1 Introduction: language variation

1.1 Synchronic variation

All languages that we can observe today show variation; what is more, they vary in identical ways, namely geographically and socially. These two parameters, along which variation occurs, are in principle independent of each other, although we shall see that there are ways in which they (and others to be discussed later) are interlinked. We shall consider each in turn.

1.1.1 Geographical or diatopical variation

It is a universal characteristic of human language that speakers of the 'same' language who live in different parts of a continuous territory do not speak in the same way.[1] Careful observation shows that such variation is usually smooth and gradual: the speech of each locality differs in some feature or features from the speech of each neighbouring locality, but without seriously impairing mutual comprehension.[2] Successive small differences accumulate as one crosses an area, and in an extensive territory this accumulation of differences may result in total mutual incomprehensibility between the speech belonging to distant parts of the territory being examined.

We shall see in Section **4.1.2** that the northern part of the Spanish Peninsula displays this kind of variation; that is, we can observe there what is known as a *dialect continuum*. A village-by-village journey from the west coast of Galicia to the Costa Brava reveals at each stage only small linguistic differences between a particular village and its neighbours on either side, these differences being few where communications are good between the villages concerned and more numerous where communications are poorer. Provided one skirts the Basque Country (where one faces forms of speech unrelated to those which surround it), there is no point on the journey where mutual comprehension between

speakers from neighbouring villages is threatened, even though speakers will often be aware, sometimes acutely, that their neighbours speak a little differently from them. The greater the distance travelled, the greater the total number of differences between the speech of one's present location and that of one's starting point, and such accumulation of differences causes a correspondingly increased degree of mutual incomprehension, to the extent that the speech of a Galician fisherman will be barely understood, if at all, by a fisherman on the coast of Catalonia.[3]

In fact, dialect continua are not only unaffected by internal administrative boundaries (such as those which divide Galicia or Catalonia from the rest of Spain), but also pay no heed either to national frontiers. The northern Peninsular dialect continuum is part of a broader Romance continuum which extends in unbroken fashion over all the European territory where descendants of Latin are spoken (with the exception of now-isolated varieties of Romance such as Rumantsch in Switzerland and the various kinds of Romanian used in Romania and other parts of the Balkans). At the level of everyday rural speech, the Pyrenees do not form a frontier; the varieties spoken on the northern and southern flanks of the central Pyrenees have long been known to be similar and, to a substantial degree, to be mutually intelligible (Elcock 1938). Similarly, in the eastern Pyrenees, there is close continuity between the speech used on Spanish territory and that used in neighbouring parts of France; we are here discussing the way in which Catalan straddles the political frontier.

It will be appreciated from this discussion that geographical variation is a two-dimensional phenomenon. Although our main example (a journey across the northern Peninsula) presents linguistic variation in one dimension only, the fact is that variation is observable in whatever direction or combination of directions one moves across a territory.

1.1.2 Social variation

It is also evident, from even casual observation, that in any one place not all people speak alike, even if they were all born there. Differences of speech are correlated with one or more social factors which apply to the speaker concerned. These factors include age, sex, race, class background, education, occupation, and income. To take an example, Spanish participles in *-ado(s)* (and some other, similarly structured, words) reveal a range of pronunciations; the final segment of words

like *cansado, pescado* may be pronounced in one or other of the follow-
ing ways: [-áðo], [-áᵟo], [-áo], [-áu̯]. But the appearance of one or other
of these variants is controlled (at least in part) by the sociological char-
acteristics of the speaker. Thus, the variant [-áu̯] is much more frequent
in working-class speech than in that of the middle classes; similarly, in
certain studies of this phenomenon (Williams 1983b, 1987: 71), women
of all classes are seen to be substantially more resistant to total deletion
of the consonant than are men.[4]

It follows from this brief account of social variation that such vari-
ation is multi-dimensional; there are many parameters which define
the social 'space' within which the speaker is located, and his or her
speech varies, in different ways, in accordance with each of these para-
meters.

We shall see shortly (**2.5**), however, that even a single individual
does not use just a single variant from the range of those available in the
community. Rather, each individual commands at least part of the
range and selects a particular variant according to the circumstances
(formal, informal, relaxed, etc.) in which he or she is speaking. And even
in the same speech environment, a speaker may alternate between two
or more variants.

1.2 Diachronic or historical variation

All languages for which we have information (e.g., written records or, in
the last hundred years, recordings) which is spread over a period of time
show more or less rapid change. The traditional view of such linguistic
change was that one variant succeeded another in the community con-
cerned, so that one could establish a chain of events in which each form
was replaced by its successor. Such a chain is typically expressed thus:
Latin LĀTUS > Hispano-Romance [ládo] > medieval Spanish [láðo] >
modern Spanish [láᵟo] or [láo]. As a summary of what has happened
over time to particular linguistic features, particularly in highly codified
languages, such a statement is not unreasonable.[5] But closer examina-
tion of recent language development has revealed that, at any moment
of time, a feature which is undergoing change is represented (in the
community and in the speech of individuals) by two or more competing
variants. Change takes the form of the addition of further informal
variants and the loss over time of the most formal variants.[6] Linguistic
change can therefore be pictured as the replacement of one state of

Stage 1	[láto]	[ládo]				
Stage 2	[láto]	[ládo]	[láðo]			
Stage 3		[ládo]	[láðo]	[láᵟo]		
Stage 4			[láðo]	[láᵟo]	[láo]	
Stage 5			[láðo]	[láᵟo]	[láo]	[láu̯]
* Stage 6				[láᵟo]	[láo]	[láu̯]
* Stage 7					[láo]	[láu̯]

Table 1.1   Model of diachronic variation

Stage 1	[nído]	[níðo]		
Stage 2	[nído]	[níðo]	[níᵟo]	
Stage 3		[níðo]	[níᵟo]	[nío]
Stage 4		[níðo]	[níᵟo]	
Stage 5		[níðo]		

Table 1.2   Regressive development

variation by another. To take the previous example, we can restate the change which leads from LĀTUS to *lado* in the (deliberately oversimplified) way shown in Table **1.1**, in which Stage 5 represents the present and Stages 6 and 7 have not yet been reached, but are tentatively predictable.

Note that it is not claimed here that change exclusively progresses through the addition of newer variants and the loss of older ones. There may be blind alleys or reversals. That is to say that variants which are added at a certain stage to the range of existing variants may be subsequently lost while older variants remain. This kind of process can be seen in the history of words like *nido*, and others whose intervocalic consonant descends from Latin -D- (see Table **1.2**).

Many words offering intervocalic -D- in Latin show the smoother development in which the variants with some internal fricative are dropped after Stage 3, leaving the variant with no internal consonant to descend into the modern language (e.g., SEDĒRE > *ser)*. Other words, however, followed the pattern outlined for *nido,* frequently appearing

without /d/ in the Middle Ages (CRŪDU > *crudo* > *crúo,* VADU > *vado* > *vao*), and then appearing to go into reverse, leaving behind only *nido, crudo, vado,* etc. Such reversals are impossible to conceive, I suggest, outside a variationist framework.

It will be evident from this discussion of diachronic variation that such variation is not independent of geographical and social variation, in the way that geographical and social variation are independent of one another. In particular, diachronic variation results from social variation (see note 6) and is inconceivable without it.

It also needs to be clarified that, since change proceeds item by item, each change occupying a different segment of time in a particular community, while the same change will occupy different segments of time in different communities, all notions of periodization are misplaced in language history. Although we are far from understanding all the factors which hasten or restrain linguistic change (but see **3.3**), it seems fairly certain that at some places and times change is more rapid than at other places and times; that is to say that in the history of a particular variety there will be changing rates of innovation. However, the way in which linguistic innovations succeed one another, without exactly coinciding, implies that there can be no linguistic basis for dividing one period of that history from another. It may be a convenience, in the interests of relating language history to political and cultural history, to refer separately to, say, Old Spanish, Golden-Age Spanish, or Modern Spanish, but such periodization can have no linguistic motivation. Linguistic development is as seamless as all other cases of linguistic variation (Penny 1998).

1.3 Variables and variants

All aspects of language (sounds, phonemes, morphemes, syntactic structures, lexemes, meanings, etc.) are subject to variation according to these parameters. A linguistic feature which displays variation according to one or other parameter is called a variable and is indicated by a symbol between parentheses. For example, the phoneme /x/ of Spanish, the *jota*, varies geographically in its articulation, being pronounced in some places as the velar fricative [x], in others as the glottal fricative [h], and in yet others with sounds intermediate between [x] and [h], or as the palatal fricative [ç]. We can therefore say that the variable (x) (or (h)) is realized (in different, specific places) as [x], [h], [h^x], [ç], etc.

1.4   Co-variation

The parameters of linguistic variation are independent, but a feature which shows variation according to one of these parameters (say, the geographical dimension) may show similar or identical variation along another (say, a social or diachronic dimension). Thus, the feature known as *yeísmo* (see **4.1.7.2.2**, **4.2.1**, **5.1.2.1**, **6.3.3(2)**) can be described as showing variation along all three.[7] For some speakers, a meaningful contrast is available between the phonemes /ʎ/ and /ǰ/ (*pollo* 'chicken' vs *poyo* 'stone bench'), while for others these phonemes have merged, and a single articulation is used for both sets of words (frequently [ǰ], but also [dǰ], [ʒ], etc.). Variation between distinction of these phonemes and their merger is, firstly, geographical: in rural areas of the northern half of Spain, in the Andean area of America, etc., distinction is found, whereas in the larger part of the Spanish-speaking world merger is the norm. However, the same variation can be observed along sociolinguistic parameters: older, middle-class, urban speakers from the north of Spain use distinction between /ʎ/ and /ǰ/, while younger speakers from the same cities, whatever their class background, allow the phonemes to merge. Likewise, the same variability can be seen over time: several centuries ago, all speakers of Spanish no doubt distinguished words with /ʎ/ from those with /ǰ/ (e.g., *pollo* from *poyo*), while at some stage in the future all speakers of Spanish will no doubt have allowed the two sets of words to merge.

The implication of this three-fold variation is that over time *yeísmo* has progressed geographically (occupying more and more territory), and socially (affecting the speech of more and more members of society in any given locality).

1.5   Register

No speaker uses the resources of his or her language in exactly the same way on all occasions; according to the social circumstances in which the act of communication occurs, the speaker may choose different variants of a particular variable. More precisely, register variation appears to be as multidimensional as social variation. Halliday (1978: 33) distinguishes three parameters of register variation: 'field'

(within which, variation is determined by the purpose and subject matter of the communication), 'mode' (which controls variations due to the channel, written or spoken, of the communication), and 'tone' (according to which, variation is determined by the person to whom the communication is addressed). Thus, in choosing particular features of language with which to communicate, the speaker/writer places himself or herself at a particular position in a complex social matrix.

Of course, the range of variants between which a speaker/writer chooses in any act of communication may be similar or identical to the range of variants strung along any of the parameters already discussed (the geographical, the social, and the historical). Thus, to take the case of *yeísmo* (discussed in **1.4** as an example of geographical, social and historical variation), the speaker who in formal circumstances (delivering a lecture, say, or speaking to people he or she is seeking to impress) distinguishes the medial phonemes of *malla* and *maya* may pronounce these two words identically one to another when speaking informally (that is, in relaxed circumstances, with friends, etc.). Similarly, the different variants discussed in **1.1.2** in connection with words like *pescado* (currently [-áðo], [-áᵟo], [-áo] and [-áṷ]) also correspond with different points in the communicative matrix: speakers who command all four variants will use the first only in formal or fully monitored speech, the second when a moderate degree of formality is felt to be required, and the last two only in unmonitored, relaxed speech.

It is this kind of register variation which gives rise to *hypercorrect* forms. For example, since the word *bacalao* shares some of the range of variants also shown by *pescado* (namely [bakaláo] like [peskáo], [bakaláṷ] like [peskáṷ]), the similarity may be extended to the full range. Thus, in communicative circumstances which require care or formality, such as speaking to a stranger, the pronunciation [bakaláðo] may be used, matching formal [pescáðo]. Since hypercorrect forms are most usually produced by the illiterate, who by definition cannot be guided in their pronunciation by the standard written forms of words, they are usually heavily stigmatized.[8]

1.6 Variation in the past

Since it is the case that all languages observable today or in the recent past show all the kinds of variation discussed here, we are entitled to conclude that such variation must be true of all languages that have

ever been spoken, in all places, at all times. This principle cannot be tested, since linguistic evidence from the past (except the very recent past) comes only in written form, and such written evidence is incapable of showing more than a small fraction of the range of variation we assume to have existed. In particular, each piece of written evidence will typically reflect the formal register (because written) of a particular user of the language concerned, a user who must, of course, reflect the variants in use only at one place, in one social milieu, at one moment. Comparison of different pieces of historical evidence can amplify the range of variation observable, but can never come close to establishing the full range of variation which must have existed at each moment in the past.[9]

2   *Dialect, language, variety*: definitions and relationships

A common perception, among those who are not linguists, is that there is some difference in kind between a 'language' and a 'dialect'. The question is often posed in the following form: 'Is *x* a language or a dialect?', where *x* is some such label as 'Valencian', or 'Asturian'. And it is a question which the linguist, as linguist, cannot answer, first because of the insuperable difficulty of defining the concepts *language* and *dialect* (see **2.1** and **2.2**), but secondly because any difference between these concepts resides not in the subject matter of linguistic description, but in the social appreciation accorded to particular codes of communication. The historical linguist will make it clear that every code to which the label 'language' is attached (e.g., 'the Spanish language', 'the English language', 'the French language', 'the Latin language') has its origins in what would usually be called a 'dialect', loosely defined in terms of geography (as the speech of a particular locality or area) and in terms of social class (as the speech of a particular social group, usually the dominant, educated, classes). Thus, the French language has its origins in the speech of upper-class Paris, specifically of the Court.[1] If 'dialects' can gradually become 'languages', it follows that there cannot be any difference of kind between these concepts, but only differences of degree.

But degrees of what? A full answer to this question would duplicate the discussion in Chapter **7**, but it is perhaps in order here to anticipate the conclusion reached there. What the non-linguist means by a 'language' is most usually what is otherwise called a 'standard language', that is, a dialect which has undergone the various processes which together constitute standardization (selection, codification, elaboration of function, acceptance; see Haugen 1972; Hudson 1996: 32–4), all or most of which are inconceivable in the absence of writing. A 'language', then, differs from a 'dialect' only in the degree to which it has been subjected to each of these processes (although the process of selection should perhaps be disregarded here, since it is not a matter of
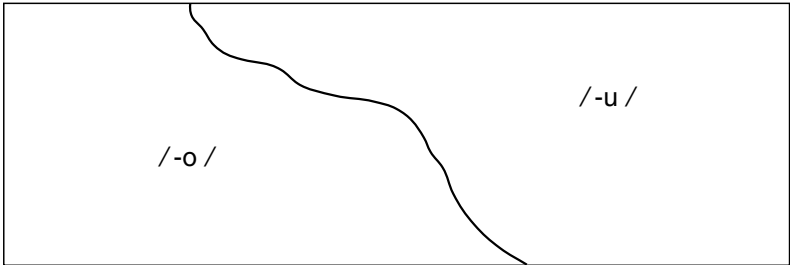
Figure 2.1   Territory divided by a single isogloss

degree). A 'language' will be more highly codified (it will possess such things as an agreed orthography, and prescriptive grammars and dictionaries), it will have an expanded vocabulary and more elaborate syntax (to allow the discussion of topics which are simply not handled by speech), and it will enjoy higher social prestige (because of its association with high-prestige activities, such as education, and with high-prestige sectors of society, such as the educated and the wealthy).

Although it is possible to define a *standard language* (along the lines of what is said in the previous paragraph), it will now be seen that there are insuperable problems in defining the concepts of *dialect* and *language* (as in *the Spanish language*, etc.). For further discussion of these concepts, see Alvar (1961).

2.1 *Dialects*

We have already seen (in **1.1.1**) that geographical dialects (that is, 'dialects' in the sense most frequently used by non-linguists) have no definable boundaries. Examination of data from linguistic atlases, such as the *Atlas lingüístico de la Península Ibérica* (*ALPI* 1962), reveals that each item (such as a word, a meaning, a sound, or an element of grammar) occupies an area which is usually continuous and almost always differs from the area occupied by any other item. To take a theoretical example, the territory represented by the box in Figure **2.1** is divided into an area where a large class of masculine singular nouns ends in /-o/, and a second area where the corresponding class of nouns ends in /-u/. The dividing line between these two areas is called an