107. The European linguistic area: Standard Average European

1. Introduction

This article summarizes some of the main pieces of evidence for a linguistic area (or *Sprachbund*) in Europe that comprises the Romance, Germanic and Balto-Slavic languages, the Balkan languages, and more marginally also the westernmost Finno-Ugrian languages (these will be called *core European languages* in this article). This linguistic area is sometimes called **Standard Average European** (abbreviated SAE), following Whorf (1941) [1956: 138]. The existence of this linguistic area is a relatively new insight (cf. Bechert et al. 1990, Bernini & Ramat 1996, Haspelmath 1998, van der Auwera 1998, König & Haspelmath 1999).

While the close syntactic parallels among the Balkan languages have struck linguists since the 19th century and the existence of a Balkan *Sprachbund* has been universally accepted, the European linguistic area has long been overlooked. This may at first appear surprising, because the members of the *Sprachbund* are among the best studied languages of the world. However, it is easy to understand why linguists have been slow to appreciate the significance of the similarities among the core European languages: Since most comparative linguists know these languages particularly well, they have tended to see non-European languages as special and unusual, and the similarities among the European languages have not seemed surprising. Thus, it was only toward the end of the 20th century, as more and more had become known about the grammatical properties of the languages of the rest of the world, that linguists realized how peculiar the core European languages are in some ways when seen in the world-wide context. From this perspective, Standard Average European may even appear as an "exotic language" (Dahl 1990).

A linguistic area can be recognized when a number of geographically contiguous languages share structural features which cannot be due to retention from a common proto-language and which give these languages a profile that makes them stand out among the surrounding languages. There is thus no minimum number of languages that a linguistic area comprises (*pace* Stolz 2001a). In principle, there could be a linguistic area consisting of just two languages (though this would be rather uninteresting), and there are also very large (continent-sized) linguistic areas (Dryer 1989a). Likewise, there is no minimum number of structural features that the languages must share in order to qualify as a *Sprachbund*. For instance, Jakobson (1931) establishes his "Eurasian linguistic area" on the basis of just two phonological features, but of course an area that shares more features is more interesting. As will be shown below, Standard Average European languages share over a dozen highly characteristic features, so we are dealing with a very interesting *Sprachbund*.

A linguistic area is particularly striking when it comprises languages from genealogically unrelated languages (like the South Asian linguistic area (→ Art. 109), or the Mesoamerican linguistic area (→ Art. 110)), but this is not a necessary feature of a *Sprachbund*. The Balkan languages are all Indo-European, but they are from different families within Indo-European (Romance, Slavic, Greek, Albanian), and not all languages of these families belong to the Balkan linguistic area, so nobody questions the validity of the Balkan *Sprachbund* (→ Art. 108). In the case of SAE, three entire branches of Indo-European (Romance, Germanic and Balto-Slavic) belong to the linguistic area. However, here too it is clear that we are not dealing with a genealogical grouping, because nobody ever proposed a branch of Indo-European that consists of precisely these three families. On the contrary, Indo-Europeanists typically assume a particularly close genealogical relationship between Italic and Celtic (and sometimes even an Italo-Celtic protolanguage), but Romance (the sole descendant of Italic) is inside SAE, while the Celtic languages do not belong to SAE. And since so much is known about the grammatical properties that Proto-Indo-European must have possessed, it is fairly easy to test whether an SAE feature is an Indo-Euro-

peanism or not. As was shown in Haspelmath (1998), most of the characteristic SAE features (also called **Europeanisms** here) are not Indo-Europeanisms but later common innovations.

Thus, what needs to be shown in order to demonstrate that a structural feature is a Europeanism is

(i)   that the great majority of core European languages possesses it;
(ii) that the geographically adjacent languages lack it (i. e. Celtic in the west, Turkic, eastern Uralic, Abkhaz-Adyghean and Nakh-Daghestanian in the east, and perhaps Afro-Asiatic in the south);
(iii) that the eastern Indo-European languages lack it (Armenian, Iranian, Indic); and
(iv) that this feature is not found in the majority of the world's languages.

Particularly the last point is not easy to demonstrate for many features because there are still far too few representative world-wide studies of grammatical structures, so to the extent that our knowledge about the world's languages is incomplete and biased, we cannot be sure about the European linguistic area. In this article, I will cite whatever information is available, and sometimes I will have to resort to impressionistic observations.

The designation "core European language" for members of SAE is deliberately vague, because the European linguistic area does not have sharp boundaries. It seems possible to identify a nucleus consisting of continental West Germanic languages (e. g. Dutch, German) and Gallo-Romance (e. g. French, Occitan, northern Italo-Romance). For this set of languages, van der Auwera (1998a: 824) proposes the name **Charlemagne Sprachbund**. Of the other languages, those which are geographically further from this center also seem to share significantly fewer SAE features, i. e. Ibero-Romance, insular Scandinavian (Icelandic and Faroese), East Slavic (Russian, Ukrainian, Belorussian) and Baltic. Even English, a West Germanic language, is clearly not within the nucleus. Of the non-Indo-European languages of Europe, the western Uralic languages (i. e. Hungarian and Balto-Finnic) are at least marginal members of Standard Average European; they are in many ways strikingly different from eastern Uralic. Maltese also exhibits a number of Europeanisms not shared by other Arabic

varieties, but Basque seems to show very few of them. Somewhat further to the east, Georgian in the southern Caucasus (and perhaps the other Kartvelian languages) shares a surprising number of features with the core European languages. These impressionistic statements should eventually be quantified, but since it is not clear how much weight should be attached to each feature, this is not straightforward.

All of the features discussed below are syntactic, or concern the existence of certain morphosyntactic categories. I am not aware of any phonological properties characteristic of the core European languages (cf. Jakobson 1931: 182: "do six por ne udalos' najti ni odnogo obščeevropejskogo … položitel'nogo fonologičeskogo priznaka [so far not a single Europe-wide positive phonological feature has been found]"). Perhaps phonologists have not looked hard enough, but at least one major recent study of word prosody in European languages has not found any phonological evidence for Standard Average European (van der Hulst et al. 1999, especially Maps 1−4) (but cf. Pisani 1969). A few generalizations are discussed by Ternes (1998), but he finds that in most respects European languages are unremarkable from a world-wide perspective. Perhaps the only features worth mentioning are the relatively large vowel inventories (no 3-vowel or 4-vowel inventories) and the relatively common consonant clusters (no restriction to CV syllables). In these respects, European languages are not average, but they are by no means extreme either.

2.   The **major** Standard Average European features

In this section I will discuss **a dozen** grammatical features that are characteristic of the core European languages and that together define the SAE *Sprachbund*. In each case I will briefly define the feature and give a few examples from SAE languages. Then a name map, which indicates the approximate location of languages by the arrangement of (abbreviated) language names, shows the distribution of the various feature values within Europe. In each case it can be observed that the nuclear SAE languages are within the SAE isogloss, and that the marginal languages tend to be outside the isogloss to a greater or lesser extent. (Part of the material presented here was already included in Haspelmath 1998.)

2.1. Definite and indefinite articles

Both a definite and an indefinite article (e. g. English *the book/a book*; → Art. 62) exist in all Romance and almost all Germanic languages plus some of the Balkan languages (Modern Greek, perhaps Albanian and Bulgarian), but not outside Standard Average European. To be sure, their forms and syntactic behavior show considerable diversity (see Nocentini 1996 for an overview), but their very existence is characteristic enough. The distribution of articles in European languages is shown in Map 107.1. (Abbreviations of language names are given in the Appendix.)
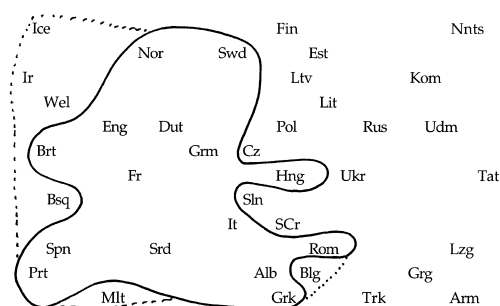
—— definite and indefinite article present
- - - - only definite article present

Map 107.1: Definite and indefinite article

In large parts of eastern Europe there are no articles at all (East Slavic, West Slavic, Finno-Ugrian other than Hungarian, Turkic, Nakh-Daghestanian, Kartvelian). Some neighboring non-SAE languages do have definite articles (e. g. Celtic, Semitic, Abkhaz, Mordvin), and Turkish has an indefinite article, but no neighboring non-SAE language has both definite and indefinite articles. The only exception among Germanic languages, Icelandic (which only has definite articles like nearby Celtic), is also the most peripheral Germanic language geographically. We can also be certain that the existence of definite and indefinite articles is not an Indo-Europeanism: The Iranian and Indic languages have generally lacked articles throughout their history.

World-wide, articles are not nearly as common as in Europe: According to Dryer's (1989b: 85) findings, "it appears that about a third of the languages of the world employ articles" (125 out of a sample of about 400 languages). Only 31 languages of those in Dryer's sample (i. e. less than 8%) have both definite and indefinite articles.

2.2. Relative clauses with relative pronouns

The type of relative clause found in languages such as German, French or Russian seems to be unique to Standard Average European languages. It is characterized by the following four features: The relative clause is postnominal, there is an inflecting relative pronoun, this pronoun introduces the relative clause, and the relative pronoun functions as a resumptive, i. e. it signals the head's role within the relative clause (cf. Lehmann 1984: 103−109, Comrie 1998). In English, a relative construction like *the suspicious woman whom I described* also displays all these features. Furthermore, in most SAE languages the relative pronoun is based on an interrogative pronoun (this is true of all Romance, all Slavic and some Germanic languages, Modern Greek, as well as Hungarian and Georgian). (Languages like German, whose relative pronoun is based on a demonstrative, or Finnish, which has a special relative pronoun, are not common.) The geographical distribution of the relative pronoun strategy is shown in Map 107.2.
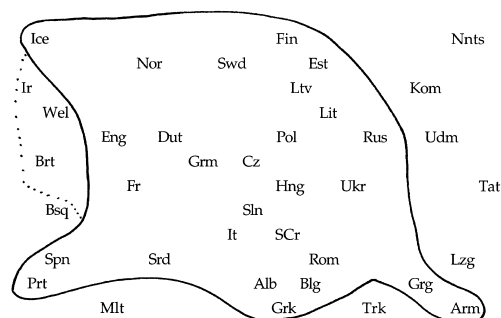
—— relative clause with introducing relative pronoun
- - - - only particle relative clause

Map 107.2: Two relative clause types in Europe

The only other type that is widespread in Europe is the postnominal relative clause introduced by a relative particle (Lehmann 1984: 85−87), which often occurs in the same language beside the resumptive relative pronoun type just described (an English example would be *the radio that I bought*). Particle relatives of this type exist in most Slavic and Romance languages, as well as in Scandinavian languages and Modern Greek, but also in Welsh and Irish (Lehmann 1984: 88−90). The relative particle is sometimes difficult to distinguish from a degenerate resumptive pronoun, and in many European languages

it developed from a relative pronoun through the gradual loss of inflectional distinctions. However, this also means that the relative clause loses its specifically European flavor, because particle relatives are also attested widely elsewhere in the world (e. g. in Persian, Modern Hebrew, Nahuatl, Indonesian, Yoruba, and Thai, cf. Lehmann 1984: 85−97).

However, the relative pronoun strategy clearly is typically European. It is not found in the eastern Indo-European languages, and as Comrie (1998: 61) notes, "relative clauses formed using the relative pronoun strategy are quite exceptional outside Europe, except as a recent result of the influence of European languages … The relative pronoun strategy thus seems to be a remarkable areal typological feature of European languages, especially the standard written languages".

2.3. 'Have'-perfect

Another well-known feature typical of SAE languages is the (transitive) perfect formed by 'have' plus a passive participle (e. g. English *I have written*, Swedish *jag har skrivit*, Spanish *he escrito*; → Art. 59). A perfect of this kind exists in all Romance and Germanic languages plus some of the Balkan languages (Albanian, Modern Greek, Macedonian), and also in Czech (Garvin 1949: 84). These perfects do not all mean the same thing, because they are at different stages in the grammaticalization process: in French and German, the perfect can be used as a normal perfective past, including the function of a narrative tense, while in Spanish, English and Swedish the perfect has a distinct present-anterior meaning. What is important here is that they all must have had basically the same meaning when they were first created. The geographical distribution of 'have'-perfects in Europe is shown in Map 107.3.
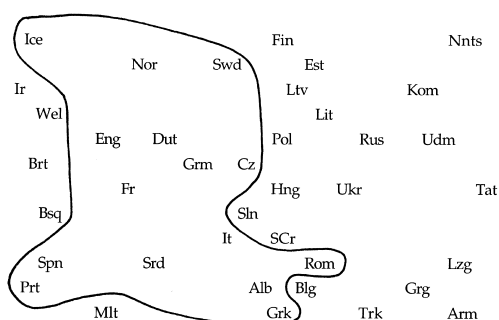
Map 107.3: 'Have'-perfects in Europe

In contrast to the languages just mentioned, in Slavic, Finno-Ugrian and Armenian the perfect is usually based on a participial construction with an active participle and a copula (e. g. Finnish *ole-n saa-nut* [be-1SG receive-PTCP] 'I have received'). Hungarian seems to lack a perfect completely. In some Nakh-Daghestanian languages (e. g. Lezgian and Godoberi), the perfect is formed on the basis of the past converb plus the copula. Georgian comes closest to the SAE prototype in that its transitive perfect is based on a passive participle, but this is combined with the copula rather than the transitive verb 'have', so that the perfect has a quasi-passive structure, with the agent in the dative case ('The letter is-written to-me', rather than 'I have-written the letter'). In Welsh, the perfect is formed with the preposition *wedi* 'after' ('She is after selling the house' for 'She has sold the house'). The eastern Indo-European languages also lack a 'have'-perfect (for instance, both Persian and Hindi/Urdu have a perfect based on a participle plus the copula, somewhat like Slavic and Armenian).

Dahl (1995, 1996: 365), taking a global perspective, notes that the 'have'-perfect is almost exclusively found in Europe. Now one might object that this is not a primitive feature of European languages. Many languages do not use a transitive 'have'-verb for indicating predicative possession at all, and it has in fact been suggested that the very existence of a transitive verb of predicative possession is a Europeanism (e. g. Lazard 1990: 246−47; Benveniste 1960 [1966: 195]: "L'expression la plus courante du rapport indiqué dans nos langues par *avoir* s'énonce à l'inverse par *être à* … Telle est la situation dans la majorité des langues.") The restriction of a 'have'-perfect to Europe would then be just a consequence of this (cf. Dahl 1990: 7). However, so far no published research has documented an areal restriction for 'have' verbs. From Heine's (1997: 47−50, 240−44) survey of predicative possessive constructions, not much support can be drawn for such a claim. Still, this is an interesting idea to be addressed by further research. If 'have'-verbs turn out to be typical of Europe, that would fit with the tendency of European languages to have nominative experiencers in experiential verbs (see the next section).

4. Nominative experiencers

There are two ways of expressing experiencer arguments of verbs of sensation, emotion, cognition and perception: The experiencer
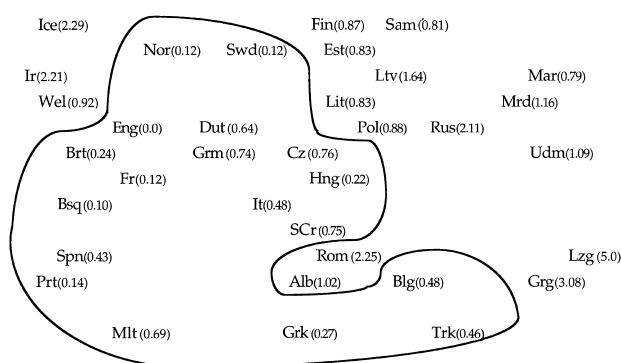
may be assimilated to agents and coded as a nominative subject (e. g. *I like it*), or it may be assimilated to a patient or goal, so that the stimulus argument is coded as the nominative subject (e. g. *It pleases me*). In Bossong's (1998) typology, the first type is called *generalizing*, and the second type is called *inverting*. Bossong studies the expression of ten common experiential predicates in 40 European languages. He computes the relation between inverting predicates and generalizing predicates, arriving at figures between 0.0 for English (where all predicates are generalizing) and 5.0 for Lezgian (where all predicates are inverting). By arbitrarily dividing the languages into those showing predominant generalization (ratios between 0.0 and 0.8) and those showing predominant inversion (ratios between 0.8 and 5.0), we arrive at the geographical pattern shown in Map 107.4.

European is fairly clear: Indic languages are well-known for their "dative subjects" of experiencer verbs, so again the feature is not genetic (see also Masica 1976, especially Map 6, for the areal distribution of dative subjects in Eurasia and northern Africa). (See Haspelmath 2001 for more discussion of experiential predicates in European languages.)
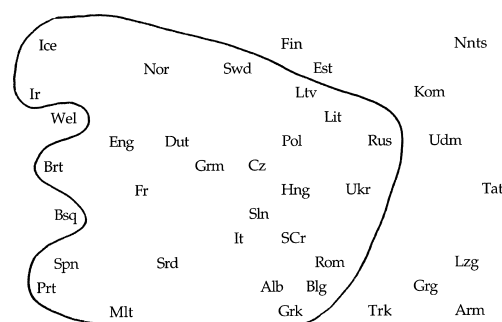
.5. Participial passive

Standard Average European languages typically have a canonical passive construction (→ Art. 67) formed with a passive participle plus an intransitive copula-like verb ('be', 'become', or the like). In this passive the original direct object becomes the subject and the original subject may be omitted, but it may also be expressed as an adverbial agent phrase. Such constructions occur in all Romance and Germanic languages, but also in

Ice(2.29)   Fin(0.87)   Sam(0.81)
Nor(0.12)   Swd(0.12)   Est(0.83)
Ir(2.21)   Ltv(1.64)   Mar(0.79)
Wel(0.92)   Lit(0.83)   Mrd(1.16)
Eng(0.0)   Dut(0.64)   Pol(0.88)   Rus(2.11)
Brt(0.24)   Grm(0.74)   Cz(0.76)   Udm(1.09)
Fr(0.12)   Hng(0.22)
Bsq(0.10)   It(0.48)
SCr(0.75)
Spn(0.43)   Rom(2.25)   Lzg(5.0)
Prt(0.14)   Alb(1.02)   Blg(0.48)   Grg(3.08)
Mlt(0.69)   Grk(0.27)   Trk(0.46)

Map 107.4: Predominant generalization (center) vs. inversion (periphery)

Thus, Bossong's study basically confirms earlier claims (Lazard 1990: 246−47, Dahl 1990: 7) that the generalizing type is characteristic of SAE, although some of the figures are perhaps a bit surprising (e. g. the fact that Hungarian turns out to be more SAE than German or Dutch, and the inclusion of Turkish, but not Romanian or Albanian, with respect to this feature). It is not possible to explain everything here, but we evidently have before us a fairly typical SAE pattern with French and English at the center, Celtic (plus Icelandic this time) at the western margin, Balto-Slavic, Finno-Ugrian and Caucasian at the eastern margin, and fairly gradual transitions within the macro-areas. No systematic world-wide studies have been made, but at least the behavior of eastern Indo-

all Slavic (including East Slavic) and Balkan languages, as well as in Irish. The geographical distribution of such participial passives is shown in Map 107.5.

Ice   Fin   Nnts
Nor   Swd   Est
Ir   Ltv   Kom
Wel   Lit
Eng   Dut   Pol   Rus   Udm
Brt   Grm   Cz
Fr   Hng   Ukr   Tat
Bsq   Sln
It   SCr
Spn   Srd   Rom   Lzg
Prt   Alb   Blg   Grg
Mlt   Grk   Trk   Arm

Map 107.5: Participial passives in Europe

No passives exist in Nakh-Daghestanian and in Hungarian, and passives of different formal types are found in Turkic, Georgian, and Armenian (stem suffix), in Basque, and in Celtic (cf. the Welsh 'get'-passive: 'Terry got his hitting by a snowball' for 'Terry got hit by a snowball'). Finnish and Irish have passives of a different syntactic type: In this construction, only the subject is backgrounded, while the direct object remains in its place.

Participial passives are very rare in languages other than Standard Average European. In Haspelmath (1990) I surveyed a world-wide sample of eighty languages and found that a passive exists only in the minority of the languages (thirty-one). Of these thirty-one languages, only four have a passive formed from a participle plus an intransitive auxiliary, and two of them are European languages (Latin and Danish). The most common formal type of passive is the stem suffix (found in twenty-five languages). Syntactically, the possibility of an adverbial agent phrase is also by no means universal, but it is characteristic of SAE languages (Lazard 1990: 246).

It must be admitted that the SAE status of this feature is less evident than that of the first two features because the eastern Indo-European languages also tend to have passives of this type. In fact, in my 1990 study, the two non-European languages with participle-auxiliary passives were Baluchi (an Iranian language) and Maithili (an Indic language). Thus, one might say that this feature is an Indo-European genealogical feature. However, at least the Celtic languages and Armenian, two non-SAE branches of Indo-European, do not have such passives, and Maltese is a non-Indo-European language with such a passive (calqued from Italian).

6. Anticausative prominence

There are three ways in which languages can express inchoative-causative alternations such as 'get lost/lose', 'break (intr.)/break (tr.)', 'rise/raise'. One is by means of a causative derivation (→ Art. 66), i. e. a derived verb based on the inchoative member of the alternation, e. g. Mongolian *xajl-uul-* 'melt (tr.)', from *xajl-* 'melt (intr.)'. The second is by means of an anticausative derivation, i. e. a derived verb based on the causative member, e. g. Russian *izmenit'-sja* 'change (intr.)', from *izmenit'* 'change (tr.)'. (The third type, in which neither member is derived from the other, i. e. *non-directed* alternations, will not

be considered further here.) In Haspelmath (1993), I examined 31 verb pairs in 21 languages and found that languages differ greatly in the way inchoative-causative pairs are expressed: Some languages are *anticausative-prominent*, preferring anticausatives to causatives, while others are *causative-prominent*. It turns out that anticausative-prominence is a characteristic feature of SAE. In my sample, German, French, Romanian, Russian, Modern Greek and Lithuanian show the highest percentages of anticausative verb pairs (between 100% and 74% of all pairs that do not belong to the third, non-directed, type). The percentage in the European languages of my sample are shown in Map 107.6.
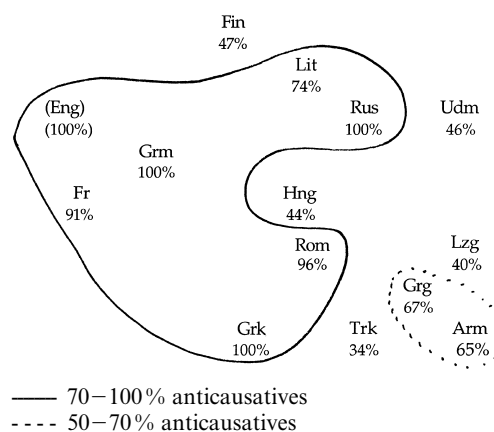
Map 107.6: Percentage of anticausative pairs

—— 70−100% anticausatives
- - - - 50−70% anticausatives

By contrast, Asian languages show much lower percentages of anticausatives, preferring causatives instead (e. g. Indonesian: 0%, Mongolian: 11%, Turkish: 34%, Hindi/Urdu 35%, Lezgian: 40%). An intermediate position is occupied by the Finno-Ugrian languages of eastern Europe (Finnish 47%, Udmurt 46%, Hungarian 44%) as well as Georgian (68%) and Armenian (65%). In a study involving more languages from Asia, Africa and Europe but less language-particular detail, Masica (1976) found a clear distinctive pattern for Europe: few causatives, heavy reliance on anticausatives (see especially his Maps 2 and 3). In a recent world-wide study of 18 verbs from 80 languages, Nichols et al. (to appear) report that in inchoative-causative pairs involving inanimate participants (i. e. the most typical subtype), the causative is generally favored worldwide and is strongly disfavored only in Europe.

Anticausative-prominence is not an Indo-Europeanism: Older Indo-European had a productive causative formation, which lost its productivity in the European branches, but continued to be productive in eastern Indo-European (cf. the low figure of 35% anticausatives in Hindi/Urdu).

2.7. Dative external possessors

In König & Haspelmath (1998) and Haspelmath (1999), we studied the distribution of external possessors in thirty European languages (→ Art. 73). We found three main language types in Europe: (i) those with dative external possessors, e.g. German *Die Mutter wäscht dem Kind die Haare* 'The mother is washing the child's hair', (ii) those with locative external possessors, e.g. Swedish *Någon bröt armen på honom* 'Someone broke his arm (lit. on him)', and (iii) those that lack external possessors and must express possessors NP-internally, e.g. English. The SAE feature, external possessors in the dative, is found in Romance, Continental West Germanic, Balto-Slavic, Hungarian and Balkan languages (Greek, Albanian). North Germanic and Balto-Finnic languages have locative external possessors, i.e. they are somewhat peripheral SAE languages with respect to this feature. The geographical distribution is shown in Map 107.7.
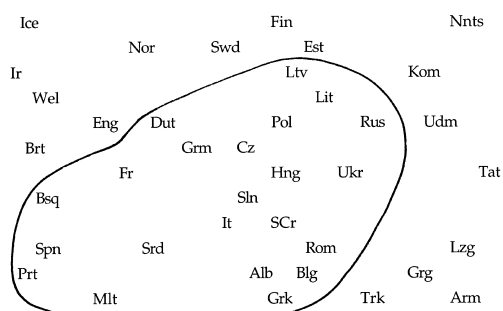
Map 107.7: Dative external possessor

In the far west (Welsh, Breton, English) and in the southeast (Turkish, Lezgian) of Europe there are languages which do not have external possessors at all. The eastern Indo-European languages Kurdish, Persian and Hindi/Urdu also belong to this type. Outside Europe a fourth type enjoys considerable popularity: the "relation-usurping" type, where he possessor "usurps" the syntactic relation of the possessum (e.g. Chichewa, a Bantu language, has 'The hyena ate the hare the fish' for 'The

hyena ate the hare's fish'). This type is not found in Europe at all. Conversely, dative external possessors seem to be very rare outside Europe (the only case I am aware of is Ewe, cf. Ameka 1996), so this is a very robust example of an SAE feature.

2.8. Negative pronouns and lack of verbal negation

The areal distribution of negation in Europe has been studied in detail by Bernini & Ramat (1996) (see also Ramat & Bernini 1990). Here I will single out just one aspect of negation, the cooccurrence of verbal negation with negative indefinite pronouns. I distinguish two main types: (i) V + NI (verb + negative indefinite), e.g. German *Niemand kommt* 'nobody comes', and (ii) NV + NI (negated verb + negative indefinite), e.g. Modern Greek *Kanénas dhen érxete* 'nobody (lit. not) comes'. A third, mixed type might be distinguished in which verbal negation cooccurs with negative indefinites only when the indefinite follows the verb but not when it precedes it, e.g. Italian *Nessuno viene* 'nobody comes', but *Non ho visto nessuno* 'Not I have seen nobody'. For our purposes we can classify this type as a subtype of (i), V + NI.

The Standard Average European type is V + NI (cf. Bernini & Ramat 1996: 184, Haspelmath 1997: 202). It is found in French (if we disregard the particle *ne*), Occitan and all Germanic languages, as well as (in the mixed variety) in Ibero- and Italo-Romance and Albanian (but not in Romanian or other Balkan languages). The geographical distribution of the types is shown on Map 107.8.
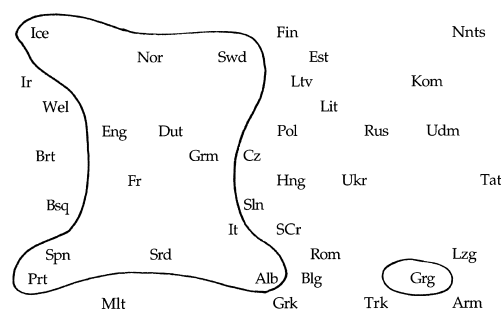
Map 107.8: Languages lacking verbal negation with a negative indefinite

All the eastern European languages (Balto-Slavic, Finno-Ugric, Turkic, Nakh-Daghestanian) with the exception of Georgian, and the Celtic languages in the west show the NV + NI type. This type is also that of the

eastern Indo-European languages (Iranian and Indic), as well as that of the clear majority of the world's languages: Kahrel (1996) has studied negation in a representative world-wide sample of 40 languages and found only five languages with V + NI negative patterns, one of which is the SAE language Dutch (the other four are Mangarayi (Australia), Evenki, Chukchi (Siberia), and Nama (southern Africa)), as against 41 NV + NI patterns, and seven others. I found a very similar pattern in my (non-representative) sample of 40 languages (Haspelmath 1997: 202).

2.9. Particles in comparative constructions

Comparative constructions were investigated by Stassen (1985) in a world-wide study of 19 languages (→ Art. 75). Stassen distinguishes six main ways in which the standard of comparison may be expressed: Three kinds of locative comparatives ('bigger from X', 'bigger to X', 'bigger at X'), the exceed comparative ('Y is big exceeding X'), the conjoined comparative ('Y is big, X is little'), and the particle comparative ('bigger than X'). The particle in this latter type is often related to a relative pronoun (cf. English *than/that*, Latin *quam/qui*), and the case marking of the standard is not influenced by the particle (so that it is possible to distinguish 'I love you more than she' from 'I love you more than her').

As Heine (1994) notes, the six types are not evenly distributed among the languages of the world. Of the 18 particle comparatives in Stassen's sample, 13 are in Europe, and of the 17 European languages in the sample, 13 have a particle comparative. The distribution within Europe again conforms to our expectations: Particle comparatives are found in Germanic, Romance, Balto-Slavic, the Balkans, Hungarian, Finnish and Basque, so this is the SAE type. The distribution is shown in Map 107.9.
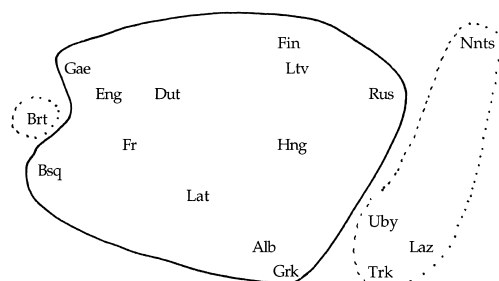
——— particle comparative
- - - - locative comparative

Map 107.9: Comparative types in Europe

The locative comparatives are all at the western fringe (Breton) or the eastern fringe of Europe (Finnish, Russian, Nenets, Ubykh, Turkish, Laz). The other two types do not exist at all in Europe − the exceed comparative is found particularly in Africa, and the conjoined comparative occurs only in the Americas and Oceania.
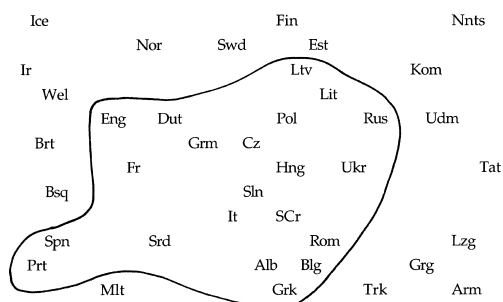
2.10. Relative-based equative constructions

Comparison of equality (equative constructions) is discussed less often than comparison of inequality, and nobody has undertaken a study of equatives on a world-wide scale. Still, there are good reasons to think that equative constructions provide evidence for Standard Average European (Haspelmath & Buchholz 1998). In Europe, many languages have an equative construction that is based on an adverbial relative-clause construction. For example, Catalan has *tan Z com X* 'as Z as X' (where Z is the adjective and X is the standard). Catalan *com* is an adverbial relative pronoun, and *tan* is a correlative demonstrative. A very similar construction is found elsewhere in Romance (Portuguese *tão Z como X*, Occitan *tan Z coma X*), in Germanic (German *so Z wie X*), in Slavic (Czech *tak Z jako X*, Russian *tak(oj) že Z kak X*), in Romani (*kade Z sar X*), in Hungarian (*olyan Z mint X*), in Finnish (*niin Z kuin X*), and in Georgian (*isetive Z rogorc X*). In the English construction, the relative-clause origin of *as* is not fully transparent synchronically, but diachronically *as* derives from a demonstrative (*eall swa > all so*) that was also used as a relative pronoun. In some Balkan languages, the correlative demonstrative is not used (e. g. Bulgarian *xubava kato tebe* 'as pretty as you'), but the standard marker is clearly of relative-pronoun origin. (There is probably some connection between the relative-pronoun origin of equative markers and the relative-pronoun origin of comparative standard markers that we saw in § 2.9.).

Non-SAE languages have quite different equative constructions. Many SOV languages in eastern Europe have a special equative standard marker (Lezgian *x̂iz*, Kalmyk *šing*; also Basque *bezain* and Maltese *daqs*), and the Celtic languages have a special (non-demonstrative) marker on the adjective (e. g. Irish *chomh Z le X* 'EQUATIVE Z with X'). In the Scandinavian languages, the word 'equally' is used on the adjective (e. g. Swedish *lika Z som X* 'equally Z as X'). The distribution of the relative-based equative con-

struction in Europe is shown in Map 107.10, following Haspelmath & Buchholz (1998: 297).

Map 107.10: Relative-based equative-constructions

Impressionalistically, relative-based equatives seem to be rare in the world's languages, and the eastern Indo-European languages do not seem to use them in general (however, a counterexample is Punjabi).

2.11. Subject person affixes as strict agreement markers

The majority of the world's languages have bound person markers on the verb that cross-refer to the verb's subject (or agent). When these subject affixes cooccur with overt subject NPs (full NPs or independent subject pronouns), they are called *agreement markers*. However, in most languages they can occur on their own and need not cooccur with overt subject NPs. For example, in the Bulgarian phrase *vie rabotite* 'you (pl.) work', we see the subject suffix *-ite* (2nd person plural) cooccurring with the independent subject pronoun *vie* 'you (pl.)', showing that *-ite* is an agreement marker. But in Bulgarian it is equally possible and probably more common to say just *rabotite* 'you (pl.) work', i. e. the subject suffix can have a referential function on its own. In German, by contrast, this is not possible: 'you work' is *ihr arbeit-et*. Since the agreement suffix *-et* does not have such an independent referential function, the subject pronoun *ihr* cannot be omitted. Languages like German are often called "non-pro-drop languages", and languages like Bulgarian are called "pro-drop languages"; better terms would be "strict-agreement languages" vs. "referential-agreement languages".

It has sometimes been thought that strict agreement, as exhibited by German, English, and French, is the norm and that referential agreement is somehow special. But in fact, referential agreement is far more widespread

in the world's languages, and strict subject agreement is characteristic of a few European languages, some of which happen to be well-known. In her world-wide sample of 272 languages, Siewierska (1999) finds only two strict-agreement languages, Dutch (an SAE language) and Vanimo (a Papuan language of New Guinea). Siewierska further notes that outside of Europe, she is aware of only two additional strict-agreement languages that are not in her sample (Anejom and Labu, two Oceanic languages). Gilligan (1987) reached a similar conclusion on the basis of a sample of 100 languages. The distribution of strict subject agreement markers in some European languages is shown in Map 107.11.
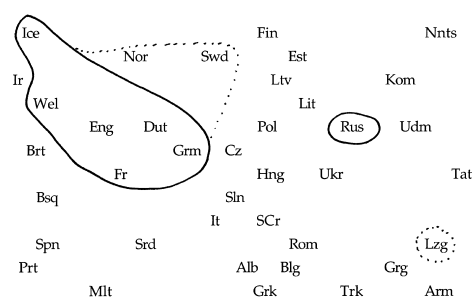
——— languages with strict subject agreement
- - - - languages with obligatory subject pronouns, lacking verb agreement

Map 107.11: Obligatory subject pronouns

The map shows two non-contiguous areas in which subject agreement suffixes cannot have a referential function: Germanic and Gallo-Romance languages with Welsh on the one hand, and Russian on the other. Perhaps only the western European area should be thought of as being relevant for SAE; in Russian, past-tense verbs do not have subject person affixes, so Russian is not a very good example of a strict-agreement language. In the eastern Nordic languages (Norwegian, Swedish, Danish), the subject pronouns are obligatory as they are in English, German or Icelandic, but the languages have lost agreement distinctions on the verb entirely (cf. Swedish *jag biter/du biter/han biter* 'I/you/he bite(s)', Icelandic *ég bít/þú bítur/hann bítur*). These languages are thus "non-pro-drop" in a sense, but they are not strict-agreement languages. English is approaching this type, as the only remnant of subject agreement is the 3rd person singular present-tense suffix *-s*. (There are also some languages of this type

in the eastern Caucasus, and indeed in many other parts of the world, but they may never have had subject person agreement marking.)

2.12. Intensifier-reflexive differentiation

Intensifiers are words like English *self*, German *selbst*, French *même* and Russian *sam* that characterize a noun phrase referent as central as opposed to an implicit or explicit periphery (e. g. *The Pope himself gave us an audience*, i. e. not just the cardinals (→ Art. 57; König & Siemund 1999). In many languages, the intensifier expression is also used as a reflexive pronoun, for instance in Persian (*xod-aš* 'himself': *Hušang xod-aš* 'Hushang himself', and *Hušang xodaš-rā did* [Hushang self-ACC saw] 'Hushang saw himself'). However, a feature that is typical of SAE languages is the differentiation of reflexive pronouns and intensifiers (König & Haspelmath 1999). For instance, German has *sich* (reflexive) vs. *selbst* (intensifier), Russian has *sebja* vs. *sam*, Italian has *si* vs. *stesso*, Greek has *eaftó* vs. *ídhjos*. Map 107.12 shows the languages in Europe with special reflexive pronouns that are not identical to intensifiers.
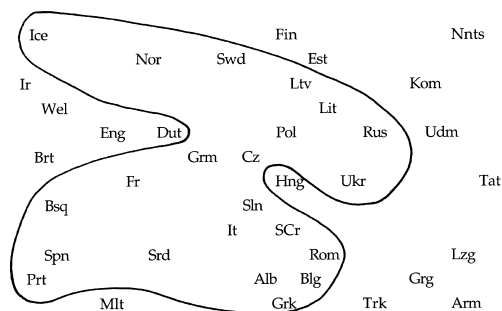
Map 107.12: Intensifier-reflexive differentiation

Intensifier-reflexive differentiation is not an Indo-Europeanism, because eastern Indo-European languages have the same expression for intensifiers and reflexives (e. g. Persian *xod-aš*, Hindi *aap*). There are no published world-wide studies yet, but it seems that non-differentiation is very common around the world, and while differentiation is also found elsewhere, it is not found in areas immediately adjacent to European languages.

3. Some further likely SAE features

In this section, I will mention a few features which are less well-documented than those in § 2, or whose geographical distribution is less

striking, but which nevertheless seem good candidates for Europeanisms. No maps will be given for these features, and the evidence will be summarized only briefly.

3.1. Verb fronting in polar interrogatives

In the large majority of languages, polar interrogatives are marked by interrogative intonation or an interrogative particle or both (→ Art. 77). In his sample of 79 languages, Ultan (1978) found only seven languages showing the alternative strategy of verb fronting (often called "subject-verb inversion"). Of these, six are European (English, French, Romanian, Russian, Hungarian, Finnish; the seventh language is Malay), so that the SAE status of verb fronting seems beyond doubt. In fact, the large majority of Germanic, Romance and Slavic languages (plus Modern Greek) appear to have verb fronting in polar questions in one form or another. The three European languages for which Ultan explicitly reports that no verb fronting occurs are peripheral: Basque, Gaelic and Lithuanian. Furthermore, SAE languages are characterized by the absence of an interrogative particle. In Ultan's data, the nine European languages exhibiting a particle in polar questions are all peripheral to a greater or lesser extent: Basque, Irish, Scottish Gaelic, Albanian, Hungarian, Lithuanian, Russian, Finnish, Turkish (and I can add Nakh-Daghestanian). Verb fronting in polar questions was suggested as a Europeanism already by Beckman (1934) (cf. Dahl 1990).

3.2. Comparative marking of adjectives

Most European languages have special forms for adjectives occurring in comparative constructions. For instance, English uses the suffix -*er* in this way (*The dog is bigg-er than the cat*). Such an inflectional marker of adjectives is not common in the world's languages outside of Europe. Some languages use some kind of adverbial particle modifying the adjective ('more'), but perhaps the most common type is represented by Japanese, where the comparative semantics is carried by the standard marker alone (e. g. *inu-ga neko yori ookii* [dog-SUBJ cat from big] 'the dog is bigger than the cat').

Special comparative forms are found in all Germanic, Balto-Slavic and Balkan languages (with the exception of Romanian and Albanian), and most Romance languages preserve at least four suppletive forms (e. g.

Italian *maggiore* 'bigger', *minore* 'smaller', *peggiore* 'worse', *migliore* 'better'). Comparative forms also exist in Basque (e. g. *haundi-ago* 'bigg-er'), Hungarian (*nagy-obb* 'bigg-er'), Finnish (*iso-mpi* 'bigg-er'), and other Finno-Ugrian languages.

Comparative forms are not completely unknown outside of Europe. Arabic has a special comparative form (e. g. *ʔakbar* 'bigger', from *kabiir* 'big'), but it is unique among Afro-Asiatic languages in this respect. Old Indo-Iranian languages had comparative forms, and the modern Iranian languages have preserved them to some extent (e. g. Persian *-tær*, Zaza *-êr*). But further east, in modern Indic, the comparative does not exist anymore, and languages like Hindi-Urdu and Bengali use a construction analogous to the Japanese example just cited. Similarly, in the Uralic languages, the further east we go, the fewer comparatives we find. For instance, Khanty (a Finno-Ugrian language spoken in western Siberia, i. e. outside of Europe) does have a comparative form in *-sək* (e. g. *jam-sək* 'better'), which is used when no standard is present. But in a complete comparative construction, no marking is found on the adjective (e. g. *naŋ ke:se:-n e:welt jam* [you knife-2SG from good] 'better than your knife', Nikolaeva 1999: 21).

Thus, although this feature is not confined to Europe, it is typical of a SAE feature in that it is robustly present in western Indo-European and Uralic languages, but gets rarer the further east we go in these families.

3.3. "A and-B" conjunction

The feature discussed in this section is less distinctive than the others mentioned so far, but I hope to show that it is not at all devoid of interest. Stassen (2000) offers the first world-wide typological study of NP conjunction strategies, based on a sample of 260 languages (→ Art. 82). He distinguishes two basic types, *and*-languages (using a symmetric particle) and *with*-languages (using an asymmetric comitative marker). Two thirds of Stassen's sample languages are *and*-languages, and since SAE clearly belongs to this type, too, it is not a very distinctive property. *And*-languages cover all of northern Eurasia, South Asia, the Middle East and northern Africa, Australia, New Guinea, and parts of Central and South America. *With*-languages are encountered in sub-Saharan Africa, East and Southeast Asia, the islands of Oceania, and large areas of North and South America.

However, within the *and*-languages there are several sub-types according to the position of the particle, which we may call "A and-B", "A-and B", "A-and B-and", and "A B-and" (of the remaining logical possibilities, "and-A B" seems to be inexistent, and "and-A and-B" occurs only as a secondary pattern). Most European languages, and in particular all SAE languages, belong to the sub-type "A and-B". The types "A-and B-and" and "A-and B" are found in some languages of the Caucasus and in some Turkic languages, as well as scattered throughout northern Eurasia and South Asia (e. g. in Abkhaz, Archi, Persian, Sinhalese, Tamil, Burmese, Korean according to Stassen; Stassen also points out that there is a correlation with verb-final word order here). Furthermore, some peripheral European languages make restricted use of the *with*-strategy (e. g. Russian *my s toboj* 'I and you', lit. 'we with you', and also Old Irish, Lithuanian, Polish and Hungarian, according to Stassen). Taken together, these data do show that belonging to the "A and-B" type is not a trivial feature of the SAE linguistic area.

3.4. Comitative-instrumental syncretism

In all SAE languages, the preposition that expresses accompaniment (= comitative) also serves to express the instrument role (e. g. English *with: with her husband*/*with the hammer*). Such languages are said to exhibit comitative-instrumental syncretism. Stolz (1996) studied comitative and instrumental markers in a world-wide sample of 323 languages and found that this kind of syncretism is typical of Europe. Non-European languages more commonly possess separate markers for these two semantic roles (e. g. Swahili *na* 'with (comitative)', *kwa* 'with (instrumental)'. As Table 107.1 shows, about two thirds of Stolz's sample languages are non-syncretic, and only one quarter is syncretic. (The remaining languages belong to a mixed type, which I ignore here for the sake of simplicity; thus, the percentages do not add up to 100%.)

Two areas diverge significantly from the general trend: Oceania has far less syncretism than the world average, and Europe has far more syncretism than the world average. When we look at the pattern within Europe, it becomes even clearer that we are dealing with an SAE feature (as Stolz recognizes, cf. 1996: 120). Of the 16 non-syncretic languages in Europe, 10 are Caucasian languages, i. e. they are clearly outside of SAE, and one is

Table 107.1: Comitative-instrumental: Syncretic and non-syncretic languages

	syncretic (e.g. English)		non-syncretic (e.g. Swahili)	
	languages	percentage	languages	percentage
Europe	25	49%	16	31%
Africa	20	31%	38	58%
Americas	16	21%	54	69%
Asia	12	18%	47	71%
Oceania	6	10%	54	86%
World	79	24%	209	65%

only politically, not anthropologically, in Europe (Greenlandic). Four of the remaining five languages are also otherwise not typical instances of SAE (Basque, Finnish, Maltese, Mari). And when we look at the 38 Indo-European languages in Stolz's sample, we see that syncretism cannot be regarded as an Indo-Europeanism: Of the eight Indo-European languages not spoken in Europe, only three show syncretism, while five show non-syncretism. Thus, in Asia Indo-European languages behave like Asian languages, and there is no general pattern for Indo-European.

3.5. Suppletive second ordinal

Most languages have a suppletive form of the ordinal numeral 'first', i.e. a form not derived from the cardinal numeral 'one'. An example is German, where '1st' is *erster* (unrelated to *eins* '1'), contrasting with other ordinals such as *zweiter* '2nd' (cf. *zwei* '2'), *vierter* '4th' (cf. *vier* '4'), and so on. In Stolz's (2001b) study of 100 languages world-wide, there are 95 languages with special ordinal numerals, and of these, 78 have a suppletive word for 'first'. Thus, languages that say (literally) 'oneth' for '1st' are not common. However, the same sample has only 22 languages in which the word for '2nd', too, is suppletive and not derived from '2' (e.g. English *second*). Thus, most languages have (literally) 'twoth' for '2nd'. The 22 languages that have a suppletive '2nd' word are heavily concentrated in Europe: 17 are European languages, and this type is clearly the majority within Europe (which is represented by 27 languages in Stolz's sample). Of the 10 European languages that do not have a suppletive second ordinal, six are clearly outside SAE (Basque, Turkish, Armenian, Georgian, Lezgian, Greenlandic). Among SAE languages, only some Balkan languages (Romanian, Albanian, Romani) and German lack a suppletive second ordinal.

This is clearly a very marginal feature in grammar, but it is intriguing that it should show such a clear geographical distribution.

3.6. Some other characteristics of SAE

The features examined so far present the most striking evidence for Standard Average European, but there are probably many more features that will turn out to be characteristic of the core European languages in one way or another. In this subsection, several such candidates will be mentioned briefly. The first few features in the following list are purely negative: At first glance, this may seem odd, but of course the lack of a category that is widespread elsewhere is no less significant than the presence of a category that is rare elsewhere.

(i) Lack of an alienable/inalienable opposition in adnominal possession (→ Art. 72). In Nichols's (1992) world-wide sample, almost half of the languages show such an opposition, but no European language does (1992: 123). More generally, this opposition is rarer in the Old World and common in the New World, but in Europe it is even less common than in Africa and Asia.

(ii) Lack of an inclusive/exclusive opposition in first person non-singular pronouns. Again, this opposition is commonest in the New World and in the Pacific region, but in Europe it is even rarer than in Africa and Asia, as was shown by Nichols (1992: 123).

(iii) Lack of reduplicating constructions. I have no systematic evidence to back up the claim that this is a characteristic feature of European languages, but reduplication is so common across languages that its almost total absence in the core European languages becomes striking. (Interestingly, reduplication existed in older Indo-European languages at least in one construction, the perfect, but even here it was lost entirely by the Middle Ages.)

(iv) Discourse pragmatic notions such as topic and focus are expressed primarily by sentence stress and word order differences (Lazard 1998: 116). Only the Celtic languages and French give a very prominent role to clefting, and particles marking discourse pragmatic notions are virtually unknown.

(v) SVO basic word order at the level of the clause. This feature is of course found elsewhere in the world, but in Europe it correlates particularly well with the other SAE features. The Celtic languages in the west have VSO order (except for Breton, which is also otherwise more SAE than Irish and Welsh), and the eastern languages have SOV word order. Interestingly, Balto-Finnic (Finnish, Estonian, etc.) and (less unequivocally) Hungarian have SVO word order, whereas the eastern Uralic languages have SOV. Similarly, the eastern Indo-European languages tend to show SOV word order. (See Dryer 1998 for more on word order in the languages of Europe.)

(vi) European languages tend to have just one converb (→ Art. 83) (cf. Nedjalkov 1998). For instance, Romance languages have the *gerundio/gérondif*, English has the *-ing*-form, and Slavic and Balkan languages have their adverbial participle. The Celtic languages in the west completely lack such a form, and the languages east of SAE tend to have more than one converb. Otherwise the core European languages tend to have adverbial conjunctions (→ Art. 63) to make adverbial clauses. According to Kortmann (1997: 344), they have "a large, semantically highly differentiated inventory of free adverbial subordinators placed in clause-initial position". More generally, they tend to have finite rather than non-finite subordinate strategies (→ Art. 100), though a multi-purpose infinitive usually exists (except for the Balkan languages).

(vii) European languages usually have a special construction for negative coordination, e. g. English *neither A nor B*, Italian *né A né B*, Russian *ni A ni B*, Dutch *noch A noch B*, Hungarian *sem A sem B*. Again, no worldwide study has been published, but such a negative coordinating construction is rarely reported from languages outside Europe (cf. Haspelmath to appear).

(viii) SAE languages have a large number of characteristic properties in the area of phasal adverbials (expressions like *already, still, no longer, not yet*) (van der Auwera 1998b). These are rather well documented, but for the detail I have to refer the reader to van der Auwera's thorough study.

(ix) "Preterite decay": the loss of the old preterite and its replacement by the former present perfect. This is a change that occurred in the last millenium in French, German and northern Italian, as well as in some other adjacent European languages (cf. Thieroff 2000: 285). Its distribution is far narrower than that of the other Europeanisms, but it is the only feature of those studied by Thieroff whose geography comes close to Standard Average European (cf. also Abraham 1999).

Quite a few additional features have been mentioned in the earlier literature as characteristic of SAE, but earlier authors have sometimes neglected to make sure that a proposed Europeanism is not also common elsewhere in the world. Most of Whorf's original examples of SAE features seem to be of this kind. For instance, he notes that in contrast to SAE, Hopi lacks "imaginary plurals" (such as 'ten days', according to Whorf a "metaphorical aggregate"). But of course, we have no evidence that such plurals of time-span nouns are in any way characteristic of European languages. It may well be that they are common throughout the world. (To give Whorf his due, it must be added that he was not interested in demonstrating that SAE languages form a *Sprachbund*. He just used this term as a convenient abbreviation for "English and other European languages likely to be known to the reader", without necessarily implying that these languages are an exclusive club.)

4. Degrees of membership in SAE

Membership in a *Sprachbund* is typically a matter of degree. Usually there is a core of languages that clearly belong to the *Sprachbund*, and a periphery of surrounding languages that share features of the linguistic area to a greater or lesser extent.

In order to quantify the degrees of membership in SAE, a simple procedure suggests itself that was first applied to areal typology by van der Auwera (1998a). In addition to individual maps in which the lines denote isoglosses (as in Maps 107.1−12), we can combine different features in a single map and show the number of isoglosses shared by the language. Map 107.13 shows such a "cluster map" in which the lines stand for "quantified

isoglosses" (or "isopleths"). The map combines nine features of § 2.: definite and indefinite articles, relative clauses with relative pronouns, 'have'-perfect, participial passive, dative external possessors, negative pronouns and lack of verbal negation, relative-based equative constructions, subject person affixes as strict agreement markers, and intensifier-reflexive differentiation. The languages in the nucleus (French and German) show the SAE value in all nine of these features. The languages in the next layer (Dutch, other Romance, Albanian) show eight features, the next layer (English, Greek, Romanian) shows seven features, and so on. In this map, the resulting picture is actually very clear, because the SAE area with at least five SAE features stands out from the remaining languages, which have at most two SAE features.
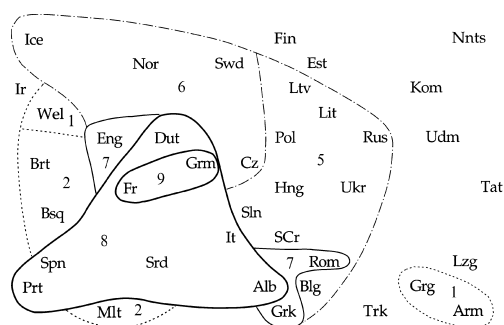
Map 107.13: A cluster map combining nine features

Such cluster maps are thus a fairly direct representation of degrees of membership in a linguistic area. But of course, the cluster map directly reflects the choice of features that are combined, and this choice is always somewhat arbitrary. Of the twelve features in § 2, only nine were selected here because information on the other three was incomplete. Ideally, the features of § 3 should have been added, too. But it seems to me that the main results of Map 107.13 would not be changed (this map can also be compared to the very similar map in van der Auwera (1998a: 823), which combines five adverbial features or feature clusters). The most striking features of Map 107.13 are:

(a) The nucleus of Standard Average European is formed by French and German (a finding that led van der Auwera (1998a: 824) to propose the term *Charlemagne Sprachbund* for the nuclear area of SAE). In view of the

historical role played by speakers of these two languages both in the early medieval history of continental Europe and in the very recent attempt at European unification, this is of course an extremely intriguing result.

(b) The southern European languages (both Romance and Balkan languages) are at least as close, if not closer to the nucleus than the northern languages and English. This means that it is misleading to call SAE features "Western European features", as is sometimes done. It is true that the Slavic languages in the east lack many SAE features, but the Balkan languages are generally more SAE than Slavic, although they are not western European.

(c) England stands somewhat apart from the European nucleus (as noted also by van der Auwera 1998a: 823), although it is closely related genealogically to German and has been thoroughly influenced by French. Since English is currently the dominant language throughout the world, it is worth pointing out its somewhat marginal status among its European sister languages.

It is important to keep in mind that the features on which Map 107.13 is based have not been selected randomly and are thus by no means representative of the morphosyntactic features of European languages. They were included precisely because they were known to show a distribution that supports the SAE hypothesis. Thus, no claim is made that all (or even the majority of) features will show a similar distribution. It is perfectly possible that we will some day discover another *Sprachbund*, based on a different set of features, that has Russian at its core and extends all the way to western Siberia in the east and central Asia in the south, but within Europe comprises only the Slavic, Balkan, and Scandinavian languages. This area would overlap with SAE, but it would not contradict it. Thus, a language may in principle belong to different linguistic areas, and different linguistic areas may coexist "on top of" each other. Since areal typology is only in its infancy, we do not know how common such situations are, but nothing in the logic of a *Sprachbund* implies that the world should be exhaustively divisible into non-overlapping *Sprachbünde*.

In fact, a number of smaller linguistic areas within Europe have been proposed in the literature (apart from the Balkan area, whose importance is not doubted by anyone),

e. g. by Lewy (1942), Wagner (1959), Décsy (1973), Haarmann (1976), and Ureland (1985) (cf. also Wintschalek 1993 on a Volga-Kama area). Currently the most thoroughly studied areas are the Circum-Baltic area (cf. Stolz 1991, Dahl & Koptjevskaja-Tamm (eds.) 2001) and the Mediterranean area (cf. Cristofaro & Putzu (eds.) 2000). However, no strong claims about a Circum-Baltic or a Mediterranean linguistic area seem to have been made as a result of these studies.

5.   How did SAE come into being?

Linguistic areas arise through language contact, but precisely which contact situation gave rise to Standard Average European is not immediately clear. And what is the source of the various Europeanisms: Who borrowed from whom? A full discussion of the socio-historical, cultural and sociolinguistic issues is beyond the scope of this article, so I will restrict myself here to mentioning just five possibilities:

(i)   retention of Proto-Indo-European structures and assimilation of some non-Indo-European languages to Indo-European language structure;

(ii) influence from a common substratum of a pre-Indo-European population in Europe;

(iii) contacts during the great transformations at the transition from late antiquity to the early Middle Ages in Europe;

(iv) the official language (Latin) and the common European culture of the Middle Ages;

(v)   the common European culture of modern times, from the Renaissance to the Enlightenment.

The fifth possibility must be rejected because a time depth of 300−500 years is not sufficient to account for grammatical commonalities of the kind discussed above. If lexical similarities between the European languages are discussed − for instance neoclassical compounding (*socio-/paleo-/ortho-/demo-*, *-graphy/-logy/-cracy*, etc.) or idiomatic structure (e. g. *ivory tower/torre d'avorio/Elfenbeinturm, as poor as a church mouse/pauvre comme un rat d'église/arm wie eine Kirchenmaus*) − then the last several centuries are the appropriate time frame for explaining the historical links, but the basic syntactic structures common to SAE languages must be older.

The first possibility must be rejected because the great majority of Europeanisms are innovations with respect to Proto-Indo-European. For instance, as far as we know, Proto-Indo-European did not have articles, a 'have'-perfect, "A and-B" conjunction, strict subject agreement, particle comparatives, or relative clauses with relative pronouns (cf. Lehmann 1974, Haspelmath 1998). With respect to Proto-Indo-European, and also with respect to the oldest Indo-European languages attested in Europe (Ancient Greek, Old Latin, Gothic), Standard Average European is clearly an innovation.

The second possibility, a pre-Indo-European substratum in Europe causing the SAE features, would be extremely difficult to demonstrate, but it might be worth pursuing. It is intriguing to note that the geographical space occupied by SAE languages coincides fairly precisely with the area of the Old European hydronymy, i. e. the homogeneous layer of river names discovered by Hans Krahe (see Vennemann 1994 for recent discussion). Vennemann (1994) proposes that these Old European hydronyms were not coined by an early prehistoric Indo-European population, but by a pre-Indo-European people which he calls Vasconic (the only surviving Vasconic language being Basque). Furthermore, the Old European hydronymy is hardly attested in the British Isles, where the Celtic languages are spoken, i. e. they could not have been influenced by the Vasconic substratum. This is in perfect harmony with the well-motivated hypothesis that the Celtic languages acquired some of their striking features from a different substratum related to the Afro-Asiatic languages (Pokorny 1927−30, Gensler 1993).

The main argument against the substratum view is that the SAE features seem to be gaining ground too late for a pre-Indo-European substratum to have caused them. Some SAE features appear only in the first millenium CE, but also the earlier features usually come fairly late, so that the earliest records of Indo-European-languages in Europe still show traces of the Proto-Indo-European patterns (e. g. causatives, relative clauses, locative comparative, "A B-and" conjunction). If these SAE features were caused by a substratum, then we should have much more evidence of the population speaking this substratum language. Moreover, a Vasconic substratum can hardly account for the SAE features because modern Basque is in most relevant ways very much unlike the SAE languages.

Of the remaining two possibilities, we can probably exclude option (iv) (the influence of Latin in the Middle Ages), because most SAE features were absent in Latin and developed only in the Romance languages. There are only two features for which Latin influence is a likely factor: negation and relative pronouns. In the case of these two features, the standard languages sometimes show deviations from the vernacular dialects, so at least the written standard languages may have been influenced by Latin, the European written language par excellence for many centuries. Thus, non-standard English has constructions like *I won't do nothing* ('I won't do anything'), and similarly in non-standard German and French (cf. Haspelmath 1997: 205). Analogously, Latin-type relative pronouns occur widely in the standard languages of Europe, but vernacular speech often prefers relative particles (Lehmann 1984: 88, 109). However, Latin probably only helped to reinforce these structures in those languages where they existed already independently as variants.

Thus, we are left with option (iii), the time of the great migrations at the transition between antiquity and the Middle Ages. This seems to be the appropriate time frame at least for articles, the 'have'-perfect, the participial passive, anticausatives, negative indefinites, nominative experiencers and verb fronting. The rise of these constructions can be observed only with difficulty because they were by and large absent in the written classical languages but seem to be well in place once the vernacular languages appear in the written record toward the end of the first millennium CE (cf. also Fehling 1980). This hypothesis derives some further plausibility from the fact that language contact must have been particularly intensive and effective during the great migrations, and in the case of French and northern Italian we have ample records of the lexical effects of these contacts. However, it is not so easy to fit features such as particle comparatives, „A and-B" conjunction and relative pronouns into this picture, because these features seem to have developed around the middle of the first millenium BC or even earlier (cf. Haspelmath 1998). Of course, we must always reckon with the possibility (or even likelihood) that different SAE features are due to different historical circumstances, and the correct picture is likely to be much more complicated than we can imagine at the moment, let alone discuss in this article.

6. Abbreviations of language names

Alb	Albanian
Arm	Armenian
Blg	Bulgarian
Brt	Breton
Bsq	Basque
Cz	Czech
Dut	Dutch
Eng	English
Est	Estonian
Fin	Finnish
Fr	French
Gae	Scots Gaelic
Grg	Georgian
Grk	Greek
Grm	German
Hng	Hungarian
Ice	Icelandic
Ir	Irish
It	Italian
Kom	Komi
Lat	Latin
Laz	Laz
Lit	Lithunian
Ltv	Latvian
Lzg	Lezgian
Mar	Mari
Mlt	Maltese
Mrd	Mordvin
Nnts	Nenets
Nor	Norwegian
Pol	Polish
Prt	Portuguese
Rom	Romanian
Rus	Russian
SAE	Standard Average European
Sam	Saami
SCr	Serbian/Croatian
Sln	Slovene
Spn	Spanish
Srd	Sardinian
Swd	Swedish
Tat	Tatar
Trk	Turkish
Uby	Ubykh
Udm	Udmurt
Ukr	Ukrainian
Wel	Welsh

7. References

Abraham, Werner. 1999. "Preterite decay as a European areal phenomenon". *Folia Linguistica* 33.1: 11−18.

Ameka, Felix. 1996. "Body parts in Ewe grammar". In: Chappell, Hilary & McGregor, William (eds.). *The grammar of inalienability: A typological*

perspective on body part terms and the part-whole relation. Berlin: Mouton de Gruyter, 783−840.

Bechert, Johannes & Bernini, Giuliano & Buridant, Claude (eds.). 1990. *Toward a typology of European languages*. (Empirical Approaches to Language Typology, 8.) Berlin: Mouton de Gruyter.

Beckman, Natanael. 1934. *Västeuropeisk syntax: Några nybildningar i nordiska och andra västeuropeiska språk*. [West European syntax: Some innovative constructions in the Nordic and other West European languages.] Göteborg: Göteborgs högskolas årsskrift.

Benveniste, Émile. 1960. "'Être' et 'avoir' dans leurs fonctions linguistiques". *Bulletin de la Societé de linguistique de Paris* 55: 113−134. (Reprinted in: Benveniste, Émile. 1966. *Problèmes de linguistique générale*. Paris: Gallimard, 187−207.)

Bernini, Giuliano & Ramat, Paolo. 1996. *Negative sentences in the languages of Europe: A typological approach*. (Empirical Approaches to Language Typology, 16.) Berlin: Mouton de Gruyter.

Bossong, Georg. 1998. "Le marquage de l'expérient dans les langues d'Europe". In: Feuillet (ed.), 259−294.

Comrie, Bernard. 1998. "Rethinking the typology of relative clauses". *Language Design* 1: 59−86.

Cristofaro, Sonia & Putzu, Ignazio (eds.) 2000. *Languages in the Mediterranean area: typology and convergence*. Milan: Franco Angeli.

Dahl, Östen. 1990. "Standard Average European as an exotic language". In: Bechert et al. (eds.), 3−8.

Dahl, Östen. 1995. "Areal tendencies in tense-aspect systems." In: Bertinetto, Pier Marco & Bianchi, Valentina & Dahl, Östen & Squartini, Mario (eds.) *Temporal reference, aspect and actionality*, vol. 2: *Typological perspectives*. Turin: Rosenberg & Sellier, 11−28.

Dahl, Östen. 1996. "Das Tempussystem des Deutschen im typologischen Vergleich." In: Lang, Ewald & Zifonun, Gisela (eds.). *Deutsch − typologisch*, Berlin: de Gruyter, 359−368.

Dahl, Östen (ed.). 2000. *Tense and aspect in the languages of Europe*. (Empirical Approaches to Language Typology-EUROTYP, 20−6.) Berlin: Mouton de Gruyter.

Dahl, Östen & Koptjevskaja-Tamm, Maria (eds.) (2001) *Circum-Baltic languages*. Vol 1−2. Amsterdam: Benjamins.

Décsy, Gyula. 1973. *Die linguistische Struktur Europas*. Wiesbaden: Harrassowitz.

Dryer, Matthew. 1989a. "Large linguistic areas and language sampling". *Studies in Language* 13: 257−92.

Dryer, Matthew. 1989b. "Article-noun order". *Chicago Linguistic Society* 25: 83−97.

Dryer, Matthew. 1998. "Aspects of word order in the languages of Europe". In: Siewierska, Anna (ed.) *Constituent order in the languages of Europe*. Berlin: Mouton de Gruyter, 283−319.

Fehling, Detlev. 1980. "The origins of European syntax". *Folia Linguistica Historica*, 1: 353−387.

Feuillet, Jack (ed.). 1998. *Actance et valence dans les langues d'Europe*. (Empirical Approaches to Language Typology-EUROTYP, 20−2.) Berlin: Mouton de Gruyter.

Garvin, Paul. 1949. "Standard Average European and Czech". *Studia Linguistica* 3: 65−85.

Gensler, Orin D. 1993. *A typological evaluation of Celtic/Hamito-Semitic syntactic parallels*. Ph. D. dissertation, University of California at Berkeley.

Gilligan, Gary M. 1987. *A cross-linguistic approach to the pro-drop parameter*. Ph. D. dissertation, USC.

Haarmann, Harald. 1976. *Aspekte der Arealtypologie: Die Problematik der europäischen Sprachbünde*. Tübingen: Narr.

Haspelmath, Martin. 1990. "The grammaticization of passive morphology". *Studies in Language*, 14.1: 25−71.

Haspelmath, Martin. 1993. "More on the typology of inchoative/causative verb alternations". In: Comrie, Bernard & Polinsky, Maria (eds.). *Causatives and transitivy*. Amsterdam: Benjamins, 87−120.

Haspelmath, Martin. 1997. *Indefinite pronouns*. Oxford: Oxford University Press.

Haspelmath, Martin. 1998. "How young is Standard Average European?" *Language Sciences* 20: 271−87.

Haspelmath, Martin. 1999. "External possession in an European areal perspective". In: Payne, Doris L. & Barshi, Immanuel (eds.) *External possession*. (Typological Studies in Language, 39.) Amsterdam: Benjamins, 109−135.

Haspelmath, Martin. 2001. "Non-canonical marking of core arguments in European languages". In: Aikhenvald, Alexandra & Dixon, R. M. W. (eds.) *Non-canonical subjects and objects*. (Typological Studies in Language) Amsterdam: Benjamins.

Haspelmath, Martin. To appear. "Coordination". In: Shopen, Timothy (ed.) *Language typology and linguistic description*. 2nd ed. Cambridge: Cambridge University Press.

Haspelmath, Martin & Buchholz, Oda. 1998. "Equative and similative constructions in the languages of Europe". In: van der Auwera (ed.), 277−334.

Heine, Bernd. 1994. "Areal influence on grammaticalization". In: Pütz, Martin (ed.) *Language contact and language conflict*. Amsterdam: Benjamins, 55−68.

Heine, Bernd. 1997. *Possession: Cognitive sources, forces, and grammaticalization*. Cambridge: Cambridge University Press.

Jakobson, Roman. 1931. "K xarakteristike evrazijskogo jazykovogo sojuza". [Characterizing the Eurasian linguistic area.] Reprinted in: Jakobson,

Roman. 1971. *Selected writings, vol. I.* The Hague: Mouton, 144−201.

Kahrel, Peter. 1996. *Aspects of negation.* Ph. D. dissertation: University of Amsterdam.

König, Ekkehard & Haspelmath, Martin. 1998. "Les constructions à possesseur externe dans les langues d'Europe". In: Feuillet (ed.), 525−606.

König, Ekkehard & Haspelmath, Martin. 1999. "Der europäische Sprachbund". In: Reiter, Norbert (ed.). *Eurolinguistik.* Wiesbaden: Harrassowitz, 111−127.

König, Ekkehard & Siemund, Peter. 1999. "Intensifiers and reflexives: a typological perspective." In: Frajzyngier, Zygmunt & Curl, Traci S. (eds.) *Reflexives: Forms and functions* (Typological studies in language, 40.). Amsterdam: Benjamins, 41−74.

Kortmann, Bernd. 1997. *Adverbial subordination: A typology and history of adverbial subordinators based on European languages.* Berlin: Mouton de Gruyter.

Lazard, Gilbert. 1990. "Caractéristique actancielles de l'"européen moyen type'." In: Bechert et al. (eds.), 241−53.

Lazard, Gilbert. 1998. „Définition des actants dans les langues européennes". In: Feuillet (ed.), 11−146.

Lehmann, Christian. 1984. *Der Relativsatz.* Tübingen: Narr.

Lehmann, Winfred P. 1974. *Proto-Indo-European syntax.* Austin: University of Texas Press.

Lewy, Ernst. 1942. *Der Bau der europäischen Sprachen.* (Proceedings of the Royal Irish Academy, Vol. 48). Dublin: Hodges & Figgis.

Masica, Colin P. 1976. *Defining a linguistic area.* Chicago: University of Chicago Press.

Nedjalkov, Igor' V. 1998. "Converbs in the languages of Europe". In: van der Auwera (ed.), 421−455.

Nichols, Johanna. 1992. *Linguistic diversity in space and time.* Chicago: The University of Chicago Press.

Nichols, Johanna & Peterson, David A. & Barnes, Jonathan (to appear). "Causativizing and decausativizing languages". Ms., University of California, Berkeley,

Nikolaeva, Irina. 1999. *Ostyak.* Munich: Lincom Europa.

Nocentini, Alberto. 1996. "Tipologia e genesi dell'articolo nelle lingue europee". *Archivio Glottologico Italiano* 81.1: 3−44.

Pisani, Vittore. 1969. "Tipi d'accento nelle lingue dell'Europa". In: *Actes du Xe Congres International des Linguistes, Bucuresti*, vol. II, 57−60. Bucuresti.

Pokorny, Julius. 1927−30. "Das nicht-indogermanische Substrat im Irischen". *Zeitschrift für celtische Philologie* 16: 95−144, 231−266, 363−394; 17: 373−388; 18: 233−248.

Ramat, Paolo & Bernini, Giuliano. 1990. "Area influence versus typological drift in Western Europe: the case of negation". In: Bechert et al. (eds.), 25−46.

Siewierska, Anna. 1999. "From anaphoric pronoun to grammatical agreement marker: Why objects don't make it". *Folia Linguistica* 33.1−2: 225−51.

Stassen, Leon. 1985. *Comparison and universal grammar.* Oxford: Blackwell.

Stassen, Leon. 2000. "AND-languages and WITH-languages". *Linguistic Typology* 4.1: 1−54.

Stolz, Thomas. 1991. *Sprachbund im Baltikum? Estnisch und Lettisch im Zentrum einer sprachlichen Konvergenzlandschaft.* Bochum: Brockmeyer.

Stolz, Thomas. 1996. "Some instruments are really good companions − some are not: On syncretism and the typology of instrumentals and comitatives". *Theoretical Linguistics* 23.1−2: 113−200.

Stolz, Thomas. 2001a. "No Sprachbund beyond this line! On the age-old discussion on how to define a linguistic area". *Sprachtypologie und Universalienforschung* 54.

Stolz, Thomas. 2001b. "Ordinalia − Linguistisches Neuland: Ein Typologenblick auf die Beziehung zwischen Kardinalia und Ordinalia und die Sonderstellung von EINS und ERSTER". In: Igla, Birgit & Stolz, Thomas (eds.) *Was ich noch sagen wollte … Festschrift für Norbert Boretzky zu seinem 65. Geburtstag.* Berlin: Akademie-Verlag.

Ternes, Elmar. 1998. "Lauttypologie der Sprachen Europas". In: Boeder, Winfried & Schroeder, Christoph & Wagner, Karl Heinz & Wildgen, Wolfgang (eds.) *Sprache in Raum und Zeit: In memoriam Johannes Bechert. Band 2.* Tübingen: Narr, 139−152.

Thieroff, Rolf. 2000. "On the areal distribution of tense-aspect categories in Europe". In: Dahl (ed.), 265−305.

Ultan, Russell. 1978. "Some general characteristics of interrogative systems". In: Greenberg, Joseph H. (ed.) *Universals of human language, vol. 4.* Stanford: Stanford University Press, 211−248.

Ureland, P. Sture. 1985. "Sprachkontakt und Glottogenese in Europa". In: P. Sture Ureland (ed.) *Entstehung von Sprachen und Völkern: Glotto- und ethnogenetische Aspekte europäischer Sprachen.* Tübingen: Niemeyer, 7−43.

van der Auwera, Johan. 1998a. "Conclusion". In: van der Auwera (ed.), 813−36.

van der Auwera, Johan. 1998b. "Phasal adverbials in the languages of Europe". In: van der Auwera (ed.), 25−145.

van der Auwera, Johan (ed.). 1998. *Adverbial constructions in the languages of Europe.* (Empirical Approaches to Language Typology-EUROTYP, 20−3.) Berlin: Mouton de Gruyter.

van der Hulst, Harry & Hendriks, Bernadet & van de Weijer, Jeroen. 1999. "A survey of word pro-

sodic systems of European languages". In: van der Hulst, Harry (ed.) *Word prosodic systems in the languages of Europe*. (Empirical Approaches to Language Typology-EUROTYP, 20−4.) Berlin: Mouton de Gruyter, 425−475.

Vennemann, Theo. 1994. "Linguistic reconstruction in the context of European prehistory". *Transactions of the Philological Society* 92: 213−282.

Wagner, Heinrich. 1959. *Das Verbum in den Sprachen der britischen Inseln: ein Beitrag zur geographischen Typologie des Verbums*. Tübingen: Niemeyer.

Whorf, Benjamin Lee. 1941. "The relation of habitual thought and behavior to language". In: Spier, Leslie (ed.) *Language, culture, and personality: Es-*

says in memory of Edward Sapir. Menasha, Wis.: Sapir Memorial Publication Fund, 75−93. [Reprinted in Whorf (1956), 134−159.]

Whorf, Benjamin Lee. 1956. *Language, thought, and reality: Selected writings of Benjamin Lee Whorf*. Edited by John B. Carroll, Cambridge/MA: MIT Press.

Wintschalek, Walter. 1993. *Die Areallinguistik am Beispiel syntaktischer Übereinstimmungen im Wolga-Kama-Areal*. (Studia Uralica, 7.) Wiesbaden: Harrassowitz.

Martin Haspelmath, MPI Leipzig
(Deutschland)

108. Aire linguistique balkanique

1.   Généralités

La linguistique balkanique est une discipline relativement récente, bien que la découverte de traits communs entre les langues balkaniques remonte à la première moitié du XIXe siècle. Les spécialistes (Asenova 1979: 5−45; Schaller 1975: 37−45) s'accordent à diviser l'histoire de la discipline en trois périodes: une période *préliminaire*, où l'on cherche à expliquer les traits communs par l'influence du substrat, une période *classique* où la linguistique balkanique acquiert ses lettres de noblesse grâce à la publication en 1930 de *Linguistique balkanique. Problèmes et résultats* de Sandfeld, qui représente la première synthèse complète, et une période *moderne*, marquée par le polycentrisme et l'internationalisation des recherches (nombreuses revues spécifiques et organisation de congrès).

La linguistique balkanique ne consiste pas à juxtaposer des descriptions de langues diverses dont le seul lien serait la contiguïté géographique: il faut que ces langues forment une « union linguistique » (*Sprachbund*). Même si certaines voix s'élèvent encore pour nier la réalité de l'union balkanique (Andriotis & Kourmoulis 1968), la plupart des linguistes sont convaincus de son existence. En

effet, les traits communs sont trop nombreux pour qu'ils soient le fruit du hasard. Il est vrai que les spécialistes discutent encore de la notion de « balkanisme », que l'on définira ici comme un trait typologique propre à au moins trois langues de l'union. Ce trait n'a pas besoin d'être unique en son genre (ainsi, l'article défini postposé existe dans les langues scandinaves, le « redoublement » de l'objet se retrouve dans les langues romanes); il doit être le résultat d'une convergence qui aboutit à un résultat identique ou quasi identique, alors qu'il n'existait pas à des stades plus anciens.

Les tâches de la linguistique balkanique sont consignées dans l'histoire de la discipline. Elles ont un triple aspect: synchronique (description) panchronique (extension) et diachronique (formation et évolution). Bien que l'essentiel du travail descriptif semble avoir été acompli (la monographie de Sandfeld a été complétée, souvent améliorée, par des centaines d'articles et d'études de détail qui ont permis d'accroître et d'approfondir les données), il reste toujours beaucoup à faire. L'étude de l'extension des balkanismes nécessite le recours à la géographie linguistique (ou linguistique aréale) pour déterminer avec exactitude le lieu d'apparition de chaque balkanisme et son extension réelle sur le terrain. Enfin, la perspective diachronique n'est jamais perdue de vue par les balkanologues, malgré les nombreuses difficultés auxquelles ils sont confrontés, faute de documents écrits. Trois aspects sont à prendre en considération: 1) La genèse de l'union linguistique balkanique; 2) La genèse des balkanismes; 3)