
Variation and change in Spanish

Ralph Penny



Contents

Preface—ix

Abbreviations and symbols—xi

1 Introduction: language variation—1

1.1 Synchronic variation—1

1.1.1 *Geographical or diatopical variation*—1

1.1.2 *Social variation*—2

1.2 Diachronic or historical variation—3

1.3 Variables and variants—5

1.4 Co-variation—6

1.5 Register—6

1.6 Variation in the past—7

2 *Dialect, language, variety: definitions and relationships*—9

2.1 *Dialects*—10

2.2 *Languages*—12

2.2.1 *Do languages have temporal limits?*—12

2.2.2 *Are languages delimitable?*—13

2.3 Relationship between *dialects* and *languages*—15

2.4 *Varieties* and *idiolects*—18

2.5 Relationship between *varieties*—19

2.5.1 *The tree model of relationships between varieties*—20

2.5.2 *Geographical discontinuity*—28

2.5.3 *Diasystems*—30

2.5.4 *Diglossia*—32

2.5.5 *The neolinguistic model*—34

2.5.6 *Other models*—36

3 Mechanisms of change—37

3.1 Dialect contact—38

3.1.1 *Accommodation*—39

PUBLISHED BY THE PRESS SYNDICATE OF THE UNIVERSITY OF CAMBRIDGE
The Pitt Building, Trumpington Street, Cambridge, United Kingdom

CAMBRIDGE UNIVERSITY PRESS

The Edinburgh Building, Cambridge, CB2 2RU, UK

40 West 20th Street, New York, NY 10011-4211, USA

10 Stamford Road, Oakleigh, Melbourne 3166, Australia

Ruiz de Alarcón 13, 28014 Madrid, Spain

www.cup.cam.ac.uk

www.cup.org

© Ralph Penny 2000

This book is in copyright. Subject to statutory exception
and to the provisions of relevant collective licensing agreements,
no reproduction of any part may take place without
the written permission of Cambridge University Press.

First published 2000

Printed in the United Kingdom at the University Press, Cambridge

Typeface Monotype Dante & Adobe Futura System QuarkXPress™ [SE]

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

ISBN 0 521 78045 4 hardback

7 Standardization

Most of the discussion in this book has been concerned with non-standard varieties of Spanish, since it is there that most cases of variation and change are to be observed. However, account must also be taken of the standard varieties, which are perhaps best regarded as highly unusual forms of language, in the sense that they are rare (all humans use language, but only a minority use a standard language) and recent (they arose only in the last few thousand years of the multi-millennial history of human language). The reason why the process of standardization has an important place in any treatment of linguistic variation and change, in Spanish and other languages, is that an essential aspect of this process is reduction in variation within certain high-prestige varieties (see 7.2). However, paradoxically, we should not ignore the fact that for speakers of low-prestige varieties, the establishment of a standard may imply an increase in the range of variation available, since variants from the standard may enter the speech of non-elite groups and be added to the competition among pre-existing variants (see 7.3).

Although we have seen (1.1–2) that variation is inherent in language, the process of standardization may, in principle, reduce variation to zero in the variety which is subject to it. This elimination of variants, naturally, applies to instances of variation which are due to the normal effect of changes which are working their way through society (3.4), as well as to instances of variation which are more stable and long-established.¹

Standardization is a process which takes place within written language, and which is indeed inconceivable in the absence of writing. However, the variants selected for use in writing may then oust their competitors from those varieties of spoken language which have most in common with written language (those based on written language and used on formal social occasions), and may go on to be similarly selected in other spoken varieties, owing to the prestige associated with written forms of language. But because standardization belongs essentially to

written language, it cannot, in principle, directly affect the phonetic and phonological levels of language. That is to say that those who use a given standard may do so while differing from one another in their pronunciation. Nevertheless, the speech of those who are responsible for the development of a linguistic standard (generally, members of powerful urban groups) is likely to be invested with the prestige which derives from their association with (including control of) the written language, so that the pronunciation features used by such groups may come to constitute an effective phonetic standard, and over time there is likely to be reduction of phonetic and phonological variation in the society concerned.

Nor should we overlook the fact that the spelling system which is used to write a language can have an effect on the phonology of spoken varieties. Although orthographical systems, when they are first applied to a language, typically match the phonological structure of (some high-prestige variety of) that language, this one-to-one relationship between letter and phoneme may be disrupted (by phonological change, etymological preconceptions, etc.), so that mismatches occur. Given the prestige that is associated with writing, users of the language may come to believe that distinctions made in spelling ought to be reflected in phonological distinctions, and may then make changes in their pronunciation. Thus, it may have been an awareness of spelling (in this case the Latin spellings *CRUDUS*, *NIDUS*, *NUDUS*, *VADUM*) which resolved the variation in early medieval Spanish between *crudo*, *nido*, *desnudo*, *vado*, etc. and *crúo*, *nío*, *desnúo*, *vao*, etc., in favour of the forms with /d/ (see 1.2). Similarly, the early modern resolution of variation between such pairs as *escrebir*~*escribir*, *recebir*~*recibir*, *vevir*~*vivir* (in favour of the forms spelt with <i> and pronounced with atonic /i/, despite the strong dissimilatory pressure which led to preference for *decir*, *sentir*, etc., from earlier *dezir*~*dizir*, *sentir*~*sintir*, etc.) may be due to spelling influence from Latin *SCRĪBERE*, *RECIPERE*, *VĪVERE* (Penny 1991a: 159–60). Again, although unsuccessful in the long run, an attempt has frequently been made by schoolteachers, especially in Spanish America, to introduce a labiodental /v/ in words spelt <v> (e.g., *vivir*), in order to make a contrast with bilabial /b/, which should supposedly be reserved for words with (e.g., *beber*); this, despite the fact that there has been only one non-nasal voiced bilabial phoneme in Spanish since the sixteenth century (see 3.1.3.3).

In the case of the standardization of Spanish, the processes concerned almost all occurred in Spain, with the prescriptive results being exported to the rest of the Spanish-using world.² Only since the

middle of the twentieth century has American Spanish been taken into account in establishing prescriptive grammars and dictionaries of the language, a change of vision which has accompanied increasing worries about the unity of the language and which has perhaps been spurred by a perceived threat of its fragmentation into mutually unintelligible varieties.³

In discussing standardization it is useful to distinguish between two contrasting but interlinked aspects of the process. On the one hand, there are the more purely social, strictly extra-linguistic, aspects; these are concerned with the way in which society selects a variety to serve as a basis for the standard, codifies and promotes it, and seeks acceptance for it. These aspects are sometimes encapsulated in the term *status planning*.⁴ On the other hand, there are the intra-linguistic aspects of the process, concerned with which variants are selected and which abandoned, and with such matters as the sources of elaboration of syntax and vocabulary. Although these aspects of standardization are often less conscious, an element of planning may be involved here, and the label *corpus planning* is sometimes applied to this part of the process. We shall consider status planning (7.1) and corpus planning (7.2) in turn.

7.1 Status planning

The processes which make up status planning (selection, codification, elaboration of function, and acceptance; see Haugen 1972; Hudson 1996: 32–4) reflect the different degrees of power exercised by different social groups. Only varieties spoken by politically and economically powerful groups are likely to be selected as the basis of a standard language. Similarly, only such groups (or individuals) are capable of imposing particular codifications of the language and of ensuring it will be used in an increasing number of domains. Likewise, only the powerful can bring about acceptance of the emerging norm, since only they enjoy sufficient social prestige to cause other groups to follow their linguistic preferences.

7.1.1 Selection

Every standard language grows out of some spoken variety or varieties, which are in competition with a much larger number of other

varieties, which are not so selected. Since the creation of a standard cannot be achieved without the devotion to it of great resources (required for writing, making, and copying books, etc.), it follows that the varieties which underpin standards are always those spoken by the wealthiest and most powerful groups. The development of standard Spanish is no exception to this rule.

A number of the features which later help to characterize standard Spanish are seen to occupy areas which, in the tenth and eleventh centuries, overlap in a small segment of the northern Peninsular dialect continuum (4.1.2), an area comprising the city of Burgos and a number of surrounding towns in northern Old Castile (Menéndez Pidal 1964: 485–6). Among the most geographically restricted of these features are:⁵

- 1 Use of glottal /h/ as a reflex of Latin F-, while surrounding districts maintained labiodental /f/ (e.g., /húmo/ vs /fumo/ < FŪMU 'smoke'; 3.1.3.2).
- 2 Use of prepalatal /z/ as the reflex of the Latin groups -c'L-, -G'L-, -L- + [j], while adjacent districts showed /ʎ/ (e.g., /abéza/ vs /abéʎa/ < APIC(U)LA 'bee'; 3.1.3.1).
- 3 Loss of the initial palatal consonant (probably /j/ in spoken Latin) when followed by an atonic front vowel, while other areas maintained a palatal of some description (e.g., /enéro/ vs /jenéro/, /zenéro/, etc. < *JENUĀRIU 'January').
- 4 Use of the prepalatal affricate /tʃ/ as a reflex of Latin -CT-, while further afield the Latin /t/ was maintained, sometimes preceded by a palatal glide (e.g., /estrétʃo/ vs /estréito/, etc. < STRICTU 'narrow').
- 5 Use of the dental affricate /tʃ/ as a reflex of the Latin groups -sc- (followed by a palatal vowel), -sc + [j], and -st + [j], etc., while surrounding districts showed the prepalatal fricative /ʃ/ (e.g., /kret'ér/ vs /krefér/ < CRESCERE 'to grow', /at'uéla/ vs /aʃuéla/, etc. < ASCIOLA 'hoe').
- 6 Retention of undiphthongized /o/ as a reflex of Latin tonic ō in the vicinity of a palatal glide, while surrounding varieties show a diphthongized result (e.g., /ózo/ vs /ueáo/, etc. < OCULU 'eye').

Since, at this time, no spelling system was in use in Spain which was capable of specifying contemporary phonology (by contrast with the traditional writing-system usually referred to as 'Latin'; see Wright 1982), there was no possibility of the development of a standard language based on the speech of Burgos. However, many of the features

of Burgos speech were extended outwards from this core district, from the eleventh century onwards. This happened partly as a result of resettlement of people from Burgos to other areas, in the wake of the military success of Castile during the early Reconquest (4.1.7), and partly as a result of accommodation processes (3.1.1) which led speakers in a wider and wider area to imitate the speech of the politically and culturally prestigious city of Burgos. Crucially, groups of speakers who used varieties which had developed in the Burgos area achieved political and social power in Toledo, following its recapture in AD 1085, and, in the dialect mixture which consequently arose in the new Castilian capital, features of Burgos speech more often found favour than those of the local Toledan varieties (4.1.1) or those brought by other immigrants (see 3.1). Consequently, when the speech of upper-class Toledo was selected as the basis of the Castilian standard, it was characterized by a significant number of features imported from Burgos.

The selection of the speech of Toledo as the foundation of the standard followed from the political and religious importance of the city (it was the headquarters of the Castilian Church, and the most usual seat of the Castilian Court), and from its cultural prestige (it was there that the major scientific and literary enterprises of the twelfth and thirteenth centuries took place). When new systems of writing Romance were introduced from France (applied sporadically from the late eleventh centuries, and regularly from the early thirteenth; see Wright 1982), these were employed to write literary and non-literary texts using varieties of Romance from a wide range of localities. But, within Castile (and Leon following the definitive merger of the Crowns in 1230), it was the features of educated Toledan Spanish which soon became pre-eminent in writing. Because of the overwhelming use of this variety in the literary, legal, and scientific output of the scriptorium of Alfonso X the Learned (1252–84), it became the model for all writing, including that of government documents, throughout the kingdom.

This state of affairs continued until the mid-sixteenth century, despite competition from the great mercantile and cultural centre of Seville (4.1.7.2). Juan de Valdés, the great arbiter of linguistic taste, writing his *Diálogo de la lengua* in about 1535, derives his self-assigned authority from the fact that he is an 'hombre criado en el reino de Toledo y en la corte de España'.⁶ Others evidently shared his perception of the prestige of Toledan Court usage, since the forms he selected for recommendation were most frequently, if not quite always, those

that came to be preferred in the written standard. His often unreasonable objections to the prescriptions of Antonio de Nebrija (see (7.1.2) were based on the fact that Nebrija was from Andalusia.

The situation only changed in 1561, with the creation of Madrid as the capital of Spain (see 3.1.3.1–3). The new dialect mixture that this event inspired, through influx of speakers from Old Castile, gave rise to a selection of features that differed to some extent from those used in Toledo. Thereafter, whatever might be claimed by other centres, such as Valladolid, it was educated varieties of Madrid Spanish that were mostly regularly reflected in the written standard.

What happened in Castile was paralleled by the selections made in other parts of the Peninsula. In the west, the early importance of Santiago de Compostela as a religious and cultural centre implied that, when characteristically Romance writing emerged in the northwest, it would be based on the speech of that city. Galician writing of this kind continued until it was replaced by the Castilian standard in the early modern period. However, although the features of northwestern speech, extended southwards during the Reconquest (4.1.7.1), contributed substantially to the dialect mixture which arose in the major cities of Portugal (independent of Castile–Leon from 1143), the Portuguese standard which later emerged, based upon the speech first of Coimbra and then of Lisbon, differed in minor ways from the norm applied in Santiago (which remained part of Castile–Leon).

Some texts written in the early thirteenth century in Castile–Leon show certain features characteristic of the speech of regions other than Toledo. Thus a number of Leonese features appear in one version of the *Libro de Alexandre*, northern Castilian features are visible in the *Disputa del alma y el cuerpo*, the poetry of Gonzalo de Berceo has some features characteristic of La Rioja, and the language of the *Poema de mio Cid* has been identified as belonging to the east of Old Castile (Lapesa 1980: 203–5). Similarly, the language of the *Auto de los reyes magos*, the subject of much dispute, most probably reflects varieties of Toledo speech distant from that of the most cultured classes, and preserving certain features of Mozarabic origin.

Aragon remained a separate kingdom until the union with Castile–Leon in 1474, and was a territory where two standards emerged. The Catalan standard was based upon the speech of the main northeastern cities, principally that of Barcelona, because of its political and mercantile importance, but came to be used throughout Catalonia, Valencia, and the Balearic Islands, as a result of the Reconquest of these areas,

during the thirteenth and fourteenth centuries. From the beginning of the sixteenth century, the Castilian standard became the normal vehicle of writing, and Catalan writing was restricted to documents of purely local significance, until the nineteenth-century recreation of a Catalan standard, which continues in use beside Castilian.

In the inland regions of the Crown of Aragon, it was the city of Saragossa that played the major linguistic role. A flourishing centre of Islamic commerce and culture, it was recaptured in 1118 by the then tiny kingdom of Aragon and immediately became the new capital, retaining this role after the merger of Aragon with Catalonia in 1137. The development of writing there seems never to have been based simply upon the speech of the city, which, in the century following its reconquest, no doubt consisted of a dialect mixture comprising varieties transferred there from the central Pyrenean valleys (the core territory of the kingdom), together with Mozarabic varieties spoken by the existing population, and other forms of speech contributed by immigrants from other Peninsular and trans-Pyrenean areas. Even the earliest Romance writing produced in the expanded Aragon shows a preponderance of Castilian features, and the same continues to be true even during the fourteenth-century expansion of Aragonese writing sponsored by Juan Fernández de Heredia. Although this Aragonese standard had developed in a state which was politically independent of Castile, it was not independent of the Castilian linguistic standard which had emerged in thirteenth-century Toledo, and scarcely survived the union of the Crowns in 1474.

7.1.2 Codification

Following selection of a base-variety for the purposes of writing, a further important stage in the process leading to standardization is the codification of this variety. The goal of this mechanism (utterly unachievable in the case of spoken language) is 'minimal variation in form' (Haugen 1972: 107). In the case of written language, it consists of the prescription of a set of unvarying orthographical, grammatical, lexical, and other rules to which writers should conform, if their writing is to carry the highest prestige.

Although explicit codification of Spanish did not begin until the late fifteenth century, the linguistic preoccupations manifested by Alfonso X in the works he sponsored (7.1.1) had the effect of producing a linguistic model which could be and was imitated by other writers.

7.1 Status planning

Whereas at the beginning of his reign, in the mid-thirteenth century, there was much variation between the language of one text and that of another (see 7.1.1), only half a century later this variation has significantly declined. By the end of the thirteenth century, as a direct result of the linguistic efforts that went into producing the vast Alfonsine literary and scientific output, it is no longer possible to recognize the regional origins of the writer, and other types of variation have been very much reduced.

Another, even more effective, type of informal codification of Spanish is that carried out by the early printers. Ray Harris-Northall (1996–7), by comparing the manuscripts of the *Gran conquista de ultramar* with the early-sixteenth-century printed version of this work, has shown how variation across a wide range of linguistic features was dramatically reduced, and how the text was extensively revised in order to project the norms favoured by a political elite.

Turning to explicit codification, we can see that Spanish was the first post-classical language to be subjected to this process. From the middle of the fifteenth century there appeared a series of small-scale lexicographical enterprises (Lapesa 1980: 286–7), followed by Alonso de Palencia's monumental *Universal vocabulario* (1490). Although the latter was a Latin dictionary, its Spanish component reveals a subtle discrimination among the words of the late-medieval lexicon. However, the most important early work of codification of Spanish, and the first grammar of a modern language to appear (in 1492), was Antonio de Nebrija's *Gramática de la lengua castellana* (Nebrija 1980).⁷ Born in Andalusia, and professor successively at the universities of Salamanca and Alcalá de Henares, he also published in 1492 his Latin–Spanish dictionary (Nebrija 1979), followed, probably in 1495, by the first edition of his Spanish–Latin dictionary (Nebrija 1973). A 1517 treatise on Spanish orthography (Nebrija 1977) expands on the discussion of spelling in the *Gramática*.

Nebrija's only clarification of the type of Spanish reflected in his books is his comment on the basis of his spelling system ('I write as I speak'), the same formula adopted by Juan de Valdés in his *Diálogo de la lengua*. With these words, these authors claim to eschew etymology as a guide to spelling and (to interpret their precept in a modern way) to match their graphical systems to the system of phonemes they used (and which was presumably also the system used by those who spoke similar varieties). That they do not quite achieve this is unsurprising, given that, like scholars for centuries before and after their time, they

did not manage to fully disentangle the discussion of sounds from that of letters. With regard to the other aspects of language, we can only assume that Nebrija, like Valdés, sought to codify the variety of language written by the educated men of Castile.

Because of the phonological changes which took place in the late medieval and early modern periods (see, for example, 3.1.3.1–3), the spelling system established in the Alfonsine texts (and adopted with little change by Nebrija) became more and more distanced from the phonology of Castilian. Concern to reform the spelling of Spanish was constant during the Golden Age, and found expression in such works as Mateo Alemán's *Ortografía castellana*, published posthumously in 1609 (Alemán 1950), and in the radical (but disregarded) *Ortografía castellana*, published in 1630 by Gonzalo Correas (Correas 1971). However, successful reform and codification of Spanish spelling had to await the establishment of the Real Academia Española in 1713. Inspired by the role of the Académie Française, one of the main elements of the Spanish Academy's watchword ('limpia, fija y da esplendor') was to codify the language, and the academicians explicitly tackled Spanish spelling in their *Ortographía* (1741). This and subsequent orthographical pronouncements made by the Academy (in the late twentieth century published after consultation with similar Academies established in every Spanish-speaking country including the United States) have been closely and rapidly followed by printers and publishers, so that the spelling of Spanish today, at least in printed texts, is highly uniform throughout the world.

Codification of morphology and syntax was a slower process. Following Nebrija's *Gramática* (see above), innumerable grammars of Spanish were published in Spain, Flanders and America, during the sixteenth and seventeenth centuries, including, in 1625, Gonzalo Correas's notable *Arte de la lengua española castellana* (Correas 1954). A milestone was reached in 1771 with the publication of the Real Academia Española's first *Gramática*, repeatedly updated down to the present. The result of these and hundreds of other grammatical publications is that the morphology of written Spanish is almost entirely uniform across the world, while its syntax varies to a minor extent (see, for example, 4.1.7.2.7, 5.1.2.5, 5.1.3.2).

Codification of vocabulary is bound to progress most slowly, and cannot be completely achieved, because of the open-ended nature of the lexicon. Alonso de Palencia and Nebrija made great strides in the late fifteenth century, and were most notably followed by Sebastián de

Covarrubias, whose 1611 *Tesoro de la lengua castellana o española* no doubt continues to give primacy to the vocabulary used in his native Toledo. Once again, it was the activity of the Real Academia Española, publicly realized in 1726–39 in its three-volume *Diccionario de autoridades*, which provided the most authoritative lexical codification of Spanish. More recent editions of the Academy's dictionary have reflected not just written Peninsular usage but, with input from the American Academies, have aspired to include items from throughout the Spanish-speaking world and belonging to various registers.

7.1.3 Elaboration of function

The goal of this aspect of standardization is 'maximal variation in function' (Haugen 1972: 107), that is, the introduction of the language concerned into the largest number of domains. In the case of Spanish, and other Romance languages, this process could begin only once there existed a consciousness of the language as a code separate from Latin. Roger Wright (1982) has successfully shown that such consciousness only grew as a result of the introduction of a spelling system which sought to specify the phonemes of the variety being written, a process which was applied to Spanish from the end of the eleventh century (Lloyd 1991). At first, use of the new spelling was infrequent and inconsistent, as seen in the Glosses written in the monasteries of San Millán de la Cogolla and Santo Domingo de Silos, but the new system was used with increasing sophistication in some documents of this period, especially in the early thirteenth century, and culminated in the almost fully consistent system adopted by the school of Alfonso X from the mid-thirteenth century.

Once full consciousness of Spanish as a separate code had been established, it could challenge Latin in an increasing number of domains. Narrative poetry was the first prestigious domain in which written Spanish asserted its independence (along with that of religious drama, where we have the isolated example of the *Auto de los reyes magos*), in the early thirteenth century, although we have seen (7.1.1) that at this stage written Spanish still revealed some geographical variation. It was, once again, the scientific and literary undertakings of Alfonso X the Learned which dramatically enlarged the domains in which Spanish could be used. It now came to be the vehicle of prose narrative, historiography, science (astronomy, astrology, mineralogy, etc.), and jurisprudence, domains from which it rapidly ousted Latin,

and to which were soon added lyric poetry and even some types of religious writing, such as the many thirteenth- and fourteenth-century translations of the Bible. From this time onwards, no domain was beyond the reach of Spanish, despite the partial return to Latin for learned discourse during the Renaissance, and the persistence of Latin as the language of the Catholic mass until the second half of the twentieth century.

This medieval and early-modern elaboration of the function of Spanish not only led to informal codification of Spanish (7.1.2), but to dramatic expansion of lexical and syntactical resources. The need to express ideas previously unexpressed in Spanish required a vast increase in its vocabulary, a need which was supplied chiefly by borrowing (from Arabic, Latin, French, Occitan, etc.), but also by frequent derivation, mostly through suffixation (Lapesa 1980: 243–4, Penny 1987–8). Similarly, although the syntax of early written Spanish was relatively simple, a quality especially evident in the rarity of subordination and in the frequency with which clauses are connected by *e* ‘and’, the linguistic sophistication required to express the complex logical relationship of historical and scientific ideas was provided by a growing syntactical elaboration, so that by the end of the Alfonsine period Spanish was capable of elegantly expressing any relationship of ideas in any domain.

7.1.4 Acceptance

Acceptance of a particular linguistic code as a symbol of a given nation is a process which, in its full sense, belongs properly to recent centuries. However, it should be remembered that all language names are political in origin, in that they begin by referring to the speech of a portion of territory whose boundaries are determined by the control exercised by some political entity. Thus the term *castellano* (‘Castilian’) is an abbreviation of *romance castellano*, approximately ‘those forms of speech (descended from Latin) used in the territory named Castile’, where *Castile* is a political entity, not originally sovereign, whose boundaries separate it from other political entities. The existence of a language name therefore implies, at least for its first users, a connection between that name and some political entity.⁸

The promotion of a language name (together with, or separately from, the promotion of the language to which it refers) is therefore an instrument of nation-building. In the case of Castilian, it is likely (but

unprovable, since he left no manifesto) that Alfonso X was in part promoting Castilian nationhood when he promoted written Castilian, for administrative as well as literary and scientific purposes. The advantages, in this respect, of Castilian over its available competitors were considerable. Each of Latin, Arabic, and Hebrew was indissolubly linked with one of the three religions adhered to in Castile, whereas almost all inhabitants of the kingdom spoke Castilian or some closely related variety of Romance.⁹ The politically unifying effects of promoting the use of written Castilian in the thirteenth century are therefore evident.

Explicitly nationalistic exploitation of language arrives only at the end of the fifteenth century. In his address to Isabella I of Castile and Aragon, placed at the head of his 1492 *Gramática de la lengua castellana*, Antonio de Nebrija famously links language and nation when he states: ‘siempre la lengua fue compañera del imperio; & de tal manera lo siguió, que junta mente començaron, crecieron & florecieron, & después junta fue la caída de entrambos’ (Nebrija 1980: 97). *Imperio* here does not refer to what we think of as the Spanish Empire; Columbus’s ships were indeed on the way to America as Nebrija wrote, but the American Empire was as yet undreamed of. Nebrija no doubt had in mind the territories governed by the Catholic Monarchs (Spain (still excluding Navarre), the Balearic Islands, much of central and southern Italy, Sicily and Sardinia, the Canary Islands) and any other territory they might acquire. This was the Empire which, he claimed, shared the same fate as the Spanish language, and whose success required the promotion of the Spanish language.

Nebrija refers to the language whose forms he prescribes as *la lengua castellana*, in keeping with earlier usage, in which the term *castellano* was the only label available for this concept. The term *español* (earlier *españón*), and the noun *España* from which it derives, were infrequent in the Middle Ages, when they referred to Islamic Spain, or to the historical concept of Roman Spain, or to the geographical concept of the Spanish Peninsula. Following the union of the Crowns of Castile and Aragon in 1469, the terms *España* and *español* were newly applied to this United Kingdom, from which it followed that the main language of this state, hitherto called only *castellano*, could also be called *español*. Since that time the two terms have been equivalent, although one or the other has been preferred at different times and in different places. See Alonso (1943) for a more detailed history of these terms.

In America, the ruling elites of the post-colonial states, which emerged from the Spanish Empire, were all Spanish-speaking (although

speakers of Native American languages no doubt formed larger proportions of the populations of these countries than they now do). Nationhood was consequently linked to the use of Spanish, sometimes explicitly so when the new constitutions named Spanish as the official language of the state, as also occurred in the post-Franco constitution of Spain.

The political acceptance of Spanish, implicit for centuries, was now explicit in almost all of the Spanish-speaking world.

7.2 Corpus planning

When we turn to the intra-linguistic aspects of standardization, we can see a progressively intensifying intolerance of variation in written Spanish from the thirteenth century onwards. As anticipated in 7.1.2, this process is earliest and most clearly observed in spelling, where it has become almost totally successful. In morphology, restriction of variation is slower but is almost completely successful by the nineteenth century, while in syntax the twentieth-century language still allows some degree of variation between competing structures. It is in vocabulary and semantics, predictably, that the greatest degree of variation persists in the written language, despite severe restriction over the centuries.

It will be convenient to divide our consideration of corpus planning chronologically, although the divisions we make are arbitrary and conventional. As we have seen (1.2), there are no linguistic grounds for periodizing the history of a language.

7.2.1 The medieval period

The emergence, in the first decades of the thirteenth century, of a considerable body of literature, written in the recently devised Romance spelling system, reveals very substantial variation, both between one text and another, and within any single text. Because this spelling system was an adaptation of the system devised for reading Latin texts aloud (Wright 1982: 126), and because those texts had no letters which demanded a palatal realization (since palatal consonants were absent from Latin when its spelling system was devised), the writing of Spanish (and other Romance languages) was bedevilled by the difficulty of finding spellings which might represent its palatal

phonemes. A similar difficulty arose in the case of affricate consonants generally, although it is less clear that affricates were *not* used in reading traditional texts aloud. A further problem was that the reading aloud of Latin required no distinction between voiced and voiceless sibilants, since, when its spelling system was created, Latin had only one sibilant (voiceless /s/).

The consequence of these difficulties was a very great variation in the Romance spelling of the phonemes belonging to either or both of the palatal order and the sibilant class. Early spelling hesitates, in the representation of /ɲ/, among <ni>, <gn>, <ng>, and <nn> (or abbreviations of the latter like <n̄> or <n̄>), /ʎ/ is reflected not only by <ll> and <l> but probably also by <pl>, etc. (e.g., *Poema de mio Cid: gallos, lamar, lorar, falir, plorando*), while the spelling <ch> for /tʃ/ was a late introduction from France, gradually displacing such ambiguous renderings as <g> and <i> (e.g., *Disputa del alma y el cuerpo: nog* (elsewhere *noch(e)*), *leio* (elsewhere *lecho*)). Likewise early writing does not discriminate consistently between /s/ and /z/, showing <s> or <ss> for both phonemes, or between /tʃ/ and /dʒ/ (e.g., *Disputa del alma y el cuerpo: amaneziēnt, lenzuelo, corazon* (where later spelling requires <ç>, for /tʃ/), *fecist* (where later spelling requires <z>, for /dʒ/)).

The Alfonsine texts are the first to rationalize the spelling of Spanish, using a system which comes close to the iconic ideal of one letter (and not more than one) for each phoneme. However, even the vigilance and linguistic good sense of Alfonso and his group could not produce a fully iconic system. Thus, <u> and <v>, <i> and <j> were used for both consonants and vowels (e.g., *viejo, vieio, uiejo, uieio* (for modern *viejo*), *uvo, vuo* (for modern *hubo*)), as elsewhere in Europe until the introduction of printing. Similarly, <g> had the double value of /g/ and /ʒ/ (*ganar, linage*), while /ʒ/ was also represented by <j> or <i>, disparities which have never been eliminated from the spelling of the descendants of these phonemes (witness MSp. *ganar, cogēr*, with <g> for both /g/ and /x/, and *coger, jengibre*, with both <j> and <g> for /x/). Nor were /s/ and /z/ consistently distinguished, since <s> (rather than <ss>) for intervocalic /s/ is frequently found in Alfonsine and later medieval texts.

Other prominent cases of variation in the emerging medieval standard are the following:

- 1 At the intersection of phonology and morphology, there was much hesitation in thirteenth-century Spanish between forms with final /e/ preserved and those without /e/, in words where

the vowel was (or had been) preceded by a consonant group or by a single consonant other than a dental or alveolar. Thus, beside invariable cases like *pared*, *paz*, *pan*, *mar*, *fiel*, *mes*, where earlier /e/ was preceded by an ungrouped dental or alveolar consonant, medieval Spanish showed variation between apocopated and unapocopated forms like *nuef/nueva*, *princep/principe*, *noch/noche*, *cuend/conde*, *mont/monte*, *part/parte*, *estonz/entonces*, etc. Rafael Lapesa (1951, 1975, 1982) charts the slow resolution of this variation in Alfonsine Spanish, in favour of the unapocopated forms, and suggests that these more closely reflect the usage of Alfonso's group. While Ray Harris-Northall (1991) finds little evidence of the decline of apocopated forms during Alfonso's lifetime, this element of variation was substantially reduced in the fourteenth century, and disappears by the end of the medieval period.

- 2 It is difficult to establish whether the variation between word-final <t> and <d> (e.g., *edat~edad*, *voluntat~voluntad*) was merely a matter of alternative spelling of a single sound (neutralization of /-t/ and /-d/ with voiceless realization, sometimes spelt <t> to reflect this voicelessness, but sometimes spelt <d> on the model of the invariable plurals *edades*, *voluntades*), or whether there was a genuine phonemic contrast here between /t/ and /d/ (unlikely, since this would be the only case of a contrast of voice in word-final position in Old or Modern Spanish). At all events, this variation of form continued throughout the medieval period and into the early sixteenth century.
- 3 In the fifteenth century, words which had inherited syllable-final /b/ (<-p-, -b-, -v-) (e.g., *cabdal*, *çibdat*, *debda*) alternated with forms in which /b/ had been vocalized to [y] (*caudal*, *çiudad*, *deuda*).¹⁰ Although this variation was still effective at the end of this century, because it is reflected in Judeo-Spanish (6.3.2(5)), it was then rapidly resolved in the standard, in favour of the forms with [y].
- 4 Although early literature, including the Alfonsine opus, prefers *-iello* as a first-choice affectionate diminutive, its competitor *-illo*, probably originating in the Burgos area, later begins to dominate, and *-iello* is ousted in the fifteenth century.
- 5 Verbal morphology was far from invariant in the thirteenth century. In the preterite we find variation between competing paradigms, drawn from separate Latin verbs (both *sove* and *fui* as past tenses of *ser*), between competing strong (stem-stressed)

preterites (e.g., *estide*, *estove*, and *estude* as preterites of the verb *estar*; *andide*, *andove*, and *andude* as preterites of the verb *andar*), between strong and weak preterites of the same verb (e.g., *fuyó* and *fujo* for *fuir*, *metí* and *mise* for *meter*, *prendí* and *prise* for *prender*, *sonrei* and *sonrise* for *sonreír*, *conquirió* and *conquiso* for *conquerir*, later replaced by *conquistar*), and between forms associated with the same grammatical person (third-person sing. *veno*, *fezo* beside *vino*, *fizo* in the case of *venir* and *fazer*). Most of these cases were resolved during the fourteenth century; few survived into the fifteenth.

- 6 Competition between rival imperfect and conditional endings in *-ía* and *-ie* (e.g., *tenía*, *tenie*, *tenié*; *cantaría*, *cantarie*, *cantarié*), frequent in the thirteenth century, was only resolved at the end of the fourteenth (Malkiel 1959), although the *-ie* forms today continue in rural use in the province of Toledo (Moreno Fernández 1984).
- 7 Medieval Spanish showed frequent examples of variation between rival lexical items with the same meaning. Sometimes it is possible to detect differences of register between the competing forms (e.g., *sinistro* 'left' and *can* 'dog' may have belonged more clearly to literary style than their synonyms *izquierdo* and *perro*, which ultimately prevailed, while the arabisms *alfayate* 'tailor', *alfageme* 'barber', *albéitar* 'vet', *alarife* 'architect', etc., similarly fell out of literary use, but later, in favour of their competitors *sastre*, *barbero*, *veterinario*, *arquitecto*, etc., borrowed from more fashionable Romance or Graeco-Latin sources). On other occasions, no such nuance is evident (e.g., *cabeça/tiesta* 'head', *pierna/cam(b)a* 'leg', *rodilla/(h)inojo* 'knee', *prender/tomar* 'to take', *quedar/remañir* 'to remain', *salir/exir* 'to go out', *mañana/matino* 'tomorrow'). Some of these cases of variation were slow to be resolved in the standard, and in a few instances (e.g., *prender/tomar*) the variation reached the early modern period.

7.2.2 The Renaissance and the Golden Age

Following the introduction of printing and the beginning of the codification process (7.1.2), pressure to avoid variation in written Spanish became more intense. However, standardization was far from being completed in the sixteenth and seventeenth centuries. Not only was the orthography of Spanish increasingly distanced from its

followed the whole verb, depending on clause-structure (*cantarélo, lo cantaré*).

- 9 Until the seventeenth century, the present subjunctive form of the verb *ir* was not standardized. We find *vayamos, vayáis* beside *vamos, vais* in this role.
- 10 The increment /g/ which had been added to the first-person singular present indicative and all present subjunctive forms like *caigo/caiga, oigo/oiga, traigo/traiga* (earlier *cayo/caya, trayo/traya*) was frequently extended to other verbs with stem-final /j/ (Penny 1991a: 154), so that literary Spanish until the seventeenth century allows *haiga, huiga*, and occasionally *vaiga*, beside *haya, huya, vaya*. The forms with /g/ were then restricted to non-standard use, where *haiga* and *vaiga* continue to thrive.
- 11 The irregular preterites of certain verbs remained unfixed until the seventeenth century. Until then, in the case of *traer*, literary Spanish allows *truxe, truxo*, etc. beside *traxe, traxo*. Similarly, the preterite of *ser* and *ir* hesitated, although only till the mid-sixteenth century, between the paradigm prescribed by Nebrija (1980: 250), *fue, fuete*, etc., and the forms which eventually became standard, *fui, fuiste*, etc.
- 12 In the early sixteenth century, the possessive adjective could still be optionally preceded by the definite article (*mi casa/la mi casa*), but thereafter the latter form was restricted to Northern rural use.
- 13 Until the mid-sixteenth century, the auxiliary used in the perfect tense of at least some intransitive verbs could be either *ser* or *haber* (*soy/he muerto, eres/has llegado*), with concord between the participle and the subject when the auxiliary was *ser* (*son llegados, somos idas*). This variation goes back at least to the thirteenth century, when, in the *Poema de mio Cid*, we find *tornado es don Sancho, el día es exido* beside *a Valencia han entrado*.
- 14 Until the late seventeenth century, the verbs *ser* and *estar* had not undergone their modern distribution of functions, so that the standard language of the Golden Age allowed both verbs to occur in locative expressions (*es/está aquí* = MSp. *está aquí*), and in resultant-state passive constructions (*es/está escrito* = MSp. *está escrito*).
- 15 It is only in the sixteenth century that the variation between the verbs *haber* and *tener* is resolved. Until that time, both verbs could occur in possessive phrases, although *haber* was by then rare

(*ha/tiene tres años de edad*), while *tener* could occasionally be found as a perfect auxiliary, without possessive sense, beside *haber* (*lo ha/tiene hecho* = MSp. *lo ha hecho*).

7.2.3 The eighteenth and nineteenth centuries

We have noted (7.1.2) the concern in eighteenth-century Spain with ridding the written language of variation, a concern best reflected in the express determination of the newly established Real Academia Española to 'fix' the language, and one which persisted into the second half of the twentieth century. Some of the main effects of the Academy's activities were the following (Lapesa 1980: 422–4):

- 1 The orthography had undergone no reform since the time of Alfonso X, and at the beginning of the eighteenth century was seriously out of alignment with the phonological structure of the language. However, spelling was progressively fixed between 1726 and 1815, by which time it achieved essentially its modern shape, where there is a fair (but far from perfect) match between phonemes and letters. The most important changes adopted were:
 - The prohibition of *y* as a representation of /i/ (thus *frayle/fraille* is settled as *fraille*), except (illogically) in word-final position (*muy, hay, rey*).
 - The allocation of exclusively consonantal value to *v* (*ave/aue* > *ave*) and vocalic value to *u* (*vno/uno* > *uno*).
 - The fixing of *h*, in the case of words whose Latin etymon contained *h-*, whether they were popular or learned (thus *ora/hora* > *hora*, *yerba/hierba* > *hierba*, *umilde/humilde* > *humilde*), in popular descendants of Latin words with *f-* (*hazer/azer* > *hacer*, but mistakenly *fazera/azera* > *acera*), and in arabisms (*alheli/aleli* > *alheli*).
 - By the early eighteenth century, all prestigious varieties of Spanish had a maximum of three phonemes descended from the six medieval sibilants (3.1.3.1, 4.1.7.2.1, 5.1.1.1). In central and northern Spain these were /θ/, /s/ and /x/, but their orthographical representation lagged behind the phonological reality, so that /θ/ was correctly spelt <z> in *hazer, dezir*, etc., and <c/> in *caça, cena*, etc., but with many confusions. Similarly, /s/ was represented 'correctly' either by <s> (in *casa, rosa*), or by <ss> (in *esse, cantasse*), and /x/ by <j> (in *viejo*,

'the friends he had known in his youth'), acquired conditional value in late Latin and maintained it in medieval and Golden-Age Spanish (a value now restricted to a handful of verbs, viz., *hubiera*, *debiera*, *quisiera*, equivalent to *habría*, *debería*, *querría*), and then from the fourteenth century gained their current value, at first only in the protasis of conditional sentences expressing improbability or impossibility (Wright 1932, Penny 1991a: 146–7). The forms in *-se* (*cantase*, etc.) go back to the Latin pluperfect subjunctive, which had already acquired imperfect subjunctive value in spoken Latin and have maintained it ever since. In Spanish America today, the *-se* forms are effectively unused, while in Spain the *-ra* forms vastly outnumber the *-se* forms (Marín 1980), the latter only being frequent in areas of coexistence with Catalan, whose sole past subjunctive is marked by /s/ (DeMello 1993). The eventual resolution of this morphological variation can be safely predicted.

- 2 The value and function of the third-person clitic pronouns is far from standardized at the present. As we have seen (4.1.2.3, 4.1.7.2.7), there are several competing systems of third-person pronoun reference in the Spanish-speaking world, originally associated with different regions, but often appearing in written Spanish from the same area. American Spanish is more uniform than Peninsular Spanish in this regard, having generalized the etymological system in which *lo*, *la*, *los*, *las* retain accusative value, while *le*, *les* retain dative value, both sets being unmarked for such values as [\pm human] or [\pm animate] (Table 4.1). This system was no doubt inherited from southern Spain, where it predominates today, in speech and writing. However, throughout Spain, there is competition in the written language between this system and others. Next, there is the semantically determined system of clitic pronoun reference (Table 4.2), in which *lo* is restricted to non-countable referents and *le*, *la*, *les*, *las* refer to countable concepts and are unmarked for case; that is, each of these latter forms is used as both a direct and an indirect object pronoun, the only relevant distinctions being gender and number. This system, one of extreme *leísmo* and *laísmo*, was dominant in Golden-Age Spanish, but is less firmly established in the modern Peninsular standard, where a third system competes with the two just outlined. This third system, described above as hybrid or interdialectal (see Table 4.3), deletes case-distinctions only for masculine singular human

referents (i.e., it displays limited *leísmo*), but maintains them for all other referents (e.g., it avoids *laísmo*). This last system, although it has been available since the thirteenth century, only found its way into prescriptive grammars of Spanish in the twentieth. There is no sign yet of which system will predominate in the long run.

7.3 The relationship between standard and non-standard varieties

Standard languages are different kinds of entities from spoken varieties. Whereas spoken language is infinitely varied, along all the familiar geographical and social parameters (1.1.1–2), the essence of a standard language is its lack of variation. Again, whereas spoken varieties have no boundaries, but merge one with another in an infinitely complex way, standard languages, because they are the creations of nation-states, can usually be mapped, their extent being determined by political geography. However, the two kinds of language are related (each standard has been developed from some set of spoken varieties, and a standard language will influence the spoken varieties used in the same territory), so how is this relationship to be envisaged? A helpful model is the one devised by Alberto Várvaro (1991) in connection with the way the various Romance standards emerged from the single Latin standard: the standard can be envisaged as a roof which covers a delimited portion of territory, and below which stretches the continuum of spoken varieties. To elaborate Várvaro's picture, we can add that the speech continuum may stop some way short of the edge of the roof (giving way to some unrelated set of varieties, as spoken Romance gives way to Basque while the standard Spanish roof continues to the French frontier), or may extend beyond the roof (as happens in the central and eastern Pyrenees, where the northern Peninsular dialect continuum seamlessly merges with the rest of the Romance-speaking world (4.1.2.4), while the Spanish standard roof stops where it abuts sharply upon the French roof).

The reality is, of course, more complex than this, since spoken varieties not only have geographical coordinates, which locate them in physical space, but social coordinates which locate them in social space. This social space can usefully be envisaged as the volume between the roof and the ground, with varieties spoken by the most powerful located nearest the roof (being most similar to the standard and

varying least among themselves), and the speech of the least powerful at ground level (where we also find maximal geographical variation).

It is worth emphasizing what has just been said: that the greatest geographical variation is to be observed at 'ground level', and this book has been in part designed to describe this ground-level variation. However, it should not be forgotten that there are certain features of non-standard Spanish, usually concentrated in the speech of the least powerful and absent from that of the educated, which are extremely widespread and sometimes universal in the Spanish-speaking world. Because they are so widespread, we can infer (even where we do not have direct evidence) that they are ancient, and in many cases we know that they have been excluded only in recent centuries from educated speech and from the standard. We have already noted (6.3.5) that certain characteristics of Judeo-Spanish belong to this category of recently non-standard features, so that a good number of the features considered here have also been mentioned there.

- 1 The vowels /i/ and /e/, like /u/ and /o/, are neutralized in atonic syllables, and their high or mid realization depends upon adjacent phonemes, such as the height of the tonic, the presence of palatals, etc. (4.2.4, 6.3.5(1)).
- 2 Atonic vowels other than /a/ (and not just /i/ and /u/, as in the standard) are reduced to glides when followed by another vowel, tonic or atonic, forming a diphthong with it: *cambiar* [kambjár], *patear* [patjár], *cuota* [kwóta], *cohete* [kwéte].
- 3 When a high tonic vowel is immediately adjacent to a lower vowel (cases of hiatus), the accent is transferred to the lower vowel (creating a diphthong): *baúl* [bául], *maíz* [máɪθ]~[máɪs].
- 4 The diphthong /ei/ may merge with /ai/, with variable results: *seis* [sáɪs]~[séɪs], *maíz* [máɪθ]~[méɪθ]~[máɪs]~[méɪs].
- 5 Clusters of consonants of which the first is a labial or a velar are not permitted, so that borrowed words (e.g., from Latin) which penetrate these varieties do so without the labial or velar, or these consonants are vocalized to [u] or [ɪ]: *apto* [áto]~[áuto], *objeto* [oxéto], *efecto* [eféto]~[eféuto], *ignorar* [inorár]. We have seen (7.2.2(1)) that reduction of these groups was permissible in the standard until the eighteenth century; since then it has become restricted to non-standard varieties.
- 6 Loss of /d/ occurs in limited circumstances (viz., where it is word-final, or in the masculine participial suffix /-ádo/), but with high frequency, in most varieties of spoken Spanish, including

those used by educated speakers, except under the most formal speech-conditions (Navarro Tomás 1961: 101).¹¹ However, loss of /d/ in other morphological circumstances (e.g., *llegada*, *venido*, *comida*, *madera*, *maduro*) is highly frequent and geographically universal among speakers at the 'ground level' of the social matrix.¹² In the same social milieu, loss of /g/ is also highly frequent, although limited to cases where the consonant is followed by a high back vowel: *agujero* /auxéro/, *aguja* /aúxa/, etc.

- 7 Loss of intervocalic /t/, in a few lexical items, is extremely widespread at these sociolinguistic levels (e.g., *parece* /paéθe/~/paíθe/, *quieres* /kiés/, *para* /pa/).
- 8 Reinforcing of morpheme-initial /ue/ to [ɣwe], less frequently [βwe], is extremely widespread at these levels (4.2.5), e.g., *ahuecar* [aɣwekár]~[aβwekár], *huerta* [ɣwérta]~[βwérta]. Just as frequent is the related interchange of [ɣwe] with historic [βwe], so that *abuelo*, *vuelta*, etc., are articulated [aɣwélo], [ɣwélta], etc.
- 9 Corresponding to standard [fwé], [fwí], in words like *fuera*, *fuenta*, *fue*, *fuiste*, there is a widespread rural realization with a voiceless bilabial or labiovelar initial, [φwé]~[φwí], [mwé]~[mwí], often regarded as allophones of /h/, although not restricted to areas where /h/ (ultimately from Latin F) is retained in words like *humo*, *hambre*, *hilo*.¹³
- 10 Interchange between word-initial *es-* and *des-* is in part a phonological matter (see point (6) of this section), but also a residue of the confusion in spoken Latin of the prefixes EX- and DIS-, which increasingly came to have the same meaning. Medieval and Golden-Age Spanish show frequent hesitation in writing between such forms as *estender* and *destender* and in popular Spanish this uncertainty has been extended to all words of this pattern, with preference for *es-*: *estrozar* (standard *destrózar*), *eslumbrar* (standard *deslumbrar*), etc.
- 11 Use of the form *mos* as the first-person plural object pronoun, e.g., *mos vieron*, *mos lo dieron* (standard *nos vieron*, *nos lo dieron*), is widespread at rural level, no doubt through interference from the corresponding verbal ending *-mos*. The reverse interference, leading to the verbal ending *-nos* (e.g., *estábanos*, cf. standard *estábamos*) is less common, but frequently reported in rural American Spanish (Espinosa 1946: 221).
- 12 Retention of *vos* (standard *os*) as the second-person plural object pronoun is widespread in rural Peninsular Spanish and in

Judeo-Spanish (but is absent from American Spanish, where the corresponding forms are *los/las/les*, consistent with the replacement of subject *vos(otros)* by *ustedes* (5.1.1.2).

- 13 Use of *le* as both a plural and singular indirect object pronoun for third-person reference (standard *les* vs *le*), is widely found at this level, no doubt owing to interference from its allomorph *se*, which is unmarked for number (e.g., *se lo dio* 'he gave it to him/her/them').
- 14 The non-standard ordering of clitic pronouns exemplified by *me se cayó* (standard *se me cayó*) is common at 'ground level', while the order Direct Object + Indirect Object (e.g., *lo me dio*, standard *me lo dio*) is a little less so.
- 15 Addition of /-s/ to the second-person singular forms of all preterite paradigms (e.g., *hicistes*, standard *hiciste*), frequently found in medieval Spanish, is all but universal at the level of uneducated speech, and frequently penetrates that of educated speakers. In America, it characterizes both areas of *voseo* (*Vos dijistes/dijites*) and areas of *tuteo* (*Tú dijistes/dijites*).
- 16 The form *caminemos*, etc., used with preterite value (cf. standard *caminamos*), was occasionally found in Old Spanish (Penny 1991a: 180–1), was excluded from the standard in the late Middle Ages but has widely survived in rural speech.
- 17 Use of *haiga*, *vaiga*, *huiga*, and occasionally *veiga*, as present subjunctive forms of *haber*, *ir*, *huir* and *ver* (see 7.2.2(10)) is found all over the Spanish-speaking world (Espinosa 1946: 244–6), at a rural level, and sometimes at other non-standard levels.
- 18 Irregular preterite forms such as *truje*, *vide* (cf. standard *traje*, *vi*) frequently appeared in written Spanish in the medieval and Golden Age periods, but have since then been restricted to non-standard, chiefly rural, use.

All of the features considered above are today in more or less fierce competition with their standard counterparts, which are increasingly penetrating rural speech, typically through lexical diffusion (see 3.5). This competition is yet to be systematically examined, although a few notably successful attempts have been made in this direction (see Borrego Nieto 1981, 1983, Holmquist 1988).

Notes

1 Introduction: language variation

- 1 Since, as we shall see later, there are no natural boundaries between descendants of a single ancestor language, it is no easy matter to define what is meant by the 'same' language in particular parts of the world. What we call 'languages' (e.g., French, Spanish, Italian, German, Dutch) are in fact distinguished one from another according to non-linguistic criteria; thus it is often the existence of a frontier between two nation-states which gives rise to the conclusion that the people living on each side of the frontier must speak different languages, while those who belong to the same nation-state (and understand one another's speech) are thought to speak the same language. However, the speech of people separated by a national frontier may be more similar than the speech of people belonging to different parts of the same nation-state. And this applies not just to territories, such as Africa, where frontiers are relatively recent, but to territories such as Europe, where the notion of the nation-state is much more ancient, as we shall see in the case of the Spanish–French frontier (4.1.2.4).
- 2 Forms of speech which differ in accordance with the locality of the speakers are traditionally labelled *dialects*, although this term is also used for linguistic systems which are differentiated according to other criteria. Thus it is possible to use the term *dialect* to indicate the set of features (system) used by a particular social group (i.e., a *social dialect*), or for a particular purpose (e.g., the *standard dialect*). For this reason, we shall use only the terms *geographical* or *diatopical* when referring to variation correlated with geographical space.
- 3 Although it is usual to say that the Atlantic fisherman speaks 'Galician' while his Mediterranean counterpart speaks 'Catalan', these terms do not, as we shall see (4.1.2), have a linguistic basis. They are terms whose justification lies in political history, and are applied (according to criteria which are for the most part politically contrastive and non-linguistic) to segments of a continuously varying and intermeshed series of dialects. Furthermore, it will be evident that in this discussion of geographical variation it has been necessary to leave out of consideration (rather artificially) all social factors. The example we are considering concerns the informal registers of uneducated rural dwellers, although the speech of other social groups can be examined in the same geographical way, usually with smaller degrees of difference between localities. Of course, our Galician and Catalan fishermen, depending on their degree of schooling, may resort to an alternative code in order to understand one another, such as the standard dialect ('Castilian') or some approximation to it.
- 4 It will be seen that the relationship between a particular variant and a particular social characteristic is a statistical one. For example, a particular variant can be shown to be