Early

Ibero-Romance

# Early
# Ibero-Romance

Twenty-one studies on language and texts from the
Iberian Peninsula between the Roman Empire
and the Thirteenth Century

by

ROGER WRIGHT
University of Liverpool

Juan de la Cuesta
Newark, Delaware

# Contents

# Introduction

T HIS SELECTION OF TWENTY-ONE studies published over the last fifteen years focusses on one topic in particular; the nature of Early Romance in the period between the end of the Roman Empire and the Twelfth-Century Renaissance, concentrating for the later centuries on the Iberian Peninsula. "Early Ibero-Romance" is the name here given to the language of the Romance-speaking areas of the former Empire in the first part of this period, before it was given geographically-based names such as "castellano." There is no clear chronological boundary between the two.

My interest in the topic began from the viewpoint of Historical Linguistics, and dissatisfaction with the usual explanations of cases of lack of expected phonetic development. This led to a reassessment of a more broadly cultural kind, presented in Late Latin and Early Romance (in Spain and Carolingian France), written between 1977 and 1981, published in 1982. The central thesis was that they did not then, and thus we should not now either, distinguish clearly between Latin and any Romance language until towards the end of the period considered. Gradually, after 1982, it has become clear that this hypothesis involves a reassessment of the data of several separate specialist disciplines. Accordingly, in the ensuing years the consequences of the hypothesis have been considered from a variety of perspectives, as the ideas have been presented and discussed with colleagues in many universities and countries, in many conferences, invited lectures and articles. The results are here collected together, presented in chronological order of subject matter rather than of composition. In each case, there is a brief postscript explaining the paper's original context and relevant later developments.

Thus the hypothesis is here seen from the point of view of sociolinguistics (Chapters 2 and 12 in the present volume), of the metalinguistics of language-naming (4, 17, 18), of modern analyses of writing techniques (13) and the teaching of spelling (14), and of the theory of invention (3); also studied are apparent infelicities of traditional philological practice (5, 9), the textual basis of historical-semantic analysis (6) with detailed examples (7, 8), the traditional assessments of Early Medieval Hispanic documentation, including legal texts (10), historiography (11, 18), glosses (15), and hagiography (17), as well as a text from a Catalan context in which Latin and Romance could genuinely be distinguished (16), and the copiously attested, but only in later texts, oral literature of at least the latter part of this age (19, 20, 21). The earliest-composed paper here (16) was prepared in 1978, and the latest (4) in early 1992.

Most collections such as this are retrospective, but this one is an interim report only. Since 1992 several further related studies have gone to press; a *Status Quæstionis* of the present state of diachronic historical studies of Spanish; a general survey and explanation of the lack of Latin-Romance translation before the twelfth-century Renaissance; a consideration of the linguistic situation in the Romance-speaking communities of Moslem Spain; a general study of the relationship between speech and writing within Latin and Castilian, plus another of the relationship between textual changes and linguistic changes; a study of the spelling of the first official undisguisedly Romance text in the Peninsula (the Treaty of Cabreros, 1206), and another comparing its orthography to that of the *Poema de Mio Cid*; a review-article of Michel Banniard's *Viva Voce* (1992); and a syntactic study of the history of the non-agentive uses of *se*. Others are planned.

The most important moral to be drawn is mentioned explicitly in some of the later-written articles; we should stop patronizing the past. The Romance-speakers of the centuries that followed the end of the Roman Empire lived in a versatile and functioning speech-community, with considerable participation in educated culture (when texts were read aloud, as they normally were), and were not obviously less intelligent or linguistically skilled than the Romans before them or modern Europeans. It is possible, advisable, and more

generous to assess the surviving evidence in such a light. It is also more plausible.

*Late Latin and Early Romance* was prepared at a time when the Arts Faculty of the University of Liverpool was a lively and encouragingly interdisciplinary place, where it was possible to discuss aspects of the topic with many specialists. Since then, the Departments of Linguistics, Latin, Italian, Medieval History and English Language have all been abolished. Many colleagues have given help and inspiration in the last fifteen years, but three deserve special thanks: Max Wheeler, now at Sussex, without whose help this enterprise would never have started; Francis Cairns, now at Leeds, without whose encouragement *Late Latin and Early Romance* would never have appeared; and the late Derek Lomax, whose help with the early medieval history of the Iberian Peninsula has proved irreplaceable (not only to me). Derek Lomax died tragically in early 1992. This collection is dedicated to his memory.

R. W.

Part I:

Early Romance

# 1

# Complex Monolingualism
## in Early Romance

## 1. Metalinguistics

MANY DIFFERENT NAMES have been applied to various hypothesized forms of language used in the Romance-speaking world after the end of the Western Roman Empire and before the invention of written Medieval Romance languages. Distinctions have been made between Classical Latin, Church Latin, Christian Latin, Late Latin, Vulgar Latin, Medieval Latin, Notarial Latin, Romance, Early Romance, Proto-Romance, Gallo-Romance, Italo-Romance, pre-literary Catalan, pre-literary Galician, pre-literary Italian, etc., etc.. Many of these distinctions have a point (e.g. in Gimeno Menéndez 1988), but their proliferation has come to cloud the main issue. It would be simpler to do what the speakers of the time seem to have done, and to regard the whole gamut of spoken and written usage as being a single language, even though a complex one; thus we should avoid making apparently strict diatopic distinctions between different spoken languages in different areas before the general spread of the fashion for using reformed writing systems (from the eleventh to the thirteenth century); and we should also avoid making apparently strict diastratic or diglossic distinctions before the general spread of the eventually clear conceptual distinction between Latin and Romance, a metalinguistic development which roughly coincides with the spread of the diatopic ones (Janson 1991, Wright 1992b). We can avoid such artificial misunderstandings by calling their language by a single name. It is not immediately obvious what name to use; at the time they still called it *lingua latina*, naturally, in the

way that Greek is still called Greek and English is still called English, regardless of diachronic developments, but if we too call the language of that speech-community Latin we run the risk of being thought to imply either or both of two dangerous misconceptions: firstly, that the language of the eighth century was essentially the same as that of the first, and secondly, that they had even at that early time the conceptually separate educated form of speech which we now call Medieval Latin. To avoid confusing this speech-community with either the earlier Latin-speaking one of the distant past or the Medieval Latin ones of the unforeseeable future, I propose that we agree to regard the whole continuum of geographically, socially and stylistically varied speech styles, in the wide area and long intervening period concerned, as being Early Romance.

## 2. The Myth of Diglossia

There are two main sources of data for the language of the period concerned; the written texts of the time, and the results of reconstruction. Since reconstructed "Proto"-Romance is in some respects not very like the language of the written texts that survive from the time when it is supposed to have been in general use, it has been tempting to regard these two sources as attesting two different languages in a diglossic relationship, either explicitly (e.g. Lüdtke, 1968:ch.5) or implicitly (e.g. Baüml 1980:237). Diglossia, however, as sociolinguists will tell you (e.g. Silva 1988:178; Fasold 1984:36), only works if a conscious conceptual distinction is made between the two languages concerned, and the high version is specifically reinforced by the education system. No scholar can find any explicit reference to such conscious awareness and teaching of a separate non-vernacular high language in the texts of the writers of the time from Romance-speaking areas, despite several occasions where such a distinction would seem to be crying out for a mention if it did indeed exist. There is, for example, no such evidence in the copious writings of the encyclopaedically well-informed Isidore of Seville; when Isidore wrote in the seventh century of the requirements of the ideal reader aloud of written texts (the *lector*; in *De Ecclesiasticis Officiis* II.11, quoted in Wright 1982:87-88), he told him not to read too fast or too slow, too high or too low, and to be sure to emphasize the appropriate words and pause at appropriate

places, etc., such as is good advice for all readers anywhere, but —despite Fontaine's (1972) and Banniard's (1985) explicit declarations that he did—he seems never to have given any specifically phonetic advice at all; he never said anything, neither explicitly nor in even the vaguest way, to encourage the reader to avoid voicing intervocalic plosives, for example, or to avoid reducing unstressed front vowels to semivowels, or to maintain a phonetic distinction between originally long /õ/ and originally short /ŭ/, etc. On the other hand, he assumed that if the *lector* read aloud with intelligence and care, the audience would understand. Gregory the Great, in Italy, expected his sermons to be largely understood when read aloud to a congregation (see Herman 1988, Banniard 1986 & 1989; 153). So did Caesarius of Arles in France, whose sermons... were intended to attract large congregations drawn from all classes of society (Wood 1990:71). So did every preacher, for the epoch is full of instructions to preachers to be sure to preach, even to the most unpromising customers. There seems to be no hint in any of the copious early medieval histories, Church councils, hagiographical texts, etc., that in these instructions they were asking for the impossible, as the common modern postulation of diglossia implies that they would have been; we should acknowledge that if they themselves expected their texts to be largely intelligible when read aloud, then it seems only rational that we should also expect that to have been the case.

The first explicit news we have that sermons were not generally intelligible is in the famous edict of the Council of Tours of 813 A.D. By that time the diglossically high Medieval Latin pronunciation for such texts certainly existed, probably as a result of the recent Carolingian reforms of Church Latin. Guerreau-Jalabert's recent study (1981) implies that for some time these reforms only applied in Church contexts; that is, not only before 800, but also for a couple of centuries after that time, except in church circles, reading aloud was generally intelligible, as McKitterick (1989) deduces on other grounds; given the normal vernacular pronunciation of all lexical items, the archaic syntax and vocabulary found from time to time in the authorized sermons of the Fathers of the Church does not really seem to have caused general confusion, despite the picture painted by Richter (1983) in which the Northern French could not

cope with it but the Southerners could. Thus I see the diglossic state as only beginning with the ninth-century reforms, whereas Lüdtke sees it as ending then; despite the fact that he claims to agree with my diagnosis, we do in fact disagree diametrically (Lüdtke 1986, Wright 1991a). My view is much closer to that of Van Uytfanghe (1989; it has changed since Van Uytfanghe 1976); or to that of Ganz (1987:41), who, working independently, refers to it as a reform which separated Latin from the vernacular.

Several studies in an interesting recent collection of historical essays entitled *The Uses of Literacy in Early Mediæval Europe* (McKitterick 1990) suggest an underlying cultural continuity in the Romance-speaking areas from the Roman Empire through to the Carolingians, in which the organization of society still depended to a large extent on the assumption that written texts—letters, laws, instructions, documents of practical kinds—would be intelligible to recipients and interested parties, at least if read aloud to them (see in particular the studies by Wood, Noble, Collins and Nelson; and Banniard 1989). Not many people could write, but several people could read, and virtually everybody, then as now, however illiterate, could understand most written texts when read aloud. That is, the inability to read does not now and did not then necessarily cut people off from literate culture. Each court, each noble house, each monastery, each church, each office, perhaps each village, had at least one person available whose official duties included the reading aloud of written texts to his neighbors, and often also, if necessary, writing on their behalf. Communication wasn't necessarily perfect then any more than it is now, but at least it happened regularly enough for the system to work. The view expressed in such influential studies as Clanchy (1979) and Stock (1983), to the effect that the Early Middle Ages depended more on memory and custom and less on literacy than later times, is currently being modified into differences of degree rather than of kind. The Carolingian and the Twelfth-Century Renaissances increased the proportion of literate people in Europe, but the practical possibilities of running Romance-speaking societies via written documents had existed all along. Furthermore, general sociolinguistic studies carried out in the last few years (such as Graff 1979, Cressy 1980, Stubbs 1980, Levine

1986) suggest that the disastrous effects of restricted literacy have been much exaggerated (e.g. by Goody 1968).

## 3. Reconstruction

Our problem, the apparent incompatibility of reconstructed Proto-Romance and the language of the texts of the time, cannot thus be solved by merely postulating early diglossia. We need not exaggerate the differences between the two types of data; most of the vocabulary and syntax, and much of the phoneme (grapheme) and morpheme inventory, is the same in both the texts and the reconstructed Proto-Romance, and in any event modern studies of all sorts of monolingual communities conclude there are always considerable statistical differences between the distribution of features in written and spoken usage. Sociolinguists see wide variation as not only normal but necessary. Historical linguists see a scenario based on hectic evolution in one part of a speech community allied to total immobilism in another as highly improbable, and, in this case at least, completely unnecessary. If we can accept that the literate spoke the same language as their neighbors, as happens in almost all literate societies anyway, the monolingual perspective "frees us from having to posit the rather rigidly-defined strata in society which the more conventional view requires us to imagine in order to lend credibility to its notion of linguistic apartheid, even in default of any other evidence that such layers ever existed " (Harvey 1990:181). This is a relief, for now we can evaluate the data provided by reconstruction within a more plausible social framework.

Proto-Romance, as reconstructed by extrapolating backwards from later attested Romance languages, is said to have contained—in comparison with the written texts—no neuter nouns, no ablative cases, no datives and genitives outside pronouns, no synthetic passives or futures, no phonemic length distinctions in vowels, no originally final consonants other than alveolars, and no velar consonants before front vowels other than those that were originally labiovelar; on the other hand, it did include extended uses of prepositions (particularly *ad* and *de*) to replace inflectional nominal suffixes, analytic passives with auxiliary *esse* and tense-indeterminate participles, extended use of grammatically-reflexive *se* with

passive meaning, analytic futures (and conditionals) formed with the infinitive and *habeo*, new analytic perfects (including future perfects and pluperfects) formed with activized participles and *habeo*, a multi-purpose complementizer in [ke], extensive use of *ille* and *ipse* with the functions of the definite article, many diminutives in -*iculum* and other newly affixed forms, the use of preposed *magis* or *plus* instead of comparative -*ior*, new palatal affricates and semivowels, and much new vocabulary from, in particular, Germanic sources. (This is just a selection.)

The first thing to say about these plausibly reconstructed features is that the advent of a new feature does not necessarily nor even usually imply nor coincide with the loss of the old feature with similar function. It is perfectly possible that for many centuries there were in wide use both types of future tense, both the older synthetic and the newer analytic, just as there are nowadays in most Romance languages, at least in Europe, where the "going to" forms coexist cheerfully with the synthetic forms evolved from the infinitive and *habeo*; there is no need to postulate that the old futures disappeared at an early date, and even if gradually speakers became statistically more likely to use the analytic than the synthetic, both usages could easily have remained intelligible. The arrival of the analytic perfect, formed with auxiliary and past participle, has still not replaced the old synthetic preterite form except in Northern France, and even here, written texts read aloud with the old preterites are usually intelligible. Then again, the fact that the written forms of later Romance languages reserve dative morphology to third person pronouns does not in itself suggest that the same semantic distinction between direct and indirect object would have been unintelligible if it was encountered incorporated into the nominal morphology of existing texts; on the contrary, it suggests that it would indeed have been understood. The distinction between active and passive competence held good then as it does now. Green's study (1991) of the replacement of the synthetic passive by the grammatically-reflexive *se* does not conclude that they rigidly said *se* and wrote *se* with synthetic passives, but that both were available in different proportions within both registers. As Stengaard (1991:5.2.1) also points out, with regard to the famous Riojan Glosses, even though it is traditional to analyze the glossing

of synthetic passive forms by reflexives as being a symptom that they no longer understood the passive, a moment's thought shows us that this analysis is exactly wrong; the presence of the glosses in *se* shows us that they did indeed understand the passive, or else they could not have come up with an effectively synonymous gloss, even if it is also the case that for active use they preferred *se*. Often, having such alternatives increased pragmatic subtlety; the use of *quod* plus indicative, rather than accusative and infinitive construc- tions, after verbs of saying, increased pragmatic subtlery by enabling a functional contrast to be established between the two, as Herman (1989) has shown, with the result that in this case the advent of the new seems to have required the concomitant continuation of the old to achieve its purpose. Thus the list of absentees from the recon- structed list of morphosyntactic features in Proto-Romance is not only an unproven but also completely unnecessary part of the reconstructionist hypothesis, for all but the latest part of the period concerned.

Many of the new morphosyntactic usages envisaged by the reconstruction technique undoubtedly did exist at the time we are concerned with, and very probably (as Herman 1990, Varvaro 1991, Lyons 1986 and others have suggested) over a wider area than merely those regions that subsequently chose to incorporate them into their standard. And yet most of these apparent innovations involve the reallocation of existing forms to different functions—in effect, semantic changes— Such reanalyses as happened to *ad*, *de*, *se*, *esse*, *habeo*, *ille*, *ipse*, *plus*, *quod*, etc., were in each case a statistical extension of the distributional frequency of forms that were intelligible in the speech of the Roman empire anyway. For example, using grammatically-reflexive *se* was not the commonest method of expressing passive meaning in the Roman Empire, but it was not unknown, in the same way as the grammatical reflexive occasionally finds itself being used for passive meaning in modern English. Even during the Roman Empire they could have without difficulty interpreted *de* meaning "of," *habeo* as future auxiliary, and also under different circumstances as a perfect auxiliary, *ille* with a demonstrative force sufficiently understated to be analyzable as a definite article, *plus* with an adjective, etc.; and most such usages are not entirely absent from the documentary evidence

anyway. Relative statistics of frequency attest changes in progress, as Erica García has taught us (e.g. 1985), but they are irrelevant to the present discussion; if a form is recognizable and intelligible with a particular function, it does not matter whether it appears once a month or twenty times a day, it is still intelligible, and the usages of written texts are not incomprehensible merely because speech tends to use the features concerned less commonly than writing does. Once again, what is being envisaged here is essentially what we would expect to find anyway, on general theoretical grounds; a stylistic difference between speech and writing, based on distributional differences between features which on the whole exist in both modes, even in modern societies with high degrees of literacy and a distinguishable dividing line between the two. I am not the only philologist to take reconstruction with pinches of salt; see for example, Dworkin's recent review (1989) of the handbook by Agard (1984).

The historians' scenario implies that despite the reservations expressed above over syntax and morphology, the pronunciation details as reconstructed for Proto-Romance are largely accurate, even for the reading-aloud of officially-written documents and texts. The written text was nearly always communicated orally by a reader rather than through silent study; this is certainly true of Saints' Lives, sermons and legal documents, probably true of all letters and orders sent from kings to vassals and generals to armies, etc., and even quite likely to be true of official histories such as the *Chronicle of Alfonso III* (Gil et al., 1985). What we can now come to reconstruct is a single wide Early Romance speech community of great diversity, versatility and vitality, where the availability of a wide variety of both older and newer alternative forms, constructions, lexical items, and even perhaps in some cases pronunciations—particularly during the transitional stages of what can be later diagnosed as diachronic sound changes—gave rise in speech to multiple possibilities of pragmatic subtlety and stylistic nuance, such as is anyway characteristic of the oral usage of complex monolingual societies, and also—as Fleischman (1990) has compellingly shown—of at least the unamended manuscript versions of early Mediaeval Romance literature. Early Romance was thus in no way inferior to, or merely a corrupt version of, Imperial Latin; Early

Romance is in no way inferior to, or merely an incipient version of, subsequent Romance languages; Early Romance was a fascinating and lively speech community in its own right, which deserves to be studied in itself, regardless of what came before and what was unforeseeably going to come later.

## 4. Texts.

But if we are to study the language of the period exclusively on its own terms, we have an immediate problem with the written evidence. Not many original texts survive from before the Carolingian reforms, but they certainly existed at the time, in tenth-century Spain, for example, Collins suspects that Leonese legal procedures did produce a multiplicity of written documents in even greater profusion than those of Catalonia (1985.502), at a time when Catalonia had undergone these reforms but León had not. After the end of the Western Roman Empire they were still writing on wax tablets and papyri, which are all too biodegradable. In the 670s the Merovingian Kings switched from papyrus to parchment (Kelly 1990:41), but the Papacy was still using papyrus in the ninth century (Noble 1990:88) and Byzantium in the tenth (Mullett 1990:158). Such early originals as survive are startling, all in capitals, usually without gaps between the words, with many abbreviations, and often requiring an act of faith to be intelligible at all. The famous graffiti from Pompeii, the Vindolanda tablets (Bowman and Thomas 1983), the Visigothic slates from Salamanca province, and the curse-stones from Bath, for example, are not remotely as clear in real life as in their printed editions. If long early texts survive, that is usually because they were subsequently copied onto the more lasting material of parchment. That represents a technological advance for which we must be grateful, but unfortunately the copyists had specific instructions not to copy their originals faithfully but instead to "correct" them according to the anachronistic prescriptions of the fourth and fifth-century grammarians. Surviving texts without such distortions are very few, and even the letters and drafts published by Adams (e.g. 1990) are not phonetic transcriptions of actual speech. In short, many texts as edited today have been significantly altered, and their originals could usually have been less remote from reconstructible Proto- Romance than the modern

editions imply. Unfortunately, most modern editors of Latin texts still refuse to allow us to see the earliest manuscript texts without "emendations" (a view most unfortunately supported by Goffart 1987:56), and unamended texts are, of course, what all linguists and philologists need above anything else. Acknowledging these problems, Herman has spent a lifetime minutely analyzing the evidence of tombstones, whose texts cannot be emended.

Written texts were prepared by Romance-speakers in accordance with the detailed requirements of the Grammars. Even at the time of their initial preparation these Artes Grammaticae were at best selective; as grammarians all over the world tend to do, they saw their task in moral terms, choosing some of the many and varied features of the speech of the day as being "correct," and excluding others as "incorrect." As the centuries rolled by, it came to be the case that many of the features positively advocated in the Grammars were rare in active speech. Thus learning to write was getting harder all the time. Important works intended to be kept for posterity therefore went through several drafts; unimportant texts that have survived by accident, such as the Leonese cheeses, have a distinctly less polished air. This fact has the unfortunate implication that in order to know how to interpret the evidence of the written texts composed in our period we need to know how their scribes were in detail trained to write. Not just the script and the physical forming of the letters (as studied by Petrucci 1972, 1986), but of knowing which letters to attempt; that is, orthography. Unfortunately, despite the valuable work of Riché (e.g. 1989), this has hardly been studied at all. For this is not one of the ages when scribes were trying to come up with fairly faithful transcriptions of speech; they were instead trying to disguise their vernacular with a sufficient veneer for it to seem correct according to traditional criteria.

So the conclusion is that this is where research should now be concentrated; we should avoid divisive metalinguistic terminology, we should accept most of the positive reconstructionist hypotheses as being valid for all Romance-speakers, we should look in detail at the surviving unamended texts we have, and then we should attempt to visualize how the scribes who spoke in the way in which we reconstruct their speech learned in detail to reach the practical stage of writing these texts in the way that they did; as in modern

English, that process was perhaps complex, but also essentially monolingual.

## 1994 POSTSCRIPT:

This paper was originally published in *Linguistic Perspectives on the Romance Languages: Selected Papers from the XXI Linguistic Symposium on Romance Languages*, ed. William J. Ashby et al., John Benjamins Publishing Co., Amsterdam/Philadelphia, 1993, pp. 377-88. It is the text of a paper given to the Symposium at the University of California, Santa Barbara, on the afternoon of Friday 22nd February 1991. This conference was essentially for specialists in Government and Binding, but even so the audience contained many who need to give lectures on the history of the language, and the subsequent discussion was useful. The central thesis of *Late Latin and Early Romance* is in need of rephrasing, in order to allow for a later survival within Early Romance of older syntactic constructions, as a consequence of the idea introduced there, and which I have developed since, that we confuse the issue and cloud the data if we insist on trying to distinguish Latin from Romance in the Early Medieval centuries.

## 2

# Modern Sociolinguistics and Early Romance

THIS PAPER CONSIDERS the sociolinguistic state of the Romance-speaking world between the fifth and eighth centuries. It has often been said that the present-day state of the French-speaking, Spanish-speaking and English-speaking worlds is comparable to that of the Romance-speaking community after the fall of the Western Roman Empire in the fifth century. There is a wide geographical area over which is spoken one sociolinguistically very complex language. Many new features of speech had already arisen during the Empire, and by the fifth century there was a wide range of alternatives of many kinds, which the speakers chose between as they spoke. It is normal for new features to arise without their predecessors necessarily disappearing straightaway. Usually, in these complex communities, what seems to a later observer to have been a straightforward change could last for a long time as a case of variation (cp. Menéndez Pidal, 1926); and in these Early Medieval centuries many ancient forms survived alongside the new ones, without the speakers necessarily knowing, in any particular case, which of the available forms was the older and which was the newer. For example, in the fifth century, they would have had to choose, when pronouncing words whose written form ended in -ITIA (such as MALITIA), whether to give that ending three fully syllabic vowels, or two vowels and a semivowel, or two vowels and a palatalized consonant; and in a community in which all three pronunciations would have been equally intelligible, the unconscious choice between the three available forms would not have been made according to which was the oldest or which was the most evolved form. They would not have known which was which, and if they had known, it would not have been an important consideration; the choice between the three variants would have been made within each individual context for pragmatic, stylistic or sociolinguistic reasons. It is quite likely that they pronounced this suffix in a different way when reading a text aloud (or singing it) from the one they used in formal speech, and both of those might well have been distinct from the way they pronounced it colloquially. For one discovery of modern sociolinguistic theory is that our own pronunciation varies in a more or less systematic way between these separate contexts of use (e.g. Labov, 1972), and that we can understand many linguistic features that we do not ourselves say. Nor is it necessarily the case that the most formal usage is going to be the most archaic. With reference to this particular word, St. Isidore of Seville tells us (Etymologiae I,27,28) in the seventh century that MALITIA is pronounced with a palatalized consonant, and we can probably conclude from this that he spoke it in that way in spontaneous speech, at least, and perhaps at the reading level also. Stylistic and sociolinguistic considerations often mean that it is useful to preserve variation in a speech community; more than that, variation is not only natural, but essential. It would be interesting to discover how far the techniques and discoveries of modern sociolinguistics can help us understand what was happening in these Early Romance centuries.

It will already be clear that I put little faith in the traditional geographical explanations of the multiplicity of available alternatives in the speech of these centuries. It is not easy to believe in the existence of clear boundaries between separate Romance languages at that time; it is not convincing to suggest that we can solve all these problems by suggesting that one alternative was unique to Italian and another to Catalan (etc.) already by the sixth century. The sociolinguist Bailey (1973) has shown that such rigid "genealogical trees" misrepresent what happens in real life. Certainly, statistical differences could have been found between different geographical areas, but at that time these differences can only have been statistical rather than diagnostic, that is, of no necessary importance (even if they happened to be going to be diagnostic later). Many specialists in the field will not agree with this assessment, citing, for example, the supposedly very early separation of Sardinian (e.g. De Dardel,

1982); but we are all aware (with Herman 1965) of the great difficulty that is encountered by those who try to locate the place of origin of a text from these centuries on the base of linguistic evidence alone. For example, Manuel Díaz pointed out (1965) the Romance features which can be detected in the texts of the Visigothic liturgists of the seventh century, but these seem to be general features of Romance rather than specifically Spanish: the femininization of originally neuter plural forms, the regularization of formerly irregular verbs, the activization of original deponents, the avoidance of dative and ablative endings, the addition of a prothetic [e-] before a word beginning in [s-] plus a consonant, the voicing of some intervocalic consonants, the use of compound prepositions, etc. There they are, features of seventh-century Romance to be found in the Spanish liturgy, but they are not particularly "Spanish." Geographical frontiers between separate Romance languages seem to correspond better to the political divisions of the ninth century and later. That is why it seems hard to believe that the choice between available variants in the spoken Romance of the previous centuries can have been made on simple geographical grounds in more than a few cases. The Irish Latinist McManus (1984) has suggested this also for the spoken Latin of the British Isles, which was not in fact as distinctive as has been claimed.

Nor does there seem much point in trying to distinguish clearly between Latin and Romance before the ninth century (an opinion explained at length in *Late Latin and Early Romance*). We should accept that the sociolinguistic situation of these years is far more complex than would be expected if we were merely dealing with a geometrical distinction between two languages, Latin and Romance (in my view, Medieval Latin, distinct from Romance, was a conceptual innovation of the Carolingian reformers of c.800). Before that there was one language, Early Romance, which was getting increasingly complicated sociolinguistically. We can compare this with Modern French, where there are differences between speech and writing, but it is unreasonable to analyze the two as being separate languages. What happens (in both cases) is that spoken language evolves but the rules taught to writers do not. Since we cannot now support the traditional view that there were two clearly separate languages in the seventh century, Latin and Romance, it is worth

reassessing from a different approach the texts composed then by Romance-speakers; for it is possible that they can give us clues to the sociolinguistic situation of the time.

This task has to be undertaken with great caution, of course. As we know from modern studies (such as those collected in Coulmas and Ehlich, 1983), and indeed from common sense, texts never reflect spontaneous speech, not even the speech of their author. As regards morphology, there is a great deal to be said for the view expressed by Politzer (1961), to the effect that when we come across suffixes that were spelled perfectly, in a text of these times, we can explain this perfection best if the ending in question was entirely absent from normal speech; for example, the ending -IBUS was more consistently spelled correctly in the seventh century than in the fifth, and the reason for that is that by the seventh century the affix had disappeared from normal spontaneous speech, so the scribes could learn the ending written -IBUS and reproduce it exactly without any interference from a spoken form in [for example] [-eßos]. This has an equivalent in Modern French; many writers always spell totally correctly the verbal affix for third person plural verbs -ENT, but none ever pronounce it at all, neither as [-ent] or as anything else. The endings of the French preterite tense, which is only used in writing, are normally spelled correctly, and can be pronounced without any hesitation when reading aloud, even though not used in colloquial speech. In the same way, we need to be aware that texts of those centuries, if written by a speaker of Early Romance, represent one stylistic level of that Early Romance. It is normal in every language for some features of texts, even when read aloud, to be ones that would never appear in spontaneous speech, and in Early Romance as in the complex speech-communities of modern times this is a normal sociolinguistic phenomenon.

It is also worth pointing out that nowadays the differences there are in the normal spelling of texts from different geographical areas do not usually correspond to phonetic distinctions. For example, the spelling differences that there are between the written English of Britain and the USA occur in words that are pronounced more or less the same on the two sides of the Atlantic. That is, spelling variation is not direct evidence of spoken variation. For example, the

British forms centre, programme, marvellous, catalogue, defence and through are of words that are pronounced with the same phonemes corresponding to the different letters of North American center, program, marvelous, catalog, defense, and thru. If there is any kind of analogy to be drawn between the two speech-communities, we have to conclude that it is not always valid to deduce the existence of geographically-based variation in speech on the evidence of geographically-based variation in texts in the seventh century either.

It is possible that the errors made in inscriptions and public notices might be more reliable, but they do not in fact give us clear evidence of syntax or vocabulary now. Constructions and words are used on posters and public signs that would never be used in normal speech, because they too are on a separate stylistic level. For example, there is in English a word ALIGHT; in speech this is an adjective meaning "on fire," in public notices it is a verb meaning "get down" (usually from a train). This latter meaning for the word is hardly ever heard in real life, but even so we understand the word when we see it, and we would be able to read it aloud if need be, and even though it is now an archaism we would read it with the normal evolved phonetics of the late twentieth century. That kind of phenomenon is found in all modern languages, and could well have occurred also in the sixth and seventh centuries. Words and forms that were falling out of colloquial usage could easily stay alive on the sociolinguistic level of public notices, and could have been understood and read aloud without difficulty by people who would never have used them otherwise. And the same applies to morphology; in the Iberian Peninsula, most nouns were probably used in speech almost always only in the form that was originally only accusative, whatever the syntactic context, but for centuries the speakers would have been able to recognize and read aloud the other case endings, with their normal evolved phonetics; and even (if it were necessary) write these latter forms not used in speech, if they had been professionally trained to write. For example, one word which is found in a number of legal texts contains the word VOLUNTAS. In normal speech, in eighth-century Spain, the population, including the lawyers, would probably have used the form [boluntáde]. When reading aloud, that is, operating at that stylistic level,

perhaps they would not have read it exactly the same as they would have spoken it colloquially; and if they happened to read it as [boluntás], or perhaps [boluntá], that is, without the normal final syllable, that would not have struck anyone as wrong, because in all societies we know that lawyers are liable to speak in an esoteric manner in their work but speak more or less normally when they get home from the office (cp. Crystal and Davy 1969: chapter 8; and Late Latin, pp. 166-71). In short, all the variations, alternatives and confusions that we can reconstruct for the Romance-speaking world are phenomena which modern sociolinguistic theory tells us that we would expect to find there anyway.

But although sociolinguistic theory can help Romance philologists to get the general problems into focus, it seems unlikely that the detailed investigative techniques used by the modern researchers will be appropriate for use when we consider the details of the Early Romance centuries. Lüdtke (1968), for example, has applied the modern concept of "diglossia" to the Early Romance world (as the word was used by Ferguson, 1959); but this cannot be accepted, for Fasold (1984) (and others) have established that the existence of "diglossia" depends on the general recognition within the speech community of the existence of two separately distinguishable languages, not merely of different stylistic levels, and this recognition cannot be identified for the period preceding the Carolingian reforms. Labov (1972) and Trudgill (1974) have developed complicated techniques to illuminate complicated data collected from the modern United States and Great Britain, but it seems improbable that we will ever have enough data, or any clear data at all, concerning linguistic variation in seventh-century Toledo, comparable with the data available to Labov and Trudgill. Nor do we have enough available data to exploit the theory of variable rules; that is, that observable variation is the result of internalized rules of statistical variation, of such a kind that at a given stylistic level 60% of our relevant utterances would be of one type and 40% of another, for example. Romaine (1982) has used this idea, among others, in her excellent historical study of Scottish English of the 15th Century even in the absence of spoken data. But in the seventh century, everything would be guesswork. Lesley Milroy (1980) has used the theory of "social networks" to explain some of the diversity there

is in Modern Northern Irish speech, and she and Jim Milroy have recently (1985) explained the implications that these networks have for our understanding of modern linguistic changes; but this theory, and the "Wave" theory preferred by Bailey, depend on a detailed analysis of who knows who in the community, which is also out of our reach for the Early Romance world. Even so, there are specialists in the "Late Latin" texts of the age who understand sociolinguistics; in particular, the works of Van Uytfanghe, Itkonen and Richter should be mentioned here.

Thus although it would hardly be feasible for modern investigators into Early Romance to make use of the detailed techniques developed by recent sociolinguistic theory, it would certainly be valuable for us to acquire a general sociolinguistic understanding of how and when people nowadays vary in speech and in reading aloud; and to realize that people can understand all sorts of uses that they would rarely actually say themselves, that variation is usually explicable according to other criteria than the merely geographical, that the same person can vary in speech according to the same stylistic parameters, that educated people vary along the same stylistic scales as the uneducated, and that textual variation does not inevitably indicate variation in speech between the authors of the texts. Such realizations should at the least prevent Romance philologists from descending into nonsense. Several of the assumptions of the Romance philological manuals now in use would probably seem absurd to a modern sociolinguist. For example, some of them still say that periods of wide contacts and communications, such as the Roman Empire, favor linguistic change less than periods of comparative isolation between communities, such as the seventh century; but the Milroys, and Trudgill, have established that the opposite is true, that linguistic innovations increase at times of mobility and good communications: It is absurd to continue to tell our students (if we do) that Latin only began to change much after 600.

It is also absurd to keep telling our students that the spelling of the Strasbourg Oaths was designed that way to help native speakers of French Romance. As we all know, reading (and writing) phonetic transcriptions to which we are not accustomed is much harder than reading (or writing) normal spelling. In the modern world, phonetic

transcriptions are only useful for foreigners who can read their own language but do not know how to pronounce the language in question (as in tourist phrasebooks); phonetic transcriptions are no help to those who speak the language anyway. That is why I suggested (in *Late Latin*) that the Oaths were written the way they were for the benefit of a Germanic-speaker who did not know Romance well; and (at the Aix-en-Provence conference) that the famous 11th-century Riojan Romance glosses were written the way they were to help a foreigner who did not know the local speech but wished to read aloud there in an intelligible manner.

To conclude: it may be true that modern sociolinguistic theory cannot help us greatly in the reconstruction of detail, partly because New York and Norwich are not very like seventh-century towns, but it can help prevent us from talking nonsense. A general reorientation of our perspectives along these lines will allow us to realize how very improbable are some of the traditional explanations of aspects of these Early Romance communities. At least we can now prepare to carry out the new kinds of research that this perspective implies; for the moment, it is a case of "reculer pour mieux sauter."

## 1994 POSTSCRIPT.

This paper was originally printed as "La sociolingüística moderna y el romance temprano" in *Actes du XVIII$^e$ Congrès International de Linguistique et de Philologie Romanes*, ed. Dieter Kremer, Vol. V, Tübingen, Niemeyer, 1988, pp. 11-18. The conference was held at Trier. That was the unamended text of the contribution to the Table Ronde entitled "Les Langues romanes—champs exemplaire pour le développement de la sociolinguistique et de la pragmatique," held on the morning of Wednesday 21st May 1986. I had been invited to contribute to the Round Table after submitting the abstract, and had no time to rewrite it; thus its early historical bias was alien to the interests of most of the audience, and, being in Spanish, of little interest to the central Franco-German axis that dominates these conferences. So far as I know, the paper has aroused almost no interest. The point about British and American spelling reflecting different teaching conventions rather than different pronunciations has eventually led to some of the other studies in the present volume, and others I have prepared more recently. Since 1986, the

idea seems to have spread that Lüdtke and I agree about diglossia; but as the next paper in the present volume shows, we in fact disagree diametrically, for Lüdtke sees Latin-Romance diglossia as ending with the Carolingian Reforms, and I see it as beginning then. Romance historical sociolinguistics has occasionally been practiced since; Lloyd (1992), for example, has reconsidered its general potential, and Penny (1992) has applied social network theory to the history of Judeo-Spanish. My views have not changed since 1986 as to the conclusion: just a little sociolinguistic understanding is all that it would take to inhibit some of the elementary misrepresentations that still get printed within our field.

I should have given in full the quotation from Isidore's *Etymologiæ* I.27.28, because it illustrates the point about the non-phonetic nature of spelling conventions in a startlingly direct manner:

28. Y et Z litteris sola Graeca nomina scribuntur. Nam cum iustitia sonum Z littera exprimat, tamen, quia latinum est, per T scribendum est. Sic militia, malitia, nequitia et cetera similia.

The word MALITIA eventually becomes Spanish *maleza*, "weeds." Instead, as an appendix to the original paper I added the text of the first Strasbourg Oath and of *Glosa Emilianense* 89, neither of which seems really necessary to reproduce here.

# 3

# The Conceptual Distinction between Latin and Romance: Invention or Evolution?

Several historians have recently been arguing that early developments in human society, such as agriculture, were often consequences of human inventiveness rather than of any kind of automatic evolutionary process (cp. Van der Leeuw and Torrence 1989). I have argued before (Wright 1983) that historians of language need to pay more attention to the philosophy of history, and this distinction between evolution and invention seems a crucial one to make within our own historical discipline.

Some diachronic linguistic developments certainly occur in an evolutionary manner, without any speakers particularly willing them into existence. But several other developments are the consequences of a decision made by one or more speakers, and these changes can sensibly be regarded as in origin cases of invention. The latter are sometimes assumed to be peripheral to historical linguistics. Lass, for example, accepted without argument that change does not involve (conscious) human purpose (1980:82). Chomsky has also exiled such phenomena from consideration *a priori*: 'each actual "language" will incorporate a periphery of borrowings, historical residues, inventions, and so on, which we can hardly expect to—and indeed would not want to—incorporate within a principled theory of Universal Grammar' (1981:8). Fortunately there is no compulsion for us all to be necessarily interested exclusively in universals, and even if we are, linguistic inventiveness can plausibly be seen as being as much of a human universal, in the sociohistorical develop-

ment of languages, as is the linguistic creativity that Chomsky himself has so often stressed, despite the fact that not every individual invention exists in every language community. This paper suggests that one such conscious decision concerns the separate establishment of systematically different levels of language for different social purposes; specifically, that it still seems most likely that the conceptual distinction between Latin and the contemporary Romance Languages of the Early Middle Ages can only have been the result of an innovation made on purpose in a particular historical context, that of the Carolingian renewal of Christian intellectual life, rather than the inevitable result of a gradual evolution. I disagree, therefore, with R. A. Hall's explicit view that 'A diglossic situation would have arisen anyhow,' Carolingian "reform" or no "reform" (1986:215); there was nothing inevitable or evolutionary about the arrival of the eventually clear diglossia between Latin and the Romance languages in the later Middle Ages. The Carolingian scholars did not merely become conscious that Romance and Latin were different (Michael 1988), as has often been suggested; they invented the difference.

Some historical developments can only have been invented on purpose, the creations of an individual genius rather than the result of unconscious evolutions by the mass of the human community. The wheel, the rowing boat, coinage, the bow and arrow, the internal combustion engine, for example, and within linguistic history, writing. Ong (1982:83-85), Harris (1986), and others argue that writing can only have been an invention. Historical linguists have tended to shy away from this conclusion, using some non-committal and indeterminate phrasing such as the 'development of writing' (as in Jeffers and Lehiste 1979:161), but writing cannot just have turned up 'unasked and unpremeditated. The idea of writing at all was a giant step for mankind, comparable to the invention of the wheel, and so subsequently were the successive elaborations of ideographic script based on the lexicon, syllabic and then phonemic scripts based on the phonology, punctuation, diacritics, the establishment of spaces between written words, shorthand, written tone curves, the International Phonetic Alphabet, the initial teaching alphabet, word processors, voice synthesizers, modems, etc. They were all inventions, which would not be here at all if some enter-

prising character had not thought of the idea. One inventor is enough, though; we can watch, in a historical atlas, how the idea of writing spread geographically from its Sumerian origins of c.3500 B.C. (e.g McEvedy and Woodcock 1967:26, 36, 44, 56). Subsequently, new systems of recording a language do not just emerge unbidden either; for example, the distinctively non-Latinate spellings of the Romance languages were intentionally elaborated for a practical purpose (as were shorthand and the International Phonetic Alphabet, etc.). In Late Latin and Early Romance (Wright 1982) it was suggested that the new writing system which we now call written Old French was consciously first elaborated for a particular purpose in a specific context; that is, to assist those speakers of Germanic who knew how to read Medieval Latin aloud to use that knowledge to read aloud in a manner that might be intelligible to speakers of Old French. Here too, as Elcock showed long ago (1961), we can see on a map how the initial invention, in this case of writing in a manner intended to give rise to an intelligible Romance reading, was imitated successively in geographical areas spreading outwards from its north-eastern French origins; one inventor followed successively by adjacent imitators.

Sinclair's 1987 article entitled 'Language: a gift of nature or a home-made tool?' suggests reasonably that language is both. Many linguistic developments are undoubtedly of an unpremeditated type that cannot seriously be thought of as invention. Most sound changes are unintentional. Indeed, if teleology is consciously invoked by speakers during the course of a sound change, the aim seems generally to be that of preventing the change taking place at all. Some phonetic changes, however, have proved to be explicable by appeal to conscious or semi-conscious phenomena such as phonosymbolism (e.g. Malkiel 1987, to conspiracies to conform to intuited phonetic templates (Pharies 1986), or to desires to escape from undesirable homonymy, and people of unusually explicit metalinguistic awareness may perhaps initiate these on purpose. But these changes are a minority, and Pagliuca and Mowrey (1987) could well be right to see most sound changes as being simply the consequences of unpremeditated relaxation of muscles round the mouth. It is thus understandable that such phenomena as the precise conditions of a conditioned change, or the detailed strength hierar-

chies that determine the chronology of related changes (as in Harris-Northall 1990, Cravens 1988), need to be painstakingly unearthed by specialist linguists long after the event, for they are neither the conscious inventions of a human mind nor accessible to native intuition.

Many grammatical changes are similarly unintentional. The replacement of the Latin case system with Romance prepositions, changes in statistically preferred word orders, the creation of compound prepositions from adverb + preposition sequences, for example, seem to be probably, though not necessarily, best regarded as unintended and evolutionary. And yet Ridruejo (1988) has argued rationally that morphosyntactic innovations can easily be intentional. Many semantic changes are also gradual and evolutionary; those that involve a shift in prototypical reference points, for example (e.g. Wright 1985), or those that occur when a superordinate term comes increasingly to be used with the reference of one of its hyponyms (e.g. Wright, 1990). Some semantic changes can be established on purpose, however. Scientists, philosophers or social reformers often have recourse to the establishment of their own definitions of words that are already in use with a related but slightly less clear or defined meaning, and if they have sufficient authority they can in time succeed in changing the meaning of the word thereby. Einstein did not invent the word *relativity*, but he invented the definition of it that is now its central meaning. New words, that is, lexical change, are generally conceded to be, at least sometimes, inventions. The only time the word 'invention' is used in Jeffers and Lehiste is in this connection (1979:130: 'the vocabulary of a language is continually being enriched by the invention of new words'). Borrowing of foreign words is similarly also in the first place an individual initiative.

Individual initiative, in short, has a higher place in most types of linguistic development than it is sometimes given credit for; as the Milroys say (1985:345), 'it is not languages that innovate: it is speakers who innovate.' For every systematic feature of language, however obvious it may subsequently appear, there must have been a first time. Hurford's (1987) discussion of the psychological history and present basis of numerals is illuminating in the present context. Numerals seem a natural linguistic feature to us now. But they were

not always there in language. Only up to the number three can the human brain perceive number without calculating it. Above that number, humans have invented their systems of counting, in a long sequence of successive small progressions whose complexity and rationale vary from community to community (and thus from language to language). Above three, all numeral systems are in origin invented, as was the wheel. The method of counting in tens, the *hundred*, the *thousand*, the *million*, are all inventions, and simultaneously conceptual and linguistic inventions. Hurford does not come up to date, but in our own time both the concept and the lexical item of the *light year*, the *parsec*, the *googol*, are all human inventions of the same kind. In a millennium's time the concept and the word *googol* (a 1 followed by a hundred zeros) will seem as commonplace to English-speakers as the *hundred* does now. They were individual inventions once, but once invented anyone can learn them, and they come to seem self-evident.

New linguistic standardization of all types, not merely the orthographic, requires a conscious standardizer (as Marcos Marín has demonstrated for Spanish: Marcos Marín 1979, Marcos Marín and Sánchez Lobato 1988). The prescriptiveness of all prescriptive rules is invented by grammarians who think they perceive a moral order in grammatical details; the demanding peculiarities of the Latin rhetorical cursus, metrical poetry, and indeed, as Norberg (e.g. 1958) has shown, Latin rhythmical poetry, are examples of this. The detail of the morphology required of written Latin was prescribed by grammarians (especially Donatus); and so, I suggest, were the details of the peculiar and artificial Late Medieval system used for reading written texts aloud even in Romance-speaking areas, that is, producing one specified sound for each already-written letter (or digraph). All reading involved reading aloud. It seems obvious and natural to read Latin that way now, as obvious and natural as it is to count in tens. But such a method of reading aloud one's native language is totally unnatural. Anglo-Saxon speakers, however, at least since Bede, had learned to read the same texts, in what was to them a foreign language, aloud in that way; they brought this system with them to the Continent, and there, as a result of using such pronunciations in speech as well, they were at times unintelligible to Romance speakers (as Boniface was to the Pope). Whenever and

wherever this reading system began to be required of native Romance-speakers, its prescriptive rules must have been in origin introduced from some non-evolutionary source. Most modern sociolinguists, including Schlieben-Lange (1982) on Romance, suggest that systems of diglossia need not only to be intentionally set up, but also to be continually reinforced subsequently—mainly by teaching the high variety in the education system—in order to exist at all, and do not arise naturally otherwise. What exist otherwise, and do indeed evolve unplanned, in a single wide speech community, are complex patterns of sociolinguistic variation.

Such patterns as modern sociolinguistic theory would lead us to expect to find anyway, particularly as concerns the relationship between speech and writing (e.g. Tannen 1982; Traugott and Romaine 1985; Pellegrini and Yawkey 1984; etc.), seem to be sufficient to explain attested phenomena from pre-Carolingian Romance Europe, largely reconciling reconstructed Romance with the (un-amended) manuscript evidence. Fontaine (1981), Varvaro (1984) and others have recently been in essence envisaging such a state, in which the many and varied registers of spoken and written language were still even so part of the same language, and read texts such as sermons were given vernacular phonetics in the ordinary way, as they always are now. Sabatini (1983:170) picked up this interpretation of the evidence with enthusiasm, since in this way the many conscientious pre-Carolingian preachers can be at last thought to have been intelligible to their audiences, the scholars who continually urged priests then to preach can be absolved from the charge of asking for the impossible, and hundreds of thousands of Early Romance-speaking individuals can recover their linguistic self-respect, their voice, their ability to understand their priests, indeed, their very participation in society in pre-Carolingian Romance Europe, as opposed to the idea, still widely held—e.g. by Coleman (1987:50)—that, 'of course,' pre-Carolingian Christian congregations found their services incomprehensible. The monolingual view of Early Romance Europe seems to be confirmed, for example, by two recent independent studies on Gregory the Great (Herman 1988, Banniard 1986) which both conclude that, unlike some of the Carolingian scholars, Gregory had no clear conceptual distinction between Latin and Romance in his mind, of the kind required for

diglossia to exist; the question just did not arise, and he cheerfully wrote sermons intended to be read aloud intelligibly to the illiterate. This question apparently never arose in the mind of Isidore of Seville either (cp. Fontaine 1981:776). If Gregory and Isidore did not know about such a distinction, it cannot have existed in the sixth and seventh centuries.

The hypothesis of Romance monolingualism also solves the problem that worried Bullough (1985:285, 287), concerning what language Charlemagne spoke with the Italian scholars at his court in the 780s: they all spoke Romance, in a mutually intelligible, if not similar, manner. Versteegh (e.g. 1986:426, 447) is the only historical sociolinguist I know of explicitly to disagree with this view. Versteegh, unlike most historical linguists (cp. Wright 1987: 621), sees lack of change as normal, and thus not in need of any explanation, and change therefore as necessarily externally caused. In his view high status, within a postulated nascent state of diglossia, accrues inevitably to those speakers who remain unaffected by the externally-caused deviations which lead to what he sees as pidginization in the speech of the socially less prestigious, such that this high level, reinforced by grammarians, naturally survives in formal situations even when the spread of the pidginized (in this case, Romance) variety means that the correct variety is hardly anyone's native speech any more. Hence, to Versteegh, many societies are de facto diglossic, without anyone having willed that diglossia into existence. But the supposed inevitability of the survival of a Latinate level of speech alongside evolving Early Romance speech, which seems to have been based on an assumption that (in the words of the Milroys, 1985:343) 'speakers who lead sound change are those with the highest status in their local communities as measured by a social class index.' That is how archaisms get stigmatized (cp. Silva 1988:164); pace Versteegh, untenable as a supposition ever since Labov and others established that educated people do not get involved in sound changes, has been archaism rarely has automatic high prestige. (Diez rejected Versteegh's approach in 1826, in fact; see Diez 1975:277-82.)

The tradition of reading Latin aloud as an artificial language, a sound for each written letter, in the Romance-speaking world as everywhere else, has the air of being obvious, and as though it had

been forever present. But someone, somewhere, had to establish that as a standardized norm, for it could not arise naturally in a native Romance community. There was a kind of continuity through the years between Carolingian and Imperial Latin in the vocabulary and syntax of the educated, for these could always be resurrected from Classical books by antiquarians (Fontaine 1981:786), but what we now think of as traditional Latin pronunciation had no such direct continuity with that of the Empire (cp. Lüdtke 1988:63, on [-m], for example). That is why the invention of the need for what we now call Latinate pronunciation (with the sounds determined by the spelling) is the key issue here. As the historian Hobsbawm (1983:1) pointed out, 'traditions which appear or claim to be old are often quite recent in origin and sometimes invented.'

In *Late Latin and Early Romance* the chapter which recounted the details of the suggested source of the Latin-Romance distinction, located in the latter years of Charlemagne's reign, was entitled "The Invention of Medieval Latin." This use of the word *invention* has been criticized (e.g by Godman, 1985:146). But I shall stick by it. That chapter argued in detail that the Carolingian scholars established the phonetic distinction around the year 800 A.D. as part of the educational reforms, in order initially to standardize the performance of the Church offices, and that the Latin-Romance distinction is only clearly felt subsequent to those innovations. Charlemagne and Alcuin knew they were introducing something revolutionary with their edict *De Litteris Colendis*, which added the study of *litterae* to the requirements of the already revolutionary *Admonitio Generalis* of 789 in order that clerics should impress their hearers by speaking well (*bene loquendo*) when reading or singing written texts (*in legendo seu cantando*) (Wallach 1959:204), and such reading proficiency becomes a requirement of the *litterati* thereafter (cp. Stock 1983:27). The onus is on Godman and Versteegh, and any other scholars who are sure that the clear conceptual distinction was established earlier, to suggest who else did it, when, where, how and why.

The elements that came to constitute Medieval Latin existed before Alcuin's arrival at Charlemagne's court—the writing system existed everywhere, the reading aloud system existed in Anglo-Saxon England—but their combination and conceptual opposi-

tion to vernacular was something new and positive. In Hurford's words (1987:12), invention typically involves a creative act of putting together existing elements (which may or may not be physical) in some novel way (also Schon 1967:87, 192). The concept and combined attributes of Medieval Latin were thus invented out of pre-existing ingredients. Rabin's study (1985) has shown how the same kind of conscious invention of a diglossic system happened also in ninth-century Byzantine Greek, in eighth and ninth-century Arabic, and in the Hebrew written in Moslem countries in the 10th century. This was an internationally felt psychological need; not confined to the Latin-Romance civilizations but consciously pursued in culturally less peripheral areas also. In other sociolinguistically comparable societies, diglossia has only existed if it was consciously established, in a particular historical circumstance, and was then educationally reinforced, as with katharevousa Greek, and does not exist in any society if nobody has invented it there (see e.g. Silva 1988:178, Rotaetxe 1988:60-61). The subsequent, and probably consequent, emergence of distinctively and intentionally non-Latinate writing systems for recording Romance vernaculars were—can only have been—experimental inventions by enterprising and innovative linguists, even if they were based on some existing approximations, and even if we do not now know who the inventors were. Perhaps they were Nithard at Strasbourg for the elaboration of the Oaths and Hucbald at St. Amand for the *Eulalie* sequence; but even if not, those advances must have been made by someone, in the same way as shorthand was invented by Isaac Pitman. The Riojan glosses were also elaborated for a purpose, and perhaps it was to aid a Catalan visitor to read aloud in local phonetics (Wright 1986). In any event, they did not just evolve.

CONCLUSION

Probably at all levels we have linguistic innovators of the past to thank for the invention of some of the structural distinctions that are subsequently taken for granted. Few people have the capacity to invent, but some do, and the capacity that we all have to learn from others is more plausibly seen as innate than is the actual concept demarcated by the inventor. After all, even a language acquisition device can only acquire things that already exist. Socially purposive

language planning is the result of an intentional initiative. It continues to seem probable that the Latin-Romance distinction of the later Middle Ages was created through such language planning, and that it would not have existed if it had not been invented. I entirely take the point made by McKitterick (1989:12-22) that through the ninth and tenth centuries the distinction took a long time to become generally felt (maybe at first it was only at Tours and centers influenced by Tours); but this is how it began.

*1994 POSTSCRIPT.*

This paper was first published in *Latin and the Romance Languages in the Early Middle Ages*, ed. Roger Wright, London, Routledge, 1991, pp. 103-13; reproduced by permission of Routledge. It is a very slightly emended version of the contribution to the "workshop" with the same title held at the Ninth International Conference on Historical Linguistics, at Rutgers University, New Jersey, U.S.A., on Thursday 17th and Friday 18th August, 1989. The way in which such a workshop came to be organized, and the remarkably wide, intellectually stimulating and mutually illuminating nature of its contents, are explained in my introductory chapter to the same volume, "Latin and Romance: a thousand years of incertitude" (pp. 1-5), not reproduced here. The present paper probably suited an audience of historical linguists better than it fitted into the resultant volume (one surprising reaction was that it had "caught the Zeitgeist"). Reviewers, however, have tended to interpret this chapter as a mere rehash of the hypotheses expressed in Chapter 3 of *Late Latin*, an assessment which seems to miss the point in the other direction, in that the purpose of this paper was to illuminate those hypotheses from previously unconsidered general historical linguistic viewpoints. The role of individual initiative, of scribe, teacher and chancery, in the earliest *scripta* in the Iberian Peninsula, is a topic I have developed since.

# Metalinguistic change in Medieval Iberia

# 4

WHEN DID THE DIFFERENT languages in the Iberian Peninsula first become different? The answer to this apparently simple question is far from obvious. Several criteria have been suggested for the establishment of this kind of conceptual distinction, but in the end the ones that turn out to be most important are political rather than linguistic; and the best answer to this question seems to be to place these metalinguistic divergences, for the Iberian Peninsula at least, in the thirteenth century.

We cannot rely on the criterion of mutual comprehensibility. Granted, if two individuals are totally unable to understand each other, then it is likely that they are not speaking the same language. But political decisions can invent distinctions of this kind between forms of speech that on the whole are indeed mutually comprehensible. Even now, the speakers of the separate Iberian Romance languages can sometimes understand each other, each of them speaking their own language, but we should not deduce from that that they all have the same language now. We can hardly deny that Portuguese, Castilian, Catalan, and now also Galician, are conceptually separate languages. It follows from this that the criterion of mutual comprehensibility is not the only one to be used when we are trying to decide if different forms of speech are in fact different languages or not.

These distinctions have been placed at the other chronological extreme; it has been suggested that some of the differences between the modern Hispanic languages go back to the times of the earliest colonization by the Romans. Under this view, the linguistic diver-

sity of the Peninsula can be traced back to seeds sown at that time, either as the result of the pre-Roman "substratum" languages or because different colonists came from Italy already speaking different dialects; and the geolinguistic frontiers that were established then would have remained, in this theory, ever since. But although there is no reason not to accept the validity of some of the differences that can be reconstructed for that time, these do not justify the postulation for that early period of a conceptual differentiation between whole separate languages of the kind that we are looking for (within the whole Roman Empire, not just in the Peninsula).
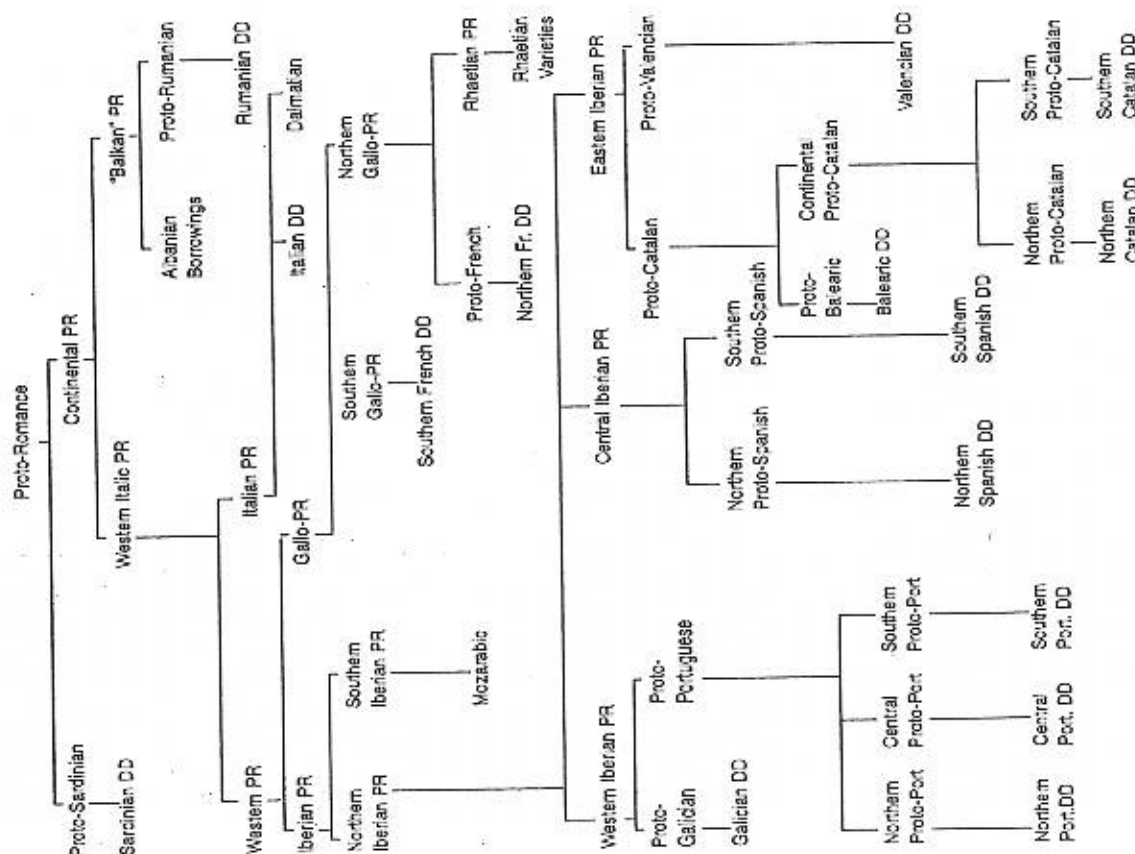
Our first task, then, is to consider the diagrams presented in the works of Robert A. Hall Jr. Here I reproduce an amalgamation of relevant diagrams from his *Proto-Romance Phonology*. Hall presents his diagrams to us as being the results of comparative reconstruction. Diagrams of this kind were originally designed on purpose to look like genealogical trees. They have often been criticized, and for a variety of reasons (e.g. recently by Craddock 1989). For example, the earliest distinction made in the diagram is shown to be that which separates Proto-Sardinian from the rest, supposedly more than two millennia ago; but it could well be that, if we want to reconstruct the time of the separation of Sardinian at all, we will come to a different answer if we consider morphology and syntax from that which we get from concentrating (as Hall does) on phonetics. Besides, the distinctively Sardinian feature, whether it is a new one or a surviving old one, could well have coexisted for centuries on the island with features which we now categorize over-definitively as being mainland Italian and not Sardinian. Sardinians and Romans were in continual contact throughout the Empire and the medieval period, and on the whole it is often possible for Sardinians and Romans to communicate successfully even now. There can be no doubt that many (or all) of the reconstructed differences between the speech of the two areas are genuine; the problem lies in the date of the conceptual metalinguistic distinction between two different languages, and the conclusion drawn from these reconstructed differences that two conceptually separate languages need to be represented at such an early stage in the genealogical tree-diagram.

These diagrams have also been criticized for a more aesthetic reason; that this tree image distorts the data in that it gives the

Diagram adapted from Hall (1976), 14-15.
("PR" = Proto-Romance; "DD" = dialects)

impression that each of these separate speech entities, once it has diverged from the rest, has continued to evolve as a separate unit in its own delimited geographical area, rather than forming part of a larger synchronic speech-community. This leaves no room for the possibility that a later innovation could have spread across the geographical frontiers implicit in the diagrammatic representation. Yet it has long been realized that at least some changes take place in such a manner, which could be better represented metaphorically through the image of a wave; if you throw a stone into a lake, you can see the waves rippling outwards to affect parts of the lake that are quite a long way from the stone. Malkiel (1983) and others have appreciated the problem of combining these two images into a single metaphor, because a tree is not very like a lake; even Hall mentions this problem, without solving it (1950:23).

In addition to criticizing these diagrams for their misleading chronological implications, and for the artistic inappropriateness of the inherent analogy, we can also criticize them for the conceptual vagueness of the labels attached to each node or branch of the tree. What sort of ontological or conceptual status should we grant to, for example, this "Western Proto-Romance" that gave birth to this "Iberian Proto-Romance"? How are we to define this "Northern Iberian Proto-Romance" which derives in turn from it? How do these three entities differ from each other, on the vertical (chrono-logical) axis? How did each one differ clearly from its "sister" languages on the horizontal (synchronic and geographical) axis? We can also reverse the implied direction of these questions and ask, for example, if this diagram is meant to suggest that there was no appreciable internal variation within each of these separate postu-lated entities. If there was no linguistic variation within the same geographical area, why not? For today, at least, there is linguistic variation inside every speech community. If such variation did indeed exist then, how are we meant to distinguish between (a) the internal variation that existed within each of these separate individ-ual units and (b) the geographical variation which encouraged the reconstruction theorists to make a conceptual distinction, between each of these supposedly separate linguistic entities on the diagram, in the first place? The more we consider such questions, the more we are likely to conclude that these diagrams, for all their apparent

neatness and clarity, raise more problems than they solve. Walther von Wartburg proposed similar theories with greater subtlety and more success, but there have been criticisms even of his fundamen-tal opposition between Western and Eastern Romance. The same kinds of problems have been seen in the work of Robert de Dardel, usually for metalinguistic reasons of the type adumbrated above.

The new element that has arisen ever stronger in the last fifty years, as the work of Alberto Varvaro in particular exemplifies, is our increasing understanding of sociolinguistics. The theory of early dialectalization—which in Hall's perspective began some time before Jesus Christ, and in that of Wartburg began in the third century A.D.—has been giving ground gradually to the realization that language-internal variation is inevitable and always to be found; a more satisfying hypothesis can now include the undoubted individual discoveries of the separatist tradition within a general picture of greater versatility and flexibility than a two-dimensional geometrical figure. For one thing, in general it now seems probable that a majority of variation phenomena, in any language, can best be explained sociolinguistically rather than merely geographically. This is partly because the new features whose presence may be going later to distinguish one daughter language from another tend to arise in other areas as well as those of the later daughter's geographical home; and also because they arise long before the disappearance of the older features that fulfilled the same function previously. Green (1991) traced the sociolinguistic and stylistic nature of this coexistence in the case of the increasing use of reflexive grammar with passive semantics, in all areas of the former Empire; this use coexisted with the old synthetic passives for centuries (and with the analytic passives still today). In the same way, all over the former Empire, future tenses formed by the combination of the infinitive with an auxiliary verb were used in some sociolinguistic levels and some stylistic registers for several centuries before the old synthetic futures disappeared entirely from speech. Outside France (and much of Catalonia) the new perfect tenses created with the auxiliary and the past participle still coexist quite cheerfully with the older synthetic preterite forms. Clear diatopic and diastratic distinctions arose late; some of the features that seem now to distinguish Old French Romance from Old Iberian

Romance, for example, actually existed in both areas for several centuries before the speakers made a subconscious choice of which one they preferred to perpetuate in the new local Romance standard; this comes out clearly, for example, from the fascinating studies by Lyons (1986, 1992) on possessive adjectives. In this way, the diversity discovered by the Reconstruction theorists can be preserved, but the theory of early dialectal divergence in separate clearly definable geographical areas between conceptually distinct languages is unnecessary. At the very least it is anachronistic to postulate such divergence actually within the Roman Empire. The results of the epigraphic analyses of Herman (and many of those by Gaeng) lead to a similar conclusion: that the Roman Empire indeed saw evolution and linguistic variation, naturally, but without there being any clear differentiation between the speech of different places.
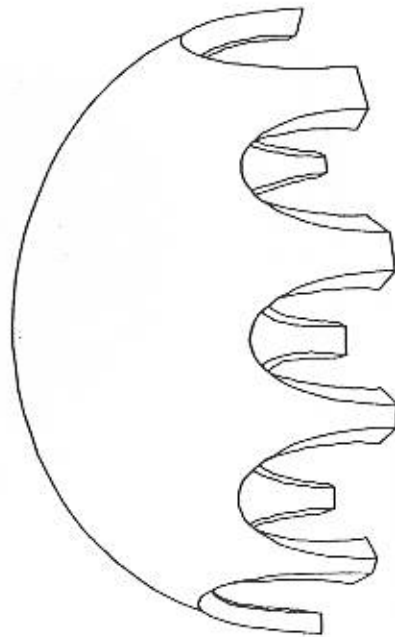


*Speech:* The Romance area (8th Century) with its universal "dome" (Varvaro 1991).

The conceptually monolingual nature of Early Romance lasted much longer than the Empire did. If we feel the need to have artistic representations of what happened in that speech-community in the post-imperial centuries, we can recommend, rather than any kind of tree diagram, the image proposed by Varvaro (1991). Varvaro does not visualize trees nor waves; he refers to the "dome" of standard (but not archaic) Latin which had social prestige, and which re-

mained spread out over and above the forces which were leading to greater differentiation between geographical areas. So far as I am aware, Varvaro has never attempted to draw this image, so I have tried to do that myself here.

In this picture we see Romania from the sixth to the eighth centuries as if it were a building. It is a single building, but each part of its dome is held up by a geographically separate column, representing the speech habits of one region. Each part of the area covered has a direct link, therefore, with the "universal" dome that unites the whole speech-community. We could, if we really wanted, give a label to each of the supporting columns, calling them "Northern Iberian Early Romance," "Southern Italian Early Romance," or names of this type, but only with the proviso that nobody should interpret these columns as being entirely separated from each other at least before the ninth century, and in many cases later still. There is geographical variation but there is also a metalinguistic unity, as each regional column provides for access to the central and higher dome. That is, every Early Romance speaker was still able to keep in linguistic touch with traditional educated culture. Another advantage of this pictorial representation is that we can preserve the normal sociolinguistic metaphor of regarding standard and geographically extensive registers as "superior" to those of lesser geographical and social extension.

Another advantage that this metaphor has over the tree (as Varvaro points out) is that between the fifth and the ninth centuries we can propose that the dome (representing the sociolinguistically "prestige" register) gradually crumbles and fragments, for practical reasons such as the decreasingly urgent necessity in speech for shifting style into that universal register, without the building thereby ceasing to seem to be a single conceptual unit within the semantics of the speakers. We can propose, for example, that the Early Romance-speaking Byzantines (speakers of "Danubian Romance," perhaps, in Hall's categorization) who came to settle on the southern Iberian coasts around the year 600 A.D. still managed to communicate successfully with the natives of the Peninsula. We can suggest, though, that by the ninth century, at least, some of the different columns were beginning to find that they had no direct communicative link with each other as a result of the crumbling of

parts of the dome that had connected them for centuries previously. The Rumanian column, for example, seems more likely to have lost these links at an early date than does the Sardinian one. Thus between the ninth and eleventh centuries some of the separate entities postulated by Hall in his diagrams for a very much earlier time (a millennium earlier) at last begin to find themselves standing as separate, conceptually distinguishable, entities. By the eleventh century, not much of the dome is left in active speech, within the Iberian peninsula it seems reasonable to suggest that although non-Catalan Ibero-Romance still has a conceptual unity, we can glimpse by then the existence of some kind of a frontier between this Ibero-Romance and the Romance of Catalonia, a frontier which coincides more or less with the border of the "Spanish March" of the Carolingian Empire. From the Carolingian era onwards, but only in the areas touched by the Carolingian reforms, we can add to our metaphorical building the image proposed by Muljačić of a newly invented "roof," to represent the conceptual innovation we now call "Medieval Latin," which was introduced into Romance Europe at the end of the eighth century; in this way the new roof can be seen to have replaced the crumbling dome in its role as a universal prestige register.

All these metaphorical representations have only been applied to speech, of course. If we still insist on pictorial metaphors, we can take a different perspective on the dome to include writing also within the same image. We can see it from the air, as in the second picture here. This is an adaptation of the image of a "nebula" which has been put forward by the French linguistic historian Michel Banniard. Banniard uses this image to represent his plausible view of the role of writing in the Early Romance communities up to the ninth century, and it is not meant to have any geographical implications. For writing, of course, there is no possibility of conceptual fragmentation at that time. From our aerial view we can see the center shining brightly, like the central part of a nebula, within our metaphor, that is the highest and most central part of Varvaro's dome.

Further out from the center we can see less brilliant areas which are even so linked directly to the shining traditional cultural centre; further out again are Romance-speakers of lesser cultural brightness

(and education) who nevertheless are a part of the same nebula, of the same "ensemble." More specifically, even in the ninth century, traditional culture was still generally accessible to all if the texts were read aloud sympathetically (cp. the studies in McKitterick 1991).
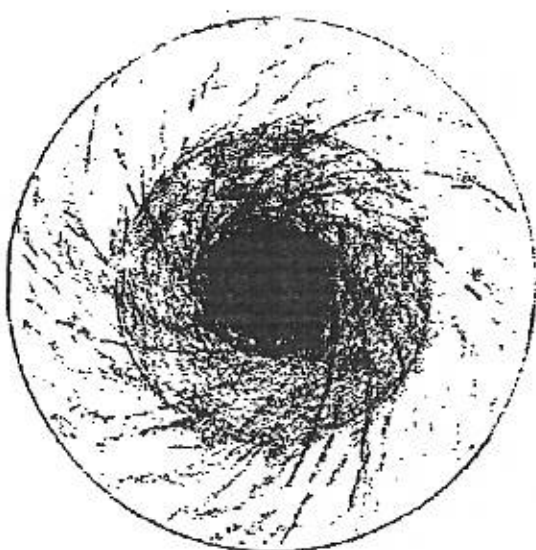
Writing: Aerial view of the dome:
= Banniard's "view."

Before the age of the Carolingian renewal of intellectual life, any pictorial image that preserves the metalinguistic unity of the early Romance area—with the exception that may have to be made for Rumanian—has greater value than an image of a tree with separate branches (or roots). The Romance world had a conceptual unity; it was monolingual. Of course, this monolingualism was heterogeneous, evolved, socially and stylistically complicated, but, being a single unit, we should use one name to refer to it (I prefer "Romance," or "Early Romance"). There is little to be gained by doing now what they did not do then, erecting clear but anachronistic theoretical distinctions between Late Latin, Vulgar Latin, Christian Latin, Romance, Iberian Proto-Romance (and Northern, Central, Southern, Eastern and Western Iberian Proto-Romances), pre-literary

Catalan, and so on, unless we are entirely clear and explicit in our own minds that with these labels we are referring to interconnected styles within one single language, and no more than that. Zumthor, 1984, and Gimeno Menéndez, 1988, are, for example; but there are also scholars who are multiplying these entities more and more. In addition to early and improbable geographically based distinctions, De Dardel (1983) postulates five separate chronological stages for Early Proto-Romance; and within these five stages there were, according to a later analysis by De Dardel (1987), no fewer than four separate successive Basic Word Orders. Historical distinctions can have validity, naturally (if we avoid the red herrings of typology, as Pinkster, 1991, points out). St. Isidore of Seville himself made such diachronic distinctions between successive stages of his own language (*Etymologiæ* IX, 1.6-7). He did not make them between geographically separate Latin/Romance languages of his own time; the only geographical distinctions he offered concern minute details such as some agricultural words of his native Baetica, but even Hall has stopped short of inventing a separate Seville Proto-Romance for the seventh century. The inhabitants of the Peninsula did make clear distinctions for centuries between their own language and Greek, Arabic and Hebrew (and, we must presumably suppose, Basque), but erected no such internal distinctions within their own Early Romance. (It is pleasing to note that the recent excellent book by Clavería Nadal makes no such internal distinctions either for these early medieval centuries).

What interpretation should be given, then, to the discoveries made with the techniques of reconstruction? Firstly, we should be happy to accept the validity of the reconstructed phonetic practice of all speakers in each area more or less as these techniques have suggested it was. We can suspect, thanks also to research from several historians, that texts read aloud could normally be generally understood, then as now. But this very fact also has implications for their morphology, syntax and vocabulary; we need to accept that within this conceptually monolingual Early Romance of the Peninsula there coexisted (a) old words, morphemes and constructions that were destined to seem archaic later, but which were still alive then in the sense that most people could still recognize and understand them when they heard them read aloud from written texts;

with (b) other words, morphemes and constructions more characteristic of active speech, which had been introduced comparatively recently, being ones we can reconstruct as forming part of Early Romance; and we should further accept that the whole apparently disparate "ensemble" was still a single conceptual unit, in the Iberian Peninsula even as late as the eleventh century. This is a millennium later than the strictest reconstructionists, such as Hall and De Dardel, prefer to believe. (The speakers of Early Romance tried to bar many normal features of their speech from their own written texts, not so much because they were new as because they were not recommended for use in the grammatical tradition deriving from Ælius Donatus; cp. Wright 1991b; Zumthor calls this filtering process a "prisme formulateur"). For present purposes, the important fact is this; at that time they did not think of their language as being a mixture of an older one and a later one (of "Latin" and "Romance"). It was their own synchronic system. Early Romance served them as well in their lives as modern Spanish serves modern Spaniards in theirs (cp. Alarcos 1982). Every synchronic state of any language could be analyzed as at least in part made up of archaic and innovating features, but is nonetheless a state of its own; there is nothing adventurous in suggesting that this applies to eighth-century Iberian Early Romance as well.

Early Romance in general was beginning to suffer geographical divergence by the ninth century, but within the Peninsula there was still unity in what had been one of the supporting columns, that is, non-Catalan Early Ibero-Romance. They did not give it such a distinctive name themselves, though; that is, it does not seem to have had yet solid conceptual independence; the most we can perceive is that at times they were aware of a distinction between the speech of the Peninsula and that of other areas, at least in some details of vocabulary and perhaps pronunciation. As I have pointed out before (Wright 1992b, 1993), it looks as though the more precise metalinguistic geographical distinctions, between Galician, Castilian, Aragonese, Leonese, had not formed in the Peninsula even by the twelfth century, and quite possibly not until the thirteenth. At the end of the thirteenth century it looks as if it is possible for us to begin to refer sensibly to Castilian, Galician, Aragonese, and Leonese, as conceptually distinct entities; although it still seems

rather unfortunate that we have to operate with labels of such an exclusively geographical origin, as if this were the only kind of variation there was. Nor should we give the impression to modern students that there was no variation at that time within each of these units. There was, of course, even in Old Castile. For thirteenth-century France we are lucky to have available the excellent studies of Dees and the Van Reenens; these show the existence of such variation with considerable clarity, which they are able to represent pictorially in simple annotated maps of Northern France, indicating the separate provinces. In these studies, variation comes over as statistical, and also as varying according to local rather than regional political units. If we look at data, rather than politics, variation usually does seem to be in this sense local rather than regional, but metalinguistic labels do not normally arise from dispassionate analysis of the data so much as being administratively imposed from above. As Janson (1991) showed, identification of separate languages tends to happen at the same time as, or even some time after, official decisions to reform writing systems. We could say that as many languages are thought to exist (in literate communities) as there are official writing systems. That is, in this case, the idea that Castilian, Gallego, Leonese and Aragonese were separate languages was inspired by the fact that each administrative area had acquired newly elaborated spelling systems in the thirteenth century. An earlier perception of these geolinguistic distinctions is not visible to us, not even, so it seems, between the Romance of Al-Andalus and the Romance of the Christian-ruled areas. I proposed (in Wright 1993) that before 1100, disregarding Basque, Arabic, Hebrew, and any surviving Berber languages, there was in the whole Iberian Peninsula (perhaps with the exception of Catalonia) a metalinguistic state similar to that of the modern English-speaking community in Britain; that is, a single language, Early Ibero-Romance, containing a great deal of variation; and that by the end of the thirteenth century there had arisen a very different metalinguistic state, more like that of late twentieth-century Spain, in which several newly-identified languages, named on a geographical basis, are coming to be thought to be different, a perception allied to and catalyzed by the invention of new spellings. For example, the modern distinction which some wish to make between

the Catalan of Valencia and that of Barcelona is reminiscent of the distinction that was made in the thirteenth century between the speech of Catalonia and that of the Occitan area, and in each case the distinction is reinforced by the elaboration of official orthographical differences. These are all political decisions, both in the thirteenth and the twentieth century. For similar reasons, the distinction between Galician and Portuguese was invented in the fourteenth century. There were differences, mostly merely statistical ones, that we can reconstruct for earlier times, but they were felt as language-internal within Romance rather than as symptoms of the existence of separate languages.

*1994 POSTSCRIPT.*

This paper is a translation of "Los cambios metalingüísticos medievales," in the *Actes du XXe Congrès de Linguistique et Philologie Romanes*, Editions Francke, Tübingen and Basel, 1993; that was the exact text of the paper given to that conference, in Zurich, on Thursday 8th April, 1992. The diagrammatic representation of his "dome" caused Varvaro, who was there, to give a broad smile, but he has still not told me how closely it fits his own image. Bannard has never drawn his own image of the "nébuleuse" either, so far as I know. All these diagrams are merely visual aids, of course, not meant as exact counterparts. The subsequent discussion was, inevitably, dominated by linguistic politicians from Spanish "autonomías" with local bones to pick. I gave the same paper the following month at the University of Valencia; the audience there fell about in laughter at Hall's diagrammatic representation of Catalan; if there is any sense in distinguishing separate "Catalans," the basic dividing isoglossic line runs North-South, that is, it is between East and West, whereas Hall (without apparently investigating the subject) regards North versus South (that is, with East-West isoglosses) as the default case. I was glad to give this paper close to the centenary of the invention of the concept of the isogloss; it has long been suspected that the individual isogloss is more real than the dialect frontier, but even many of those are now appearing as transition zones, and perhaps dialect boundaries need politically established separate standards on each side to pull the inhabitants

separate ways when they are style-shifting. Perhaps the image of competing magnets could be elaborated.

# 5

## The Asterisk in
## Hispanic Historical Linguistics

AKOV MALKIEL (1989) HAS pointed out that the use of the asterisk in diachronic Romance studies has become ambiguous. It is in danger, indeed, of becoming misleading, and Malkiel's desire to see standardization in the field should be supported. To this end, I have extended several of the arguments initiated in Malkiel (1989), and also studied what a number of specialists in Romance and Hispanic Philology have actually meant by using the device. The focus here is exclusively diachronic; specialists in synchronic syntax use the asterisk in a different way from historical linguists, in order to indicate unattested concatenations of individually attested forms (as neatly explained by Rosen, 1987), although that habit has its critics as well: most notably the scholar who claims to have started the practice (Householder 1973). The twenty authors whose works are scrutinized for the present purpose should not construe any of the forthcoming observations as being hostile—or at least, in so far as they are, please note that one of the books to be criticized is my own.

The twenty-three studies considered here are listed in chronological order at the end of this article (Appendix A). Of these authors, seven explain what the use of the asterisk is intended to convey in the study concerned; one explains why the asterisk is not used; the other fifteen give no explicit guidance to the reader on the matter. These explanations are reproduced on the next page:

1   Rickard (1974: 9); '*: postulated (i.e. unattested) form.' (44/49)

2   Elcock (1975: 32): "the linguist would indicate by the convention of an asterisk that it is a reconstructed form.' (308/334)

3   Macpherson (1975: 91): 'when the Vulgar Latin form has to be deduced from the evidence of the Romance Languages but is not documented, the postulated root is prefixed by the abbreviation V.L. and an asterisk: e.g. culebra < V.L. *COLŎBRA. (75/89)

4   Hall (1976:1): 'Forms ascribed to a Proto-language are usually marked with an asterisk and are therefore often called "starred forms," e.g. Proto-Romance */abantiare/ "to go ahead." In this work, however, the label PRom, (=Proto-Romance) always refers to a reconstructed form, and hence renders unnecessary the use of an asterisk or star.'

5   Cano (1988: 293); '*: forma lingüística no documentada pero exigible en la evolución.' (81/85)

6   Maiden (1991: 285): 'Words preceded by an asterisk are the reconstructed Proto-Italo-Romance or Proto-Ibero-Romance base of the modern forms.' (100/106)

7   Stengaard (1991: 315): 'el asterisco usado (...) por no conocer una documentación.' (27/37)

8   Penny (1991: 5): 'an asterisk indicates the lack of confirmation from written sources, and therefore the hypothetical (but not necessarily doubtful) status of the word concerned.' (137/190)

The numbers in brackets at the end of these quotations refer to the proportion of uses of the asterisk in that book which actually seem to fit the definition given. Thus for example, Rickard (1974)'s definition holds for forty-four of his forty-nine uses, but two of his asterisks accompany forms whose existence is in fact being denied, and the three other asterisked forms are in fact known to be widely attested, of Elcock (1975)'s 334 asterisks (excluding the seventeen in his hypothesized first draft of the Strasbourg Oaths), 308 fit his definition, six accompany forms whose existence is being denied, and twenty asterisked forms are attested, Macpherson (1975) uses the asterisk for purposes other than those proclaimed on fourteen of

his eighty-nine occasions, including seven that immediately follow after the definition; Hall (1976) announces at the start that he is not going to use asterisks, so the function of the twenty-seven that turn up later in that volume is thus initially baffling; Cano (1988) follows his intelligent and slightly different definition eighty-one times out of eighty-five; Maiden (1991) follows his proclaimed practice 100 times out of 106; Stengaard (1991)'s definition is presupposed rather than asserted, and applies twenty-seven times out of thirty-seven, the existence of the other ten being indeed undocumented but also denied (disregarding fourteen uses for footnotes and one to indicate a corrupt manuscript); Penny (1991)'s definition is skilfully phrased, yet his 190 asterisks include thirty-nine for forms that are known to be attested and fourteen for forms whose existence is being denied. All relevant statistics are summarized in Appendix B.

When.Malkiel or I refer to an asterisked form which is in fact attested, that is not primarily because the investigator has disregarded documentary evidence (although that is so in at least one case; Harris-Northall's *CAPU- (1990: 62) is actually attested in both the Visigothic liturgy and the Visigothic laws, as the infinitive capuisse, Díaz y Díaz 1965: 71). The references are to forms that are indisputably attested but with a different morphological case usage from that required by the etymologist's (ultimately neogrammarian) theory (Malkiel 1989: 25). Several writers, for example, give an asterisk to *LUMINE, the etymon of Sp. lumbre. Yet LUMINE is a normal attested form from Imperial times, being the ablative singular form of LUMEN. The asterisk is there to hypothesize a function rather than a form, because some scholars wish to bar a priori the possibility that a Romance noun could descend from an originally ablative form; what is being hypothesized is not the form itself but the functional use of the form as a direct verbal object. In this way using the asterisk begs an important question. Penny wrote a persuasive article (1980) arguing that the original case-functions of forms are not necessarily relevant to the choice of surviving member of a reducing paradigm; so it is surprising to see that he too stars *LUMINE (1991: 108). It would perhaps seems more logical to write *LUMINEM, with the accusative-marking -M. Asterisks have also occasionally been used to mark explicitly attested forms as being non-standard, although Romance Philologists have not usually done

so; Malkiel also points out (1989: 43) that Ernout and Meillet's etymological dictionary often stars non-classical forms while repeating an actual written attestation, e.g. (example chosen at random), *fānō is given an asterisk when being quoted from its written source in Varro (1951: 384).

When an asterisk is used in historical linguistics to deny the existence of a form at all, the scholar is adapting the device from the practice of modern synchronic syntactic study. This seems a natural thing to do (there are four such cases in Wright 1982) but it provides great potential for ambiguity and is surely inadvisable. For example, Lloyd (1987) stars 211 forms altogether, of which fifty-one are forms whose existence seems to be specifically being denied; Dworkin (1985) uses 150 starred forms altogether, of which thirty-two seem to be being denied; etc. Malkiel has made the proposal on a number of occasions (1970, 1979, 1989) that if we accept the use of the traditional superscript asterisk for undocumented forms whose existence we positively wish to postulate, then in order to avoid confusion we should use instead a subscript asterisk for undocumented forms whose existence we do not wish to postulate (e.g. $_*$MANS' as plural of MAN, 1979), yet it seems that neither Malkiel nor anyone else has actually followed this advice. Pensado, however, regularly uses a double asterisk for forms whose existence is being denied; in the two 1991 studies considered here, she double-stars **zaro (1991a: 203), **seze, **alvierto and **doice (1991b: 66, 73, 79) as denied forms, and her other eighteen single-asterisk italicized uses are all for forms postulated as existing. This use of double asterisks to deny a form's existence is an admirable practice. Henceforth I propose to follow her example.

We can thus begin to clear the air by not starring those forms such as LUMNE that are attested but whose morphological analysis offends our preexisting etymological theory, and by double-starring forms of whose existence we wish to deny the possibility. There still remains even so a great residue of ambiguity and, indeed, fog. Some of the confusion is a consequence of the fact that the device was not invented for the Romance field at all, but for pre-literate languages. The use of the asterisk for this purpose was made generally known by Schleicher, although, as Koerner (1989: 179-84) has shown, Schleicher did not actually invent it. Schleicher was not

a Romance philologist. When used by those studying prehistoric stages of a language, the implication of the use of the asterisk and the nature of the form it accompanies is unproblematic; the asterisk suggests that a lexical item of more or less the phonemic form implied by the manner of the accompanying transcription, however broad or narrow that is, existed at a period in the past for which we have no contemporary written evidence. Romance philologists mentioning languages other than Latin and Romance are at liberty to use it that way too, naturally, for reconstructions in prehistoric Indo-European, Old Germanic, Old Celtic, Old Basque, unspecified pre-Roman or what Hall calls "Mediterranean," as well as Italic or occasionally 'Ancient Latin' lexical items. In such cases what is being postulated, by definition, is a spoken form, for writing was not used in the prehistoric societies referred to. Even so, the implications of the asterisk: attached to non-Latin etyma in the work of Romance Philologists can be unclear; when scholars attach asterisks to postulated etyma of Germanic or Celtic provenance, it is rarely if ever made explicit whether the words are equally unattested in both Germanic (Celtic) and Latin, or unattested in Latin but attested in Germanic (Celtic); and conversely, when they do not attach asterisks to such etyma, it is similarly unclear whether they are attested in both languages or attested only in one, and if so which.

When Romance philologists use the asterisk for reconstructable forms in Late Latin / Early Romance / Proto-Romance, they are referring to a historical period when writing indeed existed, and the lack of written documentation is therefore significant, with different implications from those of a written form's inevitable absence in pre-literate societies. We need to appreciate that enormous confusion has been and is still being caused by the failure in our discipline to distinguish clearly between three separate phenomena: (1) a lexical item in a speaker's mental lexicon; (2) the written form of a lexical item; and (3) the spoken forms of a lexical item. (As a subset of lexical items, bound morphemes have come in for the same confusion on our part.)

Diachronic linguists from other areas of study might assume that the Romance philologist's asterisk usually postulates the existence of an item unattested in writing. But this is not the case. If we set aside the reconstructed etyma from languages other than

Latin, and those whose existence is being denied, and confine our attention for the moment to asterisked forms in capital letters leaving for later consideration the asterisked forms printed in italics, in ordinary script, in phonetic script, in phonetic script between slashes, in phonetic script between square brackets, in bold type, in phonetic script in bold type, or between inverted commas), even then, we discover that totally unattested items form only a tiny minority of the asterisked forms (see Appendix B1). Entwistle (1936) presents five out of his twenty-one asterisked capital forms in this category, Corominas (1961) seven out of his fifty-six, Menén-dez Pidal (1972) four out of his sixty-two, Elcock (1975) three out of his 296, Macpherson (1975) one out of his sixty-seven, Wright (1982) three out of seven, Dworkin (1985) none out of his forty, Lloyd (1987) none out of his 143, Meier (1987) thirteen out of his 238 (mostly, like Menéndez Pidal's, presented in elongated type rather than in capitals), Cano (1988) none out of his seventy-four, Ariza (1989) none out of his five, Harris-Northall (1990) none out of his forty-three, Pensado (1991) one out of her thirteen, Penny (1991) two out of his 126; total, thirty-nine out of 1191. Roughly, only one in thirty asterisked capital forms are asterisked in order to indicate that these are unattested items. (These statistics are not necessarily precise, for there is often a lack of clarity in the text as to why a form is asterisked, and there might be disagreement over the precise categorization of some of these examples). It would be preferable to propose that in future the asterisk is confined only to these one in thirty: that is, for use with lexical items, both words and mor-phemes, that are indeed unattested and whose existence is being postulated. The etymon for Romance words for 'small' might be a good case in point.

The other approximately 96% of asterisked reconstructed forms in capitals (Appendix B1) fall into two main further categories (in addition to those like *LUMINE, or the unattested combinations of attested words that receive the asterisk they acquire in modern syntactic studies). The largest category concerns unattested lexical-ized combinations of individually attested morphemes, very nearly always a combination of a free lexeme plus one or more reconstructably productive bound, usually derivational, affixes. In such cases, the component parts are attested, and what is being hypothesized is the lexicali-

zation of their potential union. Yet the question of whether a particular combination of live lexeme and productive affix counts as a lexical item at all is a thoroughly problematical one even in synchronic study; lexicographers vary widely over which combina-tions to include in Modern Romance dictionaries. The use of an asterisk to indicate that a particular combination of attested mor-phemes is not itself attested in lexicalized form in texts of the Late Latin / Early Romance / Proto-Romance stage would be entirely acceptable if the analyst ever stated that that was what he or she was using it for. Several scholars have intuited the awkwardness of this practice; Malkiel (1989) prefers to insert a plus sign between the morphemes in such cases, and Meier a minus sign (or hyphen), implicitly thereby demonstrating unease about using an asterisk to mark such unattested combinations of attested morphemes on the same level as totally unattested and thus far more hypothetical items. The late Professor Meier made a career out of postulating affixed forms in capital letters (or elongated script), and used the asterisk for this purpose most of the time (183 out of 238 such uses in Meier 1987); e.g. *vid-in-iare as the postulated root of Sp. guiñar, Fr. guigner, It. ghignare, Old Oc. guinhar, formed off videre plus two attested suffixes (1987: 47); or sometimes without the hyphen, e.g. *perusticare off usus as the root for the subse-quently deverbal brusca ('kindling', 1987: 23); etc. The asterisk accompanies the hyphen, in effect, even if he refrains from inserting one. Meier never seems to have proposed in these reconstructions the existence of either a totally unattested affix or a totally unat-tested lexical root. Indeed, Meier gave the impression of being constitutionally unable to prefer a non-Latin etymon, whether hypothetical or not, to the unattested combinations of attested morphemes that he almost invariably recommended (cf. Meier 1989). That is, he used the asterisk liberally, while at the same time explicitly preferring not to postulate the existence of any unattested morpheme, a conjunction of approaches that borders on the self-con-tradictory. A further complaint that can be made against the use of asterisked capitals for such reconstructed combinations of lexeme and affix, adduced also in passing by Malkiel (1989: 66), is that there is no need to postulate a Latin etymon from an early period in the many cases where both the lexical root and the affix survive into

the relevant Romance languages and no further phonetic change consequent to their lexicalization need be accounted for. For example, given Sp. *apagar* ('extinguish') and Italian *appagare* ('satisfy'), it is easier to propose that the lexicalizations happened after the separate phonetic developments of pre-consonantal AD > *a* and PACARE > *pagare*, for a separate semantic purpose in each case, rather than reconstructing an unattested, but nonetheless polysemic, asterisked capitalized *ADPACARE. The capital letters D and C in this asterisked form imply the existence of the sounds [d] and [k] in the reconstructed verb at the time of its initial lexicalization, but there is no advantage or need to reconstruct them.

This case (adduced among others by Penny, 1991; 11) demonstrates in passing the potential confusion that can be particularly caused by the other main category of asterisked capital forms; one which Malkiel [1989] refrains from mentioning, but is the most misleading of all. A surprisingly large number of asterisked forms printed in capital letters are lexical items whose existence is certain and generally agreed to be attested; the reason for the asterisk is that the form adduced (by the philologist) has been given (by him) a spelling which is not attested. Twenty-two of Lloyd (1987)'s 143 asterisked capital forms, twenty of Penny (1991)'s 126, fifteen of Macpherson (1975)'s sixty-seven, sixty-five of Elcock (1975)'s 296, and forty-two of Cano (1988)'s seventy-four, for example, fall into this category (called 'intermediate' in Appendix B). The following examples come from Cano (1988): *PRIMARU (compared with attested PRIMARIU, p. 71), *VERECÚNNIA (compared with attested VE-RECUNDIA, p. 78), *CAIPA (compared with attested CAPLAT, p. 158), etc. Twice adduced is *COLÓBRA (Cano 1988: 81, 208), compared with COLOBRA, which was also the example used in Macpherson (1975)'s definition of what an asterisk is for (see above). These lexical items are not unattested. The spellings produced by the modern author seem to be; but it should be elementary linguistic practice to distinguish carefully between a lexical item (which is constituted with phonemes, and exists in the mental lexicon), an orthographical form (which is constituted with letters, and exists in a writing), and a phonetic form (which is constituted with sounds, and exists as vibrations in the air); and in this common practice of inventing unattested spellings, the three entities are being hope-

lessly confused. This habit, of inventing and then starring unattested spellings of generally-known lexical items, struck me as so peculiar that at first it seemed not only pointless but even deliberately baffling, then a ghastly suspicion dawned, which is still with me. Some of our colleagues, and not only in the distant past, believe—or behave as if they believe—that Latin spelling was an exact phonetic transcription of the writer's speech, regardless of his time and place. This is in itself a grotesquely implausible hypothesis, but it seems that it must be the hypothesis which underlies this practice. For example, several of the scholars investigated here (including Malkiel 1989: 68) invented, starred and capitalized a form *COMPARARE, as in (here quoting Lloyd 1987: 201) '*COMPARARE "to buy" [C.L. COMPARARE] > comprar'. The implication of Lloyd's phrasing really does seem to be that the usual spelling with an 'A' (COMPARARE) can at all times and in all places only have represented a pronunciation with an [a], so that if the modern scholar wishes to postulate a different spoken vowel he needs to invent a different written form. This idea is not only strange but unnecessary. Phonetic script had not yet been invented, and even if we were to think it might have been, we know for other reasons that Proto-Romance probably had a schwa [ə] but no symbol ever biuniquely allotted to that sound in writing.

The Schleicher tradition sensibly used asterisks to reconstruct spoken forms of postulated lexical items in preliterate times; but the Romance philological tradition has developed that habit further: the asterisk is now used when mentioning lexical items whose written form, as attested, is not exactly the same as a phonetic transcription would have been of the pronunciation we wish to reconstruct for that lexeme at a Late Latin / Early Romance / Proto-Romance stage. The presence of the asterisk here thus confuses a simple issue. We might as well transcribe Modern English *ORINDZH in asterisked capitals as though it was a hypothetical entity. *orindzh* is indeed unattested in writing (other than in my own metalinguistic work, Wright 1982: 50), but asterisked capitalized *ORINDZH—written that way, as in Romance philological practice—would merely be a written representation of a pronunciation [ɔrɪndʒ] of a lexical item that is indeed commonly attested in writing, as *orange*, and does indeed exist. For what our colleagues mean by asterisking *COM-

PERARE [with a letter E after the P] is not that this is a hypothetical, but reconstructable, lexical item, but that it is a reconstructable pronunciation of a lexical item commonly attested in a different form, with the invented orthography designed in this case to represent a proposed [e] [or [a]] in the second syllable. In all the many cases in this category it is not the lexical item itself, but a particular phonetic representation of that item, which is being postulated in the reconstruction, so it is surely preferable—if we wish to write this reconstructed phonetic form on paper at all—to write this reconstructed phonetic pronunciation in the same way that we write all other phonetic entities in scholarly study, that is, in phonetic script, and preferably in the International Phonetic Alphabet. Written forms in phonetic script, of course, are never attested at all in the years before the initial invention of phonetic script, and are extremely uncommon even now, so even if we use phonetic script ourselves for such forms the use of an asterisk with the phonetic transcription of a lexical item of the past seems (as Hall says, see above) an unnecessary piece of typographically redundant hypercharacterization. What Cano means, for example (1988: 71), by asterisking *BAISU is that at some point in Late Latin / Early Romance / Proto-Romance the lexical item traditionally written BASIUM was pronounced with the offglide [i] [or [j]] preceding the sibilant [s] [or perhaps [z]]. It would be simpler and less misleading simply to write here 'BASIUM [bajsu]'—if indeed that is the pronunciation he is wishing to hypothesize, for among the demerits of the practice here being criticized is its lack of precision. Malkiel himself does not discuss this explicitly, but seems to prefer to asterisk italicized forms for such intermediate phonetic reconstructions, writing (1989: 40) *abodega for the form intermediate between apotheca and bodega, and *alicer rather than alacer as the etymon for Sp. alegre (1989: 86).

The asterisked form, then, is often meant in this way to refer to an intermediate pronunciation. Intermediate, that is, between that of ancient Latin, presumed (perhaps justifiably) to be represented isomorphically in the traditional orthography, and the earliest Romance forms, also presumed (in part justifiably) to be isomorphically represented in the Old Romance orthographies. That meaning for the asterisks in this category, however, is a deduction

on my part. None of the studies investigated here phrases the purpose of their asterisks in this way, although Cano's definition comes close. The fog caused by the failure to make this clear, and the conceptual confusion between the lexical item, its phonetic form and its written form, can lead and has led to baffling habits and comments. Consider, for example, the very common practice (which astonished me even as an undergraduate) of writing nominal etyma intended to be accusative singular in capitals without their final -M [whether or not the form is accompanied by an asterisk]; POPULU as the etymon of peuple, povo, pueblo, popolo, etc. These forms of second declension nouns are not normally attested at all; POPULU, with a final letter -U, is not the traditional spelling of any morphological case, and may never be attested as such. For the sake of consistency, the form might thereby seem to be prima facie deserving of the asterisk which should never have been allotted to *LUMINE, being in the same 'intermediate' category as *CAPA [etc]; in late manuscripts where the accusative indeed often lacks a letter -m [or the tilde] the final written letter is usually -o. The purpose of this is sometimes understandable in that the etymologist wishes to refer to a lexical item without committing himself to any particular case-form; but nearly always, at least among Hispanists, it is only intended to represent the originally accusative singular. This dropping of the written -M in capitalized etyma is still common among etymologists; Lloyd half-realized that it was inconsistent, for he devoted a note to it (1987: 72, n. 3): "When Latin words are given as the etyma of Romance words, the usual practice will be to give them in the accusative singular form of nouns and adjectives, minus the final /-m/." Unfortunately, this manner of explaining the habit compounds the confusion, for the Latin words were never found in texts with a final /-m/' in phonemic slash brackets, but instead normally with an orthographic -M, and the suspicion returns; can even the intelligent, meticulous, hard-working and near-omniscient Paul Lloyd actually believe that traditional spelling was a phonemic script? For the reason the letter is traditionally omitted from these accusatives is that we can feel sure in most cases that the final -M corresponded to no sound. Yet if philologists feel justified in reshaping the etyma unasterisked for that purpose, why not do it for all purposes? It would be best then to avoid not only the capital letters

but also the asterisk, and present the pronunciation of all etyma in unasterisked phonetic script, unless the lexical item is indeed totally unattested.

It seems likely that other laborers in this field have probably had a similar thought to this. For seven of the twenty authorities consulted in the present exercise do not use asterisked capitals at all (Hall, Rickard, De Dardel, Pharies, Ariza, Maiden, Stengaard), while nine of the twenty, at least occasionally, do indeed use asterisks with phonetic script: Hall, Lloyd, and Penny between phonemic slashes, Rickard, Elcock, Lloyd, Harris-Northall, Maiden and Penny between phonetic square brackets, Macpherson and Harris-Northall in bold type, Ariza and Maiden usually in ordinary unbracketed type (see Appendix B, 4-8). Hall's use of asterisks with phonetic script is interesting. At the start of his *Proto-Romance Phonology* (1976), as quoted above, he states that he will not use the asterisk, but in the event he does, with phonemic slashes: nineteen times to hypothesize the existence of the accompanying pronunciation, and four times to deny it. In his *External History of the Romance Languages* (1974) he uses the same device for fifty-eight Germanic etyma, one Norse etymon, and seven postulated Latin-Romance forms, plus one form whose existence is being denied. In his *Proto-Romance Morphology* (1983) the asterisk and the phonemic slashes similarly accompany fourteen Proto-Indo-European reconstructed forms, eight other reconstructed forms, and three whose existence is being denied. Presumably his asterisk is meant to imply that the form is not found in any written form at all at the time for which Hall is postulating its initial existence, although his habitual chronological vagueness makes this a far from illuminating presentation. Macpherson uses phonetic script in bold type to hypothesize the existence of six forms and deny three ([*ʎaxa, *nexát, *rwéxa', which 'clearly did not happen', 1975:157]. Harris-Northall (1990) uses an array of devices with phonetic script both to postulate and deny a form's existence; starred square-bracketed forms to postulate eight forms and deny two, bold type phonetic script to postulate six and deny and deny two, bold type in square brackets to postulate four, and unbracketed phonetic script to postulate two and deny two. In general, the hypothesized pronunciations represented in phonetic script are much more often than not intermediate between those attested in

what is assumed to be an isomorphic manner in Latin and in Romance orthographies, as in the 'intermediate' category of aster-isked capital forms, but at least there is in the case of phonetic script no doubt about what is being postulated; the pronunciation of a lexeme which we know to have existed, but which was written then in a manner different from what its form would have been had the scribes been clairvoyant experts in the modern Phonemic Alphabet. That lack of correspondence applies to all the spellings and pronunciations of the time, naturally, whether asterisked by modern philologists or not, so this asterisk is as unnecessary with phonetic script as it is with capital letters, but at least what we have here is the redundant laboring of an intelligible point (a hypothesized phonetic form) whereas the previous category (asterisked capital forms) involved the muddling of three separate points (phonetic, orthographical and lexical phonemic forms). There seems to be no nuance of meaning to correspond to the way in which asterisks are occasionally placed inside rather than outside the brackets.

One proposal that Malkiel put forward in his 1989 study seems uncharacteristically unhelpful: that is, to star forms in which a vowel has changed quantity, e.g. from ĭ ([i:]) to ī ([i]); he offers (47) *glīris (in italics, in this case) as opposed to the original glíris, the original genitive form of the word for 'dormouse'. Such quantitative deductions can often be made from verse, but Latin texts were not normally written with the macron and the micron. Probably neither form is attested with the diacritic. Glíris is attested, and we cannot presume that the first letter i never represented a short [i]. There seems little point in adding an asterisk to mark forms as undocumented if they could not have been expected to occur in that written form at all (such as in phonetic script before it was invented, or with usually absent diacritics). Let us write instead that the written form glíris came in time, particularly in this case in France, to represent a pronunciation with a short [i] rather than a long [i:] in the first syllable.

Philologists often use asterisks with italicized forms (Appendix B2), Hall's *Phonology* (1976) and Harris-Northall (1990) are the only works studied here not to use that practice at all. As with the capital forms, the implication is sometimes that a pronunciation is reconstructed to have existed in a form similar to that which would

be implied were the accompanying form written in italics actually written in phonetic script; and, as before, sometimes the existence of such a phonetic form is emphatically being denied. Rickard (1974) and Stengaard (1991), who use no asterisked capital forms at all, use italics for the purposes fulfilled elsewhere by capitals—that is, for unattested combinations of genuinely attested morphemes (such as, passing in Stengaard, *essere*—which in fact is copiously attested, but in what we now call, as a result of a change in naming practice—see Janson 1991—'Italian'), for intermediate pronunciations, and in Stengaard's case for eight denials (out of thirty-seven uses overall). Lloyd (1987) uses asterisked italics on eleven occasions to hypothesize, and on fifty occasions to deny; a form; Wright (1982) twice to hypothesize and four times to deny; Pharies (1986) forty times to hypothesize and fifteen to deny, and twice where even a specialist reader cannot be sure whether affirmation or denial is intended (it seems to be both at once: *roño*, twice on p. 15); Dworkin (1985) eighty-seven times to hypothesize and twenty-two times to deny: etc. Lloyd is again here half-aware of the ambiguity, as to what is or is not being hypothesized with an asterisked italicized form, without resolving it: e.g. 'the absence of any evidence in writing for Early Castilian forms like *faço, *yaço or *plaço or *teno and *veno merely indicates that these changes all took place at a very early stage, before there was any need to write Castilian [sc. 'to write Romance in a new and distinctively Castilian manner'] (1987: 295).

Starred italicized forms are also used for unattested concatenations of individually well attested items, as by the synchronic syntacticians. This is the only role De Dardel (1983) gives to the asterisk, that is, for items which are indeed attested. In its way, this apparently perverse decision is as it should be, for almost everything else in his book (apart from the medieval quotations) is hypothetical and reconstructed but presented without an asterisk. Asterisked italicized forms in these twenty-three studies are also found for hypothesized pre-Latin forms (Hall, Pensado), Germanic forms (Entwistle, Hall, Macpherson), Celtic forms (Entwistle, Rickard, Elcock), Basque forms (Corominas, Menéndez Pidal), and even, very inappropriately, for an Arabic form (Wright). In such usages the distinction between italics and capitals has broken down, but

generally that distinction—in the work of writers who present forms in both scripts with asterisks—seems to concern the target orthography that would have been aimed at (if this postulated form had in fact ever been written). The chronological distinction between Latin and French (etc.) is an administrators' fantasy, of course, but the distinction between trying to write in the old way and trying to write in a new way, once new ways had been invented, is a real distinction, and if for some perverse reason modern scholars do not want to use phonetic script when discussing pronunciations, that typographical distinction has at least some raison d'etre when followed consistently.

Asterisks are also used to accompany words in inverted commas by three writers (Dworkin, Meier, Penny), unambiguously to refer to a meaning which a particular word does not have. A number of the works also use asterisks as footnote signs, which seems tempting fate in an etymological study.

## CONCLUSION

As a result of this investigation I feel exactly as Malkiel (who refers to 'downright whimsical armchair reconstructions', 1989: 26) says that most people in the field come to feel over time, that the asterisk should only be used sparingly, if at all, in diachronic study. In addition, it is surely preferable to use phonetic script, rather than any other, for pronunciations, of the past exactly as it is for those of the present, and if the lexical item in question is attested in any orthographical manner at all then no asterisk need accrue to its phonetic version. In this way we can preserve the essential theoretical consistency of phonetic developments by postulating early etyma for Romance forms to derive 'regularly' from, without thereby postulating more than we need; compare Orr's distaste (1948: 82) for the 'application of sound laws to those forms of words which some literary document had happened to consecrate. Instead of writing, for example (Macpherson 1975: 109) 'LAXUS > VL. *LASSUS > O.Sp. lexos', we could write 'Late Latin' (or Early Romance or Proto-Romance, to taste) (lassius),' (if that is indeed the proposed pronunciation—it is not clear) 'written laxus, > Old Spanish (léjos), written lexos', without using an asterisk at all. (We will then need to reconstruct the pronunciation of a whole word, rather than just one

of its constituent sounds.) Then we can keep the asterisk for genuinely unattested but reconstructable lexical items (such as 'O.Sp. losa < *LAUSA', Macpherson 1975: 122—unless it is indeed attested after all), and its function is perceptibly unambiguous. Even then, though, what is reconstructed is a phonetic form, and the red herring of what its orthographic form would have been remains an unnecessary complication, so it would be preferable to write OSp. losa < *[lawsa] (in phonetic script). In both cases, a phonetic representation is usually what is reconstructed, therefore best presented in square brackets. In addition, the double-asterisked forms whose existence is being denied should also ideally be presented in phonetic script, unless it is specifically a written form whose existence is being denied. Whereupon the remaining uses of the single asterisk would be for the phonetic form of totally unattested items. Such a restricted use of the device would give it the advantages it at present does not have, of having clear reference and of neither representing nor precipitating conceptual confusion.

## APPENDIX A

The works investigated here (in chronological order), and the abbreviations used in Appendix B, are as follows:

En  Entwistle, William J., 1936. The Spanish Language (London: Faber).

Co  Corominas, Juan, 1961. Breve diccionario etimológico de la lengua castellana (Madrid: Gredos), words beginning with M..

MP  Menéndez Pidal, Ramón, 1972. Orígenes del español, 7th ed. (Madrid: Espasa-Calpe).

H1  Hall, Robert A. Jr., 1974. External History of the Romance Languages (New York: Elsevier).

R   Rickard, Peter, 1974. A History of the French Language (London: Hutchinson).

El  Elcock, W. D., 1975. The Romance Languages, 2nd ed. (London: Faber).

Mc  Macpherson, Ian, 1975. Spanish Phonology (Manchester: UP).

H2  Hall, Robert A. Jr, 1976. Proto-Romance Phonology (New York: Elsevier).

W   Wright, Roger, 1982. Late Latin and Early Romance (Liverpool: Cairns).

H3  Hall, Robert A. Jr, 1983. Proto-Romance Morphology (Amsterdam: Benjamins).

DD  De Dardel, R., 1983. Esquisse structurale des subordonnants conjonctionnels en roman commun (Geneva: Droz).

Dw  Dworkin, Steven N.,1985. Etymology and Derivational Morphology: the Genesis of Old Spanish Denominal Adjectives in -IDO (Tübingen: Niemeyer).

Ph  Pharies, David, 1986. Structure and Analogy in the Playful Lexicon of Spanish (Tübingen: Niemeyer).

L1  Lloyd, Paul M., 1987. From Latin to Spanish, I (Philadelphia: American Philosophical Society).

Me  Meier, Harri, 1987. 'Nuevas anotaciones al Diccionario Etimológico de Corominas/Pascual', Verba, 14: 5-74.

CA  Cano Aguilar, Rafael,1988. El español a través de los tiempos (Madrid: Arco).

A   Ariza Viguera, Manuel, 1989. Manual de fonología histórica del español (Madrid: Síntesis).

HN  Harris-Northall, Raymond, 1990. Weakening Processes in the History of Spanish Consonants (London: Routledge).

Ma  Maiden, Martin, 1991 'On the Phonological Vulnerability of Complex Paradigms: Beyond Analogy in Italo- and Ibero-Romance', RPh, 44: 284-305.

S   Stengaard, Birte, 1991. Vida y muerte de un campo semántico (Tübingen: Niemeyer).

Pn  Pensado, Carmen, 1991a. 'How was Leonese Vulgar Latin read?', in Roger Wright, ed., Latin and the Romance Languages in the Early Middle Ages (London: Routledge), pp. 190-204.

___, 1991b. 'Un reanálisis de la "l leonesa"', in Raymond Harris-Northall & Thomas D. Cravens, eds, Linguistic Studies in Medieval Spanish (Madison: Hispanic Seminary of Medieval Studies), pp.63-88.

Py  Penny, Ralph, 1991. A History of the Spanish Language (Cambridge: UP).

APPENDIX B

## 1 *CAPITALS

| Postulation of: | En | Co | MP | HI | R | El | Md | H2 | W | H3 | DD | Dw | Ph | LI | Me | CA | A | HN | Ma | S | Ph | Pv |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Unattested Lexeme | 5 | 7 | 4 | | 3 | 1 | | 3 | | | | | | 13 | | | | | | | 1 | 2 |
| Unattested sequence of attested lexemes | | 3 | | | 2 | | | | | | | | | | | | | | | | | |
| Unattested combination of attested morphemes | 12 | 35 | 32 | | | | | | | | | 22 | | 105 | 183 | 32 | | 28 | | | 12 | 60 |
| Different morphological analysis of attested form | 2 | 1 | | | 20 | | | | | | | | | | | | | | | | | |
| Intermediate form (between 'Latin' and 'Romance') | 3 | 3 | 12 | | 65 | 15 | | | | | 1 | | | 22 | 9 | | 5 | 6 | | | | |
| Ancient Latin form | 1 | 1 | | 2 | | | | | | | | | | 8 | 1 | | | | | | | |
| Ancient Basque form | 1 | 1 | | 1 | | | | | | | | | | | | | | | | | | |
| Pre-Roman form | 1 | | | 1 | | | | | | | | | | 1 | | | | | | | | |
| Indo-European form | 5 | 1 | | 3 | | 1 | | | | | | | | 2 | 1 | | | | | | | |
| Old Germanic form | 2 | 5 | | | 23 | 1 | | | | | | | | | | 425 | | | | | | |
| Old Celtic form | 1 | | | | | 1 | | | | | | | 6 | 3 | 29 | | 5 | | | 6 | | |
| Greek form | | | | 2 | | | | | | | | | | | 2 | | | | | | | 1 |
| Romance form | 1 | | | | | | | | | | | | | | | | | | | | | 1 |
| Denial of existence of: | | | | | | | | | | | | | | | | | | | | | | |
| Latin form | En | Co | MP | HI | R | El | Md | H2 | W | H3 | DD | Dw | Ph | LI | Me | CA | A | HN | Ma | S | Ph | Pv |
| | | | | | | | | | | | | 9 | | | | | | | | | 1 | |
| Unattested sequence of attested words | | 2 | | | | | | | | | | 1 | | | 3 | | | | | 2 | | 3 |

## 2 *Italics

| Postulation of: | En | Co | MP | HI | R | El | Md | H2 | W | H3 | DD | Dw | Ph | LI | Me | CA | A | HN | Ma | S | Ph | Pv |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Unattested lexeme | 2 | | 1 | | 1 | | | | | | | | | | | | | | | | | |
| Unattested combination of attested morphemes | | | 1 | | 15 | | | | | | | | | | | | | 19 | | | | |
| Different morphological analysis of attested form | | 1 | | | 2 | 3 | | | | | | 1 | | | | | | | | 2 | | 3 |
| Intermediate form (between 'Latin' and 'Romance') | | | | | 11 | | | | | | | | | | 3 | | | | | | | |
| 'Mediterranean' form | 2 | | | 6 | | | | | | | | | | | | | | | | | | |
| Germanic form | 7 | | | | 1 | 7 | | | | | | | | | | | | | 1 | | | |
| Celtic form | 2 | | | | 17 | 3 | | | | | | | | | | | | | | | | |
| Arabic form | | | | | | | 1 | | | | | | | | | | | | | | | |
| Basque form | 2 | 3 | | | | | | | | | | | | | | | | | | | | |
| Unattested Medieval Romance form | 5 | 18 | 21 | | 14 | 2 | | 1 | | | 87 | 39 | 11 | 18 | 6 | | | | 5 | 2 | 20 |
| Denial of existence of: Medieval Romance form | 6 | | 10 | | 2 | 6 | 4 | | 4 | 2 | | 22 | 14 | 38 | 1 | 4 | 1 | | | | 8 |
| Modern form | | | | | 1 | | | | | | | | | | | | | | | 2 | 1 | 2 |
| Modern sequence of forms | | | | 1 | | | | | | | | | 5 | | 12 | | | 3 | | | 2 |

Modern meaning of a form

| | En | Co | MP | HI | R | El | Md | H2 | W | H3 | DD | Dw | Ph | LI | Me | CA | A | HN | Ma | S | Ph | Pv |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | | | | | | | | | | | | | | 1 | |

## 3 **italics

Denial of existence of Medieval Romance form

## 4 */ / (or /*/)

| Postulation of: | En | Co | MP | HI | R | El | Md | H2 | W | H3 | DD | Dw | Ph | LI | Me | CA | A | HN | Ma | S | Ph | Pv |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Old Latin form | | | | | | | 3 | | | | | | | | | | | | | | | 4 |
| Proto-Indo-European form | | | | | | | | | | 14 | | | | | | | | | | | | |
| Germanic form | 5½ | | | | | | | | | | | | | | | | | | | | | |
| Norse form | | | | | | | 1 | | | | | | | | | | | | | | | |
| Proto-Italic form | | | | | 3 | | | | | | | | | | | | | | | | | 3 |
| Proto-Romance form | | | | | | | 3 | | | | | | | | | | | | | | | |
| Medieval Romance form | | | | | | 11 | 6 | | | | | | 3 | | | | | | | | | 16 |
| Conceivable Medieval Romance form | | | | | | 4 | | | | | | 2 | | | | | | | | | | 3 |
| Denial of existence of: | En | Co | MP | HI | R | El | Md | H2 | W | H3 | DD | Dw | Ph | LI | Me | CA | A | HN | Ma | S | Ph | Pv |
| Proto-Romance form | | | | | | | | | | | 2 | | | | | | | | | | | 2 |
| Medieval Romance form | | | | | | | 1 | | | | | | | | | | | | | | 1 |
| Modern form | | 1 | | | | | | | | | 3 | | 1 | | | | | | | | 1 |

## 5 *[ ] (or [* ])

| Postulation of: | En | Co | MP | HI | R | El | Md | H2 | W | H3 | DD | Dw | Ph | LI | Me | CA | A | HN | Ma | S | Ph | Pv |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Latin form | | | | | 1 | | | | | | | 1 | | | | | | | | | 6 |
| Medieval Romance form | | | | | 15 | | | | | | | 1 | 6 | | | | | | | | 6 |
| Modern form | | | | | 1 | | | | | | | | | | | | | | | | 100 |

## 6 *phonetic script (without brackets)

| Proto-Romance form | En | Co | MP | HI | R | El | Md | H2 | W | H3 | DD | Dw | Ph | LI | Me | CA | A | HN | Ma | S | Ph | Pv |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Postulated proto-form | | | | | 1 | | | | | | | | 1 | | | | | | | | 2 | 3 |
| Denial of existence of form | | | | | | | | | | | | | | | | | | | | | | 3 |

## 7 *bold phonetic script

| Medieval Romance form | En | Co | MP | HI | R | El | Md | H2 | W | H3 | DD | Dw | Ph | LI | Me | CA | A | HN | Ma | S | Ph | Pv |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Denial of existence of form | | | | | 3 | | | | | | | | | | 7 | | | | | | 7 |

## 8 *{ }

| Postulation of: | En | Co | MP | HI | R | El | Md | H2 | W | H3 | DD | Dw | Ph | LI | Me | CA | A | HN | Ma | S | Ph | Pv |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Medieval Romance form | | | | | 1 | | | | | | | | | | | 6 | | | | | 6 |

## 9 *ordinary type

| Postulation of: | En | Co | MP | HI | R | El | Md | H2 | W | H3 | DD | Dw | Ph | LI | Me | CA | A | HN | Ma | S | Ph | Pv |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Proto-Romance form | | | | | | | | | | | | | | | | | 1 | 1 | | | |
| Denials | | | | | | | | | 2 | | | | | | | | 1 | 2 | | | |

## 10 *'

(Assorted uses by Ariza)

## 11 *

Denial of a meaning

| | | | | | | | | | | | | | | | | | 1 | 7 | | 1 | |

* for footnotes, etc.

| 17 | 17 | 4 | 3 | 18 | 15 |

## 1994 POSTSCRIPT.

This paper was published in the *Journal of Hispanic Research*, 1, 1992, 1-16, as the journal's inaugural paper. It is reproduced here by permission of Impart Publishing Ltd, to whom we are particularly indebted for permission to reprint the final appendix neat. That published paper was a slightly expanded version of one delivered (under the title "~~") to the annual Romance Linguistics Seminar in Trinity Hall, Cambridge University, in January 1992, and later in Spanish to audiences at the Universities of Valencia (May 1992) and Oviedo (January 1993). Despite the presence in some of these audiences of practitioners referred to in the paper, it met with general approval. Two of the linguists investigated have gently pointed out that I slightly misrepresent them, and I am glad to apologize here: Robert De Dardel announces explicitly on p.31 that "Dans le texte et dans le tableau, les formes du latin et du roman commun sont données en capitales (QUANDO), sans astérisque pour les formes reconstruites (BENE-KE)," deliberately wishing not to distinguish typographically between the attested and the unattested; and Birte Stengaard indeed makes it clear in connection with her use of *ESSERE that it is only being postulated thereby as hypothetical within the Iberian Peninsula. It is probably too much to expect that the advice contained in this paper will be generally followed, but it has certainly sharpened up my own usage.

# The Study of Semantic Change in Early Romance (Late Latin)

The history of the development and divergence of the Romance languages out of the Latin of the Roman Empire is the most fully documented case of language change that we possess. This ought to mean that the development of the discipline of historical linguistics is closely interwoven with the theoretical and practical advances made by Romance philologists; that the proponents of theories of language change which are meant to have universal validity check with care the attested details of the development of Latin into the various Romance languages; and, conversely, that Romance philologists take account of the accepted discoveries of historical linguistics when interpreting their evidence. But some of them have been happy to work with assumptions that seem absurd to a general historical linguist. For example, specialists in Late Latin texts, that is, those written between the end of the Roman Empire and the Twelfth-Century Renaissance, once tended to regard the language of those texts as some kind of direct transcription of speech. It was partly a reaction to the weaknesses in this approach that encouraged the application to the Early Romance languages of reconstruction techniques originally designed for languages with no documentation at all. By working backwards from the Romance languages, scholars created a hypothetical entity called Proto-Romance, postulated to have existed at the same time as surviving Latin texts were written but not to be identified with Latin. This did lead to occasional illuminating reexaminations of the textual evidence, as in the brilliant article by Politzer (1961), but some scholars never looked at the texts, in effect priding themselves on not taking into account the copiously available evidence that is

the envy of those who study other language families. Specialists in Bantu or Athapaskan languages, for example, would be delighted to have a seventh-century text, wouldn't they? At the very least, they wouldn't ignore it.

These barriers are not impermeable. At our conference at Stanford in 1979 I pointed out that lexical diffusion theory demolished the phonetic evidence that had bolstered the view that Late Latin and Early Romance were totally separate languages. Esa Itkonen in Helsinki, Michael Richter in Dublin, Marc van Uytfanghe in Ghent and others have been working simultaneously on Romance historical linguistics and Late Latin texts. Romance historical linguists quite often now consider together Late Latin and subsequent Early Romance documentary evidence without erecting an artificial barrier between the two; e.g., among others, Timo Riiho's work on Spanish prepositions, Mario Saltarelli's paper on syntactic diffusion at the Stanford conference, Dieter Wanner's paper to this conference and Suzanne Fleischman's book on the future. But there are still scholars who ignore the Late Latin evidence entirely; for example, Saussol wrote an interesting book on the use of the Spanish copulative verbs *estar* and *ser* in the *Poema de Mío Cid*, and then included in the book's title the phrase "Origins of their functions," as if nobody in Spain had talked at all until the twelfth century, or as if their Latin etyma STARE and SEDERE had never existed.

The Romance Reconstructionists are of course right to be wary of some of the Mediaeval Latin texts: texts written in England, Ireland or Germany, because these were probably written by native speakers of languages other than Romance. In addition, texts written after the ninth-century educational reforms in France and much of Italy, or after 1080 in most of Spain, are likely to be written by authors who did regard Latin as a conceptually entirely distinct language from their own spoken Romance; but texts from before those dates in Romance-speaking areas were probably written by speakers of Old Romance, who were using the only written mode they knew for recording their language (cp. Wright, 1982). This was not a direct transcription of speech any more than modern written French or English is a direct phonetic transcription of spoken French or English. This lack of direct correspondence means that some

phonetic and morphological reconstruction is necessary, but it is even in these fields a good idea to look at texts and see if genuinely attested spellings appear in a chronological order compatible with the order required by the simplicity metric. And in lexical studies, the appearance of a new word in written form is precisely the evidence we need (as when Moralejo discovered the source of Spanish *jerigonza* written in a seventh-century text as *Ihericuntina lingua*).

In many respects the language of Late Latin texts seems to us archaic. But there is no reason to assume that the apparent coexistence of an attested, but archaic, usage in a Late Latin text with a convincingly reconstructable spoken form in the coetaneous Proto-Romance implies necessarily that the older usage did not also exist in the speech of the time. A new linguistic phenomenon can arise long before the eventual disappearance of the old form which it seems in retrospect to have neatly replaced; the author of the text, rather than indulging in antiquarianism in using an older form, might merely have been favoring one competing variant over another, the favored one being one which we know now, but couldn't have known then, was destined to die out. I've argued before [Wright, 1983] that new and old uses could regularly coexist in Early Romance until the eighth century or so. This comment is pertinent to morphological and syntactic change in Romance because, for example, of the probable lengthy coexistence of the use of old and new methods of rendering the passive (recently studied by Codoñer), it is pertinent to phonetic change, since for stylistic and sociolinguistic reasons old pronunciations are under no compulsion to die out merely because new ones have come in; and it is particularly pertinent to semantic change, since words regularly acquire new meanings without losing the old.

In general terms, the survival of the old despite the arrival of the new is not without interest. For another perspective which has shifted over the last thirty years in historical linguistics concerns the theoretical status of features that have not changed. Once upon a time it was presumed in a vague sort of way that if it were left to itself a language wouldn't change at all, so that the only interesting objects of study were taken to be those aspects which indeed had changed. More recently, people have realized that it is natural for all

languages to evolve, so that what has not changed is potentially significant. Hence the recent interest taken by historical linguists in Classical Arabic and Hebrew, Katharevusa Greek, Post-Carolingian Mediæval Latin, etc, as artificially maintained invariant systems. Hence also the realization that sounds that have not changed might have as much theoretical significance as those which have, with the resulting elaboration of theories of consonantal strength hierarchies (as in Craven's excellent paper to this conference); hence also the interest shown by lexical diffusion theorists in words that have not undergone phonetic changes as well as in those that have. This too applies to semantic change. For example, the fact that the eye is the only part of the face whose normal lexical item has not been involved in a semantic shift in the history of the Romance languages is of considerable interest, and could be interpreted as indicating that the eye has semantically real psychological boundaries in a way that other parts of the face have not. For in Spanish, for example, the word for "mouth" used to mean "cheek" (boca); the word for "chin" used to mean "little beard" (barbilla), the word for "eyebrow" used to mean "eyelids" (ceja), the word for "nose" used to mean "nostril" (nariz/ces), and the word for "cheek" used to mean "jawbone" (mejilla), but the word for eye has not changed (ojo). And now, at last, I'll turn to semantic change directly.

Semantic change in lexical items is often intelligently discussed by etymologists whose primary focus is on the phonetic: e.g. Malkiel's paper to this conference; but in general it is an underdeveloped branch of historical linguistic theory. This is not because it is in itself a small subject. Countless lexical items have acquired new meanings. Yet historical linguistic textbooks, while occasionally mentioning it, rarely even give it a separate chapter (and even these conferences average only about one paper per conference on the topic). Writers on semantic theory, at least within the European tradition, have been more likely to discuss change; Ullmann, Vilches Acuña, Waldron and Baldinger, for example; but specifically diachronic semantic principles seem not to have been elaborated in the last twenty years. Perhaps this is because of Gilliéron's famous motto, which is certainly true, that each word has its own history. Since historical linguistics, like other branches of linguistics, has recently been dominated by believers in universals, the generally

accepted presumption that there are no universals in semantic change may have led some linguists to assume that it is of no interest. This is to put the cart before the horse. There can be no objection to looking for universals, but there are strong objections to the assumption that if universals are not found the whole investigation must therefore be pointless. Roger Lass's realization that in diachronic linguistics there are no rigorously predictive principles was a great step forward; no matter that many European linguists have always taken this for granted anyway, it is an advance for an instinctive seeker after clear universals to admit that there may not be any. In diachronic semantics, no scholar has seriously been tempted to claim that there are.

How and why does the modern Spanish word verdugo, "executioner," come from a mediæval word meaning "sapling"? Schulte-Herbruggen's article will tell you how. It is hardly a predictable development. Nobody wants to predict that the English word sapling will mean "executioner" in 700 years time. But the change is explicable if we understand the difference between sense and reference, if we know a little semantic theory, and, above all, if we look at the textual evidence. In that article the intermediate stages of the semantic change are helpfully documented in chronologically intermediate texts. I am not now going to tell you what those stages are, because my point is that it is not easy to guess, simply by using general techniques of reconstruction, but that it is easy to follow if we look at the texts. My point today is that we can do the same for semantic changes known to have taken place in the Late Latin - Early Romance period.

López Pereira, of the department of Latin at the University of Santiago de Compostela, has recently explained a baffling semantic fact of Mediæval Spanish and Catalan by looking at a Late Latin text: Old Spanish and Catalan civil undoubtedly come from Latin CIVILEM, "civil," but why should they also be able to mean "cruel"? In the Crónica Mozárabe of 754 A.D. López Pereira found attested civiliter meaning "in civil war"; a ruler who behaved civiliter was treating his opponents cruelly. No amount of guesswork and reconstruction could have established that as the route of change, but because the Santiago Latin department is building up a vast

filing system of attestations in seventh and eighth century texts, López Pereira could.

The development of Latin MAXILLA "jawbone" into Spanish *mejilla* "cheek" is clarified for us by Isidore of Seville, who, in his seventh-century *Etymologiae*, regards the two *maxillae* as still being bones, but distinguished from each other as left and right, as cheeks are, rather than as upper and lower, as jawbones are; he says there are four canine teeth in the human mouth, two in the right *maxilla* and two in the left *maxilla* (*sequentes canini vocancur, quorum duo in dextra maxilla et duo in sinistra sunt*). Semantic changes in Spanish words for parts of the face are discussed at length in Wright (1985), and one of the interesting things to come out of that investigation was that even a writer as hyper-literate as Isidore of Seville could use his words this way with their developed meanings, often—as in this case—of a kind that we might be able tentatively to guess at, but without our otherwise being able to reconstruct the chronology or the precise semantic nature of the intermediate stages. As it is, we can see from this sentence that the ninety-degree shift in the perception of *maxilla*'s sphere of reference happened before that sphere of reference moved away from the bones outwards to the flesh.

If we can find semantic development in Romance attested even in the vocabulary of the scholar Isidore, does this mean that a word used by Isidore without the reconstructable developed meaning implies conversely that the semantic change in question has not yet happened in Spanish Romance? In phonetics and morphology, because all the scribes were taught from the *Artes Grammaticae* to follow the old norms, the presence of such unevolved forms in the texts of these times tells us little about pronunciation or spoken morphology. Semantics, however, was not covered by these *Artes*; yet Isidore had read many ancient works, and it is possible that he used words with an archaic sense more often than most of his contemporaries. That is a professional hazard of his being a lexicographer. But it is also true that in general semantic changes seem to take longer to complete than phonetic ones, particularly in that the old meaning can often coexist happily with the new meaning, and fail to drop out at all. So in such cases Isidore may be just favoring one meaning current in the community rather than another. Even

so it is sometimes possible to deduce that a development cannot have got far yet, if we find a sentence where the new meaning, if it was available to the contemporary reader, would have scuttified the sense of the sentence. For example, Spanish *colgar*, "to hang," comes from CONIOCARE, "to place." When we find in the seventh-century prayer-book the request that we be placed by God's right hand on the Day of Judgment (*ut... a dextris tuis tempore iudici conlocemur: Oracional*), we can provisionally assume that the liturgists did not consider it possible for the ordinary Christian to misunderstand that as being "hung by your right hand." Since other contemporary uses of this verb all seem not to mean "hang," we can for the moment date that development as starting later than the seventh century. This is what a reconstructionist would probably guess, from the fact that Catalan and Occitan *colgar*, and French *coucher*, from the same etymon, have specialized in a different way to mean "put to bed"; but it is comforting to have some documentary support.

More examples of useful attestations. Spanish *madera* and Portuguese *madeira*, "wood," come from Latin MATERIA; to follow this shift it is helpful to see Isidore say (XIX.19.4) that all pieces of wood which are destined to be made into artifacts are known as *materia* (*materia inde dicitur omne lignum quod ex ea aliquid efficiatur*). Spanish *llegar* "to arrive" derives from Latin PLICARE "to fold." It was once suggested that this development arose from ships folding their sails. In fact, the oldest Spanish vernacular texts also have *allegar*, from ADPLICARE, and the seventh-century evidence shows *adplicare* used with the meaning of "arrive"; e.g. the Visigothic Laws 8.2.3, *qui in itinere constitutus in cuiuscumque forsitam campo adplicaverit* (whoever is going on a journey and comes into someone's lands); the older meaning of "put together" is also attested. Since there seem to be no uses of *plicare* in these texts, the evidence supports the view put forward by Corominas that the direct development of these words was that of *allegar* from ADPLICARE, which could in Imperial Roman times be used of ships coming into dock (not of folding sails); we could add now that the subsequent back-formation of *llegar* is not surprising, since many Old Spanish verbs have two forms, one with a prefix *a-* and one without. Corominas (1980) quotes Classical uses, the early fifth-

century *Peregrinatio Ætheriæ,* and eleventh-century Mediæval Latin uses, but the seventh-century ones would have helped him consider-ably.

Spanish *querer,* "want" and "love," from Latin QUÆRERE, "seek," is copiously attested meaning "love" in the seventh-century Visigothic prayerbook, where the congregation are described simply as *te quærentes* (as opposed to those who hate the Lord: e.g. *Ora-cional* 765, Ne obliviscaris, Domine, vocem querentium te, ut superbia eorum qui te oderint ..); it can also mean "seek" (1117: et querere et invenire), or both at once. Spanish *domingo* "Sunday," comes from the adjective DOMINICUS: in seventh-century texts this can still be a general adjective connected with any lord, any *domi-nus,* but there are also six uses as a noun meaning "Sunday" in the Visigothic prayerbook, and eight in the full phrase *die dominico;* we can suggest that the new nominal use had arrived without implying that the old more general and adjectival use had dropped from speech yet. We also find *feria* used for "weekday," as in Modern Portuguese *feira,* developed away from the original "holiday." Other new Christian meanings include *missa* as a noun meaning "mass" in *Oracional* 550; *sermo* meaning "sermon" as well as the older and less specialized "speech"; *communicare* and *communio* with both new specialized and older general meanings; whereas there is no sign yet, for example, of the development of *incensum* from a participle meaning "burnt" to a noun meaning "incense." It has also proved possible to use this documentary evidence (Wright, 1992a) to argue that a word often thought to have undergone a semantic change did not in fact do so; Mediæval Spanish *ladino,* meaning "vernacular," comes from LATINVM; but rather than changing from meaning "Latin" to meaning "Romance," this lexical item always meant "vernacular," in the Roman Empire, in Isidore, in the ninth-century Christian writers of Moslem Spain, and in Old Spanish.

Scholars have been looking at the usage of Isidore since his own lifetime, of course: the work of Sofer and Fontaine, for example, is essential. All I want to do today is to point out that both general and Romance students of diachronic semantics have got documenta-tion available for a long period, documentation which they tend to ignore but which might well be more helpful than reconstruction. Rothwell has pointed out that discussing semantic change in Old

French from a basis of theoretical reconstruction has often led to published analyses that can be simply disproved by actually looking at Old French texts; in essence I want to move that argument five centuries earlier.

*1994 POSTSCRIPT.*

This paper was printed in the *Papers from the 7th Interna-tional Conference on Historical Linguistics,* John Benjamins Pub-lishing Company, Amsterdam / Philadelphia, 1987, pp. 619- 28; it is an exact reprint of the lecture given at that conference in Pavia, Italy, on Thursday, September 12, 1985. It has aroused no interest at all, in part because semantic change is still an undeveloped area of Romance linguistics. As regards the present paper, the only point that should be clarified is that the quotations from the Visigothic texts are not meant to be taken as necessarily the first attestations of the developed meaning, merely as illuminating data; it is, of course, entirely possible for a change to have begun at an earlier date than that.

# 7
# Indistinctive Features
## (Facial and Semantic)

FILLMORE (1975) MADE A neat distinction between two kinds of semantic theory, those involving the "checklist" and the "frame." The former has been the type most commonly used by generative linguists, who like to represent necessary conditions for the appropriate use of a lexical item as a combination of semantic units within square brackets; this technique has been at times elaborated into networks of considerable sophistication (e.g., by Kay and Samuels, 1975). The "frame" idea, on the other hand, suggests that a language has semantic structures which impose an artificial framework onto intrinsically undifferentiated aspects of human experience. This approach (which is traditional in Europe) does not construct groupings of logically necessary conditions that have to be met for an item to be used appropriately; instead, Fillmore suggests, there is at most an exemplary "prototype" in the center of each part of the frame. Other scholars following Fillmore, including Verschueren (1981) and Coleman and Kay (1981), have developed this notion, applying the word "prototype" to features which are "more or less" rather than plus or minus; Comrie (1981:101) even used it for his definition of the linguistic notion of subject. In cases where the outside world offers no clear boundaries, it may indeed be artificial for linguists to postulate them in semantic structure, and the data to be examined here also suggest that in such cases a frame may be all there is.

This note concerns semantic changes in Spanish words for parts of the face. The face seems at first sight to be composed of a number of distinguishable parts; the chin, jaws, mouth, cheeks, nose, eyes, eyelids, eyelashes, eyebrows, ears, temples, and forehead, say. Such

entities might plausibly be thought to correspond to acceptably distinctive components in a "checklist" theory of semantics; if the cow can be allotted the unit [+ BOVINE] on its list of logically necessary conditions for use, the word mouth ought to be able to require the unit [+ORAL], or something similar. We would no more predict that the word for e.g. 'cheek' would change its meaning to 'eyelash' than that the word for 'cow' would change its meaning to 'hornet.' The theory of semantic change in a checklist model finds it easy to describe the acquisition of new [...] material (semantic specialization), or its loss (semantic generalization), but it is uneasy about the replacement of such material with some apparently incompatible relative; in practice, attested changes are rarely described by these theoreticians (but see the admirable study by Werth [1973] for one of the few attempts to categorize in such a manner changes that have actually occurred, in this case using the trees of generative semantics). So what should we think of the Rumanians, who indeed have developed Latin GENA 'cheek,' 'cheekbone' into geană [dʒanə] 'eyelash'?

Such a development must surely be regarded as unpredictable. It would not be surprising to discover that this change has never happened in any other language. But one of the lessons a student of semantic change has to grasp is that explanations cannot be expected to be predictive. Lass (1980), for one, has correctly diagnosed that the study of language change is unable to follow the strict principles of natural sciences (as many European scholars have been pointing out for decades). Lass finds this conclusion depressing. Other historical linguists (e.g., Samuels 1973) seem not to expect predictability. Many people would find it exhilarating to see proof that speakers are not in the grip of inexorable logical necessities; one of the ways in which human unpredictability in linguistic activity can be demonstrated is through the presentation of data that show semantic boundaries to be porous. The meanings of words for parts of the face are so often altered in the history of Romance that it suggests that perhaps many of them are never rigid at all, and certainly not logical necessities; words for one part of the face can with apparent ease come to be used for another. This fact suit the 'frame' approaches based on Gestalt theory and field analysis

(e.g., Baldinger 1980, Ullmann 1962; some of the data here discussed were presented in Zauner 1903:355-433).

Latin had three words for 'cheek,' that is, 'the side of the face below the eye.'[1] None of their descendants mean 'cheek' in Romance. BUCCA was borrowed into other languages from Latin with the meaning 'cheek' (e.g., Welsh *boch*, Greek βούκκα, Breton *boch* and Berber *abeqqa*: Corominas and Pascual 1980-, 1860:I), but its Romance descendants do not have that meaning. In Western Romance they mean 'mouth' (It. *bocca*, Sp. *boca*, Fr. *bouche*, etc.; even the earliest attestations in Western Romance are indisputedly 'mouth'—see e.g., Smith 1977:255-61). In Rumanian *bucă* means 'buttock.' Apart from *geană* mentioned above, GENA and MALA do not survive; these disappearances may be partly due to formal similarity with a number of other words. GENA (ultimately related to English *chin*, Germanic *kinn* 'cheek,' and Sanskrit *hanu* 'jaw') may in addition have been the victim of excessive ambiguity, apparently being usable for any area between the eyelid and the chin (inclusive). The usual Latin word for 'mouth' was ŌS, ŌRIS; but the meaning of this word was also capable of spreading out beyond its boundaries. Even in ca. 200 B.C. Plautus used the phrase *ore rubicundo* to mean 'red face' (not 'red lips'):

    rufus quidam, ventriosus, crassis suris, subniger,
    magno capite, acutis oculis, ore rubicundo, admodum
    magnis pedibus. (*Pseudolus* 4.7.115-17)

It seems that OS could as easily be used to mean 'face' as 'mouth.' Ambiguity may not have often arisen, but if it did it is understandable, if not predictable, that in some circumstances BUCCA might have been considered an acceptable alternative, inasmuch as the inside of the cheek is indeed part of the mouth. OS might also have been gradually affected by impending homonymy with several words; in particular, once the neuter gender fell out of fashion, the accusative would be likely to be *OREM in speech, and

---

[1] This is the definition given in the indispensable *The Advanced Learner's Dictionary of Current English* (Oxford, 1963).

in every case except the awkwardly imparisyllabic nominative the form might have appeared unacceptably close to AUREM 'ear.' As Ullmann (1957:293-94) argued, change through "contiguity" of sense (as has happened to BUCCA) is a means of change not an explanation, for the initial impulse often arises in some perceived deficiency in the previous lexical unit. [aw] > [o] is probably quite a late change, but even before it occurred there were forms of speech in Italy that had [o] where Latin had [aw]—as in the celebrated case of *plaustrum/plostrum*—and in words with close semantic relations even near-homonymy can inspire therapeutic reactions, as Malkiel (1979) suggested has happened to Sp. *hasta* and *hacia*. Something similar is happening with the English homonyms *oral* and *aural*, both normally with [ɔ:]; it seems that in cases where the distinction is important, the form [ɔ:rəl] is usually understood to be *oral*, and the word *aural* has either to be replaced (by e.g., *auditory*) or given an ad-hoc spelling-pronunciation [awrəl]. The combination of occasional homonymy, potential ambiguity, and the availability in many cases of an intelligible substitute in BUCCA for 'mouth,' might suffice as a reasonable explanation for the loss of OS and the downward shift of the southern part of the frame for BUCCA.

The definitive adoption of BUCCA for 'mouth' probably postdates the end of the empire and predates the end of the general Old Western Romance community, in view of the Welsh and Rumanian evidence. This is also likely to be the period of the surprising adoption of a Greek word to mean 'face'; Spanish and Occitan *cara*, and Fr. *chère* were borrowed from (probably Byzantine) Greek κάρα. Corominas and Pascual (1980-, 1:839-42) cautiously accept this etymon; a Byzantine dating helps allay some of their doubts. This borrowing is intrinsically unpredictable enough to suggest that there was some kind of lexical uneasiness or deficiency that it was able to relieve (cp. Goddard 1980). If the definitive shift in BUCCA and the adoption of κάρα can be seen as roughly contemporary 6th-c. developments, perhaps the problems of OS can be seen as a contributory factor to them both. Other words for 'face' seem not to have been acceptable; VULTUS has not survived in Spain, although it has in Italian *volto* (*bulto* being a 15th-c. Latinism in Spanish, meaning originally 'image'). FACIES might have been felt even less unambiguous than OS for 'face' in view of the width of its semantic

range,' 'shape,' 'form,' 'appearance,' 'character,' 'surface,' as well as 'face'). The availability of BUCCA for the 'inside of the cheek,' i.e., 'mouth,' is occasionally attested from Classical times: a cliché *quod in buccam venerit*, meaning originally 'whatever comes into your mouth,' even developed to apply to writing as well, as in *quod in buccam venerit scribito* (Cicero, *Atticus* 1.12.4).

The diminutive form of AURIS 'ear,' AURICULA, is attested as a lexicalized form with no apparent diminutive meaning from Classical times. The suffixed form becomes the normal word for 'ear' in most Romance languages (e.g., Sp. *oreja*, Ptg. *orelha*, It. *orecchio*, Fr. *oreille*). Probably this is partly because that form was unambiguous, but AUREM in speech was often not (with HORAM and AURUM in the lexicon as well as *OREM). In writing the form continued normally to be AURIS, unaffixed. St. Isidore, in early 7th-c. Spain unhelpfully links both AURIS and AURICULA separately with AUDIO, rather than with each other, discussing a supposed relationship between LAURUS 'laurel' and LAUDES 'praise,' he adds in passing *ut in auriculis, quae initio auricula dicta sunt* (*Etymologiae* 17.7.2; Lindsay 1962); elsewhere (11.1.46) *aures* is said to be a corrupted form of *audes*. *Auriculae* are not mentioned in this latter chapter, which discusses parts of the body. The meaning 'mouth' could not have similarly transferred to a diminutive form of OS, since OSCULUM already meant 'kiss' and the rarer OSCILUM meant 'mask.' In Rumanian, where BUCCA did not move in this way, os disappeared nonetheless; the word for 'throat' shifted, GULA >*gură* 'mouth,' with the eventual result that 'throat' acquired a Slavonic neologism in *gît*. This implies that the process was a drag-chain rather than a push-chain; i.e., that the impending loss of OS led to the shift in BUCCA or GULA respectively, rather than vice versa.

If there were always strict logical necessities for the use of words, such shifts as that in BUCCA would be hard to envisage. But it could be that such distinctive semantic features as words possess tend to have been sharpened factitiously by academicians, lawyers, grammarians, and allied pedants, in their natural state the meanings of many words do not have strict intrinsic borders so much as transition zones, since the edges of the relevant frame can be shifted if the speaker wishes. The 'cheek' is an instructive example of this lability. It seems likely that during the shift in BUCCA from 'cheek'

to 'mouth' there was a period when the word could mean either or both, as if the border had been suspended; it may have been at that time that suffixed forms arose to refer unambiguously to the 'cheek,' like Portuguese *bochecha* and Corsican *buccella* (Bonfante 1951:372). *Buccella* in Latin had meant 'morsel'; the diminutive in the Corsican form is lexicalized out of significance. The Portuguese are able to use *face* (< FACIEM) as well as *bochecha* for 'cheek,' but *face* can still also mean 'face' (or 'front'), so there are no distinctive edges to the cheek there either. In general, though, once BUCCA had moved to 'mouth,' and 'mouth' had come to be considered as its primary meaning (or, in European terminology, the center of the word's semasiological field; cp. Baldinger 1980:289), the meaning of 'cheek' seems to have been left in as unsatisfactory a position as the 'mouth' had been before, having no unambiguous word available.

This indistinctiveness seems to have been the impulse for the extraordinary activity in this semantic area, both neologizing and shifting, that took place in the period before the emergence of naked written vernacular. French (*joue*) Occitan (*gauta*) and Catalan (*galta*) borrowed a Celtic word (*GAUTA), Rumanian a Slavonic word (*obraz*, which can also mean 'face'; Manoliu 1974); the Italian *guancia* is in origin a Germanic word (WANKA), and *ganascia* a Greek word (γνάθος) borrowed from the Byzantines. Italian dialects, in fact, according to the maps in Kahane (1941), have amassed a total of fifteen different etyma for their words for 'cheek.' Old Spanish did not borrow. Early Medieval Spaniards seem to have considered, moving the reflex of TEMPORA, the word for 'temples,' downwards; Dworkin (1982) has pointed out that there is no obvious extralinguistic upper boundary to the cheek, so that the shift from 'temple' to 'cheek' that seems to have been initiated in North-West Spain (e.g., *tenllera* < TEMPORA) and also in Old Castilian (*tienlla*; cp. Alvar 1980) was a simple one to perform. Castilian, however, eventually preferred instead to move the sphere of reference of MAXILLA upwards.

In Italy MAXILLA (the Latin for 'jawbone' has derivatives meaning 'cheek' as well as 'jaw,' but Standard Italian has decided to keep the word *mascella* as 'jaw.' The use of *mascella* in Dante's *Inferno* XXXII 107, for example, can only mean 'jaw,' since they chatter)

quando un altro gridò: "Che hai tu, Bocca?
non ti basta sonar con le mascelle
se tu non latri? qual diavol ti tocca?"

*Bocca* here is the man's name. In Classical Latin the word first meant the lower jaw-bone, but Pliny referred to crocodiles' upper jaws with *maxilla*: *maxillas crocodilus tantum superiores movet* (*Natural History* 11.159), and that distinction between upper and lower may not have applied colloquially. The Rumanians pulled MAXILLA in a different direction, so that *mǎseǎ* now means 'molar tooth' (although this may also involve nominalization, via ellipsis of the adjective MAXILLARE). In the West, however, Late Latin seems to be beginning to spread the domain of MAXILLA upwards and outwards. In the Vulgate, for example, the non-Classical *mandibula* is used to clarify *maxilla* in the following passage where the translator wishes specifically to refer to the 'jawbone of an ass' in Judges 15.15: *inventamque maxillam, id est, mandibulam asini, quæ jacebat, arripiens.* This renders a phrase using one Hebrew word that could mean either; the same bone is referred to four times in the next four verses, twice with each word; the implication of the phrasing is that *maxilla* is insufficiently precise, needing to be clarified with the more technical *mandibula. Maxilla* is used to translate the Greek σιαγών 'jaw,' in Christ's advice in Luke 6.29; *et qui te percutit in maxillam, præbe et alteram.* It may seem from this that Christ was not advising us to "turn the other cheek" but "offer the other jaw," i.e., that if the attacker removes some of your lower teeth you should offer him the upper ones as well. The Vulgate version of St. Matthew (5.39), however, also translating σιαγών, specifies the *dextera maxilla; quis te percusserit in dexteram maxillam tuam, præbe illi et alteram.* Here the *maxillæ* are distinguishable on either side of a vertical line, 'left' and 'right,' 'upper' and 'lower,' the change has already progressed quite a way. Isidore comments (*Etymologiæ* 11.1.44-46):

Malæ sunt eminentes sub oculis partes ad protectionem eorum suppositæ. Vocatæ autem malæ sive quod infra oculos prominent in rotunditatem, quam Græci appellant, sive quod

sint supra maxillas. Maxillæ per diminutionem a malis, sicuti paxillus a palo, taxillus a talo. Mandibulæ sunt maxillarum partes, ex quo et nomen factum.

This seems to be evidence comparable to Kahane's diagnosis (1941) based on Italian dialects: that in 7th-c. Spain *maxilla* covered at least the jaws and the fleshy part of the cheek. Perhaps, since *mala* is likely to have been absent from the vernacular, it covered the cheekbones as well. Elsewhere Isidore refers to the *dextra maxilla* and the *sinistra,* as being the place for two canine teeth each, which suggests that *maxillæ* here are distinguished as in St. Matthew, by a vertical barrier (like 'cheeks' are) rather than by a horizontal one (as 'jaws' are), but still include the jawbones within their reference; *Sequentes canini vocantur, quorum duo in dextra maxilla et duo in sinistra sunt* (11.1.52). The three uses of *maxilla* in the Visigothic liturgy suggest that the word may mean both together there too;[2] they echo the Biblical requirement to offer the other *maxilla* to your attackers, although only the third repeats it lexically. All three are from the Sacramentary (Ferotin 1912):

774    ...dorsum utique ad flagella ponens, maxillas vero ad palmas constituens, faciem quoque tuam a confusione sputorum non retrahens

803    ...Iesu...qui passionis tuæ tempore dorsum ad flagella, maxillas ad palmas ponens, nec contumax, nec contradicens, inventus es

809    Jesu, Deus Noster, cui (=qui) maxillam prebuisti &lt;percus&gt; soribus tuis...

[2] The Visigothic Latin texts are currently being given a manual concordance by the Latin department of the University of Santiago de Compostela; the huge corpus of Laws and Liturgy is completed, and the other texts have begun to follow. Since Spain is the main centre of 7th-century literacy, the concordance is of great help to Medieval Latinists and Romance linguists alike. I am very grateful to the *fichero's* organizer, Manuela Domínguez, for letting me use it; she wishes it to be known that any Romance scholar is welcome to do likewise.

It looks as if the transition period is over, and the jaws have slipped out of the potential reference of MAXILLA, by the 10th century. One of the 10th-c. Silos glossaries published by García de Diego (1933) offers MALAS: *maxillas*; the Silos glossary published by Goetz (1884:116), the unpublished San Millán glossary (Em 46), and the allied glossary in Goetz (1984:83), all offer *Manas, malas, maxillas*; although glossary equivalents can often be inexact, since they derive from the original context of the manuscript gloss, such a bald correspondence suggests that the distinction between *mala* 'upper cheek' and *maxilla* 'lower cheek,' postulated by Isidore, is not widely felt in the 10th c. even if it was in the 7th. The slightly later use of a word transcribed from Arabic script as *maxilla* (Corriente 1980: 185) in the Hispano-Arabic poetry of Ibn Guzmán apparently means 'cheek'; the lower edge of this part of the frame has gradually moved upwards. By the time of the early 13th-c. graphical reforms, which make vernacular usage easier for us to assess (cp. Wright 1982: Chap. 5), the Castilian word seems only able to mean 'cheek.' The pregnant abbess in Berceo's *Milagros de Nuestra Señora* 508b (Dutton 1971) is said to have freckles on her *massiellas*, which cannot possibly be her 'jawbone'; *fueronseli faciendo peccas ennas massiellas*, apparently a symptom of pregnancy. The mother of the converted Jewish boy in *Milagros* 364b is said to have torn her *massiellas* with her nails, and that too cannot refer to the 'jawbone': *tenié con sus oncejas las massiellas rompidas*. In the *Libro de Alexandre*, 50d young Alexander lets his feelings appear in his *maxiella*: *pareçies' la rencura del çuer en la maxiella* (Cañas 1978: 104; Nelson 1979: 162). This use of the item for 'cheeks,' now apparently the norm in Castile (and borrowed with that meaning in Basque *masailla*), is occasionally also visible to us in the Latin written by Medieval Castilians, since before Nebrija (as Rico 1978 points out) it was normal for even well-educated Castilians to write "Latin" as Castilian in Latinizing disguise. In the 14th c., St. Bridget saw a vision of the Virgin-Birth, which was put into Latin by her Spanish confessor Alfonso (formerly Bishop of Jaén); after the baby was born, *quem tunc mater suscepit in manibus et strinxit eum ad pectus suum, et cum maxilla et pectore calefaciebat eum cum magna leticia et tenera compassione materna* (Bergh 1967:7.21.16; cp. Bergh 1981). We cannot imagine that Mary comforted her baby

---

² *Maxilla* here might also be a scribal error for *mamilla*, Latin "nipple." On the use of Medieval Latin as an indicator of vernacular vocabulary, see now the excellent article by Rothwell (1980).

by pressing him to her jawbone: she pressed him to her cheek, her *mejilla*, thinly cloaked with a Latinate form in *maxilla*.² Vernacular Bibles translating Luke 6.29 now suggest that Christ told us to turn the other cheek. (The Authorized English Version of 1611 gives "to him who strikes you on the cheek, offer the other also"; English *cheek* by then means 'cheek,' although *cheek* too in the past had meant 'jawbone').

*Mejilla*, it seems, no longer meant 'jaw' in the 13th c. By this point, the Castilians had decided instead to use for 'jaw' metaphors taken from the construction industry. *Carrillo*, as Dworkin (1982) argued, originally meant 'pulley,' 'block and tackle,' as indeed it still does; and by the 13th century it had been borrowed for originally figurative, and subsequently literal, use as 'jaw' (presumably comparing the pulley to the joint of the lower jaw). Corominas (1954-57; *s.v.*) and Alarcos (1973) both commented on this meaning as attested in Alfonso X's *General Estoria* 1:222: *en el Nilo a una bestia que llaman cocadriz...quando come muerde, e non mueve el carriello de yuso si non el de suso solo...en amos los carriellos a muchos dientes e muy fuertes*. The reference to upper and lower *carriellos* assures us that it meant 'jaws' (cp. the comment in Pliny 11.159 above and Spitzer 1924). This is also the probable meaning in the *Libro de Alexandre* 469c (Cañas 1978: 151; Nelson 1979:264), where Hector tells Paris that in war it is no good looking pretty, you need strength and determination:

Paris is being told, with *carrillos denodados*, to grit his teeth and take the fighting seriously:

Non se faz la fazienda por cabellos peinados
nin por ojos fermosos nin çapatos dorados;
mester ha puños duros, carrillos denodados,
ca lanças nin espadas non saben de falagos.

It was no good. *Carrillo* could not remain within its boundaries; it soon changed to mean 'cheek.' In Nebrija's Spanish-Latin dictionary of ca. 1495 (Macdonald 1973), Spanish *carrillo de cara* meant 'jaw' despite its wanderings. *Carrillo* (*como rodaja*), 'pulley-wheel,' is a separate entry, translated as Latin *vertebra* (which, to add to the fun, meant 'joint' in Classical Latin rather than 'vertebra'), although *rodaja* is rendered as *trochlea*. Nebrija translates *mexilla de la cara* into Latin as both *mala* and *maxilla*, and ("*antiguamente*") *gena*. The Academy Dictionary now offers the thought that the *mejilla* is the top part of the cheek and the *carrillo* the lower; if so, the words are moving upwards in sequence, but although Lang (1887:72) agreed with this it seems very doubtful whether many Spanish-speakers are aware of such a distinction. This looks like a case of the Academy trying to impose a clear boundary between two words where there is none. (There may be a register distinction though; *mejillas* are more appropriate, for example, in love poetry).

Nothing daunted by the loss of *carrillo*, the 'jaw' took on another constructional metaphor. Modern Spanish *quijada*, Old Spanish *quexada* and *quexar*, Portuguese *queixo* (which meant formerly 'jaw,' but now also means 'chin'), Catalan *queix*, and *queixal* 'molar tooth,' all seem to come from an originally adjectival derivative *CAPSEUS of Proto-Romance CAPSA 'box' or CAPSUS 'framework'; their history is complex enough to have appealed to Malkiel (1945). (Perhaps the affixed form *quixada* or *quexada* was the preferred form to avoid confusion with *quexo* [< QUASSO] and *queso* [< CASEUM].) The Latin-Romance glossaries of ca. 1400 edited by Castro (1936) offer *quixada* as the Romance gloss for all of MALA (Escorial 1478), GENA (Toledo 745, Palacio 132) and MANDIBULA (Toledo 843, Palacio 132). This seems to imply that the meaning of *quijada* was in the process of following its precursors up the side of the face, but if so, that climb was reversed during the 15th century, for Nebrija offers only *mandibula* as the Latin counterpart for *quixar* (*o quixada*). Other items for 'jaw' adopted from an original meaning of some kind of 'framework' turn up in Italy (Bonfante 1951:383). The conclusion from all this movement—also adumbrated in De Witte (1950:121-27)—is that the cheek knows no boundaries, and such distinctive virtue as exists in its lexical items

works by means of polarity alone; the semantic structure of Spanish wanted to fit a framework with more or less distinct boundaries, but the external world provided no such boundaries to guide the positioning of the frame.

Affixation in these cases seems to be one of the lexical resources available for differentiation of meaning, rather than involving the addition of any intrinsic sense in the affix to the meaning of the lexeme. An *auricula* was not smaller than an *auris*, nor a Corsican *buccella* smaller than a *bucca*. In Spanish, the old form for 'heart,' *cuer*, has grown to *corazón* (perhaps to forestall confusion with *cuero* or *cuerpo*?), but the apparent augmentative force in *-azón* does not prevent the word being applied to hearts of all sizes (though Corominas claims that it originally applied to the hearts of warriors or lovers!). The 'head,' Latin CAPUT, has acquired a distinctive *-eza* (<-ITIA) ending in Spanish *cabeza*, to distinguish it from *cabo* (<CAPUT), now generally meaning 'end'—and, indeed, if applied to the anatomy, it can as easily mean the other end, so ambiguity is more possible here than it is in the normal extensions to other applications of words for 'head' to be found in all languages (cp. Cravens 1982:54-55)—; the *-eza* ending, which was usually used to form post-adjectival nouns (see Pattison 1975: 147), has here no semantic content. It may indeed be merely a reflex of a variant plural form CAPITIA for CAPITA (see Corominas and Pascual 1980-1:711), although CAPITIA as attested was the plural of CAPITIUM, 'head-covering.' (The Italian *cappezzolo*, 'nipple,' does, however, preserve some semantic point in the final extra affix).

Noses are also of interest (as Lang 1887:69-70 and others have observed). Outside Iberia NASUM survives (It. *naso*, Fr. *nez*, also Cat. *nas*). Non-Catalan Iberian noses lose NASUM for a conflated form derived from both NARES 'nostrils' (or 'sense of smell') and its late derivative NARICÆ; *NARICES is the postulated etymon of Sp. *narices*, originally meaning 'nostrils,' but now in the singular *nariz* meaning 'nose' (in both Spanish and Portuguese). Perhaps there also existed in Very Old Spanish a non-palatalized form (i.e., < NARICAS rather than *NARICES) off which to coin *narigudo* ('big-nosed'). *Narices* still means 'nostrils' as well as 'nose' (and 'noses'), but whereas other languages have kept the singular form with the meaning of 'nostril' (e.g., Cat. *nariu*, It. *narice*), Spanish now has no singular word for

'nostril' and has to say 'window of the nose,' *ventana de la nariz*. This sequence of events seems perverse; the sensible distinction between NASUM and NARIS was lost in Spain, and both forms were replaced by the ambiguous derivative NARICÆ/-ES, which eventually comes to have different meanings in the singular and plural. Perhaps the first stage lay in the expansion of the meaning of NARES to include 'nose.' NASUM is not used in Isidore 11.1.47, where NARES is the whole 'nose' as well as the 'nostrils,' Chindaswinth's Law 5.6.4. 3-4 [also from Gothic Spain] twice has NASUM and NARES co-referring in the same sentence: *qui in naribus si sic percussus est, ut nasum ex integrum perdat, C. solidos percussor exolvat...si vero nasus ita conlisus est, ut pars turpata narium pateat*. Another of these codes prescribes that women who carry out circumcision should [in theory] be punished *naribus abscisis*, which can only involve cutting off their noses, since nostrils cannot be cut off. Perhaps it was during the stage when NARES meant both 'nose' and 'nostrils' that the suffixed form came into general use, to create a semantic contrast between NARICÆ or NARICES 'nostrils' and NARES 'nose'; one of the glosses in Goetz's anthologies (1894:573) suggests that *narices* was thought to be the plural of *naris* (Cp. Corominas and Pascual 1980-, 4:213). Both the loss of NASUM in the first place and of NARES subsequently remain inexplicable. It would be desperation to see the replacement of NARES by the affixed form as in some way analogous or connected with the loss of AURES for AURICULAS.

There are other cases of meaning-shift across the Castilian face. *Barbilla* 'chin' is in origin the diminutive form of *barba* 'beard,' but even this sensible distinction is lost when *barba* is used, as it now often is, to mean 'chin.' In the *Poem of the Cid* it was unambiguously 'beard': see Bly [1978] and Smith (1977:261-62); *Barbilla*, consequently, would now be an inappropriate choice for either 'little beard' or 'little chin.' CILIUM was classically 'eyelid,' and still so to Isidore (11.1.42); its Spanish derivative *ceja* is 'eyebrow,' just as *sobrecejo* < SUPERCILIUM 'eyebrow' moved north to become the Spanish word for 'frown,' but other Romance derivatives of CILIUM, such as It. *ciglio*, Fr. *cil*, and Ptg. *celha*, mean 'eyelash.' Latin seems not to have had a lexicalization for 'eyelashes' other than merely being part of the 'eyelids,' *palpebræ*. Pliny is said by Lewis and Short to have used *palpebræ* for 'eyelashes' rather than

'eyelids,' but the most eyelash-like example can be reasonably interpreted to mean both together: *capnos fruticosa* [a plant]...*eadem evulsas palpebras renasci prohibet* [*Natural History* 25.99]. In any event, Romance derivatives of PALPEBRA, although often phonetically peculiar, such as Sp. *párpado*, succeed in referring to the 'eyelid' alone. For 'eyelash' Rumanian took over GENA (see above); Spanish *pestaña*, Portuguese *pestana*, Catalan *pestanya* and Gascon *pestagne* borrowed the Basque *pizta* [although Corominas and Pascual 1980-, 4:506-9 point out that early Castilian uses could also mean 'eyebrow,' and regional Catalan forms include both brow and lid as well]. *Pizta* did not mean 'eyelash' in Basque, but 'blear in the eye' [and might originally have been an Indo-European borrowing]. Isidore just calls eyelashes *capilli*, 'hairs,' in *Etymologiæ* 11.1.40: *in summitate autem palpebrarum locis...extant adnati ordine servato capilli tutelam oculis ministrantes*; Sp. *cabellos* (< CAPILLOS), and even *pelos* (< PILOS), would be unsuitable now for 'eyelashes.' One of the ca. 1400 Latin-Spanish glosses (Castro 1936) offers PALPEBRA *por pestanna del ojo*; and Nebrija's Spanish-Latin dictionary (ca. 1495) offered the converse in *pestaña*, [*pelo del ojo*]; *palpebra*, which might mean that the meaning of *pestañas* was threatening to creep inwards onto the eyelids in the 15th century; if so, *párpados* has effectively repelled them, for there is a clear distinction in Modern Spanish between the two words. *Párpado del ojo*, however, is rendered as Latin *gena* by Nebrija; GENA could once be used for the 'eyesocket,' but Nebrija's evidence collectively suggests uncertainty on his part. This would in a sense be entirely appropriate, since medieval uncertainty is the clearest feature of the phenomena being here examined.

The forms here examined so far collectively suggest that uncertainty about the boundaries of words for parts of the face in Spain has led to a variety of therapeutic reactions: the loss of some polyvalent words [OS, GENA, MALA]; the acquisition of derived forms [*barbilla, cabeza, nariz*]; the limitation of the reference of a word to only a part of its semasiological field, which can either correspond to the original meaning [*pestaña, quijada*], or the extended [*mejilla, boca, carrillo, ceja*]; and the acquisition of originally metaphorical uses into literal terms [*quijada, carrillo*; cp. also *coronilla* 'crown of head,' *muela* 'tooth' < MOLAM 'millstone,' Ptg. *testa* 'forehead' and

Fr. tête 'head' < TESTAM 'pot'—Cp. Cravens [1982] and German kopfe < CUPPA). Other expedients include: the humanization of animal terms (Sp. rostro 'face' < ROSTRUM 'beak'; the reverse happened in CRINEM > crin, 'woman's hair' > 'horsehair'); the adoption of the onomatopoeia (garganta 'throat' is apparently such, although the presence in other Romance areas of cognate forms meaning 'gland,' 'molar,' 'jaw' and 'cheek' [Bonfante 1951:380-81] does not inspire confidence in this word's value as a clarificatory term); and the possibility of recruiting foreign words. Source languages in this last category include not only Basque (pestaña) but also Arabic (nuca 'nape'), Byzantine Greek (cara), Technical Latin (mandibula, in the 18th century), and French (mentón 'chin,' borrowed at the end of the last century, perhaps intended to be a clearly shaven alternative to barba or barbilla). If Dworkin (1982), following Covarrubias, is right to derive sienes 'temples' from SENES 'old men,' as a reaction to the southward creep or incipient desuetude of tienlla, the source used for that acquisition seems little short of desperation as a remedy to preserve distinctiveness in linguistic features where facial features have proved to lack it.

The above data come from standard Castilian. Other Spanish dialects offer us comparable developments. Borrego Nieto (1981) for example, studying the vocabulary of Villapera de Sayago (in the province of Zamora) finds that there rostro only means 'cheek' (300); that for several speakers pestaña does include the 'eyelid' and even the 'eyebrow' (305); the pupil of the eye has metaphorically attracted the word for 'oak-apple,'-bogayo (179); and quijada has been confined to the jaws of animals, with mandibula being generally available for the human 'jaw.' Penny (1978), studying Tudanca, finds a regular use of picota for 'nose' (augmentative of pico 'beak'), gargueli rather than garganta for 'throat,' and an apparently neat distinction between barba 'chin' and barbas 'beard.' Penny (1969), studying the Montes de Pas nearby, finds papos for 'cheeks' (rather than 'double chin,' as in Standard Castilian), and both quijas, carrieras, and carrieros, probably a variant of O. Cast. unaugmented carriellos, for both human and animal 'jaws,' with mandibulas also available for human jaws. (These three works exemplify a welcome trend for dialectologists to group extensive vocabulary data into onomasiological sections, without ignoring those forms that happen

to correspond to colloquial Castilian.). Iglesias (1981) has also pointed out a colloquial Spanish liking for using proper names for parts of the body, although the head is not particularly favored with this device. A recent study on dialectal French words for parts of the body (Bimson and Thurman 1980) offers the following semantic shifts within the head, in different areas of France: bouche meaning 'lips', gaugno, 'fish gills,' also meaning 'lip,' 'jaw,' 'mouth' or 'cheek'; goule moving from 'throat' to 'mouth' to 'lip,' and gueule from 'throat' to 'mouth' or 'jaw'; gorge changing from 'windpipe' (GURGULIO) to 'throat' to 'mouth'; baras shifting from 'beard' to 'jaw'; visage from 'face' to 'cheek'; etc. The distribution of isoglosses in that study supports the idea of changes occurring in chain-shifts, as proposed here; this is to be expected in a "frame" theory, since a shift in part of the framework simultaneously affects both sides of the relevant border.

Amidst all this turmoil stands firm the eye itself. OCULUS is still with us as the Romance word for 'eye.' No other word has seriously crossed into its territory (and LUMEN has left it). Nor does it mean 'earlobe' in Meglenitic. Nonetheless ojo is a net exporter. Ramón Trives (1979) analyzes 32 lexicalizations of ojo, based on six central semes; among other things ojo means 'keyhole,' 'eye of a needle,' 'hole in cheese,' 'skylight,' 'porthole,' a 'spring of water,' 'round church window,' and even within the body it is not only 'callus on the hand' but a common word for an aperture at the other end of our anatomy, the connotations of which, combined with the literalization of originally metaphorical uses—may eventually lead to another word than ojo being preferred for the 'eye.' It seems to be losing the abstract sense of 'sight' to vista—in Old Spanish ojo was used this way (e.g., see Smith 1977: 253-54), but now the closest equivalent to the English 'with the naked eye,' for example, is probably a simple vista; prominent sights nos saltan a la vista whereas only an insect can appropriately saltarnos al ojo. (The same applies, perhaps even more so, to oreja 'ear' and oído 'hearing'). Possibly this makes vista the current favorite if any word is to take over the meaning of 'eye' from ojo (a statement of possibility, not a prediction). Classical poets could apparently use GENA to mean 'eye,' but otherwise words for other facial features have succeeded in remaining outside these particular borders.

This suggests that the facial boundaries of the eye are semantically more real than most; maybe this corresponds to extralinguistic reality. The internal parts of the eye, however, attract metaphor. *Pupila* dies out from the vernacular, and is subsequently borrowed from Latin into most Romance languages; *pupila* and *niña* 'pupil' (literally 'girl') in Spanish (cp. Nebrija: *niñilla del ojo*) are examples of what Corominas proposes (art. *niña*) is a near-universal linguistic tendency to use a word that originally represents the reflection perceived in the pupil (mentioned also by Isidore, *Etymologiae* 11.1.37). Fr. *prunelle* is, however, taken from 'sloe.' English and Spanish *iris* were taken from the Renaissance Latin *iris* 'rainbow,' ultimately from Classical Gk. ἶρις; in Greek it seems not to have meant 'iris of the eye,' and in Classical Latin the 'rainbow' was *arcus* (*pluvius*). Isidore did not use *iris* (for the iris of the eye) but *corona* (11.1.38), although in the event *corona* was adopted higher up as *coronilla* 'crown of head.' (Spanish now combines both classical words in *arco iris* 'rainbow'). *Niña* and *iris* still seem intuitively to be live metaphors; it is not clear why the adoption of a form for use as a literal word has been avoided, unless the center of the eye (as 'evil eye') attracts the same unease as weasels or other traditional recipients of euphemism.

One conclusion to be drawn from all the turbulence may be that it is worth considering whether semantic fixity is more rigid at some times than others. Times of comparatively wide education and high literacy (such as the present) may encourage professionals to impose clearer distinctions than naturally exist otherwise. The continental European tradition of semantic analysis, as represented in the work of Baldinger and many others, sees the function of the lexicon as an attempt to impose some kind of anthropocentric rationality on a reality which initially comes into the brain as a confused flux of only slightly differentiated impressions; in order for us to be able to talk intelligibly at all some kind of criteria have to be set up to enable us to-be as unambiguous as necessary about whether we are referring, for example, to the 'cheek' or the 'jaw,' the 'night' or the 'day,' etc. Without the intervention of self-conscious logical semanticists, the question of whether there is an exact boundary line, and if so, where, is not likely to seem important. Baldinger's section on legal language, for example (1980:42-61), shows

that even though we all know what 'day' and 'night' mean, the precise border depends on artificial conventions such as lighting-up times and case-law. To this extent, semantic boundaries seem comparable to dialect boundaries. Left to themselves, cognate dialects have transition zones rather than strict diagnostic lines between competing variables; they do not bundle the zones in the same place, and it requires some kind of artificial boundary to move the isoglosses (heteroglosses) closer together as each side of the transition zone acquires a separate prestige norm as a model for approximation in style-shifting (see Trudgill and Chambers 1980:10-14). Similarly, if lawyers, teachers, members of the Academy feel the urge to do so, they can decide to superimpose rigid lines on the semantic transition zones that occur between the prototypical areas of reference; so the Academy Dictionary tries to distinguish the meaning of *mejilla* as being intrinsically higher up the face than that of *carrillo*, and may have had sufficient authority to affect the usage of some people, although most Spaniards seem unaware of the distinction. Similarly, *mandíbula* has probably kept its strict meaning in Castilian only by being largely restricted to the usage of anatomists. Such scientific attempts to standardize the fit between our semantic framework and the outside world are rare, however, between the end of the Roman Empire and the Renaissance. Isidore attempts it, but his attempts are not often convincing. Many surviving glosses from that period imply the acceptability of considerable semantic leeway. Kohler suggested (1954) that Spanish is still in this state, citing *barba* 'chin' (505) as one example of several words that are often used to mean something close to, but different from, what might be expected (in this case, 'beard').

Rothwell (1962) argued that it is unproductive to apply logical semantics to Medieval France, since "in a language untrammelled by grammarians and academies the development of the vocabulary can often be quite undisciplined" (30); in an area where time, weights, coins, and other apparently objective units varied from place to place, it is hardly likely that there was sufficiently rigid agreement on other matters to permit clarity in general "structure." "Frame" Theory is less inappropriate under such circumstances in that it is only the existence of a rough boundary that is postulated,

rather than the ability to draw it precisely or position it exactly. In Medieval Spain, words of many kinds seem to have surprisingly unfixed edges in the 14th-c. *Libro de Buen Amor*, in meaning as well as in form (particularly as regards derivational morphology; cp. Adams 1970). A research task that is crying out to be done is an investigation into whether or not semantic distinctions became rigid during the Alfonsine standardizations of the late 13th century at the same time as much of the spelling and some of the grammar. Now that we have to hand the Madison Concordances (Kasten and Nitti 1978), such an investigation might be possible to carry out. My guess is that before the arrival of dictionaries they were not so standardized. If Spanish words from the same etymon (\*EXCURTIONEM) can mean 'toad' (Castilian *escuerzo*), 'hedgehog' (Mozarabic *uxcurchón*), 'scorpion' (Aragonese *escurzón*), 'viper' (Catalan *escurçó*) and even perhaps 'worm' (in the *Libro de Buen Amor* 1544c; although Hook [1979-80] argues persuasively that it means 'toad' here[4]), we need not find the gyrations of facial features surprising.

Semantic change often plays only a small part in modern discussions of historical linguistic theory (e.g., Bynon 1977). This is not because it is a small subject; a glance at any etymological dictionary shows that a surprisingly large number of words come from an undisputed etymon with a different meaning. Nor is it often discussed seriously in synchronic generative studies on semantics, although many European writers on semantics such as Ullmann (1962), Waldron (1967), Baldinger (1980:277-309), and specifically on Romance, Cremona (1959), consider it intelligently and show that it is possible to be rational and even "scientific" if we have sufficient data. Ullmann, for example, sets up a category of Semantic Change through "contiguity," which is what we have examined here. The reason for the shyness among generative theorists must be, at least partly, the difficulty in describing this kind of semantic change in easily annotated form such as could be used for an algorithmic input. But linguistics has grown out of the behaviorist belief that language has no meaning; it is currently

[4] Hook and I find our approaches interconnect here (Wright 1976:18).

growing out of the assumption (sponsored by those wishing to acquire funds from bodies wishing to invent translation machines) that natural language has the precision of computer language; with any luck it can soon grow out of the idea (despairingly proclaimed by those who instinctively want linguistics to be a natural science) that language-change runs according to some kind of predetermined typologically-induced "drift" (see Matthews 1982, Wright 1983, etc.). as part of the same tendency, checklist theories of semantics will give way to less logical varieties. "Drift," algorithmic syntax, and semantic rigidity have been ideals in the minds of linguists. If we wish to study language as it is rather than as we wish it were, we have to accept that the oppositions inherent in distinctive linguistic features at times have distinctive virtue by means of potential contrasts within the Gestalt rather than by logically necessary and referentially-precise content with fixed square-bracketable edges. This applies not only to colors. The history of Spanish words for the face suggests that facial features (other, perhaps, than the eye) have this kind of indistinctiveness, and it is probable that the history of other semantic fields in Romance may lead to comparable conclusions. Sampson (1979:59ff.) suggested that the facts of semantic change argue against the existence of semantic structure, but it would be fairer to see them as counterarguments to "checklist" theories alone, for they are easy to explain as the results of minor shifts in the fitting of a "frame" onto human perceptions.[5]

### 1994 POSTSCRIPT

[5] I am grateful to my colleagues, Max Wheeler and Andrew Hamer, for their comments on an earlier draft of this paper; their views do not necessarily connect with mine.

annual Romance Linguistics Seminar at Trinity Hall, Cambridge, in January 1983; later, subsequent to its acceptance by the journal, but before its appearance, it was delivered in slightly shortened form at the Universities of Oxford (May 1984), Dublin (November 1984), and, in Spanish, Santiago de Compostela (April 1985). Many colleagues have sent me interesting pieces of extra data as a consequence, some of which I here repeat for interest: that Romanian *bucă* can in fact be used to mean "cheek"; that Spanish *carrillos* can be used to refer to "buttocks" but *mejillas* cannot; similarly, it is possible only to *bufar los carrillos*, "blow out one's cheeks," rather than the *mejillas*; that the supposed mozarabic form *maxšila* is in the MS actually the equivalent (in the Arabic alphabet of *makhshall*) of uncertain meaning, and thus hardly a clear case, that in Portuguese "to have something in one's eye" can be *ter alguma coisa na vista*, and the speaker's own face tends to be *cara* but someone else's tends to be *rosto*; that Celtic borrowings from Latin *gena* include Welsh *gen* "jaw, chin," Medieval Breton *guen* "cheek," Old Irish *gin, grun* "mouth"; that in English all is Germanic except *face*; that, with reference to *naribus abscissis*, a princess referred to in the "Aragonese" *Cronica de Morea* had her *rostros* cur off ("lips"? Possibly "nose"), with thanks to Rodney Sampson, Lynn Williams, Ralph Penry, Richard Hitchcock, Norman Lamb and Fred Hodcroft. I have also discovered that Spanish *juanetes* "high cheekbones" really is a diminutive of the proper name *Juan*; that Berceo's *Signos del Juicio Final* tells us of serpents and/or scorpions who bite and/or sting with their *rostros* (*meter lis an los rostros fasta los corazones*, 39c); and that in *Patrologia Latina* Vol.112, cols. 1575-78, there is "B. Rabani, Glossae Latino-Barbarica de paribus humani corporis," in which the majority of the glosses are Latinate, or both Latinate and Germanic, rather than Germanic alone (e.g.: "*Palpebrae sunt sinus oculorum a palpatione dictae augbrauua*").

There is still wide and unexploited scope for pan-Romance consideration of the diachronic evolution of semantic fields of this type; much of the data were elaborated years ago (as in Zauner's article in this case), but an updated global vision of the diachronic dynamics is uncommon.

# Semantic Change in Romance Words for "Cut"

## 8

W E CAN ILLUMINATE SEVERAL semantic changes at once if we visualize some parts of the vocabulary as structured wholes, where changes in the criteria for the use of one word can have consequences for others.

Intrinsic links between lexical items arise partly because one early stage in the cognitive process that prepares us to be able to talk about the world is that which identifies separate entities as being distinct from each other; Wright (1985) studied changes that can be located at this stage. Today I consider linked changes located at a subsequent stage; if we wish to choose a word with which to refer to one of these separately delimited potential referents, one of our next tasks is to decide on the appropriate level of generality or specificity (a decision that is usually made on pragmatic grounds). For much of our vocabulary the structure is organized according to scales of increasing specificity. The standard examples of these scales are terms of natural history, sometimes represented on paper as in Table 1.

| Thing | | | | | | | |
|---|---|---|---|---|---|---|---|
| Creature | | | | | | | Object |
| Animal | Insect | Fish | Bird | | | | |
| | | | Duck | | | | Owl | Crow | etc. |
| | | | Mallard | Teal | Wigeon | Shelduck | etc. |

In this theory (e.g. Lyons 1968: ch.10) a term is said to be the 'hyponym' of the ones directly above it on these scales, and the

'superordinate' of the ones directly beneath it; the English words *duck* and *owl* are thus hyponyms of the word *bird*. Hyponyms of the same superordinate are said to be incompatible with each other; e.g. any bird describable as a *duck* cannot also be described correctly as an *owl*. The vertical lines in these scales correspond to our criteria for choosing between the hyponyms of a superordinate term, and these distinguishing criteria are based on some perceptible difference that we have learned. It seems most likely that in our search for a word with which to refer to our intended referent, we usually enter the vocabulary from the most general end, if only because it is easier for lazy speakers to find the word *thing* (*thingie*, *whatsit*) than *wigeon*; and then we are faced with increasingly specific choices to make, until we find the most pragmatically suitable place to stop searching. In this way we can find words for referents which we haven't seen before, and changes in the world do not necessitate a change in the semantic structure. For example, when the Spaniards first met pineapples in the New World, they called them *piñas*, as they still do, which till then had only been used for "fircones"; the perceptual criterion distinguishing fircones from other fruits also worked to separate pineapples out as part of the same category, and no Spanish-speaker finds this polysemy confusing. Choices from the hyponymic scales are not the only ones that need to be made, of course; words are always chosen by the speaker from several possible available items, and even though these choices are normally unconscious they are never forced on us by logical necessities in square brackets inherent in our intended meaning.

The hyponymic scales are the scene of two of the standard categories of semantic change: generalization and specialization. Generalization occurs when a word moves up the scale to a higher level. One example is the Spanish word *argolla*. In Spain that means a "large ring," as on a quayside, or at the smallest as a bracelet, as opposed to an *anillo*, a "small ring" as on a finger; but in parts of America the criterion of size is lost and *argolla* can be used for an engagement ring, which sounds *dit* in Spain. The converse is specialization: where a word only survives with a specialized sub-part of its original potential reference. An extended case

involving several words is the semantic structure of words meaning "cut" in the Romance languages.

The English word cut can be defined as "strike a successful blow with a sharp edge"; that is, it is a hyponym of the word *strike* with the criterion of "successfully with a sharp edge." In Latin the word for "cut" was SECARE, and in a few parts of the Romance-speaking world, including Sardinia (*segare*), that word still means in general "cut." There are many different kinds of cutting operation, each of which might require a separate lexical hyponym of its own. One of these is to "cut corn," that is, "reap," "harvest." With a specific direct object, SECARE could naturally always be used to refer to this. But in some areas, notably the Iberian peninsula, *segar* came to mean specifically "reap" even when used elliptically without a direct object; for example, in the thirteenth century the Riojan poet Gonzalo de Berceo intended the phrase *tiempo del segar* to convey unambiguously "harvest time" (*furtávalis las miesses al tiempo del segar, Vida de Santo Domingo 420a*), and the translator of St Matthew's Gospel used the agent noun *segador* to mean the only thing that it has ever meant, "harvester" (*La miess es mucha e los segadores pocos*: "the crop is heavy but the laborers are scarce" in the *New English Bible*, 9:37; also 9:38 and 13:30). This is specialization: SECARE, *segar*, has moved to occupy a slot to which it was originally superordinate. At this point you will be wondering whether this journey was really necessary; didn't Latin already have a perfectly good word for "to reap"? It did: METERE, which is common in the Vulgate Bible. That word survives in standard Italian *mietere* with that same meaning (and Occitan *meire*). In Spain METERE itself has gone; perhaps because of potential confusion with MITTERE > *meter*, METIRI > *medir* or METUM > *miedo*, when the rhizotonic '-ERE paradigm disappeared in Spain (as it did not elsewhere), but the so-called frequentative form, MESSARE, survived, formed off the past participle of METERE. In Sardinia it is this form that means "reap," as also in some Northern Iberian valleys; elsewhere in Spain *mesar* specialized further to mean "pull out" - that is, a form of harvesting but specifically without using any sharp-edged instrument - and then extended its referential criterion slightly so as to be applicable to pulling out hair from the head as well as grass or corn from the ground. In the *Poema de Mio Cid* (of

c.1200), *mesar* is used on five occasions, all referring to pulling pieces out of someone's beard (lines 2832, 3186, 3286, 3289, 3290), and the point would there be lost if any connotations of using a razor, knife or scissors lurked still. The result was that "cutting corn" was becoming a vacant slot, a potential squat ripe for colonization by something else. And yet the noun MESSIS, "harvest," with its meaning unchanged, as *mies* (cp. the Berceo quotation above); so *mies* came to be semantically the nominalization of *segar* rather than of its formal cognate verb *mesar*.

In France, however, the derivative of SECARE, Old French *seier*, Modern *scier*, usually means to "saw," that is, "cut with a saw," although in some areas it can also mean "reap" as in Spain. Latin hadn't had a separate hyponym of SECARE for "saw," using SERRA SECARE to convey the meaning merely syntagmatically. In Spain, Sardinia and parts of France a derivative verb *(a)serrar(e)* was formed to fill the gap; for words created via affixation can fill apparent lexical gaps as much as can semantic change or neologism—the processes are intimately linked. In Northern France the coining was in the reverse direction; they eventually created a noun *scie* off the verb, unambiguously meaning a "saw." Modern Italy is similar to France: Italian *segare* usually means to "saw," and the noun *sega* was originally only a "saw" (*segatore* being a "mower"). In France METERE disappeared, and MESSARE may well never have existed, but another verb formed off the cognate noun, in this case MESSIO, -ONIS rather than MESSIS, took that place: French *moissonner*.

Meanwhile, SECARE itself had acquired a frequentative form *SECTARE, which, where it survived, in Old Portuguese and Asturian *(as)seitar*, meant "reap." The Castilian equivalent, if it existed, would have been *sechar*, and Malkiel (1947) was surely right to argue that *cosecha*, which has in modern times become commoner than *mies* for "harvest," has some kind of connection with SECTARE, despite Corominas's disagreement (1980:121).

To sum up so far; where SECARE survives it has, outside Sardinia, specialized, that is, acquired extra criteria for use concerning the nature of the cutting concerned; where METERE has survived, mostly in Italy, it has kept its meaning of "reap," so that SECARE has there not slid down the scale to occupy that particular hyponymic slot.

SECARE seems thus not to have been determined to "push" its way into any particular lower slot, but we could reasonably call these developments a "drag"-chain; where a hyponym (e.g. METERE) is falling out of use, for whatever reason, its superordinate (here SECARE) can always be used instead, by definition. Eventually this pattern of choices can shift the distributional pattern of the reference of the original superordinate term, and, as Erica García has been arguing persuasively, such distributional shifts can lead to semantic change. In this case it has. In most of Spain *segar* came usually only to be used if the referent cut was grass or corn. The superordinate slot for "cut" was not left as a vacuum, since it was still possible to use *segar* for referring to other types of cutting; but it would have increasingly felt metaphorical to do so, as if we were now to talk in English of barbers "harvesting" their clients' hair. So it was not logically necessary, but it was nonetheless convenient, to consider using something else as the superordinate.

The French for "cut" is now *couper*, formed off the noun *coup*. French *coup*, Italian *colpo*, Spanish *golpe* and Catalan *cop* all mean "blow"; they derive from Late Latin (Early Romance) COLAPHUS, which was originally borrowed from the Greek κόλαφος, meaning "punch," but COLAPHUS had semantically generalized to mean a "blow" of any kind, losing the criterion of "with a fist." This had thus came to fill the slot being vacated by the Latin superordinate terms; the noun ICTUS was falling out of general use, and PLAGA only survived with the sense of "wound" (the results of the blow rather than the blow itself), in Spanish *llaga*, Portuguese *chaga*, French *plaie*. The superordinate Latin verb CAEDERE, "strike a blow," also fell out of use. The Old Spanish verbs *colpar*, *golpar* and *golpear* (the form that survives), Catalan *copejar* and Italian *colpire* kept close semantic contact with their cognate noun, similarly generalizing to mean "strike a blow" of any kind. In France, however, they then chose to specialize the verb, *couper*, with the new criterion of "with a sharp edge" which distinguishes the meaning of "cut" from "punch," "slap," "kick" and other fellow-hyponyms of "strike" (Cp. Lehrer 1974). The French word meaning "strike" is now *frapper*, of uncertain etymology, which had previously meant "to hurl oneself onto"; the timing of the semantic changes shows the change in *frapper* to have begun at a slightly later stage than that in *couper*,

but the change may not yet be complete. (*Frapper* has certainly become the superordinate for most of the hyponyms of "strike," but not all French-speakers seem to see *couper* also as a modern hyponym of *frapper*.) The change in French *couper* happens to none of its cognates in other languages, and can for this reason plausibly be dated to a late enough time for it to be seen as an intermediate stage there in a drag-chain, as the superordinate *couper* slid down to where *scier* would have been if it had not itself specialized, and then the loss of *couper* from the "strike" slot dragged *frapper* across in turn.

Italian *tagliare*, Catalan *tallar*, Portuguese *talhar*, Rumanian *tăià* and Old Spanish *tajar* all came to be normal superordinate words for "cut." They came from Late Latin (Early Romance) TALIARE (or TALEARE, [-l'-]). This word has generalized from being once a hyponym of SECARE, for it was formed off the noun TALEA, which was originally a "cutting," a small section cut off a bush in order to be independently planted. French *tailler*, on the other hand, from this root, remained on the same level of the hyponymic scale but enlarged its criterion for use, being now suitable for any careful cutting such as shaping precious stones, carving wood, pruning trees and cutting out clothes. In Spain, TALIARE > *tajar* has since respecialized, but it seems to have been the superordinate term in at least Early Mediaeval Spain. In a tenth-century document from León it appears to be used for slicing cheese (*quando la taliaron*: Menéndez Pidal 1926:25; Wright 1982:173). In the *Poema de Mío Cid*, *tajar* is used for cutting down orchards (line 1172: *tajavales las huertas*) rather than taking cuttings off them, and also for cutting hair (1241, *Nin entrarie en ella tigera, ni un pelo non avrie tajado*). King Alfonso X's *Siete Partidas* I.IV.99 has the phrase *mesabanse los cabellos et tajabanlos*, "they pulled out and cut their hair." The agent adjectives *tajador* (5 times in the *Cid*) and *tajante* (as in the *Libro de Alexandre* 1347d, *todos eran tajantes como foz podadera*) both meant "sharp." *Tajante* still means "sharp," mostly in a metaphorical sense, "trenchant," but the verb itself, *tajar*, has since, in Spain, specialized again, to mean usually "chop into pieces," implying strong action as with an axe, and now unsuitable as a word to use for referring to cutting hair, or to cutting a finger without cutting it off.

There are two possible reasons for the decision to respecialize the reflex of TALEARE in Castilian, unlike elsewhere. One may lie in the potential confusion, at places and times where TALEARE preserved the lateral consonant, with *talar*. *Talar* comes from a Germanic root (*talan*) and originally meant "devastate"; it tends now to be like the English *fell* and be confined for use with trees. The noun *tala* could already be used to refer to the peaceful right to cut firewood from trees on common land in eleventh- and twelfth-century law. Thus both *tajar* and *talar* came to be hyponyms of "cut." For "felling" trees Latin had tended merely to use CAEDERE, "strike," the general superordinate of SECARE, in the absence of a specific lexicalized hyponym. Later, in the sixteenth century, the Italian *tagliare* was borrowed into Castilian as *tallar* with only the meaning of "engrave"; this is a combination of borrowing and specialization that need cause no surprise, given the shape of the slot it was borrowed to fill. Thus now a Spanish tree can be felled (*talar*), chopped into sections (*tajar*) and those sections be given an engraved carving (*tallar*); a phonological minimal trio of three hyponyms of "cut."

But Old Spanish *tajar* may have been losing the battle to fill the superordinate slot anyway, to its rival, *cortar*. The origin of *cortar* was a fairly rare Latin word CURTARE, meaning "shorten, reduce," semantically related to CURTUS, "short." Latin CURTUS could mean "castrated" and "circumcised," so even then could be used to refer to the results of some cutting actions, but the adjective survives in Romance with the meaning of "short" and no cutting connotations; Spanish *corto*, Catalan *curt*, Portuguese *curto*, French *court*, Italian *corto*. Rumanian *scurt* comes from a form with the prefix Ex-. So does the Rumanian verb *scurtă* and dialectal Italian *scortare*, "shorten," and French *écourter*, "cut short," that is, "shorten with a sharp edge." In thirteenth-century Spanish the verb means specifically "cut": for example, the five uses in the *Poem of the Cid* (lines 751, 767, 2423, 2728, 3652) are for cutting through helmets, waists and heads with a sword. The semantic structure of "strike" and the semantic structure of "cut" are separate. Yet it happens at times that words with a precise hyponymic criterion in one part of the vocabulary can be adopted for use elsewhere, with the same criterion under a different superordinate. It is possible that by Very

Early Mediaeval Spanish the normal use of CURTARE was still for "shorten," but usually now specifically "with a sharp edge" [as EXCURTARE means in the Merovingian Salic Law]; and that eventually a need for a word with that specific criterion under the superordinate "strike" led the word to override structural boundaries and come to mean "cut," whether or not the cutting also involved shortening the object referent. With the subsequent specialization of *tajar*, the Spanish "cut" structure has come to be filled now as in Table 2.

cortar

| tajar | tajar | tallar | segar | (a)serrar | etc. | etc. |
|---|---|---|---|---|---|---|

Table 2.

[The etcetera in Table 2 include such words as *afeitar* "shave," *amputar* "amputate," *podar* "prune," *hender* "split," and so on.]

Once Spanish *cortar* had come to have "cut" as its central literal meaning, it could no longer be used to refer to shortening that did not involve a sharp edge, e.g. shortening sail or debates. There was emerging another potential gap. Derivational morphology came to the rescue again: the Spanish superordinate for "shorten" has always been *acortar*. *ADCURTARE did not exist in Latin, but there was nothing adventurous in this invention: many Old Spanish verbs had forms both with and without an essentially meaningless prefix *a-* [cp. the coexistence of *serrar* and *aserrar* mentioned above, or *allegar* and *llegar* referred to in Wright 1987]. In this way *corto* and its semantically related verb *acortar* have both broken off semantically from *cortar*; thus to "shorten sail" is *acortar la vela*, in which the sail remains uncut. *Acortar* is the only one of the words in this paper to come early enough in the alphabet to be in the ongoing *Diccionario Histórico de la Lengua Española*; one of its four thirteenth-century attestations [Vol.I.522-24] probably involves shortening with a sharp edge [of a wooden beam: Berceo, *Vida de San Millán* 227b], but the other three do not [shortening life-expectation and law-suits, and limiting damage in general].

(1)

| METERE. | SECARE | | CAEDERE. | |
|---|---|---|---|---|
| | etc. | £ | | etc. |

[Latin]

(2a)

| tajar ~ cortar | | golpe(j)ar | |
|---|---|---|---|
| segar. | (a) serrar+ | £ | etc. |
| | | | etc. |

[Medieval Spanish]

(2b)

| metere. | tagliare | colpire | |
|---|---|---|---|
| segare+ | £ | | etc. |
| | | etc. | |

[Italian]

(2c)

| moissoner. | couper | frapper | |
|---|---|---|---|
| scier+ | tailler@ | £ | etc. |
| | etc. | | |

[French]

Criteria: £ = with a sharp edge.
. = object being grass or corn.
+ = instrument being a saw.
@ = done carefully.

Table 3.

## CONCLUSION

We have been looking at the data of Table 3. The rigidity of these diagrams should not be taken too seriously; not everyone in the same area at the same time need have the same detail in their lexical structures, and words are used non-literally all the time. Even so, as a generalization, we see the survival of a more or less consistent substructure, intended to contain a word meaning "strike" and a hyponym thereof with the specific criterion of "successfully with a sharp edge"; this word, in turn, is superordinate to an indeterminate number of hyponyms of its own; the structure is roughly constant, though the words filling the slots have, at least