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
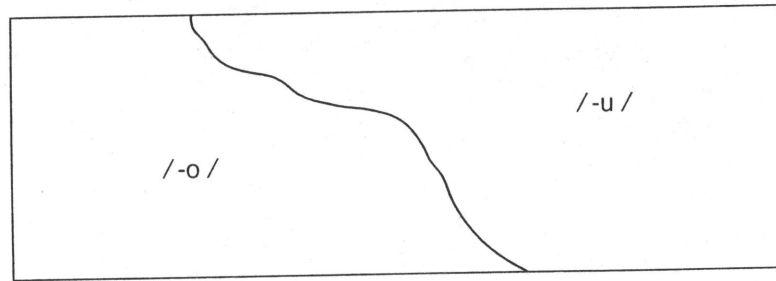
Figure 2.1   Territory divided by a single isogloss

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
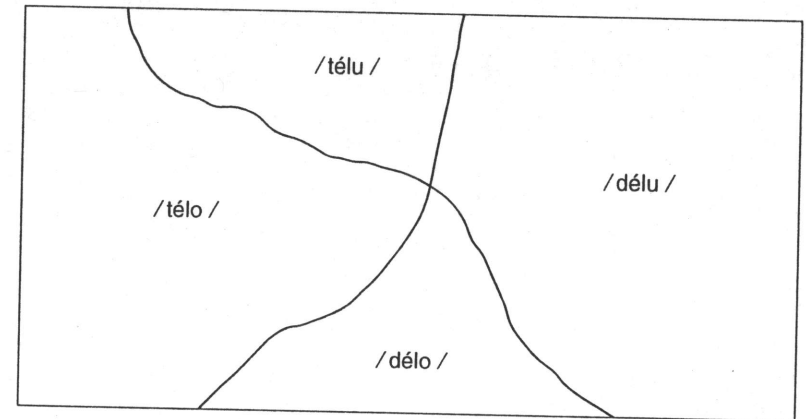
Figure 2.2   Territory divided by two isoglosses

isogloss (see **3.2.1**). Let us now imagine that the same territory is divided into localities whose speakers use the word *telo/u* while people from the remaining localities say *delo/u* to express the same concept. Since it is overwhelmingly improbable that the line separating localities showing /-o/ from those with /-u/ will exactly coincide with the line separating localities where *telo/u* is used from those where *delo/u* is used, the consequence is that our territory is divided, on the basis of only two items, into four 'dialects', as in Figure **2.2**. The mapping of each further item used in the territory will double the number of identifiable 'dialects' used there, and given that the language of any locality consists of at least several thousands of items, it follows that the number of 'dialects' identifiable in a real territory of any extent is infinite. Therefore, unless we restrict the meaning of *dialect* to 'the speech of a specific locality' (which, in turn, we shall see is unsatisfactory), we are forced to reject the notion of a *dialect* as a discrete or separately delimitable entity. To put the matter only slightly more strongly, there is no such thing as a dialect. It will be possible to talk about, say, the 'dialects of Castile', in the sense of the totality of the speech varieties used within Castile, but without any implication that there exist separate identifiable dialects within Castile, or that the dialects of Castile do not merge imperceptibly with those of surrounding areas, such as Aragon, Leon, or Andalusia.

The term *dialect* has sometimes been used to distinguish types of speech which are differentiated by social factors. In this sense, the term *dialect* (sometimes *social dialect*) alternates with *sociolect*, typically to

distinguish the speech of one social class, or one age group, etc., from another. But as in the case of the geographical dimension, the various social dimensions provide us with no basis for demarcating one social dialect from another. On the contrary, the speech of one social group merges imperceptibly with others, just as the speech of any age group fades into that of younger and older people. This observation does not deny that there are correlations, often strong, between the distribution of linguistic items and such social features as age, social class, etc. But the transitions between the speech of 'adjacent' social groups are smooth and not abrupt. In the social sense too, the term *dialect* corresponds to no objective reality.

2.2   Languages

Our problem lies not so much in defining language in general (which we might define as 'the universal symbolic activity by which humans convey meanings from the mind of one person to that of another'), as with defining what is meant by '*a* language', or, for example, 'the Spanish language'. The problem is essentially one of delimitation: what are the temporal and geographical limits of, for example, Spanish?

2.2.1   Do languages have temporal limits?

Since linguistic change proceeds item by item (at one moment a feature of pronunciation, at another moment an item of grammar, then the addition or loss of a word, etc., but in no particular order), it follows that there can be no moment in the past at which any language can be said to have begun.[2] At any specific moment, the speech in use in a given community differs only in minor ways from the speech used a generation earlier.[3] To take the specific case of Spanish, there is no point, objectively arrived at, at which it can be said that Latin gives way to Spanish; at the level of spoken communication, there is no break in the continuity (with the usual minor modifications in each generation) which leads back from the present day to what we would call 'Latin' two thousand years ago (or indeed to the Italic, Indo-European and earlier ancestors of Latin). So why do we give the name 'Spanish' to recent stages of this continuous development and apply the label 'Latin' to earlier stages? There are two answers to this question.

Firstly, and more trivially, there is a need to distinguish what are

now regarded as distinct descendants of the same ancestor. It is well known that in the Middle Ages, the spoken descendants of Latin (and eventually also the written forms of these spoken descendants) were referred to collectively by nominalized reflexes of the Latin adverb ROMĀNICE (lit. 'in the Roman manner'), e.g., Sp. *romance* ('any oral descendant of Latin'), usually by contrast to the then most prestigious (and exclusively written) form of language, namely *latín*.[4] Since people in the Middle Ages were evidently as aware as we are of geographical variation, it was often desirable to specify which kind of *romance* was under discussion. This was achieved by adding an adjective referring essentially to political entities (and by implication to geography), thus: *romance castellano, romance leonés, romance aragonés*, etc. And since, in phrases like [to speak] *en romance castellano*, the word *romance* was redundant, the phrase being perfectly unambiguous without it, it was eventually dropped, with the result that the politico-geographical adjective (*castellano*, etc.) became the name of a form of language.[5]

The second (and more important) reason why a separate name was required for some varieties of Romance springs from the fact that these varieties underwent standardization. As we shall see (Section **7.1.2**), standardization is a process which is inseparable from writing (purely oral varieties never undergo standardization), and the identification of a written code sharpens speakers' awareness of the newly codified variety as a separate entity requiring a separate name.[6] So although Castilian is a variety of Romance which (like all other such varieties) results from an unbroken series of earlier varieties stretching indefinitely into the past, the fact that at a certain point it achieved written status and underwent increasing standardization forced upon its users, and others, the need for a name by which to identify it and differentiate it from other written linguistic codes (Latin, Catalan, French, etc.).

2.2.2   Are languages delimitable?

If we mean by this question 'Do official languages have spatial limits?', then the answer is obviously yes, since languages can only be made official by political entities, such as nation-states, and their officiality is usually co-terminous with that entity. Thus the official language of the Spanish state stretches exactly to the Pyrenean frontier and there abuts abruptly on the official language of the neighbouring state. But this kind of sharp linguistic boundary, at which a piece of writing produced on one side of a frontier consists of a set of items which is different

from the set of items which characterize a piece of writing produced a few hundred metres away on the other side of the frontier, provides a very unsatisfactory basis for the delimitation of languages. Coincidence between national frontiers and boundaries which separate official languages is anyway relatively rare (and perhaps only occurs at all in the Europe of the last two centuries). So what other basis is there for delimiting one language from the rest?

One criterion that is sometimes used is that of *mutual intelligibility*; if one speaker does not understand another, then they speak different languages. But the problem with this criterion is that mutual intelligibility is a question of degree rather than being an all-or-nothing matter. Speakers of Spanish will understand a good deal of what a speaker of Catalan says, yet on the basis of the criteria outlined at p. 9 above to define a standard, both Spanish and Catalan would qualify. Similarly in the case of Spanish and Portuguese: speakers of Spanish understand at least some of what is said in Portuguese, and the Portuguese speakers will understand a good deal more of what is said in Spanish (which goes to demonstrate that mutual intelligibility can be asymmetrical: an absurd conclusion would be that this shows that Portuguese is more different from Spanish than Spanish is from Portuguese!).

There is also the matter of experience. Different speakers of Spanish will have different experience of, say, Portuguese, and will therefore understand spoken (and written) Portuguese in different measures. So that mutual comprehension, or its lack, is a quite unsatisfactory means of marking off putatively distinct languages.

What often underlies the layman's view that language A and language B are separate entities is the fact that these two 'languages' have distinct orthographic systems, especially since a common lay view is that a particular variety can only be accorded the status of 'language' if it is a variety which appears in written form.[7] However, since it is perfectly possible for distinct orthographical systems to be applied successively to the same variety, or even simultaneously as in the case of Serbian and Croatian, it is clear that the orthographical principle cannot serve as a satisfactory criterion for delimiting one 'language' from another.

In the end, we are forced to a similar conclusion to the one we were forced to reach in the case of 'dialects': there is no purely linguistic means of delimiting one 'language' from another, since closely related 'languages' form part of a continuum and any dividing line

which cuts through this continuum is drawn not for linguistic reasons but for political reasons.[8]

2.3　Relationship between *dialects* and *languages*

If dialects cannot be delimited in space and languages cannot be delimited in time or space, what are we to make of such commonly used formulations as 'X is a dialect of language Y'? What underlies such statements as 'Andalusian is a dialect of Castilian' is a significant historical misapprehension, namely that over time 'languages' are fragmented into 'dialects'.

The reason underlying this misapprehension is a failure, albeit an understandable failure, to compare like with like. It stems from the fact that our direct knowledge of past linguistic states comes to us through writing, and writing almost exclusively preserves standard languages.[9] By their nature, standard languages are the result of processes (see Chapter 7) which have reduced variation to as near zero as possible, so that the picture we receive of past language-states is one of linguistic uniformity. However, our examination of current language-states reveals a picture of variation along a host of parameters. It is therefore tempting to conclude that an earlier state of uniformity has been 'degraded' or 'debased' into a state of variation.[10] Thus, for example, the perceived uniformity of Latin is judged to have broken down into a large number of medieval Romance dialects, and in a situation where (written) Latin continued to be the standard, such dialects could be considered to be dialects 'of Latin'. Likewise, in the case of the Spanish of America, it is often thought that the present language variation results from the 'debasement' of some supposed earlier state of uniformity, although in this case such a supposition is less frequent, because we have written evidence of the variation within the Spanish that was carried to America from the sixteenth century onwards.

However, such a view of increasing fragmentation over time is clearly erroneous. If it is agreed, as argued here (**1.6**), that variation of all types (geographical, social, etc.) has always existed in human language, at all times and in all places, then it follows that linguistic development takes the form of change from one state of variation to another state of variation, even though one can argue over the different degrees or ranges of variation which apply in successive stages of a particular development. A further consequence is that it is illogical to label

any variety A as a 'dialect of language X', or any set of varieties, B, C, D ..., as 'dialects of language Y', if (as is usually the case) the unspoken assumption is that there is a historical relation between A and X or between B, C, D ... and Y, such that (in some sense) A springs from X, or B, C, D ... are developments (often 'debasements') of Y.

This view is to stand history on its head. If we take standard languages such as our X or Y, each has its origin in a specific local/social variety, which has been selected (for non-linguistic reasons) from a host of other competing varieties (see **7.1.1**). So each 'language' can be said to descend from a 'dialect', rather than the reverse. It is therefore meaningless to say that the spoken varieties used in, say, Soria or La Mancha are 'dialects of Spanish', since this implies a false historical relationship between each of these varieties and Spanish (i.e., the standard language which has its origins in the dialect of Burgos, later transferred (with modifications) to Toledo at the time of the Reconquest and later still becoming codified into the standard language of Castile and subsequently of the Spanish state).

It is equally meaningless to enquire whether a particular variety, say one used in the Pyrenees, is a 'dialect of Spanish' or a 'dialect of Catalan', since such formulations imply historical descent from either Spanish or Catalan, which is nonsense. A particular Pyrenean variety will have a certain number of items in common with standard Catalan, and a certain number in common with standard Spanish (as well as a number in common with both, and a few in common with neither), so that all one can do is to attempt to measure the degrees of affinity between the variety in question and each of the two prototypes labelled 'Spanish' and 'Catalan'. However, this is not an easy task (and is perhaps impossible) since not all the thousands of items of which any variety consists can be taken into account, and there is no agreed basis upon which to give different weight to different classes of items (say, sounds over vocabulary, or items of syntax over items of word meaning).

We have so far found no justification for formulations of the type 'variety A is a dialect of language X', but it might be thought that such justification could be found in cultural history. Might we claim that such statements are meaningful by reformulating them in the following way: 'variety A is a dialect of language X where A is spoken in a territory in which X is the standard language'? Clearly we would have to add the rider that A and X must be historically related, otherwise our formulation would claim that, say, Basque or Quechua were dialects of Spanish, a claim no one would wish to make. But even after making

this restriction, our formulation still gives unacceptable results, since it entails that we classify any variety spoken in, say, Galicia or Catalonia as dialects of Spanish, again a proposition to which few would subscribe.

So 'dialects of Spanish', we must conclude, are spoken in an area which is smaller than that within which Spanish is the standard language (or, at least, smaller than the area within which Spanish is one of the standards). But how much smaller? What about varieties spoken in, say, Zamora or Saragossa? Here it is crucial to remember that in almost all contexts the label 'Spanish' is interchangeable with 'Castilian'; so that to claim that the varieties used in Zamora or Saragossa are dialects of Spanish is also to claim that they are dialects of Castilian. Is this an acceptable statement? Dialectologists and language historians would deny that it is; the speech of Zamora represents the local development of those forms of Latin introduced into the northwestern part of the Spanish Peninsula some 2,000 years ago, a development which was at first quite independent of those other developments which led from the Latin of north-central Spain to the dialects of Old Castile (including the dialect of Burgos, which was to become 'Castilian' *par excellence*). Later, it is true, and especially from the thirteenth century onwards, the speech of Zamora underwent progressive castilianization, that is, the introduction, item by item, of Castilian features spread from central Castile in wave-like manner (see **3.2**). This process is ongoing and still incomplete, so that the rural speech of Zamora (and of other areas outside Castile) preserves many pre-Castilian features (González Ferrero 1986). It is therefore quite unsatisfactory to claim that the varieties used today outside Castile are 'dialects of Castilian'.[11]

Even within Old Castile, it remains unsatisfactory to assert that the varieties in use, say, in Palencia are 'dialects *of* Castilian'. They might be called 'Castilian dialects', but only in the sense that they are varieties spoken in Castile, i.e., using 'Castilian' in a purely geographical sense. In principle, the history and status of the speech of Castile, *vis-à-vis* the standard, is no different from the history and status of varieties used elsewhere in Spain; each locality in Old Castile has a linguistic history which is a little different from that of the cultural and political centre, Burgos, whose variety of speech influenced that of its neighbours, item by item, as the speech of Burgos increasingly became the model to be imitated elsewhere, owing to the prestige of those who spoke in the Burgalese manner, a prestige which (as we shall see in **7.1.1**) was due entirely to cultural, political and economic factors and not at all to any inherent qualities of that variety. It is true that the

influence of the speech of Burgos on that of the rest of Old Castile is deeper than its influence on the speech of areas outside Castile, but this is due only to the obvious fact that localities within Castile were in closer contact with the prestige centre than were localities in Leon or farther afield. So the speech of, say, Soria retains fewer of the features it must have had, prior to the establishment of Burgos as a prestige centre, than is the case with the speech of, say, Zamora. But the relationship between the varieties used in Soria and the standard, on the one hand, is of the same kind as the relationship between the varieties used in Zamora and the standard, on the other. Only the degree of approximation to the standard is different, and in neither case is it appropriate to speak of them as 'dialects of Castilian'.

The formulation 'variety A is a dialect of language X' therefore has no validity. The phrase 'dialect of X' should be limited to cases where X is a geographical term and not a language name, as in the 'dialects of Castile', which should be taken to mean 'that range of speech varieties used within the borders of Castile'. This formulation carries no implication that there will be any coincidence between any isogloss (see **3.2.1**) and any part of the Castilian border; it is axiomatic that the Castilian border (like all other politico-administrative boundaries) arbitrarily bisects the dialect continuum.[12]

2.4 *Varieties* and *idiolects*

The definition of the term *variety* is no easy task, since, as we have seen in Sections **1.1.1** and **1.1.2**, there is no linguistic basis upon which one geographical dialect can be delimited from others, nor are social dialects discrete entities which can be distinguished one from another. Variation in speech extends gradually and smoothly through both geographical and social space, and does not present boundaries between varieties, only more or less rapid transitions along the geographical and social parameters. Even if one limits consideration to a single locality (a town or a village, say), it is evident that each individual living there locates himself or herself at a different point in the social matrix and therefore makes use of a slightly different set of linguistic items from that controlled by any other individual; that is, every individual uses his or her own *idiolect*.

In fact, we have seen (in Section **1.5**) that the linguistic universe is even more amorphous than this, because each individual deploys a

different set of linguistic items in each different speech situation, depending on the degree of formality which is felt to be required.

So how is the term *variety* to be used? It is used here to denote any set of linguistic items used in a specified set of social circumstances. These circumstances may be broadly or narrowly defined, so that all of the following, and many more, can be regarded as *varieties*: the English language, the Spanish language, American Spanish, Mexican Spanish, middle-class Spanish, the Spanish of the oldest generation, Valencian, Andalusian Spanish, the Spanish of auctioneers, the idiolect of a particular individual, standard Spanish.

2.5 **Relationship between varieties**

From the previous discussion it will be seen that each linguistic variety consists of a set of items which differs minimally from each 'neighbouring' variety. That is, each 'neighbouring' variety will be similar to the variety in question, but will differ from it in one or more items. Here 'neighbouring' is taken to mean not just 'geographically adjacent' (i.e. adjacent along the diatopical parameter), but adjacent along any of the parameters which are correlated with linguistic variation, including the many social parameters (age, socio-economic status, education, etc.) as well as the parameter of registers.[13]

Each variety can be visualized as occupying a segment (however large or small) of the multidimensional 'area' constituted by the totality of the parameters or 'dimensions' which govern linguistic behaviour (two spatial dimensions, many social dimensions, together with the dimension of register). The segment may be small (for example, the set of items used by an individual in a given register), or large (say, the set of items used by all educated members of the Spanish-speaking world, or indeed the set of items shared by all speakers of Spanish). Each variety, except perhaps the last, shades almost imperceptibly into all neighbouring varieties.

It should not be thought, from the claims being made here, that the gradient of variation along any parameter is of necessity uniformly steep; the rate of variation may be steeper in one section of the parameter than in another. It is easiest to appreciate this point in the context of geographical variation, but the same principle can probably be applied to all parameters: equal distances do not imply equal degrees of variation (degrees of variation being measured by the number of items
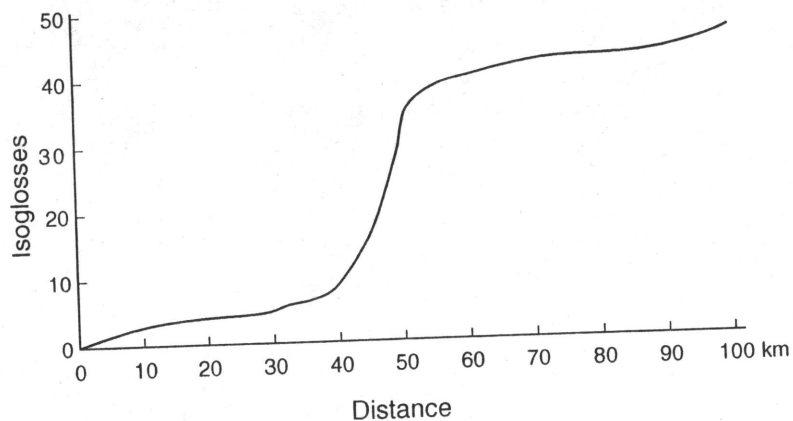
Figure 2.3 Variation gradients

that are not shared by adjacent varieties). The speech of any two points in space may be separated by more linguistic differences than those which separate another two points which are the same distance apart as the first pair. Looked at in terms of isoglosses (see **3.2.1**), which are the graphical representations of linguistic differences between places, it is not true to say that isoglosses are distributed with equal density in all parts of a territory. On the contrary, in some parts of the territory the isoglosses will be closer together (the gradient of variation will be steeper), while in other parts the isoglosses will be spread more sparsely, although in both cases it will be rare for two isoglosses to coincide exactly. This irregularity of gradient can be illustrated by the chart in Figure **2.3**, in which the vertical axis represents numbers of differences and the horizontal axis represents distance in space. And what is true of the spatial parameters of variation seems likely to be true also of the social parameters, although 'distance' here is a more problematical concept and social parameters have no agreed calibration.[14] That is, there may be more differences between the speech of two given individuals than between another two (elsewhere on the scale), even though the members of each pair are separated by the same social 'distance'.

2.5.1 The tree model of relationships between varieties

The model of the genealogical tree continues to be a frequent way of expressing the relationship between linguistic varieties, and will be
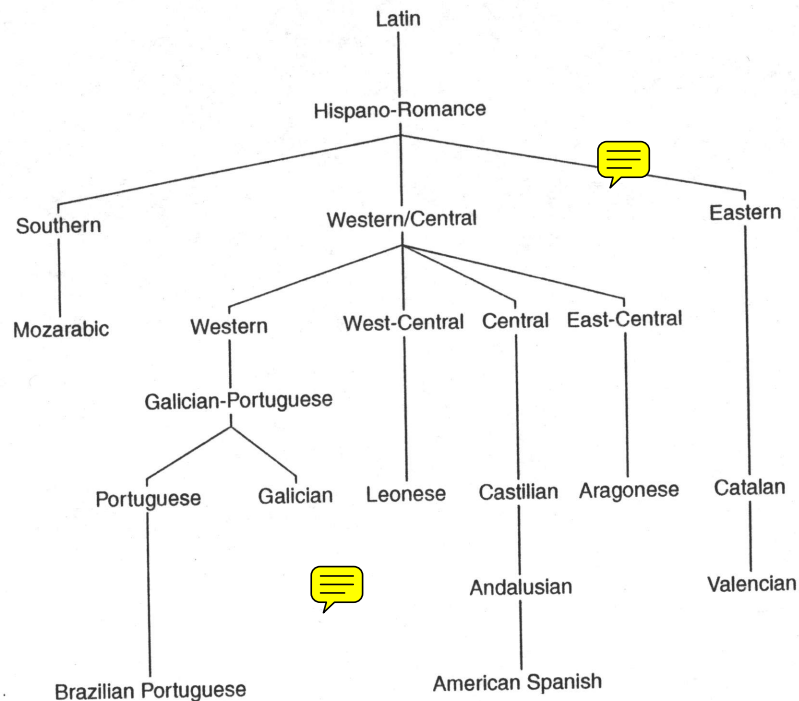
Figure 2.4 Tree model of Hispano-Romance varieties

discussed here in relation to diatopically related varieties, since there is no question (and there never has been) of using the genealogical tree to model the relationship between socially related varieties. The tree model has its origins in historical linguistics, where, since the early nineteenth century, it has been used to express the degree of historical relatedness between varieties (usually 'languages').[15]

A tree model employed to express the relationships between certain geographically distinguished varieties used in the Peninsula might have the appearance of Figure **2.4**. But the organization of such a tree is open to infinite discussion, argument, and potential re-adjustment, since a tree is capable of expressing only one spatial dimension (given that the vertical axis of such diagrams represents the passing of time), while linguistic varieties are located in two-dimensional space.[16] For example, anywhere that one places Mozarabic will be open to objection; the Mozarabic dialects developed in wide areas of the southern Peninsula and share a number of features with Leonese and Aragonese (and to a lesser extent with Galician–Portuguese and Catalan), features

which are not shared by Castilian (see Section **4.1.1**). On the other hand, although it is possible to place Galician–Portuguese, Leonese, Castilian, Aragonese and Catalan in an appropriate left-to-right array, reflecting their west-to-east distribution, the branches on which they are located in this diagram suggest degrees of relatedness which are open to challenge. Let us take a single case: Is Aragonese more similar to Castilian than it is to Catalan? This question immediately raises another: What variety of Aragonese and what variety of Catalan? And assuming that that question can be answered (a large assumption), how is one to measure the degree of difference between Aragonese and Castilian on the one hand and between Aragonese and Catalan on the other?[17] It would be theoretically possible (although actually impracticable) to list all the items over which each pair of varieties differed. But if we discovered that there were more items of difference between Aragonese and Castilian than between Aragonese and Catalan (or vice versa), would that solve the matter? Or would we wish to give greater weight to certain items than to others, since certain features strike us as more 'salient' or 'important' than others? In the absence of any principled way of assigning different weight or importance to particular features, such an enterprise is doomed to failure. And yet such a judgement of relative degrees of relatedness is inherent in the tree model. What is reflected in any given tree is a particular scholar's hunch or 'feel' for the various degrees of relatedness between varieties.

Not only is the tree model inadequate to express the relationships between diatopically related varieties, but it may seriously distort the diachronic and synchronic study of language. Some would argue that this model works well within Indo-European linguistics, where the varieties under consideration (all written and therefore partially or fully standardized) are usually well separated in space and time and where the intervening varieties have all vanished without trace, removing any possibility of viewing the Indo-European family as a continuum. However, where the object of study is a series of now-existing varieties or a range of closely related varieties from the past, the tree model is open to a number of grave objections.

2.5.1.1 Although the origins of the tree model lie in genealogy, it was its adoption by Darwinian biology that fixed this pattern so strongly in the consciousness of linguists.[18] Its use in linguistics therefore presupposes that speech varieties are like biological organisms: in placing speech varieties on the branches of a tree, we act as if we were biologists

ordering species by their degree of resemblance. This is a false analogy; biological species are sharply differentiated, because, for speciation to take place, the varieties between the surviving species have to be suppressed (because these are the varieties which are in unsuccessful competition with the variety which carries the mutation which confers advantage). But human language has nothing comparable with genetic mutation. Competition between dialects is not based on structural advantage but on non-linguistic factors like the socio-economic and cultural status of the users of those dialects. And dialects which are intermediate (geographically) between successful varieties do not usually disappear; they persist as rural or working-class varieties.

This should not be taken to imply that intermediate varieties remain unchanged; they are as subject to internally and externally motivated change as any other variety. But they continue to form a continuous chain between successful varieties. It is only through exclusive concentration on successful varieties (usually standard languages) that use of the tree model, with its denial of continuity between varieties, can be justified. What is more, awareness of this contradiction at the heart of geographical and historical linguistics is not new; it has been clear at least since the advent of linguistic geography in the last decade of the nineteenth century. But the image of the tree has exercised such a powerful pull that linguists working in this field are constantly in danger of operating simultaneously with self-contradictory models.

2.5.1.2 There is a second and more powerful reason for rejecting the tree as a model of linguistic relationships. The existence of branches presupposes the existence of a trunk, and this implies that the linguistic varieties which are located on the branches of the tree have a common, unitary origin, that they spring from a single original variety, once again as if linguistic varieties were akin to biological species. But such a pattern of development is evidently not what happens in linguistic history. To take an example, the Romance languages, like the members of any family, are the product of a language-state which must have offered all the variation (geographical and social) observable in every language we can examine in detail (see **1.1**). The range of variation may be greater now than 2,000 years ago, but it is becoming ever clearer that language history consists of the change from one state of variation to another state of variation, so that any insistence on the biological / genealogical model, with its single species/individual branching into distinct species/individuals, totally distorts linguistic reality.
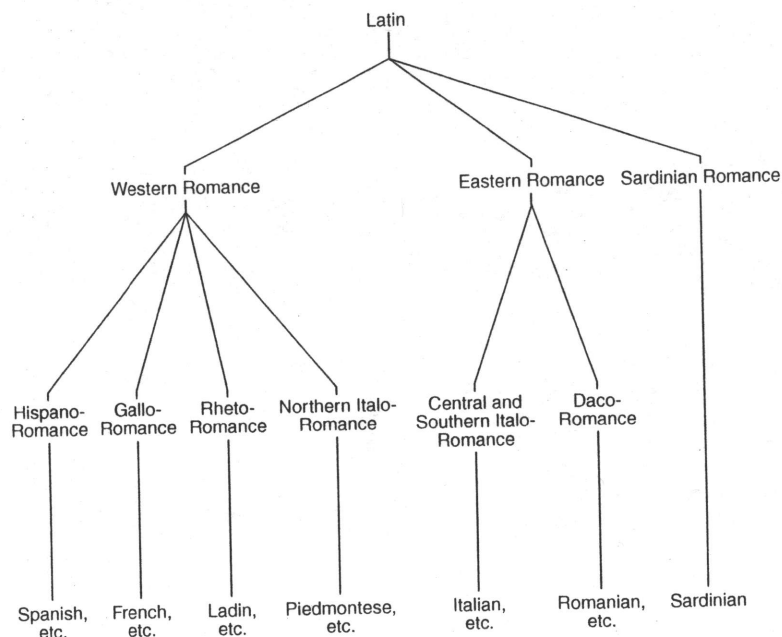
Figure 2.5   Tree model of the Romance family

A key example of this misconception can be found in traditional (but frequently repeated) classifications of the Romance language family. A common view of this family establishes a tree with three branches: a western branch, an eastern branch and a Sardinian branch, each with further branching, a simplified version of which appears in Figure 2.5. In accordance with highly respected and influential classifications, beginning with that proposed by Wartburg (1952), the division between Eastern Romance and Western Romance runs across the top of the Italian peninsula (the famous La Spezia–Rimini line). This division is based on distribution of only two features: first, the voicing of voiceless intervocalic consonants in Western Romance (but not in Eastern Romance), and second, the loss of final /-s/ in Eastern Romance (but not in Western Romance). However, if we examine the fate of the Latin voiceless intervocalics, we can see that the argument for an early bifurcation of the Latin tree into an eastern and a western branch (or its trifurcation, if one includes Sardinian Romance) is hard to sustain. One can find as least five reasons to dispute such a division:

1   Pompeiian graffiti show numbers of cases of *g* for expected *c*, and of *d* for *t* (Väänänen 1959, 1968: 102). It scarcely needs saying that

Pompeii is well south of the La Spezia–Rimini line. There is also more widespread evidence of early voicing of intervocalic plosives in the eastern Latin-speaking world (Cravens 1991).

2   The Tuscan dialects of central Italy, from one of which standard Italian principally descends, and which geographically belong to the eastern Romance branch, show frequent cases of voicing of the Latin intervocalics (*riva* < RĪPA, *grado* < GRATU, *ago* < ACU, etc.).[19]

3   Sardinian dialects, which are generally considered among the most conservative varieties of Romance, and therefore might be expected to agree in their treatment of intervocalic consonants with the Eastern Romance varieties, nevertheless show frequent voicing, at least in the south: Logudorese [neβóðe] < NEPOTE, [seɣáre] < SECĀRE, etc. (Lausberg 1965: 351).

4   Central Pyrenean dialects, despite belonging to the western branch, show frequent lack of voicing in the traditional lexis (e.g., *apella* < APICULA, *ito* < ITU) (Elcock 1938).

5   The Mozarabic descendants of Latin, spoken in Islamic Spain, most usually show retention of these voiceless consonants (a phenomenon which cannot be ascribed to orthographical conservatism, since the texts concerned are written in Arabic script).[20]

What these facts demonstrate is that we find both voicing and preservation of the Latin intervocalic consonants on both sides of the putative dividing line which, it is claimed, separates Eastern from Western Romance. And it is this dividing line which justifies the early bifurcation of the Romance tree.[21] Although these facts are well attested in standard reference books, there is a strong reluctance on the part of Romanists to abandon the tree model and the notion of an early bifurcation of the Romance tree.[22]

A potential approach to the problem of the treatment of voiceless intervocalics would be to examine the possibility of social variation in Latin between voiced and voiceless realizations of the phonemes concerned, so that what was spread from Rome to the provinces was not a set of unvarying phonemes, nor even the spread of one variant in one direction and other variants in other directions, but a variable rule whose voiced and voiceless variants were correlated with social and stylistic factors.[23] At all events, it is unacceptably simplistic to believe that a single innovation took place to the north of the La Spezia–Rimini line, which then spread to all or most of 'Western Romance', while this innovation did not penetrate south and east of the line.[24]

2.5.1.3   A third reason for rejecting the tree is that it imposes a distinction, which is often indefensible, between inherited and borrowed forms. Linguistic histories make a rigid distinction between features which are due to internal development (or to simple conservation of any earlier state of affairs) and features which have been borrowed through contact with other varieties. If the two varieties are distant in time or space, such a distinction is sound. But where the two varieties form part of the same continuum, the distinction may distort reality. To take an example, we can see that historical grammars of Spanish describe the reduction of Latin AU to /o/ (AUDIRE > oír) as a characteristic feature of the language, while the suffix -ete/-eta is labelled as foreign, borrowed from Catalan/Occitan/French. But the presence of these two features in Spanish could arguably be better explained as being due to the same process: the spread of an innovation from east to west across part of the Romance linguistic continuum. The reduction of AU certainly seems to have reached the area of Castile by diffusion (probably word by word) from the east (and to have petered out without fully affecting Portuguese). This is probably also the way in which -ete reached Castilian. Obviously there are differences of chronology: the phonological process is earlier (beginning before the break-up of the Empire) but not becoming regular in Castile until well into the Middle Ages, while the spread of the suffix is later.[25] There are also differences in the source areas of the innovations, since the reduction of AU did not affect Occitan although the suffix -et was common there from early times. However, despite these differences of chronology and geography, the process of spread is arguably the same in each case.

Why then are these two innovations in Castilian classified so differently? The answer can only be that the notion of geographical diffusion is incompatible with the tree model. The reduction of AU to /o/ can be placed before the bifurcation which separated the Castilian branch from other branches, and so can be regarded as an inherited feature;[26] but the arrival of -ete is later than this supposed bifurcation and it can therefore only have arrived by jumping from branch to branch, a process more usually labelled borrowing.

2.5.1.4   A fourth reason for abandoning the tree model lies in the need it imposes on scholars to weight linguistic features differentially. In order to decide where the nodes of the family tree should be, it is necessary to give more importance to some features than others. For example, in order to justify a classification which places Galician–Portuguese on a

branch separate from that of the central Peninsular varieties, very few features are available, and maximum attention is given to the non-diphthongization of Latin ĕ and ŏ (by contrast with their diphthongization in the centre). If such attention is not to be regarded as arbitrary, then some objective justification for the prominence of this feature must be found. However, there seems to be none; some attempts have been made to provide such justification, based on the naturalness or unnaturalness of innovations, but this approach has not met with success. Therefore, since trees depend crucially on giving prominence to certain features over others, the absence of a rationale for this selection must fatally weaken the value of the tree model.[27]

2.5.1.5   The tree model can therefore be seen to be an inadequate model of relationships between linguistic varieties. But is the matter more serious than one of inadequacy? Can the tree model be considered to be responsible for dangerously distorting reality? In the case of Peninsular Romance, at least, it can, for the following reasons.

First, it imposes a tripartite division of Peninsular varieties (into Galician–Portuguese, Castilian and Catalan) in which Leonese and Aragonese are in some sense subordinated to Castilian. For example, Corominas and Pascual (1980–91) use examples drawn from texts written in Leon or Aragon to illustrate the earliest attestation of Castilian words. The only motivation for such a procedure is the three-pronged view of Hispano-Romance: if a form does not belong to the Galician–Portuguese or the Catalan branch, it is assigned to Spanish, i.e. Castilian.[28] It goes without saying that the three-branch pattern does not express the distribution of varieties in northern Spain. We have repeatedly stressed that in this region we observe an east–west dialect continuum in which, as in the rest of the Romance-speaking world (and indeed elsewhere), all dialects are transitional and there are no dialect boundaries (see **4.1.2**).

Second, it follows that the division of this continuum into three branches, or into any number of branches, falsifies our picture and leads to such false concepts as the following: 'Galician is spoken in the extreme west of Asturias' or 'Catalan is spoken in the eastern fringe of Huesca', when all that can be meant is that the isogloss separating diphthongization from non-diphthongization of Latin ĕ and ŏ passes down a little to the east of the political boundary between Galicia and Asturias, or a little west of the boundary between Huesca and Lleida/Lérida.[29]

It is true that the organization of varieties in the southern two-thirds of the Peninsula is different; here we do find three discrete blocks of varieties, with sharp boundaries between Portuguese and Castilian and between Castilian and Catalan. But this pattern is exceptional in Romance (and elsewhere), and (as we shall see in **4.1.7**) is due to very special circumstances which arose as southern Peninsular territories were resettled following their reconquest from Islamic Spain.

In conclusion, the tree model has a limited use in expressing the relationship between standard languages which have emerged in a particular family, or between varieties which have been arbitrarily selected from a continuum. But this model is inadequate to express the subtle overlapping of features that occurs at the level of normal speech. At this level, relationships are of the gradual kind. We perhaps find it difficult to deal with relationships that are based on gradation, but language is nevertheless gradated along a number of parameters. We find it easier to use models which impose boundaries (like the colours we arbitrarily distinguish within the spectrum of visible light). This subdividing process can sometimes help, but in diachronic and synchronic language studies it more often distorts.

2.5.2   Geographical discontinuity

We have earlier emphasized that geographical variation of speech normally takes the form of a continuum of varieties which merge almost imperceptibly one with another. However, it remains true that under special circumstances we can observe lines at which sharp transition between very different varieties takes place, that is, where on either side of a geographical line there are large numbers of differences of linguistic items. There would appear to be, in principle, only two sets of circumstances under which such sharp linguistic transitions occur.

First, the wave-like spread of features from a specific prestige centre may be arrested at a political frontier, beyond which the prestige centre can offer no attraction, because those living beyond that frontier are subject to linguistic pressures coming from a different direction (see **3.1**). That is, those on either side of a frontier may accommodate their speech only to that of those living on their own side of the line, at the expense of contacts and the consequent accommodation with the speech of those living across the frontier (see **3.3**). Such circumstances have arisen in recent centuries at frontiers between European states, but may have been rare or non-existent in earlier
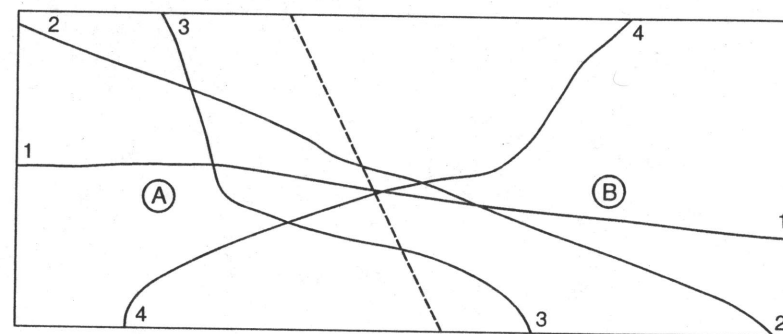
Figure 2.6   Prestige centres and isoglosses (time 1)
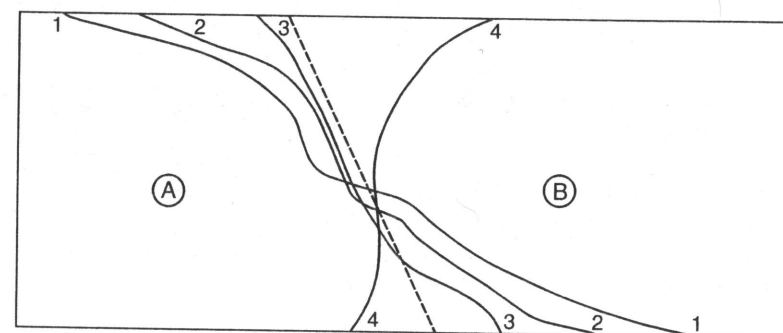

Figure 2.7   Prestige centres and isoglosses (time 2)

times, when frontiers represented no barrier to personal contacts and when prestige centres were less powerful.[30]

This process of hardening of the transition between varieties, which in an extreme case can have the effect of splitting a dialect continuum, can be envisaged as the simultaneous convergence of isoglosses (see **3.2.1**), radiated from competing prestige centres, upon an intervening political frontier. A theoretical case can be presented in Figures **2.6** and **2.7**, in which A and B are prestige centres separated by a frontier (dotted line), and where the solid lines are isoglosses moving away from points A and B. Figure **2.6** represents a time shortly after the appearance of A and B as prestige centres, at which stage isoglosses can be expected to be randomly spaced. Figure **2.7** represents the same territory, at a later time, after the isoglosses have moved towards the frontier, in some cases coming to coincide with it. Theoretically, given stability of prestige centres and frontiers for a sufficient length of time,

such a process could lead (without movement of people) to a pattern in which all the isoglosses separating points A and B (that is, each and every item of difference between the speech of A and that of B) coincided exactly with the frontier, creating an abrupt linguistic boundary. However, such stability seems rare or non-existent in the real world, where we observe some bunching of isoglosses at long-established frontiers (like that between Spain and France) but always some gradualness of transition as one moves from one country to the next.

The second way in which sharp linguistic boundaries arise is less theoretical and can be easily exemplified in the real world. This process is carried out by the movement and resettlement of groups of people in new territories, where the existing population (of course) speaks differently from the incoming group. If the movement involves sufficient people and is on a broad enough front, the result will be a sharp boundary between the speech of the old and new populations; naturally, depending on the distances involved in the population movement concerned, the difference of speech across the boundary may range from partially impeded communication to total mutual incomprehension. What is described here has of course happened repeatedly in the history of mankind and is responsible for creating major (as well as minor) language frontiers, some as striking as that between the Germanic languages and the Romance family or between Hungarian and Slavic/Germanic/Romance languages in Europe. Since the process envisaged in Figures **2.6** and **2.7** above is so slow-acting as to be effectively negligible, it is worth emphasizing that movement of population is the only real agency by which sharp linguistic boundaries are created.

Movement of population is the only explanation for the fact that the southern two-thirds of the Iberian Peninsula is divided sharply into three linguistic blocs (see **4.1.3**). In this case, the resettlement of population in new territories was the consequence of the Christian reconquest of Islamic Spain, during which each state expanded into territory defined by agreement (amicable or otherwise) with its neighbour or neighbours. The result of these movements has been the creation of linguistic boundaries which are considerably sharper than those seen in the Pyrenees.

2.5.3   Diasystems

The notion of the *diasystem* is for some just a descriptive device for expressing the relationship between adjacent varieties, while for others

it is a model of the way speakers perceive such relationships.[31] The notion was introduced by Uriel Weinreich (1954), in an attempt to bring together what were then seen as irreconcilably different approaches to linguistic description, namely classical structuralism and traditional dialectology. Although it is an adventurous idea, many scholars have found it problematical to apply (see Chambers and Trudgill 1980: 41–5, McDavid 1961), and it has not found universal favour.

As a descriptive device, and in cases of straightforward correspondences between one variety and another, some success can be claimed for the diasystematic approach. To take an example of this approach, the speech of Castile (and other central and northern Peninsular areas) displays the following range of phonemes in part of its phonemic inventory:

$/\theta/$ vs $/s/$ vs $/x/$ (e.g., *caza* 'hunt' vs *casa* 'house' vs *caja* 'box')

while the corresponding part of the phonemic inventory used by speakers in much of Andalusia (as well as the Canaries and America) offers only two phonemes:

$/s/$ vs $/x/$ (e.g., *caza* 'hunt' and *casa* 'house' vs *caja* 'box').[32]

To use the notation suggested by Weinreich, we can go on to say that these contrasting phonologies make up a single diasystem which expresses their partial similarity and their partial dissimilarity:

$$\text{C(astilian), A(ndalusian)} \;\; // \; \frac{C\;/\theta/ \approx /s/}{A\;/s/} \approx /x/ \;\; //$$

However, although Weinreich's scheme can handle differences of phonemic inventory (as in the case just examined, where varieties used in Castile have one more phoneme than varieties used in most of Andalusia), there are apparently insuperable difficulties in making it cope with differences of distribution and of incidence.

To take first the problem of distribution, it is probably impossible to reduce to a single diasystem those varieties of Spanish (e.g., those of Old Castile, Mexico, Peru) which allow the phoneme $/s/$ to occupy both syllable-initial position and syllable-final position (e.g., $/k\acute{a}sa/$ *casa* 'house' and $/\acute{a}sta/$ *hasta* 'until') from those (e.g., in eastern Andalusia) which allow this phoneme only at the beginning of syllables (e.g., $/k\acute{a}sa/$ *casa* 'house' and $/\acute{a}ta/$ *hasta* 'until').[33]

The handling of differences of phonemic incidence is also problematical. For example, all Andalusian varieties have a phoneme $/x/$,

WESTERN ANDALUSIA	EASTERN ANDALUSIA
/xámbre/ *hambre* 'hunger'	/ámbre/ *hambre* 'hunger'
/axogár/ *ahogar* 'to suffocate	/aogár/ *ahogar* 'to suffocate'
/xuégo/ *juego* 'game'	/xuégo/ *juego* 'game'
/káxa/ *caja* 'box'	/káxa/ *caja* 'box'

Table 2.1   Incidence of /x/ in Andalusian dialects

but the words which contain this phoneme in rural western Andalusian varieties do not all contain it in eastern dialects (Table **2.1**).[34] It can be seen from the data in Table **2.1** that, without the aid of non-phonological information such as spelling or historical knowledge, it is impossible to distinguish between the class of lexical items which contain /x/ in all varieties and the class of items which have /x/ in some varieties but /Ø/ in others. One has to resort simply to listing the members of each class, so that a diasystem which showed both eastern and western Andalusian dialects sharing the phoneme /x/ would fall some way short of reflecting reality.

If diasystems are to be understood as models of speakers' perceptions of language variation, as Weinreich implies, then they are surely open to even more severe challenge. The use of terms like *seseo*, *ceceo*, *yeísmo*, and *leísmo* by people other than linguists suggests that speakers are aware of differences of phonemic inventory, pronoun usage, etc., between their speech and that of others. But such consciousness would seem to be limited to a small number of very salient features, and it seems highly unlikely that awareness of variation extends to matters of distribution and incidence of features.[35]

2.5.4   Diglossia

The term *diglossia* was introduced by Charles Ferguson (1959) to refer to language situations in which two distinct varieties are used by the same community, but with very different status attached to each. In the societies originally described as diglossic (Greece, the Arabic-speaking world, etc.), the two varieties, although related, are sufficiently distinct for them to be thought of as different languages. One (referred to as the H (high) language) has high status, is highly codified, is usually a medium of literature, and is restricted to use in certain social situations,

while the second language (the L (low) language) is used by everybody in the community for all everyday purposes. The term *diglossia* was later extended by some scholars to include situations in which the two languages are unrelated. A case in point is Joshua Fishman's (1971) treatment of the language situation in Paraguay, where the H language is Spanish and the L language is Guaraní.

Diglossia, then, indicates a pattern of language use in which some or all speakers have access to two different sets of linguistic items, which either overlap little (Ferguson's original definition) or not at all (the extended sense of the term). Of course, none of this excludes variation in the L language, although since the H language is normally a standard, it will offer only very limited variation.

In the Spanish-speaking world, the concept of diglossia has not only been applied to such situations as that of Paraguay, but has sometimes been further extended to cases such as that of Galicia. Although many would agree that the early extension of the term to encompass coexistence of unrelated languages was a useful one, it is far from clear that it is helpful to use the term *diglossia* to describe the coexistence of codes we see in Galicia. It is true that until recent times, the use of Castilian in Galicia matched to a large extent the definition of an H language, while many of the everyday varieties would attract the label *Galician*.[36] However, Galician and Castilian share a large number of their linguistic items, so that it can be argued that they constitute overlapping codes, with exclusively Galician items belonging to typical L uses and exclusively Castilian items reserved for H uses, but with a broad intermediate set available for both H and L environments. Various studies of language contact in Galicia describe a situation of continuum, in which traditional Galician features predominate at rural level, but gradually diminish in intensity, in favour of typical Castilian features, as one examines the speech of small towns, larger towns, and cities, and as one moves along the social scale from uneducated to educated (see Woolnough 1988, Rojo 1981). A case in point is the degree of vowel nasalization observable in Galicia. According to Porto Dapena (1976, 1977: 23) and Sampson (1999: 207), nasalization is most intense among the least educated, that is, among those who have least familiarity with Castilian, and declines in intensity in accordance with speakers' degree of integration into Castilian-speaking sectors of society. This notion of a continuum of varieties extending from fully Galician at one end to fully Castilian at the other is supported by the apparent fact that many speakers in Galicia are unable to label

the variety they use except by some such term as *galego chapurreado*, labels which appear to indicate that the variety concerned is not fully or properly Galician (that is, presumably, that it contains many Castilian items).

Such a continuum, assuming it can be objectively verified, has been substantially altered by the re-emergence of Galician as a written language and as a spoken medium for certain educated classes, a development which began in the nineteenth century and has gathered pace in the post-Franco period. The existence of a codified version (or versions) of Galician means that items previously identified as belonging to the L varieties have come to be used as part of an alternative H code.[37]

The notion of diglossia is perhaps even less appropriate in the case of Catalonia, Valencia and the Balearic Islands. Certainly, it is arguable that the overlap between the set of items making up standard Castilian and the set constituting everyday Catalan is smaller than the overlap between Castilian and Galician items; nevertheless, it is far from true that Catalan varieties fulfil exclusively L functions; a highly codified variety of Catalan, used in writing and at least some high-prestige social circumstances, ensures that Catalan competes with Castilian in these areas for H functions.

It is perhaps only in part of the Basque Country that classic conditions of diglossia can be said to exist. Between such unrelated languages as Castilian and Basque there is, of course, little overlap of features.[38] And in those areas where Basque is used alongside Castilian, the fact that levels of literacy in Basque are so low among those who speak it ensures that Basque is used especially in L roles, while Castilian fulfils almost all H roles. But even in the Basque Country, this diglossic relationship cannot be said to be stable, for two quite opposite reasons; one the one hand, constant effort is made to introduce certain varieties of Basque into H roles (the media, education, etc.), while, on the other, the proportion of inhabitants of the Basque Country who make use of Basque (rather than some variety of Castilian) in the majority of domains for H functions is in steady decline.[39]

2.5.5 The neolinguistic model

A further way of expressing relationships between varieties was formulated in the early part of the twentieth century under the rubric of *neolinguistics* or *spatial linguistics*. This approach is particularly associated with

the work of Matteo Bàrtoli (see, for example, Bàrtoli 1945) and attempts to lay down the principles which govern the temporal and spatial relationships between varieties, especially between Romance languages. The neolinguistic approach is founded upon a codification (some would say a rigidification) of the findings of linguistic geography, combined with neogrammarian principles, and most of the tenets of this school have been dismissed by subsequent generations of linguists.[40] However, one of the central ideas of neolinguistics is still frequently appealed to, and is especially relevant to the Peninsular varieties of Romance, namely the notion that peripheral zones preserve archaic features. This notion is based upon that of linguistic waves (see **3.2**), by which innovations spread outwards from some prestige centre, but without necessarily reaching all parts of a given territory, in such a way that distant areas may be unaffected and retain an older feature. In studying the Romance lexis, this approach has a good deal of validity and a large number of cases have been unearthed in which a lexical item, thought to have once been used throughout the Latin-speaking world, has persisted in use only in peripheral areas (for example, in the central and western Peninsula, in the Alps, in southern Italy, in Dacia (approximately, modern Romania)), while speakers in more central areas (in this case, central and northern Italy, and Gaul) have replaced the term concerned with a neologism. The results of this geographical approach to the Romance lexis can conveniently be seen in the maps contained in Rohlfs (1960), where with some frequency it is possible to demonstrate that a given, older, lexical type (say FERVERE 'to boil') is found in the centre and west of the Peninsula and in Dacia (Sp. *hervir*, Ptg. *ferver*, Rom. *a fierbe*), while 'central' areas display descendants of a later replacement (in this case BULLĪRE, originally 'to bubble': Fr. *bouillir*, It. *bollire*, etc.).[41]

However, it has to be said that, while lexical data provide some limited support for the notion that territorial marginality is allied with archaism, a balanced view leads to a contrary conclusion.[42] Marginal areas, which by definition are distant from, and only loosely communicated with, prestige centres, can often be seen to develop and perpetuate innovations which the centre is powerless to obliterate. The Peninsular varieties of Romance are an excellent case in point. Spanish and Portuguese are often categorized as 'archaic' forms of Romance, on the basis of such lexical data as those reviewed here (see note 41). It is true that both have their origins in areas (Galicia, Cantabria) which are marginal both within the Peninsula and, even more so, within Romance-speaking Europe. But, looked at from any point of view but

that of lexis, one has to say that both Spanish and Portuguese, each in its own way, is a rather eccentric form of Romance.[43] And the most innovatory Romance varieties of all (those which gave rise to standard French) certainly belong to the margins of the Romance area, to its northwestern periphery. Marginality should not therefore be equated with conservatism. Quite the reverse: the marginality of the Latin varieties which underlie Castilian is one of the factors associated with its speakers' openness to radical change.

2.5.6   Other models

Our need to visualize complex relationships is intense, so that the desire for visual models (such as the genealogical tree) to help us understand the complexities of the distribution of linguistic features is acute. But no simple model is adequate. The spectrum of visible light is a possible model for geographical variation, or for any single one of the many social parameters along which linguistic variation occurs, since it consists of an infinitely gradated range of wave-lengths which is arbitrarily segmented by the human eye into the 'seven' colours of the rainbow. However, the rainbow is essentially a one-dimensional model, and language variation is multidimensional. When we come to consider the standard languages of the Peninsula and their relationship with non-standard varieties (**7.3**), we shall use the model of the roof or cupola, eloquently expounded by Vàrvaro (1991); the roof represents a standard language and covers a discrete area beneath which non-standard varieties are spread in their interlocking fashion. In modern Europe, contiguous roofs typically abut sharply upon each other, while at ground level the most unpretentious varieties usually pay no attention to the joins between roofs but interlock seamlessly across frontiers. Such a model is complex (and therefore lacks the immediate appeal of simple models). But language is multidimensional and is distorted by any one-dimensional or two-dimensional model.

3   Mechanisms of change

Language history is predicated upon the notion that most linguistic change is regular; what this implies is that all the words, phrases, or other units which are candidates for a particular change are in fact affected by it in a given speech community. Although there are great difficulties in defining what is a speech community, and although, as we shall see when we look at lexical diffusion (Section **3.5**), changes do not affect all eligible items at once and some words may not be affected at all, nevertheless it remains true that many if not most changes operate in a remarkably regular way, with all eligible units being affected, in a given place, in a measurable period of time. It may seem paradoxical that this regularity is particularly observable when there are many items eligible for change. For example, we can be fairly sure that in all the words inherited by Spanish through oral transmission from Latin and which in Latin contained an intervocalic [t] (e.g., ACŪTUS 'sharp', CANTĀTUS 'sung') there occurred the same process of voicing and fricatization which produced [ð] in Spanish (*agudo, cantado*). By contrast, it is when there are few words which display the feature which is subject to change that we find the greatest irregularity. Thus, there is only a small group of words which in Latin presented the combination NG followed by a front vowel and which have been inherited by Spanish (e.g., TANGERE 'to touch', GINGĪVA 'gum', QUĪNGENTĪ 'five hundred'), yet this small group provides evidence of three different developments: [ɲ] (*tañer*), [nθ] (*encía*), [n] (*quinientos*).[1] We shall be offering a possible explanation for this kind of fragmented development in Section **3.1.6**, but we should not lose sight of the fact that every linguistic change, however regular or irregular, presupposes a lengthy chain of imitations of one speaker by another. This chapter is concerned with this kind of imitation, the process by which change spreads through social groups, and how the composition of such groups can affect who imitates whom.

In the second half of the twentieth century, linguists reached unanimity that not only change but also variation is inherent in human