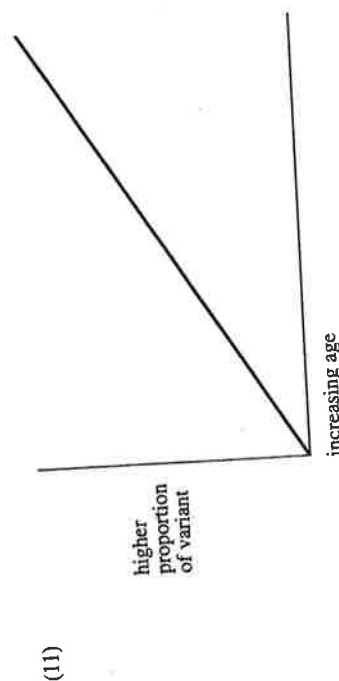
However, when a change is in progress, the variable investigated will tend to show one of the patterns of distribution in (10) or (11).



interviewed at ten or twenty year intervals, with the resulting data tracing the movement of ongoing changes through the community. There are all kinds of obvious difficulties with this procedure: informants will die, or move away, or refuse to be interviewed again; the experimenter may have to be replaced, perhaps with consequences for the methodology; and results will not be available for many years. Real-time studies using comparable populations at the different stages rather than precisely the same set of informants would not be restricted to the life-span of individual speakers, but the other problems would remain. Consequently, Labov and his followers have introduced apparent time studies, in which only one set of interviews is done, at one point in time. The set of speakers interviewed is selected to be representative of all age groups in the community, and correlations of linguistic variants with age are calculated. Variation in these correlations is interpreted as indicative of ongoing change, and the speech of older people is therefore seen as characteristic of earlier stages of the language. In other words, 'the validity of such a study hinges crucially upon the hypothesis that the speech of, say, 40 year olds today directly reflects the speech of 20 year olds twenty years ago and is thus comparable for diffusion research to the speech of 20 year olds today. Discrepancies in the speech of 40 year olds and 20 year olds are attributable to the progress of a linguistic innovation in the twenty years that separate the two groups' (Chambers and Trudgill 1980: 165). In short, apparent-time studies do not involve 'real' diachronic work at all, in the sense of explicitly comparing different periods in the history of a language; instead, they involve looking at synchrony and attempting to perceive the seeds of diachrony in it.

Various difficulties are also inherent in apparent-time work. For instance, we must assume that adult speakers maintain their language from age twenty or so throughout their lives, unless they are subject to some sound change, and this seems over-simplistic. Furthermore, and perhaps even more seriously, it is in principle impossible to distinguish age variation in apparent time which signals a change in progress, from some linguistic feature which simply varies with age, becoming apparent in the speech of people of a certain age. Younger people sometimes behave differently from older ones in other respects; why should certain linguistic features not also be restricted to particular age groups?

One compromise is to attempt to find some early survey, perhaps of a dialectological nature, relating to the variety being investigated or a closely related variety. This can then be used to confirm the apparent-time data by providing a partial real-time comparison, albeit conducted by a different investigator, using different informants and a different methodology. For instance, in the case of Martha's Vineyard, Labov was able to refer to interviews with islanders conducted in the 1930s as part of the fieldwork for the



A variant being introduced by linguistic change will occur much more frequently in the speech of younger people, as in (10). However, a variant being lost will conform to the pattern in (11), with many more occurrences in older informants.

The centralisation change in Martha's Vineyard seems to conform broadly to the pattern in (10), since the amount of centralisation is increasing from generation to generation. Labov also attempts to consider the social meaning of this change by taking into account the social context: 'no change takes place in a social vacuum' (Labov 1972: 2). He points out that Martha's Vineyard, although widely regarded as a pleasant place to live, had by the 1960s lost most of its traditional industries; many islanders had therefore left to work on the mainland, and those remaining often found it hard to make a living. Furthermore, property on the island was increasingly being bought by mainlanders, who flocked to Martha's Vineyard in the summer. Those islanders remaining understandably became resentful of this influx of wealthy summer visitors, and sought to affirm their identity as islanders. It is entirely plausible that they should have done so partly through their language; on this view, the centralisation of diphthongs is the linguistic equivalent of wearing a T-shirt which says 'I'm not a tourist, I live here'.

The next question is where this tendency towards centralisation originated. Labov's data indicated that fishermen from Chilmark, in the rural up-island, exhibited the highest amount of centralisation, which seems to have spread from their dialect. This accords well with the general island attitude towards the Chilmark fishermen, who are seen as almost emblematic of the island's history and culture, especially given their independence and the courage needed for their occupation. The feature of centralisation has therefore been adopted and exaggerated, albeit unconsciously, on the basis of their speech. It also follows quite naturally that most centralisation should be found, in Labov's survey, in the 31- to 45-

year-olds, who have generally been away at College, then made a positive decision to return to Martha's Vineyard despite the island's economic problems. This positive commitment to the island is reflected in their language.

Labov's final concern is to explain the ostensibly discrepant figures for the youngest category of speakers. In fact, young informants who wish ultimately to leave the island characteristically have very low indexes of centralisation, while those who intend to remain have extremely high levels of centralisation, rather more indeed than the average for the 31- to 45-year-olds, showing that the change is still proceeding. However, combining these two opposing groups of younger speakers produces average figures for the age group which are lower than those for the 31- to 45-year-olds, misleadingly suggesting that the change is reversing itself.

Labov's study of Martha's Vineyard, then, shows the importance of apparent-time studies in tracing change in progress, and also indicates that taking into account the social meaning of linguistic variation can allow us to pinpoint the source of the change. However, we are still really no nearer solving the actuation problem; we know that the Vineyarders adopted centralisation of diphthongs from the Chilmark fishermen, but we still have no idea why this particular feature was adopted, or why the fishermen produced centralised diphthongs in the first place.

9.3.2.3 Overt and covert prestige

Apparent-time studies based on age do not provide the only indication of ongoing change, since 'ongoing linguistic changes may often be reflected in unusual patterns of social class differentiation' (Chambers and Trudgill 1980: 95). Such unexpected patterns can also suggest which group might be at the forefront of the change. For instance, the data in (12), again from Trudgill's Norwich survey, show scores for speakers from five social classes for the variant [ʌ] of the variable (e), found in words like *hell*. The table shows that the amount of [ʌ] increases gradually from the middle middle class (MMC) to the lower working class (LWC), except that the Upper Working Class informants have far higher levels of [ʌ] than would be expected. It may well be that the Upper Working Class speakers are therefore the initiators of this change.

(12)

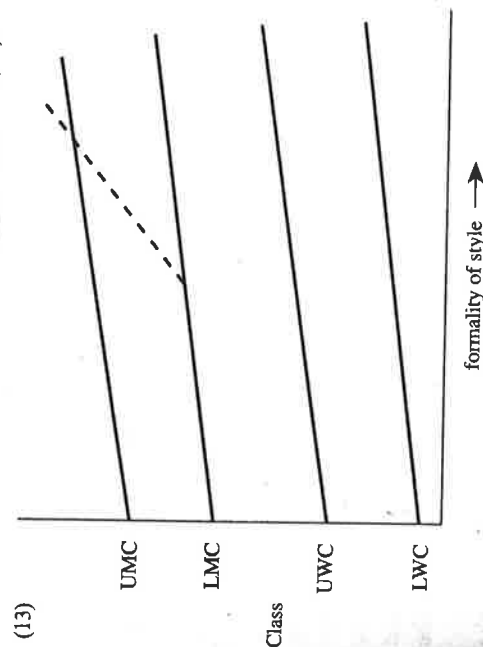
Norwich (e):[ʌ]	
MMC	2
LMC	42
UWC	127
MWC	87
LWC	77

Many unusual patterns of class-conditioning are specific to particular changes, but one such pattern has been frequently reported in the literature; this involves the phenomenon of hypercorrection. In New York City, Labov discovered that more instances of the higher-prestige variant (r):[r] were found in all speakers in formal speech, and that working-class speakers had less [r] than upper-middle-class informants, as might have been expected, the lower-middle-class informants, as might have been expected, produced levels of (r):[r] intermediate between the upper-middle-class and working-class levels; but in more formal styles, the lower-middle-class informants produced significantly higher levels of (r):[r] than those from the upper middle class. This exaggerated level of [r] production is achieved by inserting [r] even in environments where it is not historically appropriate; for example, some of Labov's lower-middle-class informants pronounced *God as* [ga:rd]. An equivalent case of hypercorrection involves some British English speakers who characteristically use (h):Ø – these '[h]-droppers' may in formal situations attempt to speak 'correctly' by inserting [h], but are unsure of the appropriate environments and therefore overshoot, producing hypercorrect forms like [hæt] for the preposition *at*. Furthermore, matched guise and self-evaluation tests indicate that lower-middle-class speakers are characteristically hypersensitive to lower-status variants, claiming virtually never to use these themselves and evaluating other speakers who do produce them extremely adversely.

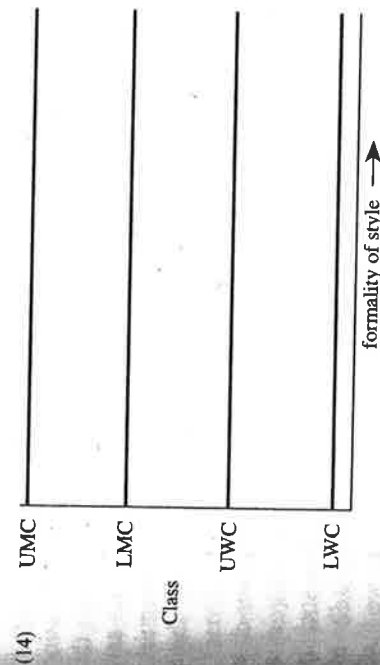
It seems that members of the lower middle class tend to be socially and linguistically insecure, perhaps due to their position between the norm-setting upper middle class, to which they aspire, and the working class, from which they may wish to distance themselves. This leads to a careful adherence to norms imposed by the upper middle class, which may become exaggerated, resulting in hypercorrection. However, lower-middle-class hypercorrection will result in a greater number of prestige variants being heard and perhaps ultimately produced by the lower classes. Because of their insecurity and social mobility, the lower middle and upper working classes are therefore of some importance in the spread of change, whereas the upper and lower extremes of the class scale are relatively conservative.

Many sociolinguistically motivated changes do involve the progressive adherence of other speakers to norms set by the upper middle classes. Such changes are often changes from above (Labov 1972), so-called 'because they operate above the level of conscious awareness' (speakers are conscious of phenomena like pronunciation versus lack of [r]), not because they involve prestige pronunciations. Such changes characteristically involve variation correlating both with style and with some other social factor such as social class or sex. Most of the variables discussed above have been of

this doubly-conditioned type, which are known as markers; the general pattern of distribution for a marker is shown in (13).



There are also variables which are conditioned by class or sex, for instance, but not by style, giving the typical pattern in (14): these variables are indicators.



The great majority of changes involve markers, but occasional changes proceed on the basis of an indicator. These are called changes from below, to suggest that the change begins below the level of conscious awareness, not that the source is a variant characteristic of a lower class, although this is often the case. The spreading medial glottal stop in words like *butter*, *bird*, *heard* in New York with a diphthong [ɜɪ], as does the pronunciation of changes from below are liable to become stereotyped; that is, speakers begin to notice them and to associate them with particular groups in the

speech community. If they are stigmatised, or regarded as indicative of low status or lack of education, they may then be lost, so that the ongoing change is reversed from above the level of conscious awareness. For example, speakers are frequently told not to say [bʌʔə] and so on, although whether this is actually reducing the incidence of the glottal stop is open to debate. Similarly, [ɜɪ] in New York is both stigmatised and stereotyped, even being represented in writing with spellings like *boïd*, and this change does indeed seem to be reversing. Interestingly, not only those variants which are stereotyped as indicative of low class membership tend to be lost, but also those which are seen as characteristic of the upper class. For instance, accurately, of would-be members of the upper class. For instance, a pronunciation of *house* as [haɪs] surfaced briefly in the 1980s among workers in the City of London, was stigmatised as a feature of the then-prevalent 'yuppie' culture, and seems to have all but gone. In cases where indicators are involved in language change, and are not stigmatised and lost, they seem to become markers and begin to vary with style before the change is completed (Chambers and Trudgill 1980).

Most of the changes discussed above involve an attempt to emulate the highest-status group, whose behaviour and language are seen as prestigious. Prestige based on norms set by the upper classes is known as overt prestige, and is reflected in the actions and attitudes of all classes in the speech community. However, not all linguistic behaviour or change is dictated by overt prestige; if it were, we would all speak similarly, using high-status variants, and language change would presumably take the form of a spiral of hypercorrections. Of course, lower-class varieties and variants are also maintained, and inspire some changes. Sociolinguists have introduced the notion of covert prestige to account for these phenomena.

Just as there are norms of upper-class usage, there are also lower-class norms; these generally have less impact on the speech community as a whole, since the speakers of low-status dialects tend to be less socially influential. These norms can also be enforced, and will tend to be adhered to by speakers wishing to emphasise their group membership; we have already seen that this is the case for lower-class males, who characteristically evaluate low-status forms favourably and are thus influenced by a second, covert type of prestige.

In their Belfast studies, the Milroys (L. Milroy 1980, J. Milroy 1992) used the notion of social network to account for the maintenance of lower-class variants. Each speaker has a social network, which is the sum of all contacts he or she has with other speakers – family, friends, acquaintances and once-met strangers at the bus-stop included. Every personal network will be unique, but some share particular characteristics. Notably, social

networks can be placed on scales of denseness and multiplexity, where a dense network is one in which any third party mentioned by a speaker is likely also to be known to her hearer, and a multiplex network includes many members who know one another in several different social spheres; for instance, two speakers might be neighbours, work for the same company, socialise together and worship in the same church. The Milroys devised various tests for denseness and multiplexity, and applied these to give a strength score for each social network. The higher this network strength score, the more integrated the speaker is likely to be into his or her community. Well-integrated, central speakers in closeknit networks are most susceptible to norm-enforcing mechanisms; although they may passively share the more overtly prestigious norms determined by upper-middle-class speech, they do not themselves produce such speech, since these norms are outweighed by conflicting standards set by peer pressure and covert prestige. Thus (L. Milroy 1987: 106), 'individuals in Belfast whose personal networks were closeknit tended to approximate closely to the stigmatized vernacular norms characteristic of the locality, which like other 'in-group' norms powerfully symbolized values of solidarity, reciprocity and to some extent opposition to standardized norms along with their associated values'.

This notion of degree of integration into a closeknit network may also account in part for our observation that men seem most influenced by covert prestige. In many working-class communities, men have traditionally gone out to work while women remained at home. Since there is often one major employer like a factory or a shipyard for a single community, men will contract relationships in the workplace which supplement ties with neighbours, producing higher network strength scores. Men will therefore lead changes inspired by features of the vernacular. Women, however, with their more looseknit social networks, will be more open to external norms set by the upper classes, so that women are characteristically in the forefront of changes based on overt prestige. These tendencies do not seem to hold in communities like India and Iran, where women have a fixed and clearly subordinate role; women seem to lead change towards the standard only where they play some part in public life (Holmes 1992: 234).

In fact, the crucial factor determining who leads changes based on vernacular norms does not seem to be gender *per se*. There is a tendency in Belfast (Milroy and Milroy 1985) for /a/ to be backed, and this variant is especially characteristic of casual speech and inner city areas; it is a core feature of the Belfast vernacular. In the Ballymacarrett area of the city, the backed variant is found predominantly in working-class men, as would be predicted in the social network model. However, in the Clonard district,

young women aged 18–25 have a far higher network strength score than young men; the closure of traditional industry in the area means that many men are unemployed, while young women typically do have jobs. These women also produce the backed variant of /a/ far more than their male counterparts, although the expected, reverse pattern is found in other age groups, showing that network strength may indeed be relevant in explaining the propagation of change and the maintenance of vernacular features.

9.3.3 Actuation

Thus far, we have largely been tacitly ignoring the problem of actuation or initiation of change, concentrating instead on its transmission through the speech community. Although variation studies do seem to produce enlightening accounts of the spread of variants, a large question mark continues to hover over the genesis and selection of the variant itself: for instance, Labov's study of Martha's Vineyard revealed that centralisation of diphthongs probably spread from the Chilmark fishermen, but not why the fishermen centralised *their* vowels or why this feature in particular should have been adopted by other islanders.

Labov (1972) provides a certain amount of encouragement on this matter, claiming that historical linguists do not fail to account for the ultimate actuation or birth of a variant, since this is simply beyond our remit. A huge number of sources for variation exist; 'variations may be induced by the processes of assimilation or differentiation, by analogy, borrowing, fusion, contamination, random variation, or any number of processes in which the language system interacts with the physiological or psychological characteristics of the individual' (Labov 1972: 1–2). However, these possible explanations are not part of our problem: the real actuation question is why some of these innovations die out and others catch on, spreading through the community, or why certain instances of variation become changes while others don't.

In fact, Labov argues that it is not meaningful for us to distinguish actuation from transmission. An idiosyncratic speech habit, or a slip of the tongue certainly introduces variation, but *change* is only initiated when the new variant is adopted by a group of speakers, becomes systematic and acquires some social significance. As Labov (1972: 277) puts it, we may 'assume that a certain word or pronunciation is indeed introduced by one individual. It becomes a part of the language only when it is adopted by others, i.e. when it is propagated. Therefore the origin of a change is its 'propagation' or acceptance by others.'

However, it seems that we cannot (yet?) predict with any accuracy which

instances of variation will lead to change (see also Chapter 12). Labov accepts that this is a function of the influence of social context on language; since every social situation is unique, linguistic behaviour in such situations is bound to be to some extent unpredictable. He does accept that variants will start off as characteristic of particular social groups and will then spread out to others, gradually acquiring some systematic social value; however, the fate of the variant, and consequently the rate, extent and direction of the change, will depend on that social value and the prestige of the group first using the form in question. Only a few tentative generalisations can be made: for instance, Labov claims that women are more active in spreading overtly prestigious variants; that the upper working and lower middle classes are most active in initiating change as their linguistic insecurity causes them to imitate upper-class speakers; and that innovating speakers will tend to be the central members of their groups, with high status within their communities and also many contacts outside them.

The Milroys (Milroy and Milroy 1985; J. Milroy 1992) adopt a subtly different approach to the problem of actuation. They see the essential goal as the explanation of why language is sometimes changed and sometimes maintained, even in very similar sets of circumstances. They note that the answer to this question has generally been sought in the dynamics of the language system but argue that we might make more progress by considering the role of the speaker, and distinguishing speaker innovation from language change. In other words, 'it is not languages that innovate; it is speakers who innovate. The reflexes of speaker-innovations are then observed in language states, where they appear as systematic and rule-governed linguistic change' (Milroy and Milroy 1985: 345). Of course, linguistic structure is still important; speaker innovation may start a change, but a chain reaction may then begin, with subsequent steps being motivated by the system. Thus, we maintain the Structuralist notions of the structure and function of linguistic systems as influences on change.

However, although 'we can describe speaker-innovation as an act of the speaker which is capable of influencing linguistic structure' (J. Milroy 1992: 169), not all innovations lead to change. Some may remain idiosyncratic features of the innovator; others may reach the speaker's immediate community but go no further; but some innovations will spread into other communities via some mediator who is connected to all the networks involved – and this is linguistic change.

The Milroys argue that dense, multiplex social networks typically retard or resist language change, instead maintaining vernacular norms. Furthermore, central, well-integrated members of such networks are liable to be more sensitive to their normalising influences, and will therefore be less

likely to innovate. This is a problem for Labov's hypothesis that innovators in language change are likely to be those with strong ties inside and outside their social group. Milroy and Milroy (1985) suggest an alternative solution: if closeknit networks and strong interpersonal ties lead to language maintenance, might not looseknit networks and weak interpersonal ties promote language change? Innovators might then be sought among speakers contracting large numbers of weak links with other groups without occupying a central position in their own, while a loosening of network structure could signal a volatile situation conducive to change.

Strong interpersonal ties might be roughly defined as those between friends, while weak ties connect acquaintances; the measurement depends on the intensity and intimacy of the relationship, and the time the two people spend together (J. Milroy 1992: 178). Characteristically, people draw their friends from their own social group, while their acquaintances may have widely differing social backgrounds; it follows that 'weak ties BETWEEN groups provide bridges through which information and influence are diffused, and ... weak ties are more likely to link members of DIFFERENT social groups than strong ones, which tend to be concentrated WITHIN particular groups' (Milroy and Milroy 1985: 364). Furthermore, since a person will generally have more acquaintances than friends, weak ties reach more people and are more numerous. Consequently, 'weak intergroup ties ... are likely to be critical in transmitting innovations from one group to another, despite the commonsense assumption that STRONG ties fulfil this role' (ibid.: 365).

The innovators crucial to linguistic change, who start the movement of a variant through society, are therefore highly likely to be socially mobile individuals who are not central enough in any group to be constrained by its norm-enforcing mechanisms, but who have weak links with enough groups to pass the variant on to their members. They are consequently liable to be members of the more mobile upper working or lower middle class, as Labov hypothesised. For the change to be successfully transmitted, however, it must be passed from these peripheral innovators to the so-called early adopters. These will be central in their social group, and subject to its norms. However, they may adopt the crucial variant from the innovators due to its (covert or overt) prestige, and because, although change is 'risky' and could lead to rejection, it bears less risk if the variant involved is already characteristic of speakers on the fringes of the group. Once the socially central, influential early adopters have begun to use a variant, other members of the group who may regard them as role-models are likely to follow suit. On this analysis, Labov fails adequately to account for the early stages of change because he does not distinguish innovators from early adopters – although, as the Milroys note, this distinction may

only be possible in principle; in the case of a particular change, a given group of speakers may not be identifiable as innovators or early adopters.

The Milroys' theory has a number of advantages. Notably, it makes the strong claim that 'linguistic change is slow to the extent that the relevant populations are well established and bound by strong ties, whereas it is rapid to the extent that weak ties exist in populations' (Milroy and Milroy 1985: 375). This may account for the existence of conservative and innovating languages, such as Icelandic and English respectively, and for periods of rapid and slow change in the history of one language. By extension, it may also explain the more innovative nature of urban dialects as compared with rural ones (and the preoccupation of traditional dialectology with the latter). Finally, this hypothesis predicts that language contact, migration and invasion should promote linguistic change, and (as Chapters 8, 10 and 11 show), this does seem to be the case.

However, there is also a considerable drawback: weak ties, by their very nature, are hard to investigate, and the initial spread of a variant from one group to another by a mediating innovator, the real actuation of the change, may therefore not be observable; we now know where to look, but can't see anything. That is, 'for the very reason that persons who actuate linguistic change may do so in the course of fleeting, insignificant encounters with others occupying a similarly marginal position in their social groups, direct observation of the actuation process may be difficult, if not impossible. What we most probably CAN observe is the take-up of the innovation by the more socially salient EARLY ADOPTERS' (Milroy and Milroy 1985: 370). For instance, as we saw above, the speakers in the Clonard area of West Belfast with the highest instances of backed /a/ are the young women. These also have a very high network strength score, are part of a closeknit community, and are therefore unlikely to be innovators; this is supported by the fact that backing of /a/ seems to be spreading from the East of the city. However, the young Clonard women may well be early adopters. They tend to be employed as shop assistants in a store with customers from East and West Belfast, and are therefore in contact with many speakers who have only weak ties to their community and to others. These weak ties can act as channels down which innovations pass; through them, the young Clonard women have adopted backed /a/ and are now passing it on to other members of their own community.

Thanks to Labov and the Milroys, we now have a general theory of the transmission of change, and at least an idea of where actuation fits in. First, variation arises, constrained by perceptual, acoustic and other general linguistic conditions on possible innovations (Harris 1986). This is not yet actuation, but speaker innovation. Some innovations spread, while others do not, at least partly for reasons of prestige; the projected link of network

theory with a general theory of social structure (J. Milroy 1992) may help us understand which changes are likely and why. Change is then actuated as it spreads from the innovator to other social groups, where it will be accepted by the early adopters, then diffuse to the other members. Gradually, the variation involved becomes orderly, the variants acquire social meaning, and change becomes apparent in the system – at which rather late stage it becomes detectable by sociolinguistic observation. Earlier stages are not yet observable or altogether explicable, and previous studies claiming to observe actuation have probably, at best, been noticing early adoption. We have, then, a framework for the transmission of gradual, socially motivated language change, and can observe and measure its progress. Actuation, the initial spread of an incipient change by innovators through weak ties with other groups, remains almost as mysterious and unattainable as ever.

10 Pidgins and creoles

10.1 Introduction

The pidgin English as spoken in these days is about the most atrocious form of speech perhaps one could find in any corner of the globe. It is neither one thing nor the other. Consisting of a mixture of Samoan and Chinese here and there, with an occasional word of Malayan, it is conglomeration truly worthy of the Tower of Babel

(*Rabaul Times*, 16 October 1925; quoted Romaine 1988: 10).

This quotation, taken from a newspaper published in Papua New Guinea, specifically involves Tok Pisin, the pidgin language of New Guinea which is now becoming a creole and which has been recognised as one of the official languages of the country. However, the criticisms contained in it are characteristic of earlier attitudes to pidgin and creole languages, both inside and outside linguistics. Popularly, pidgins and creoles are often still written off as 'inferior, haphazard, broken, bastardized versions of older, longer established languages' (Todd 1990: 1); and although the first systematic studies of pidgins and creoles were carried out in the 1880s by Hugo Schuchardt, these languages were subsequently shunned by linguists, who saw them as aberrant and uninteresting, for the best part of a century. The first conference on pidgin and creole studies took place only in 1959, and most of the key ideas in the field date from this time or later. The study of language birth is therefore a new, expanding, and often controversial one.

It may be useful to begin with some preliminary definitions. Approximately 200 pidgin and creole languages are spoken today, mostly in West Africa, the Caribbean and the South Pacific. A pidgin is essentially a contact language, developed in a situation where different groups of people require some means of communication but lack any common language. Crucially, a pidgin is nobody's first language; all its speakers learn it as adults as a second or further language, and all have native languages of their own. Pidgins are typically restricted in function, since they are used only in contact situations; for instance, in communication between masters

Published by the Press Syndicate of the University of Cambridge
The Pitt Building, Trumpington Street, Cambridge CB2 1RP
40 West 20th Street, New York, 10011-4211, USA
10 Stamford Road, Oakleigh, Melbourne 3166, Australia

© Cambridge University Press 1994

First published 1994

Reprinted 1995, 1996

Printed in Great Britain at the University Press, Cambridge

A catalogue record for this book is available from the British Library

Library of Congress cataloguing in publication data

McMahon, April M. S.

Understanding language change / April M. S. McMahon.

p. cm.

Includes bibliographical references and index.

ISBN 0 521 44119 6 (hardback) ISBN 0 521 44665 1 (paperback)

1. Linguistic change I. Title.

P142.M38 1994

417.7-dc20 93-8121 CIP

ISBN 0 521 44119 6 hardback

ISBN 0 521 44665 1 paperback

To Rob

At the still point of the turning world ...

... at the still point, there the dance is

(T. S. Eliot, *Burnt Norton*)

