

Trends in Linguistics
Studies and Monographs 163
(2005)

Editors

Walter Bisang
(main editor for this volume)
Hans Henrich Hock
Werner Winter

Mouton de Gruyter
Berlin · New York

Perspectives
on Variation

Sociolinguistic, Historical, Comparative

edited by

Nicole Delbecque
Johan van der Auwera
Dirk Geeraerts

Mouton de Gruyter
Berlin · New York

Europe's sociolinguistic unity, or: A typology of European dialect/standard constellations¹

Peter Auer

1. Preliminaries

At first sight, the diversity of the dialect-standard constellations found in Europe is enormous and seems to defy any attempt to find a common denominator. Many believe that the sociolinguistic situation, say, in England, Germany and Italy is fundamentally different and that what is meant by 'dialect' and 'standard' in these countries cannot be compared at all. In this paper, I would like to advocate the opposite position and argue for a uniform description of the European sociolinguistic repertoires. More specifically, my claim is that on a sufficient level of generalisation there is a systematicity behind the superficial heterogeneity which unfolds from a historical perspective. I will suggest a typology of speech repertoires which distinguishes five sociolinguistic types that also represent a chronological order. The typology aims at the standard-dialect dimension, while multilingual constellations will only be commented on to the degree that exoglossic standard languages enter the repertoire. There have been some previous attempts at more comprehensive typologies valid for all sociolinguistic areas in the world.² By comparison, the present attempt is a good deal less ambitious since it is restricted to Europe and since it leaves out many multilingual repertoire types; however, it is supplemented by a historical component which is usually absent.

A number of preliminary remarks are necessary. I will use the term 'dialect' as a purely relational concept (following, for instance, Coseriu 1980), which means that by definition, without a standard there can be no dialect. (The opposite is a contingent fact and an important generalisation over the European sociolinguistic history.) Also, the term will be used exclusively in order to refer to areal variability within a language (i.e., there is no 'standard dialect' of a language according to my usage of the term, but only a standard variety). On the other hand, the term 'dialect' will not be restricted to the 'base dialects', i.e. the most ancient, rural, conservative

dialects, but will be used such as to include regional and urban varieties with a larger geographical reach as well.

More controversial is the term 'standard'. Here, it is employed in order to designate a *variety* of a language (which follows a 'norm' or 'codex', i.e. 'standard' does not designate the norm itself), which is characterised by the following three features: (a) it is orientated to by speakers of more than one vernacular variety (which does not necessarily imply that it is mastered by everybody), (b) it is looked upon as an H-variety and used for writing,³ and (c) it is subject to at least some codification and elaboration (Haugen 1966) or *Ausbau* (Kloss 1967).⁴ A standard variety therefore is more than a 'common variety' (*Gemeinsprache*), which only satisfies the first criterion, and more than an H-variety (*Hochsprache*), which only satisfies the second criterion. The third requirement excludes from consideration, for instance, the various H-varieties of Middle High German as used in the medieval literature. But taken alone, it would not be sufficient to define a standard variety either, since there are codifications of dialects as well (for instance, there is a *Zürichdeutsche Grammatik* in which people look up what *Zürichdeutsch* dialect should 'really' be like, although no Swiss person would think of *Zürichdeutsch* as a standard variety).

My focus will be on the emergence of the *endoglossic* national standard varieties in Europe in the second millennium A.D. and their relationship to the dialects in their respective geographical area. These standard varieties are closely linked to the emergence of the European nation states; since the codification of a national language has often been looked upon as an important if not an essential step in nation building, the history of the standard variety and its status today usually reflect the way in which nation building has proceeded. In some areas, there has been a straightforward development from the late Middle Ages to the present day (such as in the prototypical European nation states, e.g. in England or France), in other cases, the process set in much later (as in Romania, Luxembourg, etc.), or was restarted after a period of destandardisation (Ireland, Greece, etc.). Note, however, that not all national standard languages (e.g. those whose emergence is due to a nationalistic ideology) are state languages in Europe. Some have failed to reach this status and are used as minority languages within one or more European states (such as Standard Basque/Euskara Batua). There are also some recent legally recognised standard languages which only came into being in the post-nationalist era and therefore have never played a role in nation-building (such as Reto-Romance). However, the recent developments in east and south-east Europe after 1990 provide ample evidence for the fact that the nationalist equation of one language = one nation = one state still survives. Finally, it should be mentioned that

there is at least one national language for which it is at least doubtful whether it fulfils my criteria for a standard language, since it is not a H-variety in Ferguson's sense; this is *Lëtzebuergesch*, which is the national language of Luxembourg (and together with French and German one of the official languages of the state), but is almost exclusively spoken (Gilles 1999:8-9).

The opposite is of course also true: there are standard languages in Europe which were no *national* languages (while *Lëtzebuergesch* is a national language which is no standard). They fall outside the scope of this paper. Among them is Latin, which, at least from the times of the Carolinian renaissance onwards, was not only used as an H-code, particularly in writing, but was also subject to codification. Other candidates for standard languages which were no (or not everywhere) national standards are the Attic-Ionic ('Hellenistic') koine of the fifth century, and Old Church Slavonic, at least in some periods and in some of the east and south-east European countries in which it was used.⁵

2. Type Zero repertoires: No endoglossic standard (exoglossic diglossia)

In the historically primordial situation in medieval Europe, if a non-vernacular variety was used at all, it was an exoglossic⁶ high (standard) variety: Latin, Old Church Slavonic⁷ or Arabic.⁸ This repertoire type can be represented as in Fig. 1:⁹

(exoglossic standard)

•

vernacular varieties
(spoken/written)



Figure 1. Type Zero Repertoire

Since the standard variety was exoglossic, no levelling through standardisation (levelling out of non-standard features) could take place. This does not imply, however, that the exoglossic standards had no influence on the vernaculars; the opposite is true. But here, we are dealing with language contact, not with dialect/standard levelling. Also, it should be noted that the vernaculars sometimes underwent horizontal levelling (koineisation) *before* the onset of standardisation. Examples are the dialect koinai of southern, especially Andalusian Spanish which developed in the late Middle Ages (thirteenth to fifteenth century) before standardisation set in (Villena 1996). It is possible that a similar pre-standard koineisation took place in the course of the German settlements in the Slavic east, particularly in the Upper Saxonian region, in the twelfth and thirteenth century (Frings 1957).

(exoglossic standard)

vernacular varieties
(spoken/written)



Figure 2. Koineisation in a Type Zero Repertoire

Inter-vernacular levelling is related to the development of a *Gemeinsprache*¹⁰ (common language), and it may be a helpful step on the way to a national standard variety.

In Europe today, non-standardised varieties are rarely written (and if so, only in personal genres such as in e-mail, a conceptually half-oral, half-written text type). The same is not true for the history of the European languages. Examples range from the dialects of Old High German and the Proto-Romance dialects/languages to the Greek demotic literature under Ottoman rule.

Over the centuries, when the European nation states came into being, many *national* standard languages also assumed the status of an exoglossic standard variety in Europe *outside* the nation from which they had originated. The addition of an exoglossic standard language to the repertoire took place in various guises, from the imposition of the language

of the conquerors as a standard on top of the structurally distant local vernaculars or even in competition with an endoglossic standard (examples among many: German in Czechia, Ottoman-Turkish on the Balkan and in Greece/Cyprus, English on Malta or in Gibraltar), the marginalisation or suppression of linguistic and cultural minorities within the territory of the newly developing nation states (examples: the Celtic languages in France or Great Britain, Swedish in Finland, German in South Tyrol) to the voluntary acceptance of a foreign standard language as the language of writing, ceremony and/or educated speech (examples: the Russian variant of Church Slavonic in the eighteenth century in Serbia, a variant of Church Slavonic as a literary and ritual language in Romania from the fourteenth to the eighteenth century, French in the era of enlightenment in most of Europe, French by the Flemish speaking bourgeoisie in Flanders in the nineteenth century, or French in Luxembourg today). In some areas, a combination of these factors was in play, or the motivation for the use of an exoglossic variety changed over the course of time (such as in the case of German in the Baltic).

Type Zero repertoires continued to exist until well into the twentieth century for many of the smaller (minority) languages of Europe. They resorted to the standard varieties of the nations to which the respective minority belonged as a (pseudo-) 'roof'.¹¹ Most of these language minorities have recently (with more or less success) developed standard varieties of their own, in a conscious attempt to retain the vitality and competitiveness of the respective languages against the larger European languages, usually the languages of the states to which they belong. The new endoglossic standard varieties in this case compete with the well-established exoglossic one. Examples for recent standardisation are the Sami languages (with Finnish, Swedish, or one of the Norwegian standards as the overarching heteroglossic roof), Basque with Spanish or French as the exoglossic roof, Reto-Romance with German, the Frisian varieties with Dutch and German, *Lëtzebuergesch* with French and German, or the Bretonic varieties with French as the respective roofs. There are only very few European languages which have never developed an endoglossic standard; an example would seem to be the varieties of Romani (overarched by a dozen or more European languages, with very minor and unsuccessful attempts at standardisation in the Balkan states).

Some varieties of European languages have lost their (endoglossic) standard variety because the speech community has become completely detached from its 'homeland' as a consequence of migration, and therefore from the larger part of the speech community. This holds, for instance, for Csángó-Hungarian spoken in Romanian Moldavia which has been isolated

3.1. Historical development

Apart from the extraordinary case of Old Church Slavonic (Old Bulgarian) in Bulgaria, perhaps apart from the case of Old Irish in Ireland and apart from the somewhat special case of Greek in Byzantium, we have no reason to believe that the formation of a high, codified, common language set in on the vernacular languages anywhere in Europe before 1300, and in many areas not until much later. Common languages before 1300 (local koinai) were not codified, not linked to a national discourse and usually accepted in small areas only.

The transition from Type Zero to Type A goes through a (sometimes prolonged) phase in which the exoglossic standard competes with it.

I call this first phase in the development of an endoglossic standard medial diglossia (following a terminology used in Swiss sociolinguists), because new standard varieties were used for the medium of writing only in the beginning. An important role for the spread of these written standards was played by the printing technology, and in the Protestant areas in particular, the bible written in an endoglossic standard was often the most successful print product in the post-reformation period which contributed to the standardisation of the written variety. In general, it can be said that the spread of the endoglossic written standard was to a large degree achieved through the written medium; cf. the spread of the High German Standard to the Low German areas; the spread of Castilian Spanish to the Aragonese and Leonese areas in the late Middle Ages; the spread of London English as the dominant printers' variety to other areas of England in the fifteenth century; the use of Edinburgh Scots English as the written variety in Scotland during the same period; the spread of standard Polish to the periphery under the influence of the Cracow printing offices.

The generalisation that new standard varieties are first introduced in writing also holds for the nineteenth and twentieth century standardisation processes (such as in Norway, Finland, Greece re. the *katharévoussa*, Czechia¹⁵/Czechoslovakia, Bulgaria, etc.), in which philologists/linguists took a central part. The difference is that these 'late' standard languages often were construed by conscious reference to some older, sometimes quite distant variety which these philologists considered to be more prestigious than the vernacular language actually spoken at their time (under the roof of an exoglossic language associated with some external power). It is therefore typical for these late standardisation processes that they had a strong archaising component. The resulting standard variety is bound to be restricted to written usage first since it is perceived by the population as being far removed from their ordinary language practices.

Of course, written standards were sometimes also spoken (for instance, read out, or used in oratorical style); they were then pronounced according to the written form. This *Sprechen nach der Schrift* often led to hyperarticulate forms, and even to the introduction of phonemic differences into the phonology of the language which had not been there before. Some of them had a lasting impact. Examples are the (nowadays phonemic) distinction between /e:/ and /ɛ→:/ in spoken standard German (which at least in some areas is absent in the dialects and was introduced on the basis of the written distinction between <e> and Umlaut <ä>), in standard Danish the reintroduction of /k/, /g/ before front vowels (to keep previously collapsed minimal pairs such as *kære* 'dear' ~ *tjære* 'tar' distinct)¹⁶, or in standard Finnish the 'invention' of a phoneme /d/ (a reading pronunciation of <d> taken over from sixteenth century Turku speech into nineteenth century Finnish standard; in Turku <d(h)> represented a dental fricative¹⁷).

3.2. Europe today

The oldest type of a repertoire in Europe in which standard and dialects are structurally and genetically closely related was similar¹⁸ to what we find today in German-speaking Switzerland and in Norway, particularly in areas where *Nynorsk* is used instead of *Bokmål* as the standard variety (i.e. with the exception of the eastern region around Oslo).¹⁹ In both cases, there is no strict historical continuity.²⁰ Swiss standard German and *Nynorsk* today are used exclusively in (non-private) writing and very formal oral situations.²¹ In informal and semi-formal situations, everybody speaks his or her dialect; and since the diversity of the local dialects is enormous (both in Norway and in Switzerland), and modern Norwegians or Swiss are as mobile as any other Europeans, face-to-face interaction is often 'polylectal'. Accommodation to the other speaker's dialect seems to be minor, but levelling towards regional koinai has set in in recent times.²² Some dialects carry more prestige than others, but speaking dialect as such is highly appreciated, while the standard variety has the connotations of a formality and stiffness. Since the written/spoken distinction is so important, the occasions for conversational code-switching between standard and dialect are relatively restricted (i.e. to those situations in which the standard can be used in oral communication at all).

4. Type B repertoires: Spoken diglossia

Some European standard varieties never reached beyond stage A and presumably disappeared before they were spoken by larger segments of the

population. (Examples would seem to be the Low German standard variety based on the city dialect of Lübeck, which was used in writing in the rest of the Hanse as well, but failed to develop into a spoken standard variety, or Occitan, which started to develop a written norm very early but lost its function as a standard due to French influence.) Others consolidated and developed into Type B repertoires.

In a type B repertoire, usage of the standard language extends into spontaneous oral speech, particularly in formal, out-group interaction. Standard and dialect have their strictly allocated and seldom overlapping domains of usage, which, however, are no longer backed up by different media (writing/speaking). Therefore, code-switching tends to occur, particularly of the situational type.

The shift from Type A to Type B implies a far-reaching re-orientation of behaviour, which goes far beyond the mere 'articulation' of a written language (cf. van Marle 1997). In order to be suited for spontaneous face-to-face interaction, the codified written standard language has to undergo a fundamental transformation; often, the existing oral varieties (the dialects) are used as a prop in this process. The introduction of a spoken standard variety also raises new problems regarding codification, since it extends normative control to aspects of language not coded in writing and therefore difficult to enforce in a large area. It is therefore not surprising that codification of speech often has remained imperfect. For both reasons – the need to make the written standard suited for spontaneous oral communication by resorting to already existing spoken varieties, and the difficulties to exert normative control on spoken varieties – , the spoken standard usually shows more inherent variability than the written one, and the two do not match completely.

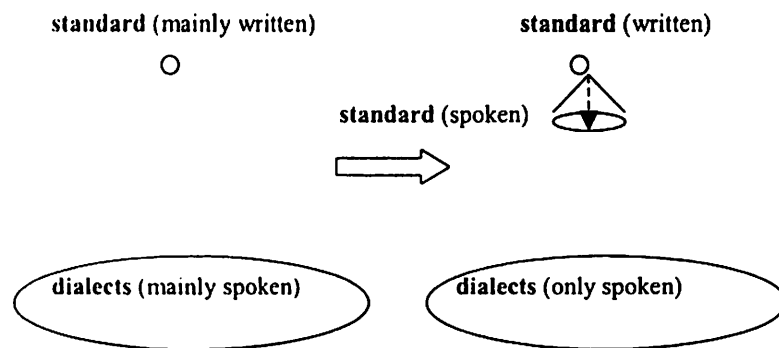


Figure 4. Transition from Type A to Type B Repertoire

4.1. Historical development

The transition from Type A to Type B may have occurred in the fifteenth till seventeenth century in England,²³ and it stretched out over at least 400 years in Germany (fifteenth – nineteenth century). In Scandinavia, it was as late as in the early 1800s that the previous 'oratorical style' closely modelled on writing, and the informal conversational style of the educated upper middle classes merged to form a spoken standard variety (Sandøy, in press). This merger took place first in Denmark and led to a complete congruence of the standard variety with the Copenhagen model of educated middle class speech (Pedersen, in press a and b), while Sweden followed but retained some difference between the standard and Stockholm speech. In Norway, the Danish written standard became norwegeianised in the process of being used for spontaneous speech, first by a Norwegian phonetic/phonological substrate, later by active and conscious Norwegianisation of the whole language, modelled on the Oslo dialect, to form one of the two modern standard varieties, i.e. what is known today as *Bokmål* (Widmark, in press).

The newly emerging endoglossic oral standard varieties continued to compete, as did their written predecessors, with the old exoglossic standard varieties (mostly Latin), but sometimes it also had an additional competitor: a newly introduced exoglossic prestige language. An example among many is English on Cyprus which competed with the two Greek standard varieties during the British rule 1878–1960 (cf. Karyolémou 2001).²⁴

The spoken standard was usually propagated and spoken by the cultural elites (the intelligentsia) first.²⁵ Until around 1900 (in some places until the present day), the amount of standard spoken in everyday life by the large majority of the population was minor, since the standard was a sign of social and cultural distinction and access to it depended on social class membership and education. The lower classes, certainly in the rural parts of Europe, continued to use the dialect for face-to-face interaction. Some sectors of the population, particularly those who migrated from the countryside into the cities, acquired some kind of (fossilised) learners' variety of the standard, such as *Missingsch* in Northern Germany, *italiano popolare* in Italy, *kyökki* ('kitchen') Finnish in Helsinki²⁶, or *trasjanka* (lit.: a mixture of hay and straw) in Belorussia. On the other hand, the dialects were often known and spoken by the upper (bourgeois and feudal) classes in adequate contexts in addition to the standard, i.e., in their case, we find diglossia with 'bi-dialectalism'.

In many areas of Europe, the emerging spoken standards differed (and continue to differ) from the written standard, sometimes considerably.²⁷

This is easy to explain. First of all, a spoken standard is subject, as all spoken varieties are, to the requirements of face-to-face interaction and on-line language processing under conditions of almost perfect speaker/hearer synchronisation. Structures which may be possible in writing/reading do not always survive such a transfer to orality (e.g. in syntax).

More important in the context of the present discussion is the fact that the written standard, once established, quickly becomes more conservative than the spoken standard. The reason is that literacy and codification restrict the amount and the speed of innovations in the written standard, while the spoken standard will develop much faster, particularly when it is closely related to the vernacular of a certain region (cf., for Hungarian, Kontra 1995). A case in point, and an example for the continuing influence of a regionalised group of speakers on an already fully established standard, is Copenhagen. Innovations in the urban vernacular often come from below (from 'Lower Copenhagen speech'). They then make their way up into High Copenhagen speech, which is more or less identical with the spoken standard variety which changes accordingly.²⁸

As a consequence, the standard variety spoken in the more peripheral parts of the language area where the vernacular is highly divergent from it may appear to be more 'norm-conforming' and therefore more 'pure' to the ear of a conservative observer than the innovating standard in its region of origin. This is the reason for which the German standard, believed to have originated in the Upper Saxonian area was considered to be more elegantly observed in northern Germany from some point in time in the late eighteenth century onwards, where it was completely detached from the Low German dialects and not subject to change 'from below' as in Saxony. Equally, the Danish spoken in the Norwegian capital Christiania (Oslo) in the early nineteenth century was often taken to be closer to the 'true standard' than Copenhagen speech.

The relationship between written and spoken standard will be slightly different in a community in which the written standard was introduced on the basis of language planning. Here, the spoken standard may profit from and older tradition (common language). The archaising written Greek *katharevousa* before 1976 is an example; instead of developing a spoken variant, this function was largely taken over by the *demotiki*, a historically developed common/standard language (Mackridge 1999).²⁹

4.2. Europe today

Stable diglossic situations in which the base dialects and the spoken national standard are structurally clearly separated from each other and used in separated domains do not seem to be frequent in Europe today, although they surely were in the nineteenth century and before (cf. Widmark, in press, for Skandinavia). A possible example are Trumper/Maddalon's (1988:232-233) so-called "micro-diglossic" or "pseudo-diglossic" areas of Italy, such as Calabria, Lucania, Emilia-Romagna, or the Marche, where, according to these authors, we find no regiolects, but highly diversified base dialects, and strictly separated domains of usage for (regional) standard and these dialects (also cf. Sobrero 1996).³⁰ Another example would seem to be the surviving dialects (*langues régionales*) of French (such as the Vallonnais which is in a diglossic relationship with [standard] French).

However, there are areas in which what might be called attenuated forms of diglossia have emerged. There are two forms of such attenuation:

(a) the old dialects have been levelled out, partly in the direction of, and very often under the influence of the standard variety; in some cases, the standard has been de-standardised as well. However, no standard dialect continuum has (yet) emerged (as in diglossia, see below, Type C). Rather, there is a dialect continuum and a separate standard continuum. Such double continua have been claimed to exist in northern Italy, in parts of Spain (particularly the Aragonese and Leonese area; cf. Martín Zorraquino et al. 1996), and in the more peripheral parts of the Dutch language area such as the Limburg province (which, however, has a tendency to shift towards Type D³¹).

(b) dialect and standard are two separate systems, but these are in close contact with each other. Contact between standard and dialect in this case follows a route which is well known from studies on bilingualism and language contact, i.e. it goes from code-switching over code-mixing to the fusion of the two varieties (Auer 1999). In the latter case (fusion/hybridisation), intermediate forms and thus a dialect/standard continuum emerge (Type C, below). Characteristically, areas which display this development show frequent turn-internal code-switching, but also frequent (non-functional) code-mixing between standard and dialect, which are often combined in the same syntactic frame and even within a word (cf. Berruto, in prep.). This means that the allocation of standard and dialect to separate usage domains is abandoned, and both of them are used within the same conversation in many situations (with a lesser degree admixture of dialect in more formal ones). This development is typical of many parts of Italy (cf. Alfonzetti 1998 for an application to Sicily) and of the south of

Germany (particularly Bavaria), as well as of Austria (but probably excluding the cities).

Since, in the long run, frequent code-switching or code-mixing usually has the effect of making the languages in contact converge, double continua and mixing/conversational switching tend to co-occur (cf. Berruto, in prep., for Italy).

In attenuated forms of diglossia, both varieties of the repertoire are structurally and attitudinally (ethno-dialectologically) kept apart, and can usually be identified by speakers and linguists; they have their own prestige, one attached to formal, official language use and writing/literature, the other to regional identity. However, there is a strong tendency for these repertoires to develop into Type C.

Excursus: The role of the capitals in the emergence of written and spoken standard varieties

Since nation-building in Europe usually implied the consolidation of a centralistic power system with the capital as the focal point, it is not surprising that the emergence of a (spoken) national standard language became linked to the capital as well in these cases: the consolidation of a centralistic society went hand in hand with the consolidation of a centralised ('centripedal') language norm. Often, the emerging national language was (roughly) *identical* with the variety spoken by the élites of the capital: Copenhagen speech for standard Danish (*rigsmål* being identified with the High Copenhagen variety), the Stockholm area for standard Swedish, the *Île de France* for standard French, the London area for standard English and the Edinburgh area for standard Scots, the middle-Russian Moscow dialect for standard Russian, or much more recently, the dialect of Luxembourg City for standard *Letzebuërgesch*.

In other cases, the emerging national language was a regional koiné in which the dialects of a larger area around the capital were levelled out. In this case, the standard is not identical with (or just a written version of) the variety spoken in one urban centre. An example is the koiné formed on the base of Castilian dialects with elements from the other Romance dialects spoken in the area, which developed into standard Spanish, or the modern Greek standard language – *Koini Nea Eliniki* – which, as its predecessor, the *Dimotiki*, is based upon (non-Albanese) varieties of Athens and/or³² the Peloponnes; other examples are the Holland dialects for Dutch and the eastern dialects of Norway for *Bokmål*. Standard English is surely centred

on educated London speech but it can also be described as a koiné of the south-eastern dialects of the Londoners and immigrants from the midlands.

In both cases, the developing standard variety spread from the political and cultural centre to the periphery over time; for instance, Castilian spread to the peripheral areas of Spain, e.g. to the south, from the fifteenth century onwards, and the process continued until the Franco era when it threatened to take over the Galician and (to a lesser degree) the Catalan area as well.³³ However, the spread did not occur in the wave-like way of a neogrammarian sound-change, but followed the city-hopping pattern, i.e. it reached the urban centers first and only (much) later the countryside.

Given the link between political centralisation and the formation of national standard varieties, it follows (a) that the dislocation of the political, cultural, economic etc. centre may also dislocate the linguistic model area, and (b) that a politically/culturally/economically polycentric area may result in a polycentric or geographically neutral standard.

An example for the first case is the Portuguese standard which originated from the vernacular of Coimbra but later developed on the basis of the standard of Lisboa. For the second case, consider Poland: here, the standard emerged as a mixture of dialect koinai each of which originated from one of the two competing political and intellectual centres, i.e. the (more often dominant) Greater Poland dialects around Posnan and the (sometimes dominant) Little Poland dialects around Krakow (Mazur 1996). A similar example is Estonia, where Tallinn and Tartu competed for influence and developed two different written norms from the sixteenth century onwards (North and South Estonian), until the latter succumbed to the first (Keevallik and Pajusalu 1995). In Germany and Italy, where standardisation stretched out over a long period of time during which the centre of power was polymorphous, there is no one area on which the present-day standard is founded – despite the association of various areas with the standard language at different times in history (Tuscany and later the north Italian cities Milan, Turin and Genova in Italy, the Eastern Middle German and Upper German area, later the Hanover area in Germany). It is a remarkable reflex of the decentralised power structure in these states that the capitals (Rome and Berlin, respectively) never became the linguistic models.

An interesting case is the modern Bulgarian standard language, an archaising fabricated variety which was modelled in the nineteenth century on urban varieties spoken in the north-east of the country (mainly the dialect of Veliko Turnovo, the ancient Bulgarian capital of the second pre-Ottoman Bulgarian Kingdom). After independence, Sofia in the west of the new state became the capital, where a very different variety of Bulgarian

was spoken by the population. For a long time, the new standard variety used by the upper classes in Sofia therefore contrasted sharply with the Sofia vernacular, and it is only in the later part of the last century that regional Sofia features became prestigious and entered the standard variety as spoken in the capital and elsewhere (Videnov 1999).

5. Type C areas: Diaglossia

Presumably the most wide-spread relationship between dialect and standard in Europe today is what (following Bellmann 1997) I will call a *diaglossic* one (Fig. 5). A *diaglossic* repertoire is characterised by intermediate variants between standard and (base) dialect. The term *regiolect* (or regional dialect) is often used to refer to these intermediate forms, although the implication that we are dealing with a separate variety is not necessarily justified. More usually, the space between base dialect and standard is characterised by non-discrete structures (standard/dialect continuum). A good degree of levelling between the base dialects is involved, which at the same time implies advergence to the standard. The internal cohesion of the repertoire is achieved by one-sided restrictions of co-occurrence between the linguistic variables (Auer 1997; Berruto in prep.).

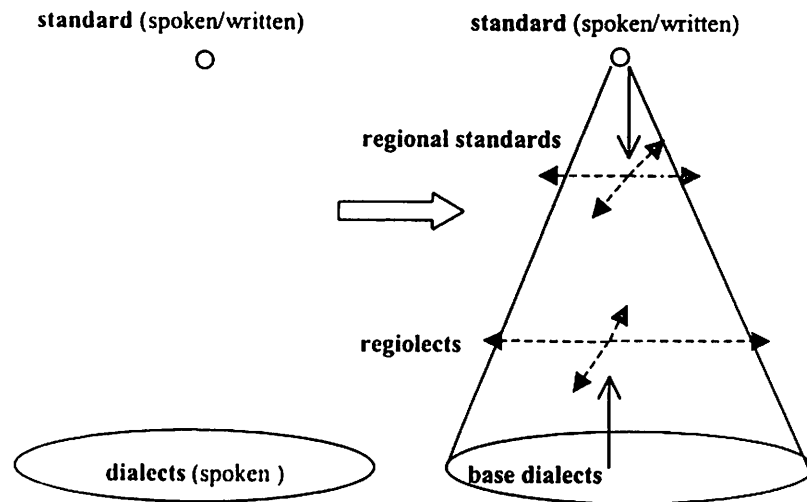


Figure 5. From Type B (diglossia) to Type C (diaglossia) repertoires

In discourse, code-switching between standard and dialect is complemented by processes of step-by-step dialectalisation and standardisation, i.e. speakers can change their way of speaking without a clear and abrupt point of transition between dialect and standard. These changes nevertheless depend on and index subtle situational changes (Auer 1988).

The intermediate forms often fulfil a sociolinguistic function by enabling their users to act out, in the appropriate contexts, an identity which could not be symbolised through the base dialects (which may have rural, backwardish or non-educated connotations) nor through the national standard (which may smack of formality and unnaturalness and/or be unable to express regional affiliation).

5.1. Historical development

The emergence of C-type repertoires is a relatively late phenomenon, usually of the late nineteenth or early twentieth century.³⁴ The exact nature of the transition from diglossia to diaglossia is not yet clear. The most straightforward model would start from the assumption that the (spoken) national standard variety was the target of dialect changes which in the end led to the filling up of the space between base dialects and standard. However, this may not be the appropriate picture everywhere. For Germany, Mihm (2001) has suggested an alternative model of phonological standardisation by at least two processes of *Überschichtung* (superimposition of acrolectal strata) in the repertoire which leads from (what here is called) type A to type C. In a first process of *Überschichtung* in the fifteenth to seventeenth century, a spoken standard came into being which affected many parts of morphology and syntax and some parts of phonology, while other features, particularly in phonology and in the lexicon, remained dialectal. At that point, the A-type repertoire became diglossic (type B). In a second process of *Überschichtung* in the nineteenth and early twentieth century the modern standard was formed by superimposing another variety on the previous repertoires which was much less regionalised, particularly in phonetics. The second-order superimposition of a (new) standard variety on an already existing standard-dialect repertoire redefined the former standard variety as an intermediate, non-standard variety in terms of its areal and social reach, and its prestige. Thus, the first-order standard variety remained in use, but was 'lowered' to a regional standard or a regiolect. The result of the second superimposition of a national standard variety is a continuous, non-diglossic repertoire structure (diaglossia, type C) (see Fig. 6).

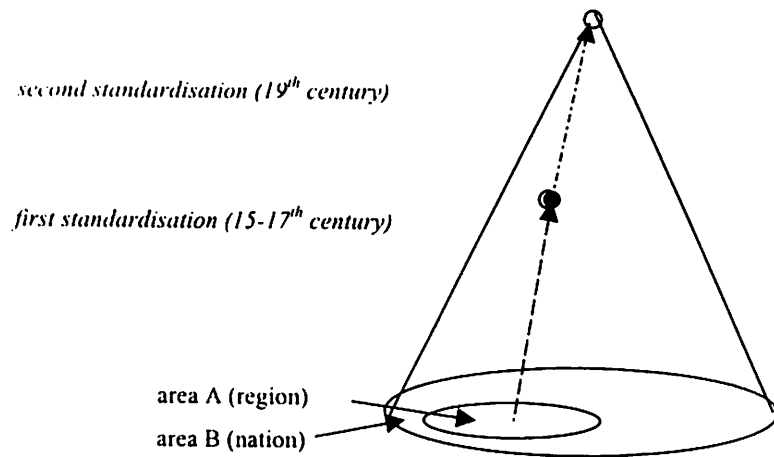


Figure 6. Superimposition of various layers of standards in a repertoire (for Germany, following Mihm 2001)

In diaglossic repertoires just as in diglossic ones, the endoglossic standard variety (or more precisely, the standard pole of the repertoire) may compete with an exoglossic standard. Standard Catalan vs. standard Castilian in Catalonia/Spain is an example. Here, the relationship between dialectal and standard Catalan is diaglossic, while the relationship between Castilian and dialectal Catalan is diglossic.

5.2. Europe today

Diaglossic repertoires are found in many parts of Europe today: in those areas of Norway in which *Bokmål* is the standard variety (particularly in the south-eastern part), in some parts of the Netherlands excluding the Randstat area, in most of Flemish-speaking Belgium, in many southern parts of Germany (Alemannic and Eastern Franconian, as well as the Rheno-Franconian and Moselle-Franconian area), in the larger Vienna area in Austria, in Scotland, in Moravia (cf. the regiolects „interdialects“ of Czech sociolinguistics), in Poland with the exception of the previously German parts, in which the loss of the dialects has further progressed (Mazur 1996), in the south of Spain (with the old koiné mentioned on p. XX above as the one extreme, and Castilian standard as the other; cf. Villena 1996), in the

Catalan language area, in Cyprus (where according to Karyolémou 2000 a mediating variety between base dialects and standard Greek, called common Cypriot by some, has been gaining force since the 40s¹⁵ of the last century), in Bulgaria (with tendencies to D/I; cf. Videnov 1999), in Sweden³⁶, in Finland, in most parts of England and presumably in areas of Italy, etc. The larger cities have played an important role in the emergence of these intermediate forms of speaking. In the process of industrialisation large segments of the population migrated from rural areas into the urban conglomerates, which favoured the levelling of the base dialects, but also the re-interpretation of what used to be dialectal variation as social variation. The new urban substandards have in turn, and in a second step, influenced the surroundings of the cities.

The main process in regiolect formation is dialect-to-standard advergence, but there are two caveats to this statement. First, a minority of the forms that win out in interdialectal levelling and in the emergence of urban sociolects fail to correspond to those of the standard; examples can be taken from the spreading urban regiolects in the areas around London, Vienna³⁷, Turin³⁸, Munich³⁹; Sevilla⁴⁰. Also, regiolects may develop linguistic innovations of their own which have no basis in the standard variety, nor in the dialects (cf. the spread of the coronalisation of the palatal fricative (*ich-Laut*) in the middle German regiolects; Herrgen 1986).⁴¹ Typically, but not exclusively, we find simplification towards more unmarked forms.

The second caveat is that complementing dialect-to-standard advergence, the standard variety may also increasingly tolerate regional features, leading to regional standards with a dialectal substratum (as in Sweden⁴², Germany, France, England, Poland, etc.) which in some countries enjoy more (and growing) prestige than in others. Destandardisation towards the urban sociolects is presently observed in many former communist countries as well, particularly in Bulgaria, Russia, Belorussia and Ukraine.⁴³

A special case of de- and subsequent restandardisation is the splitting off of a new national standard variety from another national standard variety. For instance, the spoken Dutch standard variety in the Netherlands is slightly different from the one used in Belgium (although the codified norm is the same), and has diverged from it over the last decades in phonology and phonetics, but not in the vocabulary (cf. van der Velde 1996; Geeraerts, Grondelaers and Speelman 1999); the German standard variety spoken in Austria and the one spoken in Switzerland also slightly differ from the one codified for Germany, and at least in the Austrian case, the divergence seems to increase; and the standard Swedish spoken in Finland represents a

specific norm of Swedish standard which can be called national (according to Ivars 1996). Interestingly, 40 years of political separation and a severe reduction of communication did not lead to the formation of two different standard varieties in the German Democratic Republic and the Federal Republic of Germany. More recent examples include Croatian/Serbian/Bosnian diverging as new national standards from each other.⁴⁴

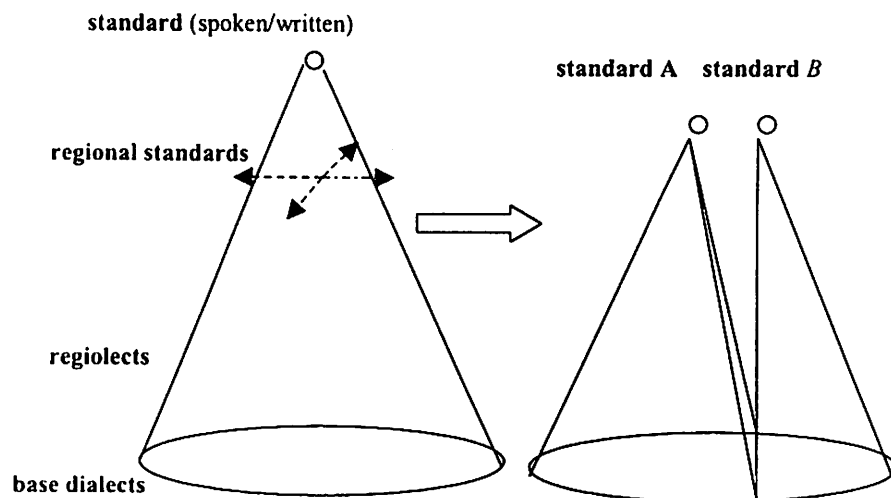


Figure 7. Destandardisation and the development of new national standard varieties

In these cases, we witness de-standardisation, regionalisation and (sometimes) re-standardisation in another national context. As in all cases in which language is used as a symbol of national identity, the matter has little to do with structural distance, but a lot with language ideologies. From these cases, one needs to distinguish divergent repertoires which emerged as a consequence of the political separation of two speech communities, only one of which developed a standard variety (Romania vs. Moldavia, Bulgaria vs. Macedonia, Portugal vs. Galicia).

6. Type D areas: Dialect loss

Both diaglossic and diglossic repertoires can lead to (base) dialect loss, or Type D repertoires. In the first case, the dialect is more and more infiltrated by standard features, in the second case, it is retained in a more conservative fashion, but it is used less and less (cf. Thelander 1983). In both cases, the base dialect features receive a strongly negative evaluation.

6.1. Type D/1: From diaglossia to dialect loss

In a diaglossic context, the 'lowest' level, i.e. the most dialectal pole of the standard-dialect continuum, may be lost; the linguistic forms with the most restricted geographical reach then disappear. As a consequence, the repertoire is restructured: regiolectal forms (intermediate forms – sometimes the standard varieties of previous times!) are now looked upon as the most basilectal way of speaking. The base dialect loses its function as a marker of local belonging (*Ortsloyalität*), which is transferred to the regiolect (Fig. 7). This type of repertoire is found, for instance in the Ruhr, in the Berlin area and also in Upper Saxony in Germany, in Bohemia in the Czech Republic, and in the more conservative areas of Denmark (Funen, southern Jutland, cf. Kristensen 2003).

But in a diglossic repertoire, there is no clear-cut separation of standard and dialect. Therefore, the breaking away of the most local, most dialectal forms is a matter of degree, and it can gradually affect large parts of the dialect/standard pyramid until almost nothing is left but the standard variety (Fig. 8). This is the stage reached in most of Denmark⁴⁵, the Randstad area (Amsterdam/The Hague) in the Netherlands, and, somewhat lagging behind, some parts of Flanders, particularly the Brabant, with Brussels and Antwerp spearheading the development⁴⁶ (cf. Willemys 1997); it is also true for parts of Poland (particularly the west). Another good example is France where the final stage of this development was already initiated shortly after the French revolution but only came to completion in the middle of the twentieth century.⁴⁷ Despite some recent trends to recover the dialects, they are only used (on a dialect/regiolect continuum) by older people in some rural areas of France today. What is left are regional traces of the former dialects in the vernacular, which are stronger in the countryside than in the cities, and more accentuated among male, less mobile, less educated and older speakers (Gadet, in prep).

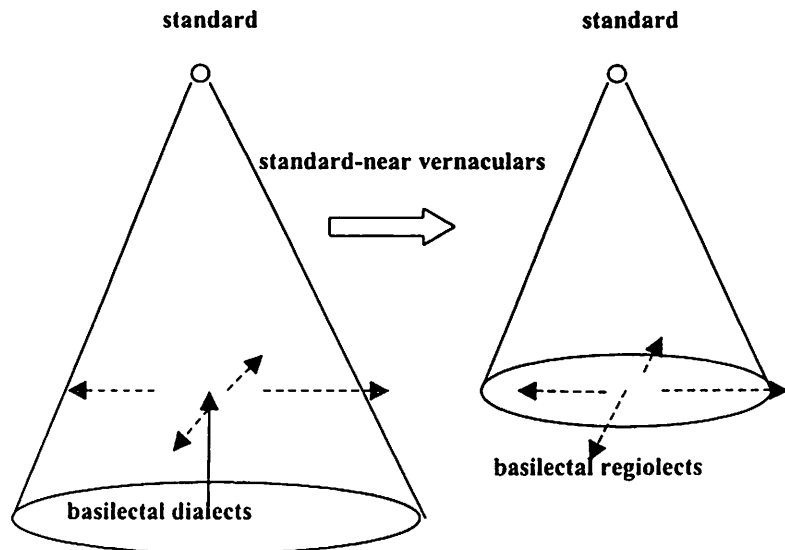


Figure 8. From Type C (diaglossia) to Type D/1 (dialect loss) repertoires, first stage

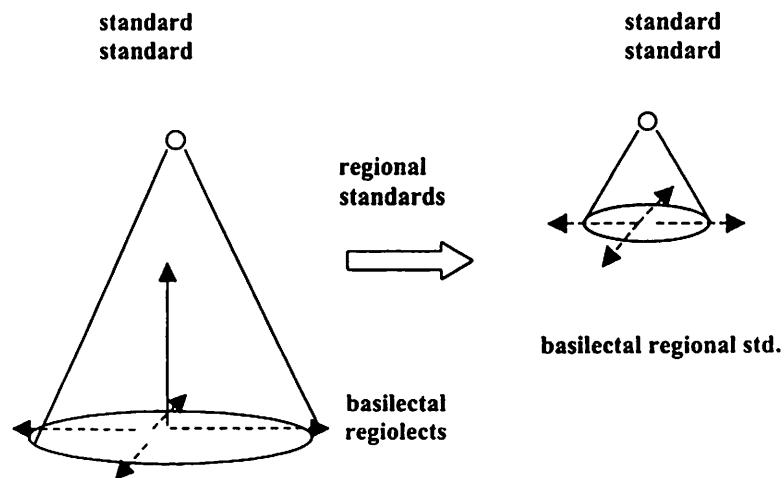


Figure 9 From Type C (diaglossia) to Type D/1 (dialect loss) repertoires, second stage

Areas in which the process of dialect loss after previous diaglossia may be on its way are parts of England as well as presumably some areas of Greece (areas of Athens, Thessaloniki) and Russia/Belorus/Ukraine.⁴⁸

6.2. Type D/2 areas: From diglossia to dialect loss

There is also a direct path from Type B (diglossia) to Type D in which case the old dialects are not handed on to the next generation, i.e. they are lost by shift. In social terms, it is the middle classes and the young women who avoid the (base) dialect first. This sets in motion a mechanism by which the dialect is associated with lower social class, manual (industrial) or farm work, lack of education, etc., first of all in the cities, later in the countryside as well. As a consequence, the base dialect loses prestige and domains of usage; most notably, parents avoid dialect with their children. Since the base dialect is seldom used, speakers' competence in that variety also diminishes, which leads to insecurity and reluctance to speak dialect in more out-group contexts. In the final stage before loss, the attitudes towards the now almost extinct base dialect are usually positive again, and folkloristic attempts at rescuing the dialect may set in – usually without success.

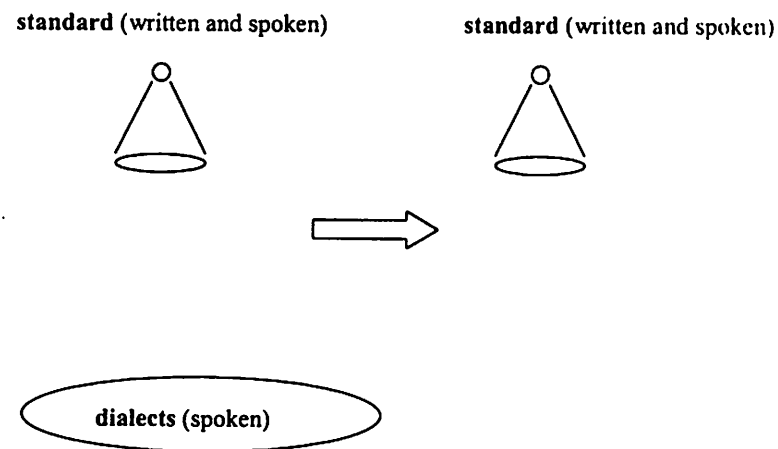


Figure 10. From Type B (diglossia) to Type D/1 (dialect loss) repertoires

This scenario of dialect loss captures the Low German dialect areas in northern Germany and the (northern and eastern) middle German dialects.

particularly in most of Hesse (with the exception of the Frankfurt area which leans towards D/1) and Thuringia (whereas Saxony is D/1).⁴⁹ It can be assumed that it equally applies to Hungary, where the remaining dialects are highly stigmatised and only used in in-group situations,⁵⁰ and perhaps to the remaining Spanish base dialects spoken in isolated areas of Aragonese and Asturian-Leonese parts (with decreasing vitality and tendencies of dialect loss).

What remains in such a case is again the standard variety alone, with some variation due to the usual lack of stability of any spoken language.

Interestingly, processes of divergence from the national standard often set in after the dialects and regional dialects have vanished, usually beginning in the large cities (cf. Røyneland 2001). In other words: the victory of the standard language does not imply a lack of linguistic heterogeneity. On the contrary, speakers seem to feel the need to sound different from the codified standard. The so-called *français avancé* may be mentioned in this context as well as ethnolectal urban varieties (such as Rinkeby Swedish in Stockholm).

7. Some possible exceptions and consequences

The four types of repertoires cover almost of all Europe, diachronically and synchronically. The historical starting points for the development of national standard varieties are, in Coseriu's sense, "primary dialects" (1980) which are older than the national standard varieties that have developed out of them.

However, there are some few areas which raise questions since they are not easily (or not at all) accommodated by the model. A first counter-example might seem to be those large parts of the Slavic language area, in which geographical diversification set in late. For instance, diversification in and between the east Slavic varieties (Russian, Belorussian, Ukrainian) does not seem to be older than the fourteenth century, and dialect diversification in these languages continues to be low compared to most other European languages. At the same period when the present-day dialects of east Slavic emerged, there were already tendencies to develop a common high variety based on a power centre (the Moscow Rus'). Different as this development may be from the older Germanic and Romance varieties (where geographical diversification set in much earlier and the vernacular varieties are older), it is no counter-example to the model presented here since the early Russian common language certainly was no standard variety (in this function, Old Church Slavonic was used).

Characteristic for these areas is that the emergence of an acrolectal common variety and dialect diversification on the basilectal level went hand in hand.

However, there is at least one nation-state and one national standard language for which it is claimed that no dialects have ever existed, and that a national standard variety has been codified directly on the basis of a common language. This is Iceland, which presents a true counter-example to my typology.⁵¹ Since Iceland was colonised by settlers who came to the island around 900 from a well-circumscribed part of west Norway, the vernacular varieties brought along by them would seem to have been relatively uniform, and it may also be assumed that there was levelling induced by the migration process itself (as is well-attested for modern migrant communities as well). What is remarkable is the fact that little geographical variation developed over a thousand years after the colonialisation. Milroy and Milroy (1985) explain this exceptional situation by dense networks in a closely knit, but socially non-stratified society inhabiting a small territory; these networks would not only have kept language change at a minimum, but also prevented symbolic barriers between social or geographically defined groups from being erected.⁵² Other factors such as a high level of literacy already during the middle ages may have played a role as well.

Finally, an interesting and also deviant case is that of (Old) Greek and the Hellenistic koiné. With some justification, it can be claimed that Greek had already gone through the development from Type A to Type D once when the other European languages had not even started to develop an endoglossic standard. The Old Greek dialects became roofed by a common standard variety based on Attic, which started out as some kind of levelled regional variety and spread to more and more Greek-speaking areas until it was made the standard variety under Philipp and Alexander the Great (the Hellenistic koiné). The new prestige variety gradually led to the disappearance of the old dialects. The present-day dialects of Modern Greek are not the offsprings, of the Old Greek dialects, but the result of a secondary/tertiary dialectalisation on the basis of the Hellenistic koiné.

In sum, on a map of Europe with the four types of standard/dialect relationships drawn out, the most wide-spread types would surely be Type C, the diaglossic one, and Type D, in which the base dialects have disappeared. The very fine-grained dialect differences which have been documented on the European dialect atlas based on materials which are now 50-100 years old are vanishing. On the other hand, many diaglossic repertoires are quite stable, with the regiolects even gaining the upper hand over the standard varieties, and sometimes introducing new variation.

This empirical generalisation has some impact on social dialectology (method and theory building) in a European context. If it is true that the European sociolinguistic situation has been characterised until very recently (Type D) by the co-existence of the old local dialects and the national standard varieties (with the latter emerging from, but also influencing the former), then models for the description of linguistic variation in the European languages are called for which sufficiently integrate these dynamics. Whatever such theories would look like in detail, they would need to start from the assumption of two or more linguistic systems „in contact“ and therefore (by empirical necessity) differ considerably from the dominant variational paradigm in linguistics, the latter being conceived in order to meet the needs of a completely different sociolinguistic context (i.e., that of secondary and tertiary dialects).⁵³ It is only for the last stage of repertoire development, i.e. Type D (dialect loss), that the variational paradigm seems to be suited.

Notes

1. Many colleagues have been of help in the preparation of this paper. In particular, I would like to thank Angel Angelov (Sofia), Ludo Beheydt (Louvain-la-Neuve), Françoise Gadet (Paris), Peter Gilles (Freiburg), Frans Hinskens (Leipzig), Jeffrey Kallen (Dublin), Maria Kakridi (Athens), Marilena Karyolemou (Cyprus), Miklos Kontra (Budapest), Jiří Nekvapil (Prague), Bengt Nordberg (Uppsala), Inge Lise Pedersen (Copenhagen), Unn Røyneland (Oslo), Anne-Catherine Simon (Louvain-la-Neuve), Marja-Leena Sorjonen (Helsinki), Mats Thelander (Uppsala) and Juan Villena (Malaga) for their partly extensive comments. The paper continues to be marred by my own ignorance on many European sociolinguistic contexts, and all mistakes are of course made entirely on my account.
2. Relatively close in scope and intention to my own is Berruto's typology (1989a) whose "dilalia" seems to be similar to my "diaglossic" Type C.
3. Note that this does not imply that the non-standard varieties are void of (overt) prestige, although this happens to be the case quite regularly.
4. This last criterion is an attitudinal one; it is not the fact of codification (such as the existence of a grammar and a dictionary) which makes a standard variety, but the fact that its speakers think that such things should exist and that, where they exist, they should determine how members of that society ought to express themselves in situations in which the standard is required.
5. Old (Church) Slavonic was a 'commissioned language' for the missionary activities in the West Slavic area, 'constructed' by Cyril and Method in the ninth century on the basis of Old Bulgarian. Since it played a role as the

national standard of Bulgaria, and in its Russian, Romanian etc. variety also as a national standard in these countries, the status of Old Church Slavonic is ambiguous.

6. The terms endoglossic vs. exoglossic correspond with Kloss' (1976) distinction between *Binnendiglossie* and *Außendiglossie*.
7. Old Church Slavonic (Old Bulgarian) started out as an exoglossic standard variety in Russia, Romania, etc. despite local adaptations, particularly in orthography and phonetics. However, the use of Old Bulgarian in the late period of the first Bulgarian kingdom (863-1018) is an early case of a European H-variety (*Hochsprache*). Since this language also underwent some *Ausbau* as the language of christianisation, it may also qualify as a standard.
8. I am referring to the Arabic part of Spain, of course.
9. This and the following figures are inspired by Hugo Moser's representations of the German language repertoire through history (1950). One should be careful not to lose sight of the simplifications which are inherent in any model of this type; in our case, this caveat refers in particular to the distance between the base dialects (ground line) and the standard variety (top point) which is systematically ambiguous between an attitudinal and a structural interpretation. Also, the model is simplified since it presupposes that the distance between any variety in the lowest circle and the standard is the same. Obviously, the opposite is usually true: some dialects are structurally and attitudinally closer to the standard variety than others. See below, section 4.
10. This *obecná čeština* ('Common Czech') has never fully submitted to the new, artificially created (written) Czech standard variety of the nineteenth century (the *spisovný jazyk*, 'literary language'); cf. Nekvapil 2000, Sgall et al. 1992:174 for details.
11. The term roof is used in a rather loose way here, deviating from Kloss' original version (Kloss 1978) where the notion *Überdachung* is restricted to dialects and 'their' standard language (cf. Berruto 2001 for a valuable discussion).
12. J. Kallen (pers.comm.). Kallen (1997) points out that the notion of (English) dialects is disputed in the Irish case, since it is not clear whether the base dialects were a uniform variety with a strong Irish substratum (or even a post-creole variety), or a rural dialect of English extended from British (plus Scottish?) English.
13. Cf. Angelov 1999 for a discussion.
14. Cf. Martin Zorraquino 1998.
15. Standard Czech was fabricated by linguists such as Dobrovský (around 1800) and Jungmann on the basis of written Czech of the late sixteenth century. The emerging Bohemian standard of the time around 1800 was rejected since it was taken to be associated with one region (Bohemia, leaving out Moravia, Silesia and possibly Slovakia) and since it was taken to be 'degenerated' by dialect and German influence. In addition, it lacked the prestige of a literary language since

- most serious written publications were in German. For details cf. Sgall et al. 1992:171.
16. Pedersen (in press).
 17. Cf. Paunonen 1994.
 18. Of course the degree of codification of Standard Swiss German or of *Nynorsk* is considerably higher than that of High German or English in the fifteenth or sixteenth century when graphical variation was still a marker of distinction (cf. Mihm 2000).
 19. The majority of Norwegians use Bokmål as their standard variety which is structurally very similar to the urban dialect of Oslo.
 20. The present-day Norwegian and Swiss situation is the result of a process in which the dialects have been upgraded in attitudinal terms; compared to the period 100 years ago, they are on the one hand seen in more positive terms today than they used to be then, on the other hand they are, at least in the case of German-speaking Switzerland, more strictly separated from the standard variety.
 21. For instance, Swiss standard German is used in university lectures, in the parliament, or in the TV news broadcasts, but also when talking to foreigners, service personnel in hotels and restaurants, etc.
 22. Cf. Siebenhaar 2000 and Christen 1998 for empirical studies on dialect levelling in Switzerland, Sandøy (in press) and Røyneland (in press) for a discussion of the Norwegian situation.
 23. Cf. Cheshire/Milroy 1993:10.
 24. English continued to be used even after independence, particularly in jurisdiction.
 25. When an exoglossic standard variety was already part of the repertoire, these elites sometimes were in the paradoxical situation of having to learn the endoglossic standard without the support of a dialect since their everyday language was this exoglossic standard. A good example is Finland in the nineteenth century where the elites propagating the codification and Ausbau of a Finnish standard language usually spoke Swedish and were not competent in a Finnish dialect (cf. Paunonen 1996, 1997).
 26. Here, the case is somewhat more complicated since Swedish-speaking immigrants played a role in the formation of *kyökki* Finnish as well.
 27. This difference may be regionalised, leading to regional standard varieties, in a larger language area such as the German one, while regional traits may be absent from it in a smaller language area such as the Danish one.
 28. Cf. Pedersen 1996, Kristensen (MS). It should be added that Low Copenhagen speech also has a direct impact on the vernacular speech forms in Denmark today.
 29. The *katharévoussa* was given up in 1976 as the official standard language and replaced by a demotic variety, which, however, has integrated many structural elements of the *katharévoussa*. This suggests that, contrary to what the term „diglossia“ used in the literature on modern Greek for the relationship between *demotiki* and *katharévoussa* implies, the archaising written standard (*katharévoussa*) and the spoken (common and also written) (standard?) language used in everyday life (*demotiki*) formed a continuum rather than two strictly separated varieties. The term diglossia is misleading also because the ‘lowest’ stratum of the Greek repertoire, the dialects (and ‘idioms’), are by no means to be identified with the so-called demotic variety, the latter having been used in (literary) writing a long time before it was promoted to the national standard (in the version of the *Koini Nea Elinikí*). The *demotiki* was also codified by a grammar as early as 1938.
 30. If this description is accurate, we are dealing with diglossia in the best sense of the word, and the addition of a „pseudo“ seems unnecessary. It should be mentioned though that other Italian dialectologists such as Berruto (1989a, b) postulate (my) Type C for the whole of Italy. Berruto (1989b) presents a picture of Italy as a whole in which the dialects are levelled (*koineizzazione*) and where there is convergence towards the standard (*italianizzazione*), while the standard variety destandardises (*dialettalizzazione* with the result of the creation of the *italiano popolare*) and standard and dialect combine into a hybrid (*ibridazione*). Basically, the discussion in Italian social dialectology has centred around the issue of how vast an area is covered by the emerging koinai, and this issue also offers the possibility of defining areas of diglossia (or *dilalia* in Berruto’s terms).
 31. Cf. Map 1 in Hagen 1989.
 32. The Greek scholars seem to disagree here; cf. Delveroudi 1999:562 for the first view (Peloponnesian koiné), Alexiou 1982:162 for the latter.
 33. Cf. Villena (in press).
 34. Pedersen, in press (a) tracks this development through Danish history, showing the emergence of regiolects under the influence of the standard variety during the first part of the twentieth century.
 35. The process has been strengthened by the displacement of a large segment of the Greek speaking population from the northern Turkish part of the island in 1974 which led to the dissolution of traditional network structures that used to shelter the traditional dialects.
 36. Sandøy, in press. In the 70s of the last century, Thelander in his investigation of Burträsk in northern Sweden found a C-type repertoire (cf. Kristensen and Thelander 1984), and Nordberg/Sundgren (MS 1999) in a recent follow-up (real time) study on Eskilstuna also found a persistence of diglossia.
 37. Moosmüller/Vollmann 2001.
 38. Berruto (in prep).
 39. Bücherl 1982.
 40. Villena (in press).
 41. For this reason, the horizontal arrows in Fig. (5) reach beyond the outer limits of the pyramid linking the base dialects to the standard.

42. Sandøy (in press).
43. For Poland, Mazur (1996:97) suggests that de-standardisation of this type may have been promoted by the purposeful recruitment of the new intelligentsia among the workers and farmers during communist times. For Belorussia, cf. Woolhiser (2001).
44. As is well known, the 'Serbocroatian language' had a fragile status and was itself an invention of the late nineteenth century.
45. See Pedersen, in press, for a detailed account of these developments in Denmark in the second half of the twentieth century.
46. In Brussels, the Dutch standard variety is dominant (and competes with French, the traditional H-variety used by the bourgeoisie), in Antwerp, an urban regiolect has taken over which is looked upon as standard Dutch by the speakers; cf. Willemyns 1997:140. Willemyns dates the transition from „dialect dominance“ to „standard dominance“ to the period between World War I and II in the Netherlands, but to the period after the second World War in Belgium.
47. Gadet (in prep.). According to estimates, hardly a quarter of the population of France spoke (standard) French as their mother tongue around 1914. Today, the old dialects have more or less disappeared, while a kind of regional standard has survived.
48. Tzitzilis 2000:88.
49. Cf. Dingeldein 1996, particularly the map on p. 130.
50. Miklos Kontra (pers. comm.) draws my attention to the fact that the Hungarian dialects survive nevertheless in the Hungarian minorities in Slovakia, Romania, Ukraine etc. (cf. for details Lanstyák and Szabó Mihály 1996, Csermicskó and Fenyvesi 2000).
51. There is some variation in Icelandic which, however, is restricted to phonetics and vocabulary (cf. Þráinsson and Árnason 1992, Árnason 1987).
52. In Old Irish, no dialectal variation is attested but since dialect differences do show up at later stages of Irish history, it is usually presumed that they were covered under 'standardised' Old Irish which was already powerful and accepted enough to represent the language in writing alone.
53. See Auer (1995) for an overview, as well as Thelander (1976), Moosmüller (1987) and Auer (1997) for further details.

References

- Alexiou, Margret
1982 Diglossia in Greece. In *Standard Languages*, W. Haas (ed.), 156-190. Manchester University Press.
- Alfonzetti, Giovanna
1998 The conversational dimension in code-switching between Italian and dialect in Sicily. In *Code-Switching in Conversation*, P. Auer (ed.), 180-214. London: Routledge.
- Angelov, Angel G.
1999 The political border as a factor for language divergence. Unpublished manuscript.
- Árnason, Kristján
1987 Icelandic dialects forty years later: the (non) survival of some northern and south-eastern features. In *The Nordic Languages and Modern Linguistics* 6, P. Lilius and M. Saari (eds.), 79-92. Helsinki.
- Auer, Peter
1986 Konversationelle Standard/Dialekt-Kontinua (Code-Shifting). *Deutsche Sprache* 2: 97-124.
1995 Modelling phonological variation in German. In *Verbale Kommunikation in der Stadt*, I. Werlen (ed.), 9-37. Tübingen: Narr.
1996 Co-occurrence Restrictions between Variables: A Case for Social Dialectology, Phonological Theory and Variation Studies. In *Variation, Change and Phonological Theory*, F. Hinskens, R. van Hout and L. Wetzels (eds.), 71-102. Amsterdam: Benjamins.
1999 From code-switching via language mixing to fused lects: toward a dynamic typology of bilingual speech. *International Journal of Bilingualism* 3 (4): 309-332.
- Bellmann, Günter
1997 Between Base Dialect and Standard Language. In *Folia Linguistica* (Special Issue: ed. P. Auer), XXXII (1-2): 23-34.
- Berruto, Gaetano
1989a On the typology of linguistic repertoires. In *Status and Function of Languages and Language Varieties*, U. Ammon (ed.), 552-569. Berlin: de Gruyter.
1989b Tra italiano e dialetto. In *La dialettologia italiana oggi - Festschrift for Manlio Cortelazzo*, G. Holtus, M. Metzeltin, M. Pfister (eds.), 107-122. Tübingen: Narr.
2001 Dialetti, tetti, coperture. Alcune annotazioni in margine a una metafora sociolinguistica. In *Die vielfältige Romania (FS H. Schmid), Vigo di Fassa, Ist. Cultural Ladin "Majon di Fascegn"*, M. Iliescu et al. (eds), 23-40.
in prep. Dialect convergence, mixing, and models of language contact: the case of Italy. To appear in: *Dialect Change*, P. Auer, F. Hinskens, P. Kerwill (eds). Cambridge: Cambridge University Press.
- Bücherl, Rainald
1982 Regularitäten bei Dialektveränderung und Dialektvariation. *ZfD*. 11. 1: 1-27.
- Cheshire, Jenny and Milroy, Jim
1993 Syntactic variation in non-standard dialects: background issues. In *Real English*, J. Milroy and L. Milroy (eds.), 3-33. Harlow: Longman.

- Christen, Helen
 1998 *Dialekt im Alltag: eine empirische Untersuchung zur lokalen Komponente heutiger schweizerdeutscher Varietäten*. Tübingen: Niemeyer.
- Coseriu, Eugenio
 1980 'Historische Sprache' und 'Dialekt'. In *Dialekt und Dialektologie*, J. Göschel, P. Ivić, K. Kehr (eds.), 106–116. Wiesbaden: Steiner.
- Csernicskó István and Anna Fenyvesi
 1999 The sociolinguistic stratification of Hungarian in Subcarpathia. *Multilingua* 19: 95–122.
- Delveroudi, Rea
 2001 The language question and the modern Greek dialects (1880-1910). In *'Strong' and 'weak' languages in the European Union*, A.-F. Christidis (ed. Thessaloniki: Centre for the Greek Language, Vol II, 561–568.
- Dingeldein, Heinrich J.
 1996 Sprachvarietäten in ‚Mitteldeutschland‘. In *Varietäten des Deutschen*, G. Stickel (ed.), 109–141. Berlin: de Gruyter.
- Ferguson, Charles
 1959 Diglossia. *Word* 15: 325–41.
- Frings, Theodor
 1957 *Grundlegung einer Geschichte der deutschen Sprache*. Halle: VEB Niemeyer. (1948³)
- Gadet, F.
 in prep. France. In *Soziolinguistik – ein Handbuch*, U. Ammon et al. (eds.). Berlin: de Gruyter (second edition).
- Geeraerts, Dirk, Grondelaers, Stefan and Speelman, Dirk
 1999 *Convergentie en divergentie in de Nederlandse woordenschat*. Amsterdam: Meertens Instituut.
- Gilles, Peter
 2000 *Dialektausgleich im Lëtzebuergeschen*. Tübingen: Niemeyer (=Phonai No 44).
- Hagen, Anton
 1988 The Netherlands. *Sociolinguistica* 3: 61–74.
- Haugen, Elmar
 1966 Dialect, Language, Nation 'Dialect, nation, language'. *American Anthropologist* 68. (Reprinted in: J. B. Pride/J. Holmes (eds.), 1972, *Sociolinguistics*. Penguin Books, 97–111.)
- Herrgen, Joachim
 1986 *Koronalisierung und Hyperkorrektur: das palatale Allophon des /ch/-Phonems und seine Variation im Westmitteldeutschen*. Stuttgart: Steiner.
- Ivars, Ann-Marie
 1996 Stad och bugd. Finlandssvenska stadsmål i ett regionalt og socialt perspektiv = *Folkmålsstudier* 37, Helsinki: Föreningen för nordisk filologi.
- Kallen, Jeffrey L.
 1997 Irish English, In *Focus on Ireland* idem (ed.), 1–33. Benjamins.
 2001 Two languages, two borders, one island: some linguistic and political borders in Ireland. *IJSL* 145: 29–63 (*Dialect convergence and divergence across European borders*, J.L. Kallen, F. Hinskens and J. Taeldeman (eds.)).
- Karyolémou, Marilena
 2000 La variété chypriote: dialecte ou idiome? In *La langue grecque et ses dialectes*. A.-Ph. Christidis (Ed.), 111–115. Athens: Greek Republic/Center of the Greek Language.
 2001 From linguistic liberalism to legal regulation: the Greek language in Cyprus. *Language Problems and Language Planning* 25 (1): 25–50.
- Keevallik, Leelo and Pajusalu, Karl
 1995 Linguistic diversity and standardization in Estonian: the history of the active past participle. *DiG* 3: 13–41.
- Kloss, Heinz
 1967 ‚Abstand Languages‘ and ‚Ausbau Languages‘. *Anthropological Linguistics* 9 (7): 29–41.
 1976 Über ‚Diglossie‘. *Deutsche Sprache* 4: 313–323
 1978 *Die Entwicklung neuer germanischer Kultursprachen seit 1800*. Düsseldorf: Schwann. (1952²)
- Kontra, Miklos
 1994 On current research into spoken Hungarian. *IJSL* 111: 9–20.
- Kristensen, Kjeld
 Standard Danish, Copenhagen Sociolinguistic, and Regional Varieties in the 1900s. Unpublished manuscript.
- Kristensen, Kjeld, and Thelander, Mats
 1984 On dialect levelling in Denmark and Sweden. *FoL* XVIII/1 2, 223–246.
- Lanstyák István and Gizella Szabó Mihály
 1995 Contact varieties of Hungarian in Slovakia. A contribution to their description. *IJSL* 120, 111–130.
- Mackridge, Peter
 1998 Le discours public en Grèce après l'Abolition de la Diglossie. In *'Strong' and 'weak' languages in the European Union*, A.-F. Christidis (ed), Vol II, 133–136. Thessaloniki: Centre for the Greek Language.
- Marle, Jaap van
 1996 Dialect versus standard language: nature versus culture. In *Taming the vernacular. From dialect to written standard language*, J. Cheshire/D. Stein (eds.), 13–34. London: Longman.
- Martín Zorraquino Maria Antonia
 1997 Sociolinguistic attitudes and beliefs towards dialectal and standard varieties. In *La Franja de Aragón* (Spain). *FoL* XXXII/1 2 (*Special*

- Issue on Dialect Levelling and the Standard Varieties in Europe*, P. Auer (ed.), 131–144.
- Martin Zorraquino, Maria Antonia, Fort, Maria Rosa and Arnal, Maria Luisa (et al.)
1995 *Estudio sociolingüístico de la Franja Oriental de Aragón*. Zaragoza.
- Mazur, Jan
1996 Konvergenzen und Divergenzen in den polnischen Sprachvarietäten. *Sociolinguistica* 10: 53–74.
- Michelena, Luis
1981 Normalización de la forma escrita de una lengua: el caso vasco. In Demonte V. / Varela S. (eds.), *El bilingüismo: problemática y realidad*. Madrid (=Revista de Occidente 10–11), 55–75.
- Mihm, Arend
2000a Zur Deutung der graphematischen Variation in historischen Texten. In *Vom Umgang mit sprachlicher Variation*. (=FS H. Löffler), A. Häcki-Buhofer (ed.), 369–390. Tübingen: Franck.
2000b Die Rolle der Umgangssprachen seit der Mitte des 20. Jahrhunderts. In *Sprachgeschichte - ein Handbuch*, Ammon U. et al. (eds.), Berlin (second ed., vol. II), 2107–2137.
- Miller, Jim
1993 The Grammar of Scots English. In *Real English*, Milroy and Milroy (eds.), 99–138. Longman.
- Milroy Jim and Milroy Leslie
1984 Linguistic change, social network and speaker innovation. *Journal of Linguistics* 21: 339–384.
- Moosmüller, Sylvia
1984 *Soziophonologische Variation im gegenwärtigen Wiener Deutsch*. Wiesbaden: Steiner.
- Moosmüller, Sylvia and Vollmer, Ralf
2002 'Natürliches' Driften' im Lautwandel: die Monophthongierung im österreichischen Deutsch. *Zeitschrift für Sprachwissenschaft* 20 (1): 42–65.
- Moser, Hugo
1950 *Deutsche Sprachgeschichte*. Stuttgart: Schwab.
- Nekvapil, Jiří
2000 Language management in a changing society. Sociolinguistic remarks from the Czech Republic. In *Die sprachliche Situation in der Slavia zehn Jahre nach der Wende* B. Panzer (ed.), 165–177. Lang.
- Nekvapil, Jiří and Chloupek, Jan
1986 Introduction: On Czech Linguistics and Sociolinguistics. In *Reader in Czech sociolinguistics*, idem (eds.), 7–18. Benjamins.
- Nordberg, Bengt and Sundgren, Eva
1999 On the rate of change of some variables in present-day Swedish. Paper presented at the NewWAVE 28, Toronto. (Unpublished manuscript, University Uppsala).
- Paunonen, Heikki
1994 The Finnish language in Helsinki. In *The sociolinguistics of urbanization: the case of the Nordic countries*, B. Nordberg (ed.), 223–245. Berlin: de Gruyter.
1996/97 Finland. In *Kontaktlinguistik - ein Handbuch*, H. Goebel et al. (eds.), 993–1007. Berlin: de Gruyter.
- Pedersen, Inge Lise
1996 Regionalism and linguistic change. *Sociolinguistica* 10: 75–89.
in press a Traditional Dialects of Danish and the De-dialectalization 1900–2000. In *International Journal of the Sociology of Language*
in press b Processes of standardization in Scandinavia. To appear in *Dialect Change*. P. Auer, F. Hinskens and P. Kerswill (eds), Cambridge University Press.
- Røynealand, Unn
2001
in press Dialects in Norway - catching up with the rest of Europe? to appear in: *International Journal of the Sociology of Language*
- Sandøy, Helge
in press Nordic language history and current trends in dialectology. In *The Nordic Languages*, O. Bandle et al. (eds.), Ch. 35. Berlin: de Gruyter.
- Sandór, Klára
1999 National feeling or responsibility. The case of the Csángó language revitalization. *Multilingua* 19(1–2): 141–168.
- Sgall, Petr, Hronek, Jiří, Stich, Alexander and Horecký, Ján
1992 *Variation in Language. Code Switching in Czech as a Challenge for Sociolinguistics*. Amsterdam: Benjamins.
- Siebenhaar, Beat
2000 *Sprachvariation, Sprachwandel und Einstellung: der Dialekt der Stadt Aarau in der Labilitätszone zwischen Zürcher und Berner Mundart*. Stuttgart: Steiner.
- Sobrero, Alberto
1995 Italianization and variations in the repertoire: the koinai. *Sociolinguistica* 10: 104–111. P. Auer, F. Hinskens and K.J. Mattheier (eds.).
- Thelander, Mats
1976 Code-switching or code-mixing? *Linguistics* 183: 103–123.
1982 De-dialectalization in Sweden. In *Linguistica e Antropologia = Atti del XIV congresso internazionale di studi*, 175–195. Roma.
- Trumper, John and Maddalon, Marta
1987 Converging divergence and diverging convergence: The dialect-language conflict and contrasting evolutionary trends in modern Italy. In *Variation and Convergence*, P. Auer and A. di Luzio (eds.), 217–259. Berlin: de Gruyter.

- Tzitzilis, Christos
 2001 Dialectes et dialectologie du Grec moderne. In *Variation and Convergence*, République Hellénique/Ministère de l'éducation nationale et des cultes (ed.), *La langue Grecque et ses dialectes*, 83–90. Athens.
- Þráinsson, Höskuldur and Árnason, Kristján
 1992 Phonological variation in 20th century Icelandic. In *Íslenskt mál* 14: 89–128.
- Van de Velde, Hans
 1996 *Variatie en verandering in het gesproken Standaard-Nederlands (1935-1993)*. Ph.D. thesis, Katholieke Universiteit Nijmegen.
- Videnov, Michail
 1998 The Present-Day Bulgarian Language Situation: Trends and Prospects. *International Journal of the Sociology of Language* 135/137: 11–36.
- Villena Ponsada, and Juan Andrés
 1996 Convergence and divergence in a standard-dialect continuum: networks and individuals in Malaga. *Sociolinguistica* X: 112–137. P. Auer, F. Hinskens and K.J. Mattheier (eds.).
 2001 Identidad y variación lingüística: prestigio nacional y lealtad vernacular en el Español hablado en Andalucía. In *Identidades lingüísticas en la España Autónoma*, Bossong Georg/Báez de Aguilar, Francisco (eds.), 107–150. Frankfurt.
 in prep. The Iberian Peninsula, in: *Soziolinguistik – Ein Handbuch* (U. Ammon et al., eds.). Berlin: de Gruyter (second edition).
- Widmark, Gün
 in print Standard and colloquial languages in Scandinavia in the 19th century. In *The Nordic Languages*, O. Bandle et. al. (eds.). Berlin: de Gruyter.
- Willemyns, Roland
 1996 Dialektverlust im niederländischen Sprachraum. *ZDL* 64 (2): 129–154.
- Woolhiser, Curt
 1999 Language ideology and language conflict in post-soviet Belarus. In *Language, Ethnicity and the State*, Camille C. O'Reilly (ed.), Vol. 2, 91–122. London: Palgrave.

From phonetic similarity to dialect classification: A principled approach

Paul Heggarty, April McMahon and Robert McMahon¹

1. Language versus dialect comparison

In many ways, dialects are like related languages: in both cases, we are dealing with synchronic similarities, and an underlying history of divergence. One might therefore expect that any tools developed for classifying language families would generalize naturally to the dialect level. Yet in fact the preoccupations of classificatory historical linguistics – particularly following the recent controversies over Greenberg (1987) and Dixon (1997), for instance – lie mainly with very large-scale comparison, and dialects tend to be ignored. Even if we did try to include dialects, the questions to be asked are not the same: we are not worrying about *whether* varieties are related, but *how closely*, and the usual models in historical linguistics simply do not work at the required level of resolution, or of detail.

To illustrate this, consider what is perhaps the best known approximately quantificational method for calculating degrees of similarity between languages, namely lexicostatistics. This involves selecting some basic word meaning; identifying (or choosing – and this begs the significant question of just how we determine what the ‘normal’ translation is) its corresponding lexemes in a number of languages; and seeing if they are cognate. With the meaning ‘castle’, for instance, lexicostatistics would identify these corresponding forms in various Romance languages: Italian *castello*, Spanish *castillo*, and French *château* (*fort*); and all of these are derived from Latin *castellum*. It follows that a lexicostatistical quantification of the ‘linguistic distance’ between some Romance language varieties on the basis of this set of words would produce the outcome illustrated in Figure 1 below.

Of course, Figure 1 is rather obvious, and might be thought a little unfair. However, it does make clear one most important point, which is that lexicostatistics is blind to what it scarcely takes a linguist to tell: namely that Italian *castello* [kas'tello] is far more like Spanish *castillo* [kas'tiʎo] than either is to French *château* [ʃato]. There is a great deal of comparative data in these words, awaiting quantification; but this cannot be achieved on