A History of the Spanish Language

Second edition

Ralph Penny

Content

Contents ix

Maps

Tables

1   Introduction

Time-line

218 BC Roman troops enter northeastern Spain.

206 BC Defeat of the Carthaginians and capture of Cádiz, their Peninsular capital.

19 BC Conquest of the remaining parts of the Peninsula (now Galicia, Asturias, Santander and part of the Basque Country), hitherto outside Roman control.

AD 76 Hadrian (Roman emperor 117–38) born, probably in Italica, near Seville.

410 The Visigoths establish (as *foederati)* a semi-autonomous kingdom in southwestern Gaul, with their capital at Toulouse.

early 6th c. Visigoths expelled from Gaul by the Franks.

c. 560–636 Lifetime of St Isidore, archbishop of Seville and author of *Origines sive etymologiae.*

585 Swabian kingdom of the northwest absorbed into Visigothic Spain.

711 The Islamic invasion of Spain.

711–18 Muslims establish control over approximately three quarters of the Peninsula.

884 Reconquest of Burgos.

1035 Creation of the kingdom of Castile.

1080 Council of Burgos.

1085 Reconquest of Toledo.

1086 First Almoravid invasion.

1137 Merger of Aragon with Catalonia.

1154 Almohads gain control of Islamic Spain.

1212 Christian victory in the battle of Las Navas de Tolosa.

1236 Reconquest of Córdoba, followed by that of Jaén (1246), Seville (1248) and Cádiz (1250).

1244 Castile gains control of the kingdom of Murcia.

1252–84 Alfonso X the Learned, king of Castile and León.

1479 Union of the Crowns of Castile and Aragon.

1492	Granada captured by the Catholic Monarchs.
1492	Expulsion of Jews from Spain.
1492	Publication of Antonio de Nebrija's *Gramática de la lengua castellana*.
1492	Discovery of America.
1519–21	Conquest of Mexico.
1532–5	Conquest of the Inca empire.
c. 1535	Juan de Valdés completes his *Diálogo de la lengua*.
1521	Discovery of the Philippine Islands, incorporated into the Spanish Empire later in the sixteenth century.
1561	Madrid becomes capital of Spain.
1605–15	Publication of Miguel de Cervantes's *El ingenioso hidalgo don Quixote de la Mancha*.
1713	Establishment of the Real Academia Española.
1726–39	Publication of the Academia's three-volume *Diccionario de autoridades*.
1800–36	Wars of American Independence deprive Spain of almost all its American colonies.
1898	Spain loses its last American and Asian colonies (Cuba and the Philippines).
1936–75	Franco controls the Spanish state.
1978	Publication of the post-Franco constitution.

This history of Spanish is conceived as an account of the 'internal' development of the language, a discussion of the way in which its phonology, its morpho-syntax, its vocabulary and the meanings of its words have evolved, and of the reasons for these developments (insofar as they can be established). It is therefore what used to be called a 'historical grammar' of the Spanish language. However, although it follows that the book is not essentially concerned with the social contexts in which Spanish is and has been used, it is appropriate to give a brief account of these contexts, by way of introduction to the main matter which follows. More detailed accounts of the 'external' history of Spanish (Lapesa 1980, Penny 2000) are available to the reader, and what is discussed here consists of an outline of the circumstances under which Spanish and its antecedents have been spoken over the centuries, an outline which is sufficient, one hopes, to explain the chronological and social terms used in later chapters.

1.1   Indo-European, Latin and Romance

Spanish is a member of the Indo-European family of languages, whose earliest reconstructible ancestor was spoken approximately 5,000 years ago in the area of the Black Sea (either, in keeping with the traditional view, to the north of that Sea, in the steppes of southern Russia, or, according to a more recent view (e.g. Renfrew 1998), to its south, in what is now Turkey). Speakers of

Indo-European gradually spread (perhaps in conjunction with the spread of farming) in various directions, and varieties of their speech came to be used in enormously broad areas: almost all Europe (where only Basque, Finnish, Sami (Lapp) and Magyar (Hungarian) do not have this ancestry), the greater part of the Indian subcontinent as far east as Bangladesh and Assam, and many territories in between (e.g. Armenia, Iran, much of Afghanistan). During this migration process, each group of Indo-European speakers inevitably lost contact with other groups, so that innovations and losses originating in one group could not spread to others and fragmentation was the natural result. However, the family resemblance has persisted over the millennia and the surviving members retain many structural similarities and a significant proportion of their core vocabulary in common. Scholars typically recognize nine surviving branches of the Indo-European family (Indo-Iranian, Slavonic, Germanic, Italic, Baltic, Hellenic, Armenian, Albanian and Celtic), while two branches (Tocharian and Hittite) have left substantial written records but are no longer spoken.

The most prominent member of the Italic branch of Indo-European is Latin. However, other members of this branch were spoken for centuries in Italy beside Latin. These included Oscan (in much of the centre and south of the Italian peninsula), Umbrian (in the area northeast of Rome), and Faliscan (immediately to the north of Rome), and were gradually replaced by Latin, as the political and cultural power of Rome spread from Latium to encompass the rest of the Italian peninsula. This process of assimilation was, naturally, long lasting, beginning in the fourth century BC, and not becoming complete until at least the first century AD. For example, some of the graffiti preserved on the walls of Pompeii (and therefore written shortly before the eruption of Vesuvius in AD 79) are best described as reflecting Oscan rather than Latin speech.

The spread of Latin outwards from Rome was not limited to the Italian peninsula, but continued into adjacent and eventually distant parts of the Mediterranean world and its hinterland. As a result of varying development in different parts of this territory, Latin evolved into the family of related dialects (some of which achieved the status of standard languages) known as the Romance language family. No classification of the Romance languages can be fully satisfactory because they form a continuum of overlapping varieties, which is broken only in the eastern Balkans, where the link between Italian and Romanian varieties has been severed by the incursion of Slavonic speech (modern Serbo-Croatian, etc.), and in the Alps, where northern Italian varieties are separated from Romansh by the expansion of German. The most prominent members of the family (those that came to be written and which constitute the standard languages within the group) are French, Occitan (once the literary language of southern French society, now a series of rural varieties), Italian, Romanian, Romansh (the fourth national language of Switzerland, spoken in the southeastern canton of the Graubünden), Catalan, Portuguese and Spanish. Dalmatian, spoken on the eastern Adriatic coast, has been extinct since the nineteenth century.

Latin is the ancestor of Spanish (and, by definition, of all other Romance languages) in the sense that there is an unbroken chain of speakers, each learning his or her language from parents and contemporaries, stretching from the people of the Western Roman Empire two thousand years ago to the present population of the Spanish-speaking world. An alternative way of expressing the relationship between Latin and Spanish is to say that Spanish *is* Latin, as Latin continues to be spoken in parts of Europe, Africa and America. Similar claims are of course justified in the case of Portuguese, Catalan, French, Italian, Romanian, etc., and the main reason the term 'Latin' is not used for these various kinds of speech and writing is one of convenience: some forms of contemporary Latin (i.e. some Romance languages) have become mutually unintelligible and it is inconvenient to use a single label for mutually unintelligible forms of language. Another, more powerful, reason for the use of distinctive labels such as 'Spanish', 'French', etc., is that the rise of nation-states in medieval and modern Europe led to the development of separate written standards and has come to demand a separate linguistic identity for each state, as an expression of its cultural and political identity.

It is self-evident that contemporary 'Latin' speech (in the sense used here to embrace what are otherwise referred to as 'the Romance languages') is not uniform, but it is equally important to recognize that Latin speech can never have been uniform. All language displays variation (and Latin speech can have been no exception) in three main ways: it varies diatopically (i.e. in space), diachronically (i.e. over time), and sociolinguistically (i.e. in the same place and at the same time it varies in accordance with factors such as the age, sex, education, occupation, etc. of the speaker). In addition, variation is inherent not only in speech-communities, but in individuals, in the sense that individuals normally vary their speech according to the circumstances in which they are speaking. The fact that we are deprived of the opportunity (for the most part) of observing such variation in the case of Latin should not blind us to the fact of its existence in the Latin-speaking world of two thousand years ago.

Evidence of diatopic variation in Latin is scarce, owing to the fact that those who wrote were trained to do so in a variety of Latin (an educated, literary variety, traditionally called 'Classical Latin') which by its nature rejected merely local characteristics. However, some evidence is available and, insofar as it refers to the Latin of Spain, it will be discussed in the following section (1.2). Evidence of diachronic variation is more plentiful, and comes from comparison of the language used by writers at different periods, and from comments made by Latin grammarians on the antiquated or obsolete status of certain features of the language.

It is the evidence of sociolinguistic variation that has received most attention from students of the Romance languages, although this aspect of variation is not traditionally described as 'sociolinguistic'. Since at least the nineteenth

century, it has been known that the Romance languages do not descend from Classical (i.e. literary) Latin, but from non-literary varieties, often referred to collectively as 'Vulgar Latin'. To take a simple and well-known lexical example, the word meaning 'horse' in literary Latin is EQUUS, a form which is clearly not the ancestor of the Romance words for this concept (Sp. *caballo*, Ptg. *cavalo*, Fr. *cheval*, It. *cavallo*, Rom. *cal*, etc.). The latter forms descend from CABALLUS, which, where it appears in literary Latin, means 'nag; workhorse', but which in non-literary language was evidently used in the generic sense 'horse'.

Definitions of 'Vulgar Latin' have abounded, and many have rested on historical models that can now be seen to be mistaken. Thus, Romance linguists have long since rejected the notion that Vulgar Latin is a later form of Latin than the Classical variety, despite the fact that much of the evidence for Vulgar Latin comes from the later centuries of the Empire and despite the fact that many (but not all) of the features of Vulgar Latin are revealed as more 'advanced' than the corresponding features of Classical Latin.

Harder to die is the notion that Vulgar Latin and Classical Latin are sharply different codes, and that the two terms represent mutually exclusive concepts. This view cannot be sustained, since all varieties of Latin of which we have knowledge share most of their vocabulary, most of their morphology and most of their syntactical rules. The model adopted here is that 'Latin', like any language observable today, represents a gamut or spectrum of linguistic styles, ranging from the codified, literary register at one end to the raciest slang at the other, with a smooth gradation of intermediate styles. On this model, 'Classical Latin' occupies one extreme of the spectrum, representing essentially written varieties (unspoken except in 'performance' or 'reading aloud' mode), while Vulgar Latin represents almost the whole of the remainder of the spectrum, perhaps with the exception of the spoken language of the educated classes (for which a separate term is required) and with the exception of the language of marginal social groups at the other extreme, since the slang of such groups is known to be unstable and therefore unlikely to have affected the speech of the great mass of the population in any consistent way.

This view of Vulgar Latin, although expressed differently, is in broad agreement with one of the more satisfactory current definitions of the term, that adopted by Herman (2000: 7) and some predecessors: 'the set of all those innovations and trends that turned up in the usage, particularly but not exclusively spoken, of the Latin-speaking population who were little or not at all influenced by school education and by literary models'. However, it is important to make clear certain corollaries which flow from the definition of Vulgar Latin adopted here.

First, Vulgar Latin has no implicit chronological limits. It is contemporary with Classical Latin and as soon as it is meaningful to refer to Classical Latin (i.e. from the first century BC) it is also meaningful to use the term 'Vulgar

Latin', despite the fact that most evidence of the nature of Vulgar Latin comes from later centuries. At the other extreme, the term 'Vulgar Latin' ceases to be useful once locally divergent forms of language begin to be recorded in writing (the ninth century AD in northern France), and the term 'Romance' is then used for any or all vernacular descendants of Latin, written or spoken. However, some scholars also use the term 'Romance' to refer to the spoken language of earlier centuries, while other scholars use the term 'Proto-Romance' to indicate those forms of spoken language which constitute the ancestor of the Romance languages, and which by definition belong to a period prior to the appearance of texts written in Romance.

Second, there can be no such thing as a 'Vulgar Latin text'. Texts of all kinds are composed, by definition, by the educated and therefore in the codified or 'standard' variety of Latin in which such writers have inevitably been trained. This is not to say that textual evidence of spoken registers of Latin is unavailable (it will be outlined below); what we do find is that certain types of text contain a greater or lesser proportion of forms (spellings, words, constructions, etc.) which differ from the standard and which reveal particular features of spoken Latin. Such information is inevitably incomplete and cannot amount to a 'full' picture of Vulgar Latin.

Third, like 'Latin' considered in its entirety, Vulgar Latin is inherently variable. The term includes reference to all the chronological, local and social varieties of Latin as spoken by the majority of the relevant populations. It cannot therefore be described in the 'grammar-book' way that is appropriate to codified or standard varieties.

What then are the sources of information about the features of Vulgar Latin? Full discussion is inappropriate here (and can be found in works on Vulgar Latin such as Väänänen 1968: 39–49), but may be summarized as:

- Literary writing purporting to reflect popular speech (dramatists such as Plautus (*c.* 254–184 BC), Ennius (239–169 BC), Terence (*c.* 195–159? BC); satirists such as Petronius (d. AD 66)).
- Informal letters, such as those written to his father by Claudius Terentianus in second-century AD Egypt (see Adams 1977).
- Christian writings, which generally rejected the exclusivist standard language in favour of a style more suited to a proselytizing religion, especially those written for an unsophisticated audience (such as the late fourth-century account by a Spanish nun of her pilgrimage to the Holy Land, usually referred to as the *Peregrinatio ad loca sancta* or the *Peregrinatio Etheriae*).
- Technical writing, which because of its practical intent and the modest education of its intended readership was usually unpretentious in style and allowed the use of vocabulary and expressions belonging to speech; such writing includes works on cookery, farming, building, medicine, veterinary science, etc.

- Writing for various purposes, literary and non-literary, from the late Roman period (third to fifth centuries AD) and from the following centuries, when standards of education and culture among the literate were lower than they had earlier been and when writers consequently may lapse into non-Classical modes of expression.
- Informal inscriptions, including gravestones but especially including painted graffiti (such as those of Pompeii and Herculaneum, which include advertisements, announcements, slogans, obscenities, etc.) and *defixiones* (metal plaques on which magical spells are scratched).
- Writings of grammarians, especially insofar as they condemn forms as incorrect, since this assures us of their existence in speech; particularly noteworthy is the so-called *Appendix Probi*, a sixth- or seventh-century (see Robson 1963) list of 227 forms to be avoided in writing, in which each recommended expression is placed alongside its condemned equivalent (e.g. BACULUS NON VACLUS, AURIS NON ORICLA, GRUS NON GRUIS, TRISTIS NON TRISTUS); of almost similar importance, especially for Spain, are the linguistic observations of St Isidore, bishop of Seville (*c.* 560–636), in his *Origines sive etymologiae*.
- Glosses of various dates from the first century AD on, where some reader has inserted into a text interlinear or marginal equivalents for words or expressions which were obsolete and therefore posed difficulty for readers, the replacements sometimes being drawn from spoken registers.
- Borrowings made by Latin from other languages, and vice versa, in which the manner of adaptation of the borrowed word to the borrowing language may reveal features of pronunciation (e.g. German *Kaiser* reveals that when Germanic borrowed the Latin word CAESAR its initial consonant was articulated [k]).

Alongside this testimony drawn largely from ancient texts is to be placed the evidence deducible from the Romance languages themselves. We have already seen that by comparing certain Romance forms it is possible to deduce that in spoken Latin the word CABALLUS had the generic sense 'horse', and it is possible to apply this procedure to any linguistic feature, on the hypothesis that if the same feature is observed in a broad range of Romance languages then that feature belonged to spoken Latin. Thus, by comparing Romance words for 'green' (e.g. Sp., Ptg., It., Rom. *verde*, Fr., Cat. *vert*), it is possible to make the minimal deduction that their spoken Latin ancestor had no more than two syllables, despite the fact that the Latin word for 'green' that we find in writing suggests that it has three: VIRIDIS. On this occasion, our deduction is confirmed by the author of the *Appendix Probi*, who prescribes VIRIDIS NON VIRDIS. However, in a large number of cases, such confirmation from written sources is not forthcoming, and many forms have been deduced (on the basis of the comparison of Romance forms) as belonging to spoken Latin without their existence being

confirmed by any written source. Thus, a comparison of the Romance verbs meaning 'to be' (e.g. Sp., Ptg. *ser*, Cat. *ésser/ser*, Fr. *être*, It. *essere*) reveals that their spoken Latin ancestor is likely to have had three syllables and that the last syllable was -RE, by contrast with the Classical Latin form ESSE 'to be'. On the basis of known facts about the development of each Romance language, it is possible to refine the deduced Vulgar Latin form to *ESSERE. It will be noted that in such cases an asterisk indicates the lack of confirmation from written sources, and therefore the hypothetical (but not necessarily doubtful) status of the word concerned.

1.2   The Latin of Spain

Latin came to be used in Spain as a result of the gradual incorporation of the Peninsula into the Roman Empire and of the consequent romanization of its diverse peoples and cultures. Romanization began in 218 BC, at the beginning of the Second Punic War, when Roman troops were disembarked in northeastern Spain to prevent reinforcement of Hannibal's army in Italy (following his famous march across the Alps) by Carthaginian troops from southern Spain, then part of Carthaginian territory. After the defeat of the Carthaginians and the capture of their Peninsular capital, Cádiz, in 206 BC, what had begun as a military enterprise became a process of colonization and settlement. This process was relatively slow, progressing in a westerly and northwesterly direction over the next two centuries and culminating in the conquest of the northern coastal area (now Galicia, Asturias, Santander and part of the Basque Country) in 19 BC (see map 1.1).

In the wake of conquest and settlement came latinization. The use of Latin was not enforced (and scarcely could have been), but was learned by the local populations, as a matter of convenience and prestige, from Roman settlers, administrators, soldiers, traders, etc. This process was rapid in some areas (the east and south, where Roman immigration was earliest and most frequent), slower in others (the centre, west and north), and is still incomplete in one area (the Basque Country). Any such language-change implies bilingualism over at least several generations, and since bilingualism persists today in the western Pyrenees it is likely that it persisted in other areas remote from the major Roman cities (that is, in parts of the north and west) at least until the end of the Roman period, in the fifth century, and in the remotest areas probably later. Such bilingualism, between Basque and Latin and between Celtic and Latin, has often been cited as the cause of certain changes which are evident in the Peninsular descendants of Latin (see 2.5.3.2, 2.5.6, etc., for discussion), and it is certain that it allowed the borrowing of certain words by Latin from the languages with which it coexisted (see 4.2). Latinization was evidently much more rapid in the east and south, where Iberian and Greek (in what is now

Map 1.1 Roman Spain in the first century AD

Catalonia and Valencia) and Tartessian (in Andalusia and southern Portugal) appear to have been displaced entirely by Latin by the first century AD at the latest.

The pace of latinization is probably correlated with geographical distance from the 'educated standard' of the 'average' Latin spoken at any given date. The factors which encouraged rapid latinization (close contact with central Italy, urbanization, good road communications, the consequent fostering of trade, etc.) are the same factors which encouraged the use of forms of Latin which were closer to the prestigious end of the sociolinguistic spectrum (see 1.1). It is therefore likely that the 'average' Latin spoken by people in the remoter, less developed, parts of the Peninsula was considerably further from the prestige norm (that of upper-class Rome) than was the speech of the eastern and southern cities. This factor is particularly relevant to the history of Spanish, since Spanish has its geographical roots in what is now the northern part of the province of Burgos, an area of the northern meseta which was remote from the centres of economic activity and cultural prestige in Roman Spain, which was latinized fairly late, and where the Latin spoken must consequently have been particularly remote from the prestige norm (that is, particularly 'incorrect') at the time of the Roman collapse. With the end of the Roman state came the effective removal

of the linguistic model towards which, however distantly and ineffectually, speakers strove to adhere, so that any 'incorrect' features of local speech were likely to be perpetuated (unless challenged by some other prestige model, which was not to be the case in the Burgos area). Spanish has often been described as a rather idiosyncratic form of Peninsular Romance (even of Romance *tout court*)*,* a view associated with Menéndez Pidal (1964a: 472–88) and developed in Penny (2000). Such linguistic idiosyncrasy can plausibly be accounted for in part by the conditions under which the northern meseta was latinized.

It is also appropriate to consider here the ways in which the Latin spoken in Spain differed from that spoken in other provinces. Such a consideration must not assume that the Latin of Spain was in any sense uniform; we have just seen that it was probably far from uniform. But it is at least arguable that there are some characteristics shared by all or most of the surviving varieties of Peninsular Romance (and which therefore belonged to the Latin spoken in most if not all of the Peninsula), which may be contrasted with the corresponding features of Gallo-Romance, Italo-Romance, etc. The characteristics which have been assigned to the Latin of Spain, at different times by different scholars, are its archaism, its conservatism and its Osco-Umbrian dialectalism. Paradoxically, there are a number of features which allow the Latin of Spain to be described as innovatory. Each of these characteristics will be considered in turn.

1.2.1   Archaism

The early date at which the latinization of Spain began (the end of the third century BC) implies that the Latin carried to Spain by the earliest soldiers, traders and migrants represents an earlier phase in the development of Latin than that represented by the language carried to other areas. For example, the latinization of northern Italy and southern Gaul begins in the second century BC, at a time when all of Spain but the northwest was under Roman rule, while the latinization of the rest of Gaul does not begin until the first century BC, and that of Dacia (approximately modern Romania) does not begin until the second century AD. On the hypothesis that colonized areas often retain features of speech which are abandoned in the parent-state (a hypothesis which finds some support in the history of English and Spanish in America, as elsewhere), it is predictable that Hispano-Romance will retain some features of third- and second-century BC Latin which were then abandoned in the Latin of Rome and other, more recently latinized, provinces. Such an argument may apply to the widespread appearance in Peninsular speech of bilabial [ɸ] (corresponding to the spelling F; see 2.5.6) rather than its successor, the labiodental [f], which is used in most of the rest of the Romance-speaking world. However, it is in the field of vocabulary that such archaism has been most closely studied. The following expressions are ones whose antecedents appear in pre-Classical

writers (Plautus, Ennius, Terence, etc.) but not in the works of those writing in
Rome from the first century BC onwards, facts which suggest that the words
concerned had fallen out of use there (while continuing in use in the Latin of
Spain):

- Sp., Ptg. *cansar* 'to tire' < CAMPSĀRE 'to bend, to round (a headland)', an
  early borrowing from Greek not found in literature after the second century
  BC.
- Sp. *cueva*, Ptg. *cova* 'cave' < the pre-Classical adjective COVA 'hollow', by
  contrast with CL CAVA 'id.', whence Fr. *cave* 'cellar', etc.
- Sp. *cuyo, -a,* Ptg. *cujo, -a* 'whose' < CŪIUS, -A, -UM '-id.', a form already
  obsolescent in the first century BC.
- Sp. *(a)demás*, Ptg. *demais* 'besides' < DĒMAGIS, not found in writing after
  the second century BC.
- Sp. *hablar*, Ptg. *falar* 'to speak' < pre-Classical FABULĀRĪ 'to converse'.
- Sp., Ptg. *querer* 'to wish' probably reflects the pre-Classical sense of
  QUAERERE 'to wish', found in Terence (early second century BC), but whose
  sense later became 'to seek'.

1.2.2   *Conservatism*

Conservatism cannot be sharply distinguished from archaism, since both terms
refer to the retention of forms which elsewhere disappear. What is meant by
the conservatism of the Latin of Spain is the retention of forms which appear in
Classical Latin (and which were presumably once current in the spoken Latin of
many areas besides Spain) by contrast with their eventual rejection in those areas
which formed the cultural centre of the late Roman Empire (central and northern
Italy and Gaul). Thus, the Latin numerals QUADRĀGINTĀ . . . NŌNĀGINTĀ
'forty . . . ninety', retain the stress on the penultimate vowel ĭ (later > /e/) in
their Spanish and Portuguese descendants: *cuarenta~quarenta . . . noventa* (see
3.6.1), whereas in other Romance areas a stress-shift to the preceding syllable
produced forms with tonic /a/: Fr. *quarante*, It. *cinquanta*, etc. But it is again in
vocabulary that most evidence of conservatism is forthcoming; in the following
cases, Spanish (together usually with Portuguese) retains a form which is normal
in Classical Latin but which, if it appears outside the Peninsula, appears only in
similarly 'remote' areas (e.g. the Alpine area, southern Italy, Sicily, Sardinia,
Romania):

- Sp. *arena*, Ptg. *areia*, Rom. *arină* 'sand' < CL ARĒNA 'id.' (cf. Fr. *sable*, It.
  sabbia).
- Sp. *ciego*, Ptg. *cego*, Cat. *cec*, C. It. *cieco* 'blind' < CL CAECU 'id.' (cf. Fr.
  aveugle, N. It. *orbo*).
- OSp., Sard. *cras*, S. It. *crai* 'tomorrow' < CRĀS 'id.' (cf. Fr. *demain*, It.
  domani, Rom. *mîine*).

- Sp. *hervir*, Ptg. *ferver*, Rom. *fierbe* 'to boil' < CL FERVERE 'id.' (cf. Fr. *bouillir*, It. *bollire*, Cat. *bullir*).
- Sp. *hombro*, Ptg. *ombro*, Rom. *umăr* 'shoulder' < CL UMERU 'id.' (cf. Fr. *épaule*, It. *spalla*, Cat. *espatlla*).
- Sp., Ptg. *ir*, OSp., OPtg. *imos*, S. It., Sic. *immu*, OSp., MPtg. *ides*, Sp., Ptg. *ido*, forms of the verb 'to go' which descend from corresponding forms of CL ĪRE 'id.' (cf. Fr. *aller, allons*, It. *andare, andiamo*, Cat. *anar, anem*, etc.).
- Sp., Ptg. *mesa*, Rom. *masă* 'table < CL MĒNSA 'id.' (cf. Fr. *table*, It. *tàvola*, Cat. *taula*).
- Sp. *queso*, Ptg. *queijo*, C. It. *cacio*, S. It. *caso*, Rom. *caş* 'cheese' < CL CĀSEU 'id.' (cf. Fr. *fromage*, It. *formaggio*, Cat. *formatge*).
- Sp., Ptg. *rogar*, Rom. *ruga* 'to beg' < CL ROGĀRE 'id.' (cf. Fr. *prier*, It. *pregare*, Cat. *pregar*).
- Sp., Ptg. *sanar*, S. It., Sard. *sanare* 'to cure' < CL SĀNĀRE 'id.' (cf. Fr. *guérir*, It. *guarire*, Cat. *gorir*).
- Sp. *yegua*, Ptg. *égua*, Cat. *egua*, Rom. *iapă* 'mare' < CL EQUA 'id.' (cf. Fr. *jument*, It. *cavalla*).

It can be seen from these examples that there is a correlation between those varieties of Romance which preserve older forms and those which are located in peripheral parts of the Romance-speaking area, that is, those that were remotest from the trend-setting centres of the late Roman period. However, this correlation is not solely evident in the preservation and distribution of forms which also appear in Classical Latin. It is also evident in the distribution of Vulgar Latin innovations, where earlier innovations are typically found in peripheral regions and later innovations are observable in the central territories of Romance-speaking Europe. This distribution can be seen in the Vulgar Latin replacements of the synthetic forms of the comparative adjective (see 3.3.2), where the earlier innovation MAGIS (+ adj.) is preserved in Sp. *más*, Ptg. *mais*, Cat. *mes*, Rom. *mai*, by contrast with the later type PLUS (+ adj.) seen in Fr. *plus*, It. *più*. In vocabulary, this pattern is frequently repeated; e.g.:

- Sp. *hallar*, Ptg. *achar*, S. It. *acchiare*, Rom. *afla* 'to find' < AFFLĀRE 'to breathe out' (see 5.3.1) (cf. Fr. *trouver*, It. *trovare*, Cat. *trobar* <*TROPĀRE).
- Sp. *hermoso*, Ptg. *formoso*, Rom. *frumos* 'beautiful' < FŌRMOSU 'shapely' (cf. Fr. *beau*, *bel*, It. *bello* < BELLU).
- Sp. *pájaro*, Ptg. *pássaro*, Rom. *pasere* 'bird' < VL PASSAR (CL PASSER) 'sparrow' (cf. Fr. *oiseau*, It. *ucello*, Cat. *aucell* < AVICELLU).

For further details, see Rohlfs (1960).

1.2.3   Dialectalism

At the time that the latinization of Spain began, at the end of the third century BC, Latin was far from having ousted its Italic competitors (Oscan, Umbrian, etc.)

from central and southern Italy; there is evidence of the use of Oscan until at least the first century AD (see 1.1). And since it seems likely that many Roman soldiers and settlers who came to Spain were drawn from areas of Italy where Latin was spoken bilingually with Oscan or Umbrian, it has been claimed that the Latin of such speakers was likely to have contained non-standard features resulting from this bilingual contact. A detailed case of this kind can be seen in Menéndez Pidal (1960), where phonological changes such as MB > /m/ (see 2.5.3.2) and -LL-, -NN-, -RR- > /ʎ/, /ɲ/, /r/ (see 2.5.3.2.9) are assigned to this origin. Similarly, the tonic vowels of the ancestors of *nudo* 'knot', *octubre* 'October' and *cierzo* 'north wind' have sometimes been explained on the basis of interference between Latin NŌDU, OCTŌBER and CIRCIU and cognate Oscan or Umbrian forms with tonic ū and ĕ (namely hybrid *NŪDU, *OCTŪBER, *CĔRCIU), an interference which did not arise outside southern Italy and Spain (cf. NŌDU > Fr. *noeud*). The distribution of forms cognate with Sp. *dejar* 'to leave' (Ptg., Cat. *deixar*, Gasc. *dechà*, Sic. *dassari*, S. It. *dassare*, OSard. *dassare*), by contrast with descendants of LAXĀRE (OSp. *lexar*, Fr. *laisser*, It. *lasciare*) has sometimes been explained on the basis of a dialectal Latin form *DAXĀRE, whose D- would be due to interference from Oscan. A similar distribution of the meaning 'to arrive' associated with descendants of PLICĀRE (CL 'to fold'), such as Sp. *llegar*, by contrast with those Romance forms which retain the Latin sense (e.g. Fr. *plier*, It. *plegare*, as also Sp. semi-learned *plegar*), is also cited as a case of the dialectal nature of the Latin of Spain. However, it cannot be said that there is general agreement on the origin of any of the instances of putative Osco-Umbrian influence so far adduced.

1.2.4   Innovation

Despite the general characterization of Hispanic Latin as archaic and conservative, there are a number of features displayed by its descendants which reveal innovatory changes which were evidently limited to the Peninsula. Among these innovations can be counted the total merger of the Latin second and third verbal conjugations (see 3.7.6), so that infinitives like DĒBĒRE and VENDĔRE, originally distinct, became identical in type (Sp. *deber*, *vender*, Ptg. *dever*, *vender*), rather than remaining separate as they do in other varieties of Romance (e.g. Fr. *devoir*, *vendre*).

Some Hispanic innovations consist of new cases of word-formation, as in:
- CIBU 'food' → CIBĀRIA > *cibera* '(animal) feed, etc.', now only in rural use.
- CIBU 'food' → CIBĀTA > *cebada* OSp. 'feed', later 'barley'.
- AMĀRU 'bitter' → AMĀRELLU 'yellowish' > *amarillo* 'yellow'.
- ARGENTU 'silver' → ARGENTEU 'of silver' > OSp. *arienço* 'a (specific) coin, unit of weight'.
- CATĒNA 'chain' → CATĒNĀTU 'chained' > *candado* 'padlock' .

- CENTĒNI 'hundredfold' → CENTĒNU 'rye' > *centeno* 'id.'.
- COLUMNA 'column' → COLUMELLU 'canine (tooth)' > *colmillo* 'id.'.
- FŌRMO 'shape, mould' → FŌRMĀCEU 'mud-brick wall' > *hormazo* 'id.', now antiquated.
- PĀCĀRE 'to pacify' → ADPĀCĀRE 'to extinguish' > *apagar* 'id.'.

On other occasions the innovation consists of a change of meaning which is peculiar to the Latin of Spain and its descendants:

- CAPTĀRE 'to seize' > *catar* 'to look'.
- FRĀTRE GERMĀNU 'true brother (i.e. one who shares both parents)' > GERMĀNU 'brother' > *hermano* 'id'; thus also GERMĀNA > *hermana* 'sister'.

Other innovations of course include the borrowing of words from the pre-Roman languages of the Peninsula (see 4.2).

1.3   Conquest and Reconquest

1.3.1   *The Visigoths*

From the fifth to the early eighth century, Spain was controlled by a Visigothic monarchy and aristocracy. The Visigoths had forced an entry into the Roman Empire in the late fourth century and following their sack of Rome in 410 established (as *foederati*), a semi-autonomous kingdom in southwestern Gaul, with their capital at Toulouse. While remaining subjects of the Roman state, they expanded their territory to include much of the Peninsula, which, together with their lands north of the Pyrenees, became an independent kingdom on the collapse of Roman administration in the west (see map 1.2). Expulsion from most of Gaul by the Franks (early sixth century) was followed by the successful absorption (completed in AD 585) of the Swabian kingdom of the northwest (in modern terms, Galicia, northern Portugal, and the provinces of Asturias and Leon), and by the eventual expulsion (in the early seventh century) of the Byzantine forces who dominated parts of eastern and southern Spain on behalf of the Eastern Roman Emperor.

The Visigoths were partly romanized before their entry into the Peninsula and it is likely that from the first they spoke Latin, bilingually with their East Germanic vernacular. The latter never achieved the status of written language in Spain and Latin continued to be the language of culture and administration throughout the Visigothic period. The influence exercised by Visigothic upon the Latin of Spain was therefore small. Apart from a number of lexical loans (see 4.5), such influence is limited to a few morphological features:

- The introduction of a new noun-declension type in nominative -ā, oblique -ĀNE (plur. -ĀNES), alongside the three types already existing in late spoken Latin (see 3.2.3). This pattern was mostly restricted to personal names of Germanic origin (e.g. OSp. *Froilán* < FROILANE, beside *Fruela* < FROILA,

Territory lost in
early 6th century AD

Toulouse ●

(Independent)

*Swabian
Kingdom*

●
Toledo

*Visigothic Spain
in AD 526*
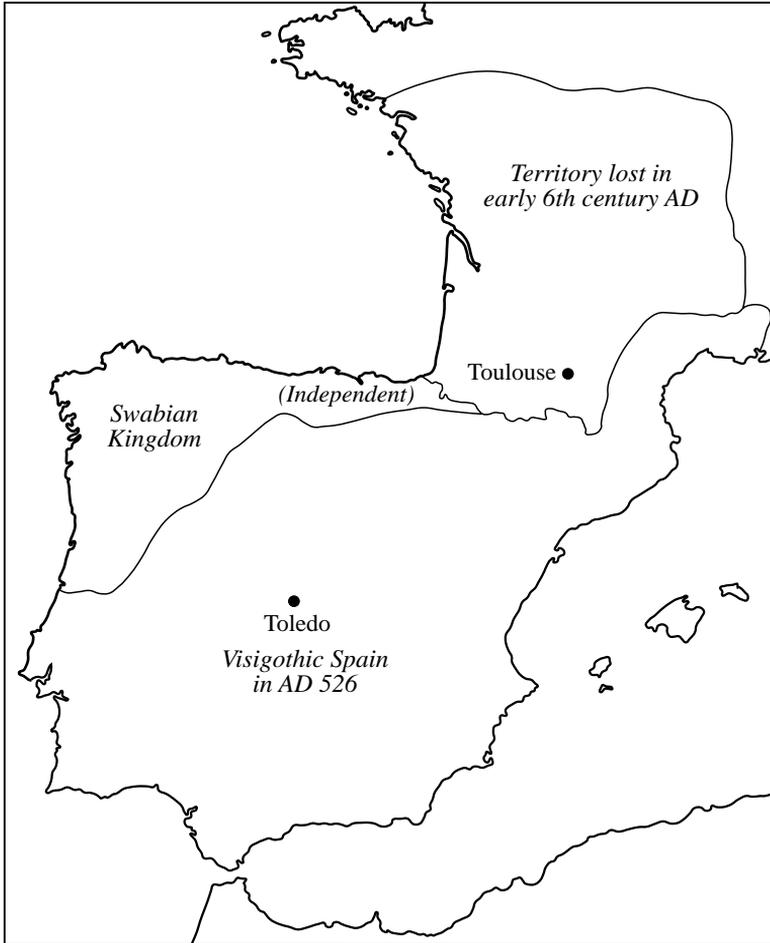
Map 1.2 The Visigothic kingdom in AD 476 and AD 526

both names applied to the same Visigothic monarch), but was occasionally ap-
plied to common nouns (usually personal, usually borrowed from Germanic).
In one instance, Spanish shows descendants of both the nominative and
oblique forms of this paradigm: *guardia* 'guard, policeman' < WARDJA
'guard(sman)', *guardián* 'guardian' < *WARDJĀNE 'id.'.

- The introduction of the suffix *engo* (< Gmc. -ING), for deriving adjectives
  from nouns. This suffix has always been of low productivity and is found
  in: *abadengo* 'belonging to an abbey', *realengo* 'belonging to the Crown',
  and, now substantivized, *abolengo* 'ancestry' (originally 'pertaining to one's
  ancestors').

• The possible introduction of the suffix *-ez, -oz,* etc., found in names which were once patronymic and are now surnames (e.g. *Rodríguez, Fernández, Muñoz*). The genitive of the latinized form of certain Germanic names in -ɪᴋs, e.g. ʀᴏᴅᴇʀɪ̄ᴄɪ̄ '(son) of Roderick', may explain certain patronymics (e.g. ʀᴏᴅᴇʀɪ̄ᴄɪ̄ > *Rodriz* > *Ruiz*). By comparison with the short form of the corresponding given name (e.g. *Ruy*), it was possible to extract an element *-z* with patronymic value, which could then be applied to other given names, including their 'full' forms: *Rodrigo → Rodríguez, Fernando → Fernández,* etc.
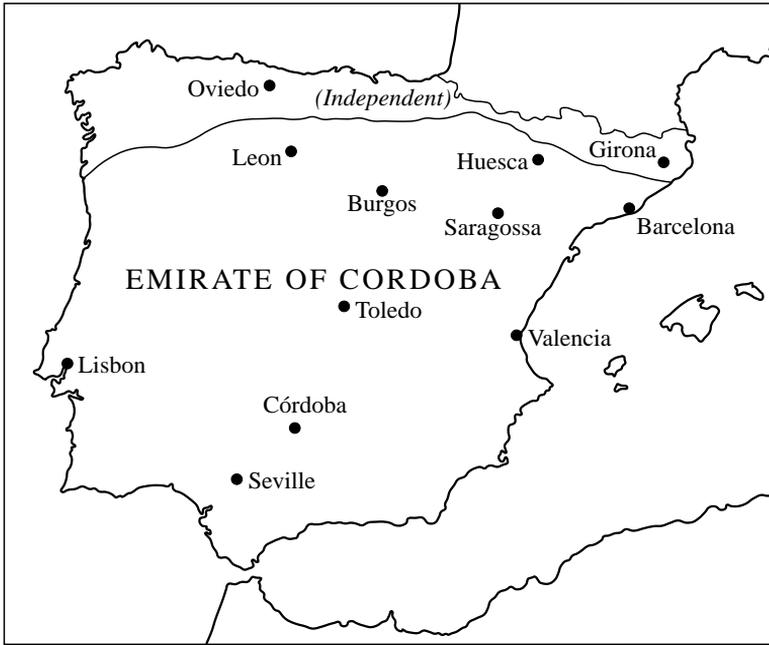
The ruling Visigothic group constituted a small fraction of the total population of the Peninsula, and despite their political supremacy, they sooner or later abandoned bilingualism and their speech became entirely assimilated to that of their subjects, who were not only numerically superior but, even in these 'Dark Ages', enjoyed a culture which was more prestigious than that of their rulers. Throughout this period, the large majority continued to speak Latin, no doubt with considerable and increasing variation between one locality and another.

It was probably this divorce between political power and cultural prestige which allowed centrifugal, linguistically diversifying, forces to gain the upper hand over centralizing and linguistically unifying forces. Despite the fact that the Visigoths eventually ruled the whole Peninsula, they presided over a period in which diatopic variation of speech was increased rather than diminished. However, there is one political event of this period which was to have great linguistic significance at a later date: the establishment of Toledo as the centre of government. For the first time in Peninsular history, the seat of political power was situated in the central meseta and, after the collapse of Visigothic Spain and the Moorish conquest of the early eighth century, Toledo therefore assumed great symbolic importance to the northern Christians, who to some extent saw their mission as the reestablishment of Christian Visigothic Spain. The fact that Toledo fell (in 1085) to Castilian reconquerors endowed Castilian speech with a prestige it might otherwise not have enjoyed, and can therefore be seen as an important factor in the rise of Castilian to national status (see 1.4).

1.3.2   *Moors and Christians*

The Islamic invasion of 711 had enormous linguistic consequences. It was not merely that it brought Hispanic Latin and its successors into contact with the language of a culture which was soon to be more developed and prestigious than that of Christian Europe, thereby creating the conditions for substantial lexical and semantic borrowing from Arabic (see 4.6, 5.1.5), for the modification of the syntax and phraseology of Hispano-Romance (see Galmés 1956; also Lapesa 1980: 156–7 for the Arabic origin of phrases like *que Dios guarde/que Dios mantenga, si Dios quiere, Dios le ampare, bendita sea la madre que te*

Oviedo ● *(Independent)*

Leon ●   Girona ●
                        Huesca ●
Burgos ●
              Saragossa ●   Barcelona ●

EMIRATE OF CORDOBA
            ● Toledo

● Lisbon   Valencia ●

        Córdoba ●

    ● Seville

Map 1.3 Spain in the late eighth century AD

parió, etc.), and for occasional morphological borrowing (e.g. the suffix *-í*; see 4.14.2.1). The linguistic effects of the Moorish conquest were even more profound, since the dialectal map of Spain was entirely changed, and importance was given to varieties of Romance which, in the absence of this political upheaval, would have remained insignificant and peripheral. The reason is, of course, that the Moorish armies failed to conquer the entire Peninsula. Between 711 and 718 they established control over approximately three-quarters of its territory, but allowed the survival of Christian nuclei in the extreme north and northwest (see map 1.3). These were precisely the areas which had been remotest from standardizing influences during the Roman period and from such linguistic levelling processes as obtained during the period of Visigothic rule. It can therefore be argued that they were the areas of the Peninsula where speech was most distant from the 'norm' of eighth-century Hispano-Romance speech. This was no doubt particularly so in the case of Cantabria (modern Santander, northern Burgos and adjacent areas), the southern part of which is the area where Castilian has its origins and which was especially resistant to Roman and Visigothic rule and whose language in the eighth century is likely to have been particularly 'abnormal'. (It is recognized that there can

have been no single accepted prestige-norm for speakers of eighth-century Hispano-Romance, and the term 'norm' here is a means of referring to those linguistic features which were common to most varieties of Hispano-Romance speech.)

The linguistic effects of the Christian Reconquest of the Peninsula are similarly great. Features of Hispano-Romance speech which had hitherto belonged to geographically peripheral and linguistically unusual varieties are extended southwards at the expense of those features which one can presume were previously the most prestigious and the most similar to those of the Romance spoken outside the Peninsula. And among these peripheral features of Hispano-Romance, it was those belonging to the most 'abnormal' variety, namely Castilian, which were to achieve the greatest territorial and cultural spread. At first typical only of the speech of the Burgos area of southern Cantabria, Castilian linguistic characteristics were carried south, southeast and southwest, in part by movement of population, as Castilians settled in reconquered territories, and in part by the adoption of Castilian features by those whose speech was originally different. The creation of the kingdom of Castile in 1035 no doubt sharpened awareness of the separate identity of Castilian speech and the capture of Toledo in 1085 (by Alfonso VI, king of both Castile and Leon) has already been noted as having considerable linguistic significance, by reason of the prestige that this success afforded to Castile and to Castilian speech (see map 1.4).

After what proved to be temporary setbacks at the hands of Almoravid and Almohad reformers of Islamic Spain in the late eleventh and twelfth centuries, the Castilian advance continued with the capture of the major cities of northern and western Andalusia (Córdoba 1236, Jaén 1246, Seville 1248, Cádiz 1250) and with control over the kingdom of Murcia (1244). By the mid-thirteenth century, then, Castile had expanded to comprise something over half of the Peninsular territory and Castilian speech was on the way to displacing its competitors, Arabic and Mozarabic, the latter term indicating those varieties of Hispano-Romance which had continued to be widely spoken in Islamic and ex-Islamic Spain. The contact between Castilian and Mozarabic produced some effects upon Castilian, largely restricted to borrowing of Mozarabic vocabulary (see 4.7), but perhaps including the development of the sibilant consonants in Andalusian (and, later, American) varieties of Castilian (see 2.6.3). However, it is likely that Mozarabic speech was assimilated to Castilian patterns (or was abandoned in favour of Castilian speech) during the thirteenth and fourteenth centuries. (For further discussion, see Penny 2000: 75–80.)

Between the mid thirteenth century and the end of the fifteenth, Islamic Spain consisted only of the mountainous southeastern parts of Andalusia, namely the kingdom of Granada. When this area was captured in 1492 by the Catholic Monarchs, Ferdinand and Isabella, it was largely resettled by speakers of Andalusian varieties of Castilian, so that in the course of six centuries Castilian

Map 1.4 Spain in AD 1150

had come to occupy a territory stretching from the Cantabrian coast to the Atlantic and the Mediterranean.

However, it should be made clear that Castilian speech characteristics were spread not simply to those central and southern Peninsular territories into which the kingdom of Castile expanded. At the same time as this southward development was taking place, people in neighbouring kingdoms were adopting Castilian manners of speech. In the case of Leon, the westward spread of Castilian features is firmly attested, in literary and non-literary writing, well before the definitive union of Castile and Leon in 1230. Unattested, but presumably no less real, was the northeasterly advance of Castilian at the expense of Basque. Similar encroachment of Castilian features upon Aragonese territory is observable in texts written in Saragossa in the fourteenth and fifteenth centuries, that is, before the union of the crowns of Castile and Aragon in 1479. At this stage, only Galicia and the Catalan-speaking areas (Catalonia, Valencia and the Balearics) remained, for the most part, outside the Castilian sphere of linguistic influence.

The reasons for this lateral spread and imitation of Castilian features lie in the political prestige of Castile, stemming from its increasingly predominant role in the Reconquest, and in the development of its literature (see 1.4), which

had no comparable counterpart in the kingdoms of Leon and Aragon. The castilianization of these kingdoms was of course not rapid (although it was undoubtedly more swift among the educated than among the majority) and it is still incomplete today, in rural areas of Asturias, western Leon, northern Huesca, etc.

1.4   Standard Spanish

The creation of early standard Spanish is arguably the result of the work of one man, Alfonso X the Learned, king of Castile and Leon (1252–84). Writing by means of a spelling system which was able to specify vernacular pronunciation, by contrast with writing in Latin, goes back to the period following the reforms of the Council of Burgos in 1080 (see Wright 1982), and vernacular writing in the kingdom of Castile, both literary and non-literary, becomes ever more frequent in the late twelfth and early thirteenth centuries. However, until the period of Alfonso X, all writing can be seen to be dialectal, in the sense that the language used shows some features characteristic of the writer's region, rather than representing any supraregional variety. Thus, the late twelfth-century *Auto de los reyes magos* reveals features of the speech of Toledo (perhaps due to contact with Mozarabic) not shared with the rest of the kingdom, while the *Poema de mio Cid* displays a number of characteristics which locate its language in the east of Castile. Non-literary writing is no different in this respect; the *Fuero de Madrid*, which reached its final form in 1202, is recognizably from New Castile.

Such regional characteristics disappear, for the most part, in the later thirteenth century, as a result of the scholarly activities of the king and his collaborators. On the one hand, the use of Castilian as the vehicle of an enormous output of scientific, historiographical, legal, literary and other work, was bound to lend great prestige to the chosen medium, Castilian, by contrast with other varieties of Hispano-Romance, such as Leonese or Aragonese, which enjoyed little literary cultivation. On the other hand, the king's express concern over the 'correctness' of the language of his scholarly output is a witness to the creation of a standard form of Castilian. Certainly, as just stated, by the end of Alfonso's reign it is no longer possible to identify a specific regional flavour in the writing of Castilians. It is reasonable to assume that the new supraregional literary standard was based upon the speech of the upper classes of Toledo, a form of speech which, as we have seen, owed many of its features to varieties spoken in the Burgos area, which had become dominant in the speech of Toledo following the Reconquest of New Castile.

A further important aspect of Alfonso's activities was the consistent use of Castilian as the language of administration. Latin had been partly abandoned in the previous reign, but was now definitively superseded by Castilian, which had the culturally unifying advantage of being religiously neutral, by contrast with